ELSEVIER

Preventive Veterinary Medicine 61 (2003) 59–70

PREVENTIVE VETERINARY MEDICINE

www.elsevier.com/locate/prevetmed

Application of phylogeny reconstruction and character-evolution analysis to inferring patterns of directional microbial transmission

Tony L. Goldberg*

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

Received 24 May 2002; accepted 12 June 2003

Abstract

I used phylogenetic analyses to reconstruct patterns of directional interspecific transmission during a pseudorabies virus outbreak in Illinois, USA, in 1989. Isolates were recovered from five species: cattle, sheep, goats, pigs, and raccoons (*Procyon lotor*). I generated DNA sequences for 16 isolates of pseudorabies virus at the glycoprotein C gene, from which I constructed phylogenetic trees. I then used these trees, in combination with parsimony-based analyses of character evolution, to infer the frequency and direction of interspecific transmission events. Comparing inferred frequencies and directions of transmission to null expectations based on 10,000 random trees indicated a significant excess of transmission events from pigs to pigs and a corresponding lack of transmission events from non-porcine species. These results are concordant with the know biology and natural history of pseudorabies virus, and they demonstrate that retrospective phylogeny reconstruction and analyses of character evolution can be used to investigate the transmission ecology of pathogens.

Keywords: Phylogenetics; Character evolution; Transmission ecology; Pseudorabies virus

1. Introduction

Retrospective studies of microbial transmission are notoriously difficult to conduct. When transmission events cannot be observed directly, inferences about the direction and pattern of transmission must be made indirectly.

* Tel.: +1-217-265-0297; fax: +1-217-244-7421.
E-mail address: tlgoldbe@uiuc.edu (T.L. Goldberg).

0167-5877/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0167-5877(03)00161-2

Traditionally, transmission has been inferred on the basis of certain implicit "proximity criteria". When cases occur close together in time and in space, transmission is suspected (Ward and Carpenter, 2000; Carpenter, 2001). Proximity of the associated microbes in other, biological dimensions (e.g. genetic, serological, antimicrobial drug resistance patterns) provides further support. Circumstantial evidence (historical records or plausible mechanisms of transmission) can strengthen the hypothesis that transmission has occurred. These lines of evidence together often are used to infer patterns of transmission from the spatio-temporal clustering of cases (e.g. Austin and Weigel, 1992; Paré et al., 1996; Doherr et al., 1999).

The proximity-criterion approach has proven useful and informative, but nevertheless has some inherent biases (for example, see Glynn et al., 1999; Murray, 2002; Murray and Alland, 2002). Inferring transmission only when cases are temporally and/or spatially associated reduces the probability of detecting transmission over long distances or protracted time periods. Inferring transmission only when the associated microbes are biologically similar decreases the probability of detecting transmission for microbes that evolve rapidly. If external, circumstantial evidence is considered prerequisite to inferring transmission, instances in which such evidence is lacking will be discounted. Finally, traditional methods generally do not allow the direction of transmission to be inferred.

Complementary analytical techniques are needed that can allow us to make inferences about transmission that are independent of proximity. Phylogenetic analyses and analyses of character evolution can easily be adapted to the study of microbial transmission (Holmes et al., 1995), and address many of the shortcomings of other approaches in this regard. Although mainstream in systematics and evolutionary biology, these techniques have been slow to gain popularity in epidemiology (Holmes, 1998; Page and Holmes, 1998). The techniques are varied, but all make use of the information contained in phylogenetic trees.

Phylogenetic trees are graphical representations of the history of the relatedness of organisms (Page and Holmes, 1998; Maddison and Maddison, 2000). Once a phylogenetic tree has been constructed, any characteristic of the related organisms can be analyzed in the evolutionary framework specified by the tree. Changes in such characteristics ("characters") can be inferred between descendent organisms and their ancestors—even if those ancestors were never observed directly (Maddison and Maddison, 2000). A list of selected phylogenetic terms used in this paper is given in Table 1 (for further definitions and expanded explanations, see Page and Holmes, 1998; Maddison and Maddison, 2000).

In the case of microbes, characters can be selected for analysis that are pertinent to the biology or ecology of transmission. Such characters might include the type of host from which the microbe was recovered, or the geographic location where the microbe was found. In such cases, inferred changes in the states of characters would represent transmission events from one host type to another, or from one geographic location to another. The inherent directionality of rooted phylogenetic trees (Maddison and Maddison, 2000) makes it possible to infer the direction of transmission.

Fig. 1 outlines the analytical methodology used in this paper. The approach is similar to that described by Slatkin and Maddison (1989) for inferring patterns of gene flow from allele phylogenies. The hypothetical phylogenetic tree depicted would be constructed from data independent of microbial ecology, such as molecular genetic or serological data. However such a tree is obtained, character states representing parameters pertinent to transmission

Table 1
Definitions of selected phylogenetic terms used in the text

Term	Definition
Branch	A line connecting two nodes on a phylogenetic tree. The length of a branch may or may not be drawn in proportion to measures such as genetic distance, evolutionary change, or time.
Character	Any trait of an organism or set of organisms that can change over evolutionary time. Examples of characters related to microbes might be "virulence", "location of origin", or "number of pili".
Character state	Any of the alternative conditions that a character can take. Examples of character states corresponding to the characters listed above might be "benign", "Europe" or "24".
Cladogram	A phylogenetic tree drawn such that the only information contained is the order of branching. Branch lengths are not proportional to genetic distance in a cladogram.
Node	The point at which a lineage on a phylogenetic tree splits or ends. Nodes that split are "internal" nodes, whereas nodes that end (usually representing extant taxa) are "terminal" nodes.
Outgroup	A taxon assumed to be more distantly related to the taxa in a group than any are to each other.
Parsimony	The principle by which the most-accurate distribution of character states on a phylogenetic tree is assumed to be that which minimizes the total number of character-state changes.
Phylogenetic tree	A graphical representation of the history of relatedness of a group of organisms. Phylogenetic trees typically contain information about the order of evolutionary divergences and relative divergence times.
Resolution	The degree to which the order of branching in a phylogenetic tree is known. Trees containing only bifurcating internal nodes are fully resolved, whereas trees containing "polyfurcating" internal nodes are less resolved.
Root	The most ancestral node in a phylogenetic tree. Directional evolutionary change occurs from the root towards the tips of a rooted phylogenetic tree.
Taxon	Any named organism or group of organisms about which phylogenetic inferences can be made. Microbial taxa might be, for example, "*Salmonella*", "isolate #2377" or "*Enterobacteriaceae*".

can be assigned to the microbes represented by the terminal nodes of the tree. The principle of parsimony then can be applied to reconstruct the character states of ancestral microbes (represented by the internal nodes of the tree). Directional character-state changes can be inferred in a straightforward manner, and the relative frequencies of such events can be calculated. These frequencies can be compared statistically to null expectations to detect transmission biases.

My goal was to demonstrate the use of phylogenetic analyses and analyses of character evolution for inferring the directional transmission of microbes. I analyzed an outbreak of pseudorabies virus (*Alphaherpesvirinae*) in which isolates were recovered from multiple species. The specific character that I analyzed was the species from which the isolates were recovered. I reconstructed the frequency and direction of inferred transmission events within and between species during the outbreak and tested the significance of the patterns observed statistically.

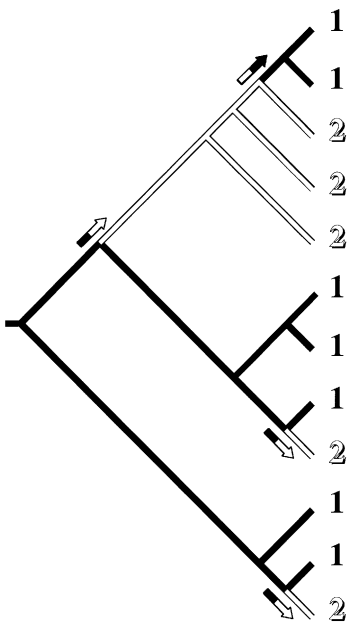

Fig. 1. Analysis of character evolution. The evolution of a single, reversible character with two states (1 and 2) is shown on a hypothetical phylogenetic tree. The tree itself would be built from independent data, such as genetic data. Character states along branches (1: solid branches, 2: open branches) were reconstructed from the states of the terminal taxa (1 or 2) using the principle of parsimony. Directional character-state changes (represented by arrows) then were inferred by tracing the evolution of the character from the root of the tree (left) to the tip of each branch (right). Four character-state changes were thereby inferred (three from 1 to 2, and one from 2 to 1). The analysis was performed using the computer program MacClade, Version 4 (Maddison and Maddison, 2000).

2. Materials and methods

Sixteen isolates of pseudorabies virus were obtained during an outbreak in western Illinois, USA, over a 10-month period in 1989. One isolate was recovered from a cow, three from dogs, three from sheep, eight from pigs and one from a raccoon (*Procyon lotor*). The isolates were obtained from all 14 positive farms identified during the outbreak. Positive farms were geographically clustered, and were all within approximately 50 km of each other. (Details of the outbreak and of the virological methods used are given in Scherba et al., 1999.)

A 798 bp segment of the viral glycoprotein C (gC) gene was PCR-amplified and sequenced in all isolates. The pseudorabies gC gene encodes an immunologically important glycoprotein and contains within it some of the most genetically variable regions of the pseudorabies virus genome (Ishikawa et al., 1996). (Details of the molecular methods used are given in Goldberg et al., 2001.)

Phylogenetic trees of viral isolates were constructed from the DNA sequences. Phylogenetic trees based on nucleotide sequence data are ideally suited to the study of microorganisms such as viruses (Haas, 1997; Hungnes et al., 2000). Because different methods of

phylogeny reconstruction can yield different phylogenetic trees, three common methods were used: the neighbor-joining (NJ) method of Saitou and Nei (1987), the maximum parsimony (MP) method (see Sober, 1989; Swofford et al., 1996 for reviews) and the maximum-likelihood method (Felsenstein, 1981).

The NJ tree was constructed using the computer program PAUP*, Version 4.0b10 (Swofford, 2000). Various models of nucleotide change available in PAUP* were used to correct pairwise nucleotide distances in the NJ analysis. The degree of genetic differentiation among the 16 isolates was small, however, and the choice of model did not affect the topology of the final tree. The MP tree was found using the Branch and Bound algorithm of PAUP*, with transitions (nucleotide changes from purine to purine or from pyrimidine to pyrimidine) weighted twice as heavily as transversions (changes from purine to pyrimidine or vice versa) and single-codon insertion/deletions weighted twice as heavily as point substitutions. These weighting schemes reflected the relative frequencies of each event, inferred by comparing the most closely related pairs of sequences in the data set. The ML tree was found using the heuristic search algorithm of PAUP*, with 100 random-addition-sequence replicates (a standard number). Parameter values for the model of molecular evolution used in the ML search were assigned manually, and were estimated from the data through an iterative likelihood-ratio approach implemented with the aid of the computer program Modeltest, Version 3.06 (Posada and Crandall, 1998). Bootstrapping (Felsenstein, 1985) was used to assign estimates of statistical confidence to all clades in all trees. One thousand bootstrap replicates of the data (a standard number) were run for each analysis (NJ, MP, ML), with search parameters identical to those used in the original search. Trees were rooted by selecting as an outgroup that isolate linked to the others by the longest branch.

Analyses of character evolution were performed using the computer program MacClade, Version 4 (Maddison and Maddison, 2000). MacClade uses the principle of parsimony to reconstruct ancestral states in phylogenetic trees, and to analyze the evolution of characters. Parsimony is a generalizable and widely accepted method for reconstructing the history of character-state change (Sober, 1989; Maddison and Maddison, 2000). Although other methods (most notably, likelihood methods) might be preferable in some situations, they are only suitable for system with well-understood mathematical properties (Swofford et al., 1996; Maddison and Maddison, 2000). Because likelihood models for describing the interspecific transmission of pathogens do not exist, parsimony is currently the best method available.

A single, unordered, discreet-valued character with five states was coded to reflect the host species from which an isolate was recovered. The evolution of the host species of origin character was traced along the reconstructed phylogenetic trees using the “Chart State Changes and Stasis” option of MacClade. MacClade also was used to tabulate the number of inferred directional state changes in the host species character from the root to the tips of each tree (see Fig. 1 for an example of this technique).

Both character-state change and character-state stasis were quantified. Character-state changes were interpreted as interspecific transmission events, and character-state stases (when no character-state change was inferred between ancestor and descendent virus) were interpreted as intraspecific transmission events. To avoid interpretive difficulties, only unambiguous character-state changes and stases were counted. Transmission events were estimated for all 25 possible classes of transmission (inter- and intraspecific, i.e. between and within each of the five host species of origin, in both directions) for each reconstructed tree.

The sampling of isolates during the pseudorabies virus outbreak was neither random nor systematic with respect to host species of origin. Therefore, different classes of transmission could be over- or underrepresented simply due to uneven sampling. To estimate the expected frequency of different classes of transmission given the empirical pattern of host species sampled, 10,000 random, equiprobable phylogenetic trees were generated and analyzed using MacClade. These trees were of the same size (16 taxa) as the NJ, MP and ML trees. A host species of origin character with five states was coded, and taxa were assigned randomly by computer to one of the five character states (proportionally to the numbers in the original data set). Changes and stases in the host species of origin character were traced on all 10,000 trees. Expected (null) frequencies of transmission for each class of transmission (inter- and intraspecific) were quantified as the maximum number of unambiguous character-state changes (or stases) observed for that class in any tree of the 10,000 random equiprobable trees.

Observed frequencies of transmission were compared to null expectations for each class of transmission using binomial-distribution probability tests implemented with the computer program Epi Info, Version 6.04b (Centers for Disease Control and Prevention, 1997). Probabilities were two-tailed, $\alpha = 5\%$.

3. Results

Within the 16 pseudorabies viral gC sequences, 30/798 (3.8%) nucleotide positions were variable. Two single-codon insertion/deletions also were observed. Nucleotide-level genetic differences between isolates varied from 0 to 1.9%, with a within-population mean pairwise difference of 0.8%. By comparison, the 16 Illinois sequences differed from epidemiologically unrelated published PrV gC sequences (GenBank accession numbers: AF403051, AF158090, SH1GGIII1, SH1GGIII2 and SH1GGIII3) by between 1.1 and 6.2%, with a between-population mean pairwise difference of 3.6%. The 16 Illinois DNA sequences are available through GenBank (accession numbers AF176479–AF176495).

The phylogenetic trees reconstructed are depicted in Fig. 2 as rooted cladograms. Branches in such cladograms are not scaled in proportion to the actual amount of genetic change inferred along them. Trees were drawn this way to facilitate the visualization of isolate clusters. Actual branch-length values are given numerically.

For the NJ and ML analyses, a single tree was recovered (as these methods necessitate; Fig. 2). For the MP analysis, four equally parsimonious trees (each of length 27) were found. The MP tree in Fig. 2 represents a 50% majority-rule consensus tree computed from the four equally parsimonious trees found during the MP search. Bootstrap values for individual clades ranged from 57 to 100%. The trees reconstructed in all analyses had very similar topologies, and differed mainly in their degree of phylogenetic resolution. The MP and ML trees had low resolution (many polyfurcating internal nodes) whereas the NJ tree was fully resolved (i.e. all internal nodes were strictly bifurcating).

Twenty-three unambiguous transmission events were inferred for the NJ tree, six for the MP tree, and four for the ML tree. The low number of inferred transmission events for the MP and ML trees resulted from their low phylogenetic resolution, in combination with the fact that only unambiguous events were counted.

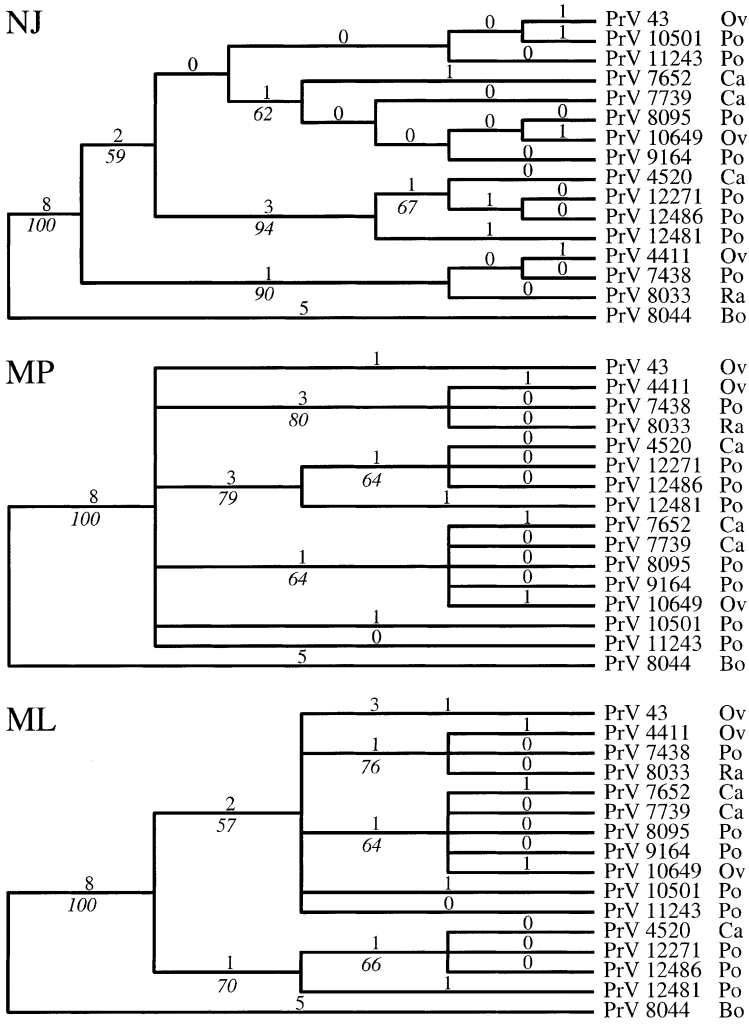

Fig. 2. Phylogenetic trees of 16 Illinois, USA, pseudorabies virus isolates recovered during an outbreak in 1989, based on 798 base pairs of the gC gene. NJ: neighbor-joining; MP: maximum parsimony; ML: maximum likelihood. Trees are drawn as cladograms, such that horizontal distances are not proportional to genetic distance. Taxon names correspond to Goldberg et al. (2001), and to GenBank entries AF176479–AF176495. Species of origin—Bo: bovine; Ca: canine; Ov: ovine; Po: porcine; Ra: raccoon. Numbers above branches are inferred numbers of nucleotide changes (branch lengths). Those below branches (in italics) are bootstrap values, indicating the percentage of trees (out of 1000 replicates) in which the grouping to the right of that branch was observed. The log-likelihood of the ML tree is −1182.9.

Inferred directional inter- and intraspecific transmission events for each of the 25 possible classes of transmission are shown in Fig. 3 for the NJ, MP and ML trees, as well as for the 10,000 random equiprobable trees. In all three analyses, the number of inferred transmission events from pig-to-pig was highest. In all three analyses, the observed proportion of

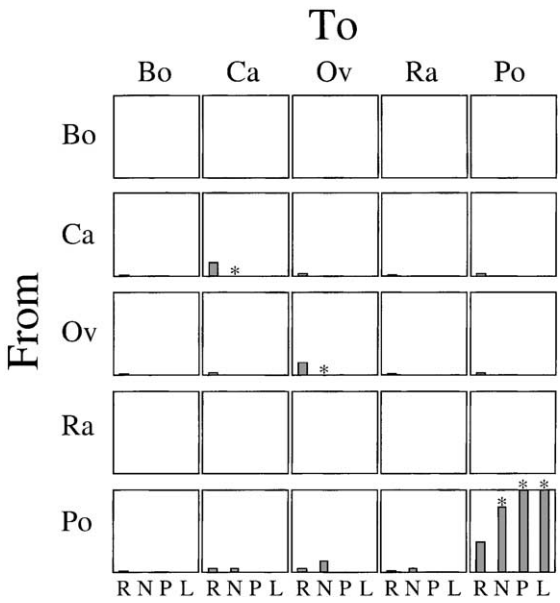

Fig. 3. Inferred directional interspecific transmission events for all 25 possible classes of transmission. Bars on unit squares represent proportions of inferred unambiguous transmission events for 10,000 random, equiprobable trees (R), and for the neighbor-joining (N), maximum parsimony (P), and maximum-likelihood (L) analyses. Asterisks indicate significant differences ($P \leq 0.05$) in the proportion of transmission events from the null expected proportion generated from the random-tree analysis (R).

pig-to-pig transmission events was significantly higher than expected by chance alone (all $P < 0.04$).

The random-tree analysis also predicted that some transmission events should have been observed from dogs to other species and from sheep to other species. No such transmission events were observed in any of the analyses. In the NJ analysis, the lack of inferred transmission events between dogs and between sheep was significant (both $P < 0.02$).

In all analyses, the proportion of inferred transmission events from non-porcine species (0% in each case) was lower than the expected proportion of such transmission events (50.8%, based on the 10,000 randomly generated trees). This trend was statistically significant in the NJ and MP analyses ($P < 0.0001$ for the NJ tree, $P = 0.014$ for the MP tree). In the ML analysis, this trend was present but non-significant ($P = 0.06$), probably because only four transmission events were inferred. Transmission in this outbreak therefore was biased towards transmission from pigs and away from transmission from other species.

The number and direction of inferred transmission events can, however, be affected by how a tree is rooted. To assess the effect of rooting, all possible rootings were examined separately for the three phylogenetic trees presented in Fig. 2. This is equivalent to examining the unrooted tree corresponding to each case (Maddison, 1990; Maddison and Maddison, 2000). Results were identical to those presented above in 30/30 alternative rootings of the

NJ tree, 19/21 alternative rootings of the MP tree, and 20/22 alternative rootings of the ML tree. The MP and ML rootings that did not agree with the consensus results inferred between one and two canine-to-canine transmission events. In no case, however, did the proportion of such events exceed or fall short of the expected null proportion.

4. Discussion

The analyses described above show that the frequencies of certain classes of transmission during the 1989 pseudorabies virus outbreak deviated significantly from null expectation. Specifically, there was a statistically significant excess of inferred transmission events from pigs to other pigs and a correspondingly significant lack of inferred transmission events from non-porcine species to other non-porcine species. These results were robust to variations in the method of phylogeny reconstruction, and to alternative rootings of the phylogenetic trees.

Pseudorabies virus is the causative agent of Aujeszky's disease, a neurologic disorder of swine that first was described over a century ago (Aujeszky, 1902). In young pigs, Aujeszky's disease is marked by incoordination, opisthotonus and convulsions, as well as non-neurological signs such as pyrexia, anorexia, and respiratory difficulties (Baskerville et al., 1973). Adult pigs tend to suffer respiratory difficulties and infertility. Pigs can be convalescent and chronic carriers of PrV. In contrast, Aujeszky's disease is rapidly and almost-uniformly fatal in non-porcine species. Pseudorabies virus can infect some wildlife species, but transmission from these species has not been documented (Thawley and Wright, 1982; Glass et al., 1994; Zanin et al., 1997). By far, the most common class of pseudorabies virus transmission in nature is from pig-to-pig (Christensen, 1995).

The present analyses were able to reconstruct this overall pattern of pig-to-pig-biased viral transmission even in the absence of external data. Although no information about the biology of the pathogen was used other than the DNA sequences of the isolates in question and the species from which these isolates were recovered, the propensity of pseudorabies virus to be transmitted almost exclusively from pigs was inferred accurately. This observation suggests that the phylogenetic approach used here is both powerful and accurate for inferring patterns of directional pathogen transmission.

The techniques described have several interpretive limitations, however. First, inferred transmission events are not necessarily direct. The parsimony criterion used to infer character-state changes reconstructs the *minimum* number of changes necessary to explain the observed pattern of character states. The actual number of changes could be higher. Inferred transmission events could represent direct transmissions from one animal to another, or they could represent a series of transmissions through intermediate hosts. A transmission event can be considered direct only when substantial external evidence supports such a conclusion. The use of phylogenetic analyses to infer direct transmissions of HIV (human immunodeficiency virus) in legal cases is an example (Albert et al., 1994; Birch et al., 2000). In these cases, corroborating testimony was necessary to establish the direct nature of transmission. Without such evidence, the techniques described here are best suited to reconstructing general patterns of transmission that can be interpreted in the light of microbial ecology.

The power and accuracy of these techniques also depend highly on the quality of the phylogenetic trees. Inaccurate trees can lead to false inferences. Trees with low resolution can yield relatively few unambiguous inferred transmission events, reducing statistical power (as occurred in the MP and ML analyses). Techniques that yield trees with greater resolution are not necessarily preferable, however—especially if phylogenetic resolution is gained at the expense of accuracy. This is illustrated by the NJ tree in Fig. 1; the neighbor-joining algorithm resolved all clades fully, but did so even when clades consisted of isolates with identical DNA sequences (e.g. in the case of isolates PrV 4520, 12271 and 12486). Ideally, trees generated using several complementary methods of phylogeny reconstruction should be analyzed before general conclusions are drawn. Where possible, likelihood-based methods of phylogeny reconstruction should be included.

Because molecular phylogenies are also sensitive to the specific gene regions chosen for study, ideally several independent genetic loci should be analyzed. If this is impossible or impractical, any single gene region chosen for study should be evaluated in terms of its utility for reconstructing phylogeny. In the case of the present analysis, for example, it was demonstrated previously that phylogenies based on pseudorabies gC DNA sequences are highly correlated with phylogenies based on whole-genome restriction fragment-length patterns (Goldberg et al., 2001). This is encouraging—especially because all the PrV gC sequences used in this study were very similar genetically to each other (and differed by only 1–6% from epidemiologically unrelated gC sequences). Despite this low degree of genetic differentiation within and between outbreaks, and despite the low sample size of isolates, the phylogenetic trees generated were informative.

One noteworthy component of the present analysis to me was that, in the NJ tree, the number of inferred transmission events (23) was actually higher than the number of isolates included in the study (16). Parsimony analyses of character evolution make use of all the topological information contained in a phylogenetic tree, including information about the character states of reconstructed putative ancestral organisms (represented by the internal nodes of the tree)—even though these ancestors were never observed. The use of phylogenetic information therefore also has the advantage that it enhances statistical power.

The techniques described in this paper cannot eliminate problems associated with sampling bias. Although the random-tree analysis attempted to correct for uneven sampling across host species, this correction was not perfect. Phylogenetic analyses only can infer transmission from or to host species that were sampled. Other species could play important but undetected roles.

Despite these shortcomings, phylogenetic analyses of character evolution appear to be powerful and accurate tools for inferring directional pathogen transmission. They avoid many of the biases inherent in traditional approaches that rely exclusively on criteria of spatial, temporal and biological proximity, and on the availability of circumstantial evidence. It would be particularly interesting to apply the techniques described in this paper to investigating the movement of zoonotic microbes within and between human and non-human hosts, to studying the transmission ecology of agents of importance to food safety and public health (such as food-borne bacteria and transmissible antimicrobial resistance elements), and to reconstructing patterns of spread of emerging pathogens within and among host species and geographic areas.

Acknowledgements

Many thanks to E. Hahn, G. Scherba, R. Weigel, T. Croix and B. Paszkiet for assistance in the collection of samples and epidemiological information, and with laboratory analyses. R. Singer provided comments on the manuscript, and T. Disotell provided advice on phylogenetic terminology. This project was supported by the Illinois Department of Agriculture, the University of Illinois Research Board, and the Illinois Council on Food and Agricultural Research.

References

Albert, J., Wahlberg, J., Leitner, T., Escanilla, D., Uhlen, M., 1994. Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 pol and gag genes. J. Virol. 68, 5918–5924.

Aujeszky, A., 1902. Uber ein neue Infektionskrankheit bei Haustieren. Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. 1, Orig. 32, 353–357.

Austin, C.C., Weigel, R.M., 1992. Factors affecting the geographic distribution of pseudorabies (Aujeszky's disease) virus infection among swine herds in Illinois. Prev. Vet. Med. 13, 239–250.

Baskerville, A., McFerran, J.B., Dow, C., 1973. Aujeszky's disease in pigs. Vet. Bull. 43, 465–480.

Birch, C.J., McCaw, R.F., Bulach, D.M., Revill, P.A., Carter, J.T., Tomnay, J., Hatch, B., Middleton, T.V., Chibo, D., Catton, M.G., Pankhurst, J.L., Breschkin, A.M., Locarnini, S.A., Bowden, D.S., 2000. Molecular analysis of human immunodeficiency virus strains associated with a case of criminal transmission of the virus. J. Infect. Dis. 182, 941–944.

Carpenter, T.E., 2001. Methods to investigate spatial and temporal clustering in veterinary epidemiology. Prev. Vet. Med. 48, 303–320.

Centers for Disease Control and Prevention, 1997. Epi Info: A Word Processing, Database and Statistics Program for Public Health, Version 6.04b. CDC, Atlanta.

Christensen, L.S., 1995. The population biology of suid herpesvirus 1. APMIS 103, 5–48.

Doherr, M.G., Carpenter, T.E., Wilson, W.D., Gardner, I.A., 1999. Evaluation of temporal and spatial clustering of horses with Corynebacterium pseudotuberculosis infection. Am. J. Vet. Res. 60, 284–291.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Glass, C.M., McLean, R.G., Katz, J.B., Maehr, D.S., Cropp, C.B., Kirk, L.J., McKeirnan, A.J., Evermann, J.F., 1994. Isolation of pseudorabies (Aujeszky's disease) virus from a Florida panther. J. Wildl. Dis. 30, 180–184.

Glynn, J.R., Vynnycky, E., Fine, P.E., 1999. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. Am. J. Epidemiol. 149, 366–371.

Goldberg, T.L., Weigel, R.M., Hahn, E.C., Scherba, G., 2001. Comparative utility of restriction fragment length polymorphism analysis and gene sequencing to the molecular epidemiological investigation of a viral outbreak. Epidemiol. Infect. 126, 415–424.

Haas, L., 1997. Molecular epidemiology of animal virus diseases. Zentralbl. Vet. Reihe B. 44, 257–272.

Holmes, E.C., 1998. Molecular epidemiology and evolution of emerging infectious diseases. Br. Med. Bull. 54, 533–543.

Holmes, E.C., Nee, S., Rambaut, A., Garnett, G.P., Harvey, P.H., 1995. Revealing the history of infectious disease epidemics through phylogenetic trees. Philos. Trans. R. Soc. London B 349, 33–40.

Hungnes, O., Jonassen, T.O., Jonassen, C.M., Grinde, B., 2000. Molecular epidemiology of viral infections. How sequence information helps us understand the evolution and dissemination of viruses. APMIS 108, 81–97.

Ishikawa, K., Tsutsui, M., Taguchi, K., Saitoh, A., Muramatsu, M., 1996. Sequence variation of the gC gene among pseudorabies virus strains. Vet. Microb. 49, 267–272.

Maddison, D.R., 1990. Phylogenetic inference of historical pathways and models of evolutionary change. Ph.D. Thesis. Harvard University, Cambridge, MA.

Maddison, D.R., Maddison, W.P., 2000. MacClade 4: Analysis of Phylogeny and Character Evolution, Version 4.0. Sinauer Associates, Sunderland, MA.

Murray, M., 2002. Sampling bias in the molecular epidemiology of tuberculosis. Emerg. Infect. Dis. 8, 363–369.

Murray, M., Alland, D., 2002. Methodological problems in the molecular epidemiology of tuberculosis. Am. J. Epidemiol. 155, 565–571.

Page, R.D.M., Holmes, E.C., 1998. Molecular Evolution: A Phylogenetic Approach. Blackwell Scientific Publications, Oxford.

Paré, J., Carpenter, T.E., Thurmond, M.C., 1996. Analysis of spatial and temporal clustering of horses with *Salmonella krefeld* in an intensive care unit of a veterinary hospital. J. Am. Vet. Med. Assoc. 209, 626–628.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Scherba, G., Wiemers, J., Siegel, A.M., Jin, L., Austin, C.C., Bowman, L., Redeford, B., Johnston, N.A., Weigel, R.M., 1999. Application of a quantitative algorithm to restriction endonuclease analysis of Aujeszky's disease (pseudorabies) virus from a geographically localized outbreak. J. Vet. Diag. Invest. 11, 423–431.

Slatkin, M., Maddison, W.P., 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123, 603–613.

Sober, E., 1989. Reconstructing the Past: Parsimony, Evolution and Inference. MIT Press, Cambridge, MA.

Swofford, D.L., 2000. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MA, pp. 407–514.

Thawley, D.G., Wright, J.C., 1982. Pseudorabies virus infection in raccoons: a review. J. Wildl. Dis. 18, 113–116.

Ward, M.P., Carpenter, T.E., 2000. Techniques for analysis of disease clustering in space and time in veterinary epidemiology. Prev. Vet. Med. 45, 257–284.

Zanin, E., Capua, I., Casaccia, C., Zuin, A., Moresco, A., 1997. Isolation and characterization of Aujeszky's disease virus in captive brown bears from Italy. J. Wildl. Dis. 33, 632–634.