Epidemiol. Infect. (2001), **126**, 415–424. Printed in the United Kingdom

Comparative utility of restriction fragment length polymorphism analysis and gene sequencing to the molecular epidemiological investigation of a viral outbreak

T. L. GOLDBERG*, R. M. WEIGEL, E. C. HAHN AND G. SCHERBA

University of Illinois, Department of Veterinary Pathobiology, 2001 South Lincoln Avenue, Urbana IL 61820, USA

(*Accepted 17 January 2001*)

SUMMARY

Restriction fragment length polymorphism (RFLP) analysis and partial-genome DNA sequencing are commonly used to infer genetic relationships among pathogens. This study compares the application of both techniques to the analysis of 16 pseudorabies virus isolates collected during a 1989 outbreak. Genetic distances derived from RFLP and DNA sequence data were not significantly correlated with geographic distances between farms from which isolates were collected. RFLP-based genetic distance was, however, strongly correlated with temporal distance between isolates (days separating time of isolation). Sequence-based genetic distance was significantly correlated with temporal distance only when synonymous changes (nucleotide changes not leading to amino acid changes) were considered separately. Conversely, non-synonymous changes were correlated with the host species of origin of the viral isolate. These results indicate that selectively-neutral genetic changes most accurately reflect historical relationships, but that non-neutral changes most accurately reflect the biological environment of the viral isolate (e.g. host immune system).

INTRODUCTION

Defining genetic relationships among pathogenic organisms for the purpose of inferring patterns of transmission constitutes one of the principal goals of molecular epidemiology. Molecular genetic data have the distinct advantage in this regard that they can be collected after an outbreak has occurred, if appropriate biological samples have been saved. The spatial and temporal course of an epidemic can therefore be inferred without historical bias, even in the absence of direct observation or traditional retrospective epidemiological data.

Two molecular techniques commonly used for defining genetic relationships among pathogenic organisms are restriction fragment length polymorphism (RFLP) analysis, and direct nucleotide sequencing. RFLP is fast and inexpensive, and can detect variation across large genomic regions [1, 2]. However, RFLP detects only a small proportion of the total genetic variation present [3, 4]. Sequencing yields more complete information about genetic variation at the locus sequenced [5]. However, its greater expense and technical requirements generally limit its application to small genomic regions (usually $\leqslant$ 1000 bases [2, 5]). These regions may or may not be representative of the genome as a whole.

Despite widespread use and acceptance of these two techniques, few studies have compared the utility of RFLP directly to that of gene sequencing in reconstructing the pattern of spread of an infectious disease agent. The goal of this study is to assess the utility and comparability of RFLP analysis and sequencing in an epidemiologic setting such as might commonly be encountered in the field.

The study uses as a model pseudorabies virus (PrV), an alphaherpesvirus. PrV is the causative agent of

* Author for correspondence.

Aujezky's disease, an economically significant disease of commercial swine, which can also fatally infect a variety of domestic and wild mammalian species [6]. PrV is currently the focus of eradication programmes worldwide, which have dramatically decreased the incidence of the disease [7]. In many countries where they now occur, outbreaks of PrV are short-lived and geographically localized; inferences about the courses of these outbreaks must therefore be made retrospectively.

The specific focus of this study is an outbreak of PrV which occurred over a 10-month period in a limited geographic region of Illinois, USA, between January and November 1989. Sixteen PrV isolates from 14 farms clustered in 5 geographic areas were obtained during this time [8]. Because the actual pattern of spread of PrV among the 14 farms in 1989 remains unknown, this study focuses on the internal comparability of the two genetic techniques. The study tests the hypothesis that both techniques are equally informative for reconstructing the spread of PrV across space and time.

METHODS

The 16 PrV isolates were obtained by direct virus isolation from infected tissues of 5 species (bovine, canine, ovine, porcine, raccoon). Whole genomic DNA was extracted from each viral isolate for subsequent separate digestion with four restriction endonucleases (*Alw*44I, *Bam*HI, *Sal*I, and *Xho*I). Laboratory techniques involved equal molar ratio end labeling of DNA fragments to improve sensitivity and to allow overlapping bands to be detected [8].

A matrix of RFLP distance between all isolates was calculated using the square root variant of a distance metric described by Weigel and Scherba [9]. This metric calculates genetic distances between pairs of isolates based on a formula that compares the distribution of restriction endonuclease fragment sizes, and has shown to be highly informative for inferring genetic relatedness between PrV isolates [8, 9].

The glycoprotein C (gC) gene of PrV was the locus chosen for sequencing. gC is among the best studied proteins encoded by the 150 kb viral genome. *g*C is a 98 kDa monomeric membrane-associated glycoprotein which is considered non-essential to viral replication in cell culture [10, 11]. However, *in vivo*, gC plays a critical role in attachment of the virus to target cells via binding with heparan sulfate receptors [12, 13], although the binding properties of gC are not absolutely required for infection [14, 15].

Immunologically, gC is a major target for both cellular and humoral immunity in pigs [16–19]. Sequence-level genetic variation in gC has previously been described as 2–3% among strains from diverse geographic areas [20].

Gene sequencing was conducted on a 798 bp fragment of the 5′ end of the approximately 1440 bp gC gene, where much of the variation in gC is concentrated. This fragment is appropriate in that its size is representative of the length of DNA that a typical molecular epidemiological field investigation might practically be able to examine. Laboratory techniques included gene amplification using the polymerase chain reaction and specific oligonucleotide primers, and are described in detail elsewhere [21].

Observed nucleotide changes were divided into synonymous and non-synonymous changes based on their predicted effect on the resulting amino acid sequence. Specifically, those sites at which a change from any base to any other base would not alter the amino acid at that position (fourfold-degenerate sites) were analysed separately, because of their assumed selective neutrality [22]. Non-synonymous changes (those which do lead to changes in the amino acid sequence) were also treated separately for some analyses, because these sites are assumed to be those acted upon most strongly by selective forces. Matrices of pairwise sequence distance between isolates were calculated as uncorrected proportions of differing nucleotide positions using the computer programme PHYLIP [23]. Three separate matrices were calculated: one for all sequence changes together, one for neutral changes only and one for non-neutral changes only.

Epidemiologic data of three types were collected: geographic coordinates of the farms from which the isolates were obtained; dates on which the isolates were collected from the field; and the species of origin of each isolate. Distance matrices between isolates were created for each type of data as follows. Geographic distances (km) between isolates were calculated directly from the longitudes and latitudes of their farms of origin. Temporal distances (days) between pairs of isolates were calculated as the number of days separating their collection. Species distances between isolates were assigned a value of zero if the two isolates were collected from the same species, or one if they were collected from different species.

Table 1. *Nucleotide sequence variation for 798 base pairs of the PrV gC gene in 16 isolates from Illinois and 4 reference sequences*

					Variable nucleotide position*
					0 0 1 1 1 1 2 2 2 2 2 2 3 3 4 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7
					0 3 6 6 6 8 2 2 2 4 5 5 4 9 8 9 9 9 0 0 0 0 1 2 2 3 5 6 9 9
					5 6 5 6 7 2 4 5 6 5 3 6 5 5 1 2 4 8 2 3 4 5 0 1 9 0 7 4 0 8
					Condon position
					3 1 1 2 3 3 3 1 2 3 2 2 1 3 2 2 1 2 3 1 2 3 2 1 1 2 2 3 2 1
Isolate	Date†	Latitude‡	Longitude‡	Species§	Sequence¶
PrV 43	1/5/89	41·27	−90·08	Ov	GCGGCAGGGGCTACGTGTG--AGAGGCAC
PrV 4411	1/12/89	41·53	−90·08	Ov	.......CC....G.........G..G..
PrV 4520	11/27/89	41·46	−89·95	Ca	..---.A..C...GA.............C.
PrV 7438	1/3/89	41·56	−90·12	Po	.......CC....G.........G.....
PrV 7652	1/19/89	41·19	−89·95	Ca	.G.......A...G..............
PrV 7739	1/14/89	41·54	−90·07	Ca	.G..........G..............
PrV 8033	1/14/89	41·55	−90·07	Ra	.......CC....G.........G.....
PrV 8044	1/3/89	41·21	−89·95	Bo	.....C.CCCGCGGACCACGACG...T..A
PrV 8095	1/19/89	41·19	−89·95	Po	.G..........G..............
PrV 9164	2/14/89	41·28	−90·05	Po	.G..........G..............
PrV 10501	3/14/89	41·25	−90·27	Po	............G..........A....
PrV 10649	2/15/89	41·24	−90·09	Ov	CG..........G..............
PrV 11243	4/5/89	41·16	−90·01	Po	............G..............
PrV 12271	4/20/89	41·32	−90·22	Po	..---.A..C...GA.............C.
PrV 12481	4/23/89	41·31	−90·33	Po	..---.A..C...GA........A.....
PrV 12486	4/26/89	41·33	−90·29	Po	..---.A..C...GA.............C.
Ea (China)					.....C....G..GAC..CGACG...A..A
Yamagata S-81 (Japan)					.....C.CCCGC.G.CCACGACG......A
Indiana S (USA)					.....C.CCCGC.GA..............
NIA-3 (Northern Ireland)					.......CC....GA..............

* Position numbers are read vertically. Only variable positions within the Illinois sequences are shown. Underlined positions are fourfold-degenerate sites. Complete Illinois sequences are available through GenBank (accession numbers AF176479-AF176495).

† Dates refer to the dates on which samples were collected in the field.

‡ Latitudes and longitudes are expressed in decimal degrees.

§ Species from which virus was isolated (Bo, bovine; Ca, canine; Ov, ovine; Po, porcine; Ra, raccoon).

¶ identity with first (reference) sequence, -, gap in sequence.

Table 2. *Genetic similarity between PrV isolates based on distributions of restriction fragment sizes. A value of zero indicates genetic identity. Data are taken from Scherba* et al. [8]

Isolate	PrV 10501	PrV 12481	PrV 12486	PrV 12271	PrV 7438	PrV 7739	PrV 8033	PrV 4411	PrV 7652	PrV 8095	PrV 8044	PrV 11243	PrV 9164	PrV 10649	PrV 43	PrV 4520
PrV 10501	0·00	0·78	1·20	0·28	0·38	0·34	0·96	0·39	0·63	0·63	2·22	0·67	0·85	1·20	1·24	0·74
PrV 12481	0·78	0·00	0·14	1·20	0·85	1·20	0·49	0·94	0·39	0·43	1·43	1·52	0·37	0·88	0·75	1·12
PrV 12486	1·20	0·14	0·00	1·76	1·14	1·76	0·52	1·37	0·67	0·73	1·18	2·08	0·45	1·12	0·82	1·16
PrV 12271	0·28	1·20	1·76	0·00	0·38	0·27	1·35	0·40	0·73	0·71	2·49	0·49	1·39	1·05	1·23	0·79
PrV 7438	0·38	0·85	1·14	0·38	0·00	0·31	0·92	0·26	0·64	0·65	1·78	0·73	0·97	1·06	1·09	0·67
PrV 7739	0·34	1·20	1·76	0·27	0·31	0·00	1·20	0·30	0·63	0·62	2·40	0·62	1·29	0·73	0·93	1·09
PrV 8033	0·96	0·49	0·52	1·35	0·92	1·20	0·00	0·69	0·30	0·33	0·62	1·17	0·50	1·15	1·07	0·81
PrV 4411	0·39	0·94	1·37	0·40	0·26	0·30	0·69	0·00	0·39	0·39	1·67	0·38	0·99	1·18	1·23	0·70
PrV 7652	0·63	0·39	0·67	0·73	0·64	0·63	0·30	0·39	0·00	0·03	1·06	0·71	0·53	0·66	0·66	0·92
PrV 8095	0·63	0·43	0·73	0·71	0·65	0·62	0·33	0·39	0·03	0·00	1·07	0·65	0·57	0·71	0·71	0·90
PrV 8044	2·22	1·43	1·18	2·49	1·78	2·40	0·62	1·67	1·06	1·07	0·00	2·20	1·59	2·04	1·83	1·82
PrV 11243	0·67	1·52	2·08	0·49	0·73	0·62	1·17	0·38	0·71	0·65	2·20	0·00	1·40	1·59	1·73	1·03
PrV 9164	0·85	0·37	0·45	1·39	0·97	1·29	0·50	0·99	0·53	0·57	1·59	1·40	0·00	0·89	0·81	0·91
PrV 10649	1·20	0·88	1·12	1·05	1·06	0·73	1·15	1·18	0·66	0·71	2·04	1·59	0·89	0·00	0·12	2·03
PrV 43	1·24	0·75	0·82	1·23	1·09	0·93	1·07	1·23	0·66	0·71	1·83	1·73	0·81	0·12	0·00	2·03
PrV 4520	0·74	1·12	1·16	0·79	0·67	1·09	0·81	0·70	0·92	0·90	1·82	1·03	0·91	2·03	2·03	0·00

Correlations between matrices were performed using Mantel tests of matrix correlation [24], conducted with computer programme *The R Package* [25]. A standardized form of the Mantel test statistic (r) was used [26]. For each matrix correlation performed, a null sampling distribution of r was created using 10000 Monte Carlo random permutations of the observed distance matrices [27]. Empirical r values were compared to their respective null distributions, and probabilities were calculated as the proportion of null values as extreme or more extreme than the observed values of r. One-tailed probabilities were calculated when correlations were in the predicted direction (positive in all cases).

Four additional PrV gC sequences of known geographic provenance were obtained from the literature for purposes of comparison with the newly-generated sequences [20, 28]. Sequence data were analysed using the maximum parsimony search algorithm and bootstrap analysis technique of the computer programme PAUP [29]. Dendrograms of relationship for sequence, RFLP, spatial and temporal distance were created using the UPGMA algorithm of the computer programme PHYLIP [23, 30]. All statistical results were considered significant at the $\alpha = 0{\cdot}05$ level.

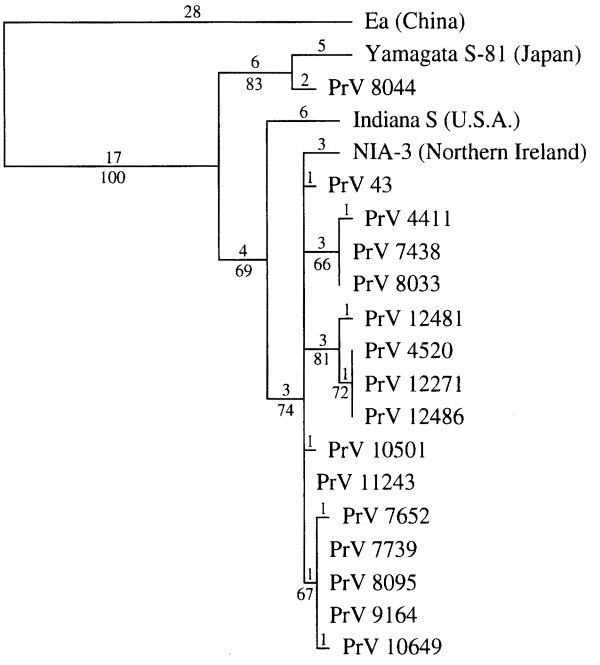

Fig. 1. Maximum parsimony phylogenetic tree of Illinois pseudorabies virus isolates (beginning with 'PrV'), and previously-published sequences of isolates from known geographic locations around the world. Numbers above branches represent branch lengths (minimum numbers of nucleotide changes). Numbers below branches are bootstrap values (%), based on 100 heuristic bootstrap replicates of the data. Single-codon insertions/deletions were weighted twice as heavily as point substitutions.

RESULTS

Geographic, temporal, species and sequence data for all 16 PrV isolates, and for 4 additional sequences from the literature, are presented in Table 1. Farms were separated by distances ranging from 2·2 to 43·5 km. Temporal distances (days separating the collection of isolates) ranged from 0 to 297 days. Sequence-level genetic distances between isolates ranged from 0 to 1·94%, with a mean pairwise sequence distance of 0·82% (gaps excluded). The 30 variable nucleotide positions were approximately equally divided between first (9), second (12), and third (9) codon positions. Six of the 9 variable third codon positions were fourfold-degenerate sites.

Results of RFLP analysis of these isolates are presented in Table 2. This matrix contains cells with values representing the genetic distances between pairs of isolates based on a formula that compares the distribution of restriction endonuclease fragment sizes [9]. In the present sample, genetic distance values ranged between zero (identity) and 2·49.

Figure 1 shows a maximum parsimony phylogenetic tree based on gC sequence data depicting the relationships among the 16 Illinois isolates and the 4 previously-published isolates from around the world. Fifteen of the 16 isolates, together with the NIA-3 isolate from Northern Ireland, form a single well-supported clade (bootstrap values of 74%) of closely-related sequences, within which appear three sub-clades of very similar or identical sequences from Illinois. The previously-sequenced Indiana S strain from the United States appears as an outgroup to this clade. Of the newly-sequenced isolates, only PrV 8044 appears genetically divergent, clustering with the Yamagata S-81 strain from Japan. The Ea strain from China appears highly divergent in this tree.

Figure 2 shows four dendrograms constructed from sequence, RFLP, geographic and temporal data using the UPGMA clustering algorithm [30]. The species of origin of each strain is labeled on this tree. In 3 of the 4 dendrograms (RFLP, Geography and Time), isolate 4520 appears divergent. Isolate 8044 is also divergent in both genetic dendrograms. Non-porcine species of origin do not form exclusive clusters in any tree. This would be expected, since PrV is rapidly fatal in non-porcine species and is, therefore, most likely to have

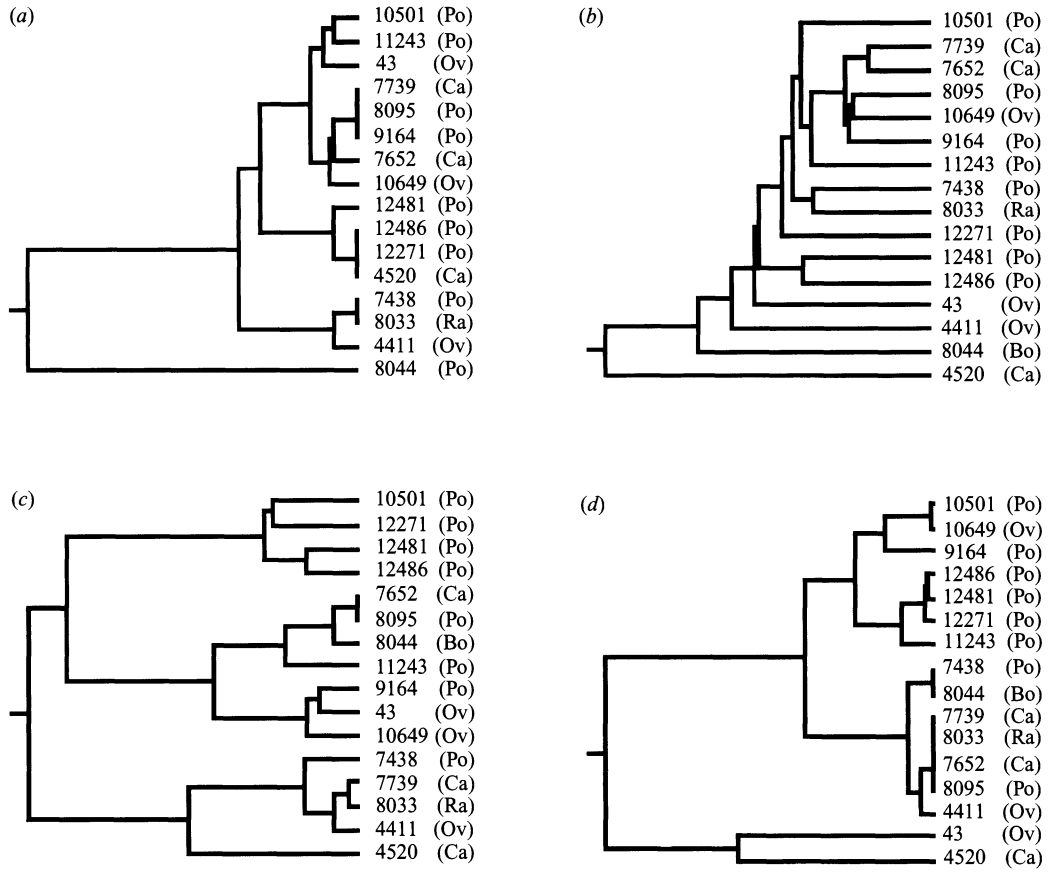

Fig. 2. UPGMA dendrograms showing relationships among the 16 Illinois PrV isolates. Isolates are identified by number (see Table 1), with species of origin in parentheses: (*a*) dendrogram based on matrix of pairwise nucleotide sequence distances between isolates; (*b*) dendrogram based on matrix of pairwise RFLP distances between isolates; (*c*) dendrogram based on matrix of pairwise geographic distances between isolates; (*d*) dendrogram based on matrix of pairwise temporal distances between isolates. Dendrograms are scaled to unit length to facilitate comparisons.

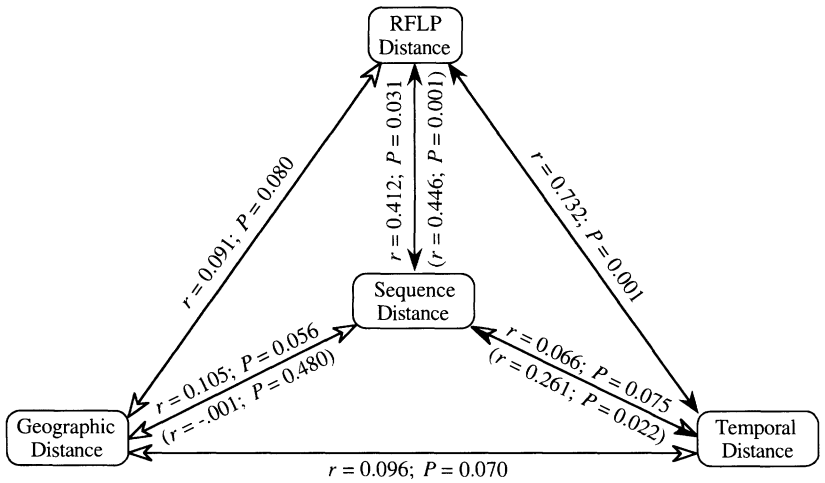

Fig. 3. Mantel correlation coefficients (*r*) between four similarity matrices for 16 Illinois PrV isolates. Lines with filled arrows represent statistically significant associations. Values in parentheses represent correlations run using only fourfold degenerate sites. (see Table 1).

been transmitted to these other species from pigs. Indeed, there is evidence of localized subclustering by species of origin on both genetic trees, particularly in the case of the porcine isolates, which appear to cluster most closely with other porcine isolates. However, this observation may be an artifact of non-

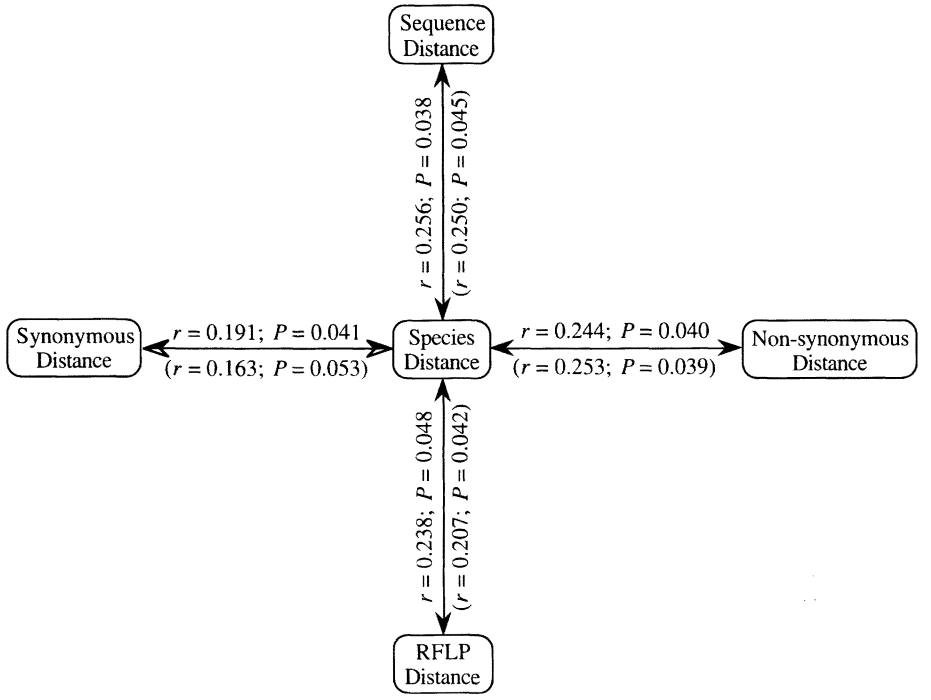

Fig. 4. Mantel correlation coefficients (r) between four different genetic distance matrices and a matrix of 'species distance.' Lines with filled arrows represent statistically significant associations. Values in parentheses represent partial Mantel correlations, with temporal distance held statistically constant.

random sampling, since more porcine isolates were analysed than were isolates from any other species.

Figure 3 shows the results of the Mantel correlation analysis used to quantify the genetic, spatial and temporal associations depicted qualitatively in Figure 2. RFLP and Sequence distance were significantly correlated. No significant correlations were found between sequence or RFLP distance and geographic distance, or between geographic distance and temporal distance. However, RFLP distance was highly correlated with temporal distance between isolates; sequence distance was not.

When selectively-neutral changes were considered separately, the strength of association between RFLP and sequence distance increased slightly ($\Delta r = 0{\cdot}034$). However, the strength of association between sequence distance and temporal distance increased considerably ($\Delta r = 0{\cdot}195$), and the correlation became statistically significant. Sequence-level genetic distance at fourfold-degenerate sites was only moderately correlated with sequence-level genetic distance at non-synonymous sites (Mantel $r = 0{\cdot}286$; $P = 0{\cdot}013$).

Both RFLP and sequence-level genetic distance were significantly correlated with identity/non-identity of species of origin (Fig. 4), indicating that, in general, isolates from the same species tend to be more closely related genetically to other isolates from that species than would be expected by chance alone. However, the strength of this association was higher when non-synonymous changes were considered separately than when synonymous changes were considered separately ($\Delta r = 0{\cdot}053$).

As shown in Figure 3, there is a strong association between genetic distance (especially RFLP) and temporal distance between isolates. The correlation between species-of-origin and temporal distance was not significant (Mantel $r = 0{\cdot}134$; $P = 0{\cdot}114$), but was strong enough to warrant concern that time could be a confounding variable in the analysis of species distance and genetic distance. Therefore, Figure 4 presents the results of partial Mantel correlations [26] between species distance and the four types of genetic distance, contingent on temporal distance (values in parentheses). With temporal distance held statistically constant, the correlation between species distance and synonymous sequence distance was slightly lower and non-significant. All other correlations were only minimally affected.

DISCUSSION

The level of genetic variability documented in PrV gC, while low, was nevertheless sufficient for drawing epidemiologic conclusions about the 1989 outbreak.

Both RFLP and sequence data were able to distinguish among isolates. Both techniques allowed hierarchical relationships to be defined, even though the isolates were sampled from a small geographic region over a short period of time. These results are encouraging, since DNA viruses such as PrV are notoriously invariant genetically relative to viruses with RNA genomes [31]. A previous study of genetic variability in Illinois isolates of porcine reproductive and respiratory syndrome virus (PRRSV; an RNA arterivirus) documented approximately seven times as much variation at the nucleotide level as was observed in the current sample of PrV [21].

As demonstrated by phylogenetic analyses, the Illinois PrV gC isolates display less variability than exists within a sample of gC from geographically diverse locations around the world. In fact, the Illinois samples (with one exception) clustered into a single clade. This would be expected if the Illinois samples did indeed represent an outbreak, in the sense that all samples arose recently from a common ancestor. However, this relationship is not entirely supported phylogenetically. Sample 8044 is divergent phylogenetically, clustering with the geographically unrelated Yamagata S-81 strain from Japan (see Fig. 1). The divergent position of sample 8044 might accurately reflect an international origin for this sample. However, sample 8044 was also the only sample obtained from a bovine host. The associated farm, while it did not raise pigs, was close to swine farms which vaccinated for PrV using a vaccine closely-related (but not identical) genetically to sample 8044 [8]. This sample's divergent phylogenetic position may therefore reflect an origin from a vaccine strain.

Statistical analysis using Mantel tests of matrix correlation substantiated the absence of any strong association between geographic and genetic distance. Neither RFLP nor sequence-level genetic distance between isolates was correlated with geographic distance.

The lack of a genetic/geographic correlation weakens the hypothesis that PrV during this outbreak tended to spread form farm to nearby farm. 'Distance-limited' processes of inter-farm transmission such as wildlife vectors or aerosols carried in wind [32–34] would lead to the prediction that isolates collected from nearby farms should be closely related genetically. Given the current sample size (16), an α level of 0·05, and β of 0·8, this study could have detected a strength of correlation of geographic proximity on genetic similarity of 30·6% or higher. That such an effect was not detected suggests that any such association must be weak. Most likely, long-distance processes of inter-farm transmission, such as the inadvertent transportation of viruses between farms by people, must have predominated during this outbreak [8].

In contrast, a strong association was documented between genetic distance (RFLP) and temporal distance between isolates. Again, this is consistent with an outbreak situation, if indeed the movement of virus from one location to another tends to be marked by progressive genetic change. This same association was not documented for sequence-level genetic distance until fourfold-degenerate sites were analysed separately. Apparently, selectively neutral genetic variation was more 'clock-like' than non-neutral variation in that it more accurately tracked the temporal course of this outbreak. Indeed, the low correlation between non-synonymous and synonymous sequence-level genetic distance supports the assumption that different evolutionary forces may be driving each type of genetic change.

The pattern of correlation between four measures of genetic distance and identity or non-identity of host species of origin (see Fig. 4) provides support for the hypothesis that one force driving sequence-level genetic change may be natural selection. When Mantel tests of partial correlation were used to control statistically for the influence of time, non-synonymous sequence changes more accurately reflected host species of origin than did synonymous sequence changes. Biologically, PrV samples of a particular genetic constitution may be predisposed to establishing a productive infection in a particular host. Inherent differences in the ability of certain variants to adhere to host cells or to evade the host's immune system probably underlie this phenomenon. Such a conclusion assumes that PrV in swine exists as a diversity of genetic types, akin to the 'quasispecies' typically used to describe RNA viruses [31, 35]. Only those variants genetically suited to a given host species will establish an infection, other variants being selected against.

Comparing RFLP and sequencing directly indicated that both techniques yielded similar results with regard to reconstructing the epidemiologic pattern of this outbreak. Genetic distances derived from the two techniques were significantly correlated (see Fig. 3). Both techniques indicated a lack of correlation between genetic and geographic distance but significant correlation between genetic and temporal dis-

tance and between genetic distance and host species of origin. Importantly, however, the congruity between the two techniques was strongest only when the sequence-level data were filtered to exclude changes which were selectively non-neutral. This observation implies that the changes leading to different RFLP patterns in the current sample must have been predominantly neutral selectively. This result may not be generalizable to other RFLP analyses, however, since the precise locations of the genetic changes underlying different RFLP patterns is not generally known. The proportion of neutral to non-neutral variation detected in a particular RFLP analysis will therefore vary depending on the number and type of restriction endonucleases used.

The observed difference between neutral and non-neutral sequence-level changes also implies that, to be most useful for epidemiological analysis, sequence data should be selectively filtered. The exclusion of structural variation appears to strengthen the ability of sequence data to reconstruct historical patterns of relatedness. Similarly, the exclusion of neutral variation appears to enhance the ability of sequence-level data to discern epidemiologic patterns that are closely linked to the structural properties of the virus.

Unfortunately, the actual history of farm-to-farm transmission of PrV in the 1989 Illinois outbreak remains unknown. It was therefore impossible to test the utility of RFLP and sequencing in this case against a known outbreak history. There was, in other words, no epidemiological 'gold standard' available for measuring the accuracy of our genetic analyses. As a result, the accuracy of the geographic and temporal data with respect to patterns of transmission must be assumed.

Were no statistically significant trends to have emerged from this investigation, the utility of either genetic technique for epidemiologic investigation would have been left open. However, several significant associations were documented, and all of these were in the predicted direction (even despite a limited sample size of 16 isolates). This observation provides support for the assumption that spatio-temporal associations are useful epidemiological indicators of the progression of this outbreak across time and space.

Comparing RFLP and sequencing directly indicates that both presented unique advantages for reconstructing the course of this viral disease outbreak. RFLP genetic distance correlated most strongly with temporal distance in this dataset. To the extent that temporal distance between isolates is an accurate indicator of the progression of PrV during this outbreak, RFLP analysis appeared to yield the most accurate epidemiologic information. While cost and ease of execution would also currently favour RFLP, progress in the technology of nucleic acid sequencing is likely soon to negate this advantage.

Sequence distance, filtered in the light of ancillary molecular biological data, yielded a variety of metrics of distance, each relevant to a different epidemiologic question. RFLP, by contrast, yielded only a single metric, which could not be subdivided into biologically-relevant components. Furthermore, the general ability of RFLP analysis to reconstruct outbreaks will likely vary depending on the population sampled and the restriction endonucleases used, and will not be predictable *a priori*. Given these differences, partial-genome sequencing should be considered at least as informative as RFLP analysis for reconstructing historical relationships between viral isolates, and also the more generally adaptable of the two techniques to specific epidemiological questions of biological relevance.

ACKNOWLEDGEMENTS

We wish to acknowledge L. Jin, T. Croix and B. Paszkiet for assistance with laboratory analyses.

This study was supported by the University of Illinois Research Board, and by Animal Health and Disease project ILLU-70-0916 through the Illinois Agriculture Experiment Station.

REFERENCES

1. Dowling TE, Moritz C, Palmer JD, Rieseberg LH. Nucleic acids III: Analysis of fragments and restriction sites. In: Hillis DM, Moritz C, Mable BK, eds. Molecular systematics 2nd edn. Sunderland: Sinauer Associates, 1996: 249–320.
2. Arens M. Methods for subtyping and molecular comparison of human viral genomes. Clin Microbiol Rev 1999; **12**: 612–26.
3. Nei M, Li W-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. PNAS 1979; **75**: 3359–62.
4. Nei M. Molecular evolutionary genetics. New York: Columbia University Press, 1987.
5. Hillis DM, Mable BK, Larson A, Davis SK, Zimmer EA. Nucleic acids IV: Sequencing and cloning. In: Hillis DM, Moritz C, Mable BK, eds. Molecular systematics, 2nd edn. Sunderland: Sinauer Associates, 1996: 321–81.

6. Christensen LS. The population biology of suid herpesvirus 1. APMIS 1995; **103**: 5–48.
7. Siegel AM. The Illinois pseudorabies eradication program and factors associated with selection of intervention strategies and success in eradication [Ph.D.]. University of Illinois, 1999.
8. Scherba G, Wiemers J, Siegel AM, et al. Application of a quantitative algorithm to restriction endonuclease analysis of Aujeszky's disease (pseudorabies) virus from a geographically localized outbreak. J Vet Diag Invest 1999; **11**: 423–31.
9. Weigel RM, Scherba G. Quantitative assessment of genomic similarity from restriction fragment patterns. Prev Vet Med 1997; **32**: 95–110.
10. Hampl H, Ben-Porat T, Ehrlicher L, Habermehl K-O, Kaplan AS. Characterization of the envelope proteins of pseudorabies virus. J Virol 1984; **52**: 583–90.
11. Wathen MW, Wathen LMK. Characterization and mapping of a nonessential pseudorabies virus glycoprotein. J Virol 1986; **58**: 173–8.
12. Mettenleiter TC, Zsak L, Zuckermann F, Sugg N, Kern H, Ben-Porat T. Interaction of glycoprotein gIII with a cellular heparinlike substance mediates adsorption of pseudorabies virus. J Virol 1990; **64**: 278–86.
13. Karger A, Mettenlieter TC. Identification of cell surface molecules that interact with pseudorabies virus. J Virol 1996; **70**: 2138–45.
14. Karger A, Saalmüller A, Tufaro F, Banfield BW, Mettenleiter TC. Cell surface proteoglycans are not essential for infection by pseudorabies virus. J Virol 1995; **69**: 3482–9.
15. Karger A, Schmidt J, Mettenlieter TC. Infectivity of a pseudorabies virus mutant lacking attachment glycoproteins C and D. J Virol 1998; **72**: 7341–8.
16. Zuckermann F, Zsak L, Mettenleiter TC, Ben-Porat T. Pseudorabies virus glycoprotein gIII is a major target antigen for murine and swine virus-specific cytotoxic T lymphocytes. J Virol 1990; **64**: 802–12.
17. Kimman TG, De Bruin TGM, Voermans JJM, Bianchi ATJ. Cell-mediated immunity to pseudorabies virus: cytolytic effector cells with characteristics of lymphokine-activated killer cells lyse virus-infected and glycoprotein gB- and gC- transfected L14 cells. J Gen Virol 1996; **77**: 987–90.
18. Mettenleiter TC. Immunobiology of pseudorabies (Aujeszky's disease). Vet Immunol Immunopathol 1996; **54**: 221–9.
19. Katayama S, Okada N, Yoshiki K-i, Okabe T, Shimizu Y. Protective effect of glycoprotein gC-rich antigen pseudorabies virus. J Vet Med Sci 1997; **59**: 657–63.
20. Ishikawa K, Tsutsui M, Taguchi K, Saitoh A, Muramatsu M. Sequence variation of the gC gene among pseudorabies virus strains. Vet Microbiol 1996; **49**: 267–72.
21. Goldberg TL, Hahn EC, Weigel RM, Scherba G. Genetic, geographical and temporal variation of porcine reproductive and respiratory syndrome virus in Illinois. J Gen Virol 2000; **81**: 171–9.
22. Li W-H. Molecular evolution. Sunderland, Mass.: Sinauer Associates, 1997.
23. Felsenstein J. PHYLIP: Phylogenetic Inference Package. 3.57c ed. Seattle: Department of Genetics, University of Washington, 1995.
24. Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Res 1967; **27**: 209–20.
25. Legendre P, Vaudour A. The R Package: Multidimensional analysis, spatial analysis. Université de Montréal: Département de sciences biologiques, 1991.
26. Smouse PE, Long JC, Sokal RR. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 1986; **35**: 627–32.
27. Hope ACA. A simplified Monte Carlo significance test procedure. J Roy Stat Soc Ser B 1968; **30**: 582–98.
28. Xiao S, Chen H, Hong W, Fang L. Cloning and sequence analysis of the gC gene of pseudorabies virus strain Ea. Wuhan, Hubei, China: Laboratory of Animal Virology, 1999: GenBank Direct Submission AF158090.
29. Swofford DL. PAUP: Phylogenetic analysis using parsimony. v. 3.1.1. Illinois Natural History Survey, Champaign, 1993.
30. Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 1958; **28**: 1409–38.
31. Domingo E, Escarmís C, Sevilla N, et al. Basic concepts in RNA virus evolution. FASEB J 1996; **10**: 859–64.
32. Scheidt AB, Rueff LR, Grant RH, et al. Epizootic of pseudorabies among ten swine herds. JAVMA 1991; **199**: 725–30.
33. Banks M. DNA restriction fragment length polymorphism among British isolates of Aujeszky's disease virus: use of the polymerase chain reaction to discriminate among strains. Brit Vet J 1993; **149**: 155–63.
34. Christensen LS, Mortensen S, Bøtner A, et al. Further evidence of long-distance airborne transmission of Aujeszky's disease (pseudorabies) virus. Vet Record 1993; **132**: 317–21.
35. Eigen M, Schuster P. The hypercycle. A principle of natural self-organization. Berlin: Springer, 1979.