

Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus

Leslie A. Real^{†*}, J. Caroline Henderson[†], Roman Biek[†], Jennifer Snaman[†], Tracy Lambert Jack[§], James E. Childs[†], Eli Stahl^{†¶}, Lance Waller[¶], Rowland Tinline^{††}, and Susan Nadin-Davis^{**}

[†]Department of Biology and Center for Disease Ecology, 1510 Clifton Road, Emory University, Atlanta, GA 30322; Departments of [§]Epidemiology and [¶]Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA 30322; ^{††}Queen's Geographic Information Systems Lab, Queen's University, Kingston, ON, Canada K7L 3N6; and ^{**}Animal Disease Research Institute, 3851 Fallowfield Road, Ottawa, ON, Canada K2H 8P9

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved June 24, 2005 (received for review January 5, 2005)

Infectious disease emergence is under the simultaneous influence of both genetic and ecological factors. Yet, we lack a general framework for linking ecological dynamics of infectious disease with underlying molecular and evolutionary change. As a model, we illustrate the linkage between ecological and evolutionary dynamics in rabies virus during its epidemic expansion into eastern and southern Ontario. We characterized the phylogeographic relationships among 83 isolates of fox rabies virus variant using nucleotide sequences from the glycoprotein-encoding glycoprotein gene. The fox rabies virus variant descended as an irregular wave with two arms invading from northern Ontario into southern Ontario over the 1980s and 1990s. Correlations between genetic and geographic distance suggest an isolation by distance population structure for the virus. The divergence among viral lineages since the most recent common ancestor correlates with position along the advancing wave front with more divergent lineages near the origin of the epidemic. Based on divergence from the most recent common ancestor, the regional population can be partitioned into two subpopulations, each corresponding to an arm of the advancing wave. Subpopulation A (southern Ontario) showed reduced isolation by distance relative to subpopulation B (eastern Ontario). The temporal dynamics of subpopulation A suggests that the subregional viral population may have undergone several smaller waves that reduced isolation by distance. The use of integrated approaches, such as the geographical analysis of sequence variants, coupled with information on spatial dynamics will become indispensable aids in understanding patterns of disease emergence.

infectious disease | isolation by distance | landscape genetics

The past two decades have witnessed an expanding list of infectious diseases characterized as emerging threats to human and/or wildlife health, agricultural production, or public security; e.g., pathogens used for bioterrorism (1). Recent reviews (2, 3) of the causal mechanisms leading to disease emergence most often point to the simultaneous effects of ecological and evolutionary genetic changes in patterns of first appearance and spread. For example, the recent worldwide distribution of severe acute respiratory syndrome–coronavirus can be understood only as a combined ecological and evolutionary phenomenon, the appearance of a novel genetic type of coronavirus coupled to the movement and dispersal of infected individuals over transportation networks (4). Similar combined influences of genetics and ecology are associated with patterns of emergence in malaria, tuberculosis, Lyme disease, hantavirus, hemolytic uremic syndrome, and cholera; in fact, it is difficult to imagine any infectious disease that will not be under the simultaneous influence of both genetic and ecological factors (5–7).

Epidemic expansion of RNA viruses may prove especially amenable to exploring the linkage between evolutionary and ecological dynamics, because these processes often occur on the same temporal scale (8). Here, we use the spatial analysis of

rabies virus evolution over the course of epidemic expansion across eastern and southern Ontario, Canada, as a model system to illustrate the combined forces of ecological and evolutionary dynamics in shaping the spatial distribution of an important world-wide pathogen.

Rabies virus is a negative-sense single-stranded RNA virus belonging to the genus *Lyssavirus*, family Rhabdoviridae. Three major global epidemics of rabies in animals have occurred since World War II, one in Europe, one within the eastern United States, and one in Canada. The European rabies epidemic is largely associated with virus spread in the red fox (*Vulpes vulpes*) and seems to be largely restricted to this host. The U.S. rabies epidemic is associated with expansion of the raccoon rabies variant after a presumed translocation of rabid animals from Florida to the West Virginia/Virginia border (9). Rabies has been endemic in the southeastern U.S. among raccoons for decades (10). The Canadian epidemic is associated with the expansion of rabies in the arctic fox (*Alopex lagopus*) into the southern provinces of Canada, especially Ontario. The southern expansion of arctic fox rabies through Canada has shifted to the red fox as host after reaching Ontario. The Canadian epidemic meets the northern wave of raccoon-specific rabies expansion from the U.S. along the Canadian and U.S. border (11).

The compartmentalization of distinct variants of rabies virus to single species of mammalian hosts has been only recently appreciated and made possible through the use of monoclonal antibodies and genetic sequencing (12–15). Spatial data on rabies have been extensively collected and compiled at the Centers for Disease Control and Canadian laboratories for >30 years. Using these data, we are now in a unique position to compare the spatial molecular evolution of viral populations across regions of rapid epidemic expansion.

In a previously reported study, Nadin-Davis *et al.* (16) explored the relationship between rabies virus evolution and local environmental conditions in red fox virus variants collected from 83 spatially distinct geographic locations across southern Ontario. The exact spatial coordinates for all samples were known. They established the phylogenetic relationship among the samples based on 20 distinct genotypes determined by restriction site polymorphisms and sequencing of the glycoprotein (G) gene-coding region (details discussed below).

Phylogenetic reconstruction revealed considerable clustering of sequences in evolutionary time and geographic extent. Nadin-Davis *et al.* (16) suggested that the geographic clustering of sequence types is attributable to ecological structuring of the

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: G, glycoprotein; MRCA, most recent common ancestor; ML, maximum likelihood; SR, single rate; SRDT, SR dated tips.

*To whom correspondence should be addressed. E-mail: lreal@emory.edu.

¶Present address: Department of Biology, University of Massachusetts Dartmouth, North Dartmouth, MA 02747.

© 2005 by The National Academy of Sciences of the USA

Table 1. Test of the molecular clock and sampling time heterogeneity

Model	ln L	K_a	Δ	df	χ^2
DR	-2920.855	37	—	—	—
SR	-2935.420	19	14.57	18	$P = 0.047$
SRDT	-2934.946	20	14.09	17	$P = 0.043$

ln L, log likelihood; K_a , parameters; Δ , likelihood differences from different rates (DR).

locations (geographic distance) extracted from the digitized image. The digitized image was analyzed on a millimeter scale and calibrated against geographic landmarks. A single digitized map unit corresponds to 85.7 m. The 83 samples generate $83 \times 82/2 = 3,403$ possibly unique pairs of geographic distances, which were then binned into 45 distance classes of width 200 map units (17.15 km).

Correlations between pair-wise genetic and pair-wise geographic distances were performed by using standard regression techniques as is commonly performed in landscape genetic analysis (17, 24). Spatial associations between geographic location and genetic distance revealed through regression analysis were confirmed by using Mantel statistics. Following Epperson (17), Mantel tests were performed by permuting the genetic distance matrix while maintaining a constant geographic distance matrix and computing the correlation coefficient between matrices. The permutation was executed 5,000 times, and the distribution of correlation coefficients within each subpopulation was compared with the observed correlation coefficient.

Each sample location is associated with a particular genotype. These 20 previously identified sequence types (16) were treated as categorical variables distributed over the sample area. The spatial clustering and geographic boundaries around clusters of sequence types (Fig. 1C) were then determined by categorical wombling (25) by using BOUNDARYSEER (26).

Spline regression, which fits partitioned data by using best linear regressions through each subset joined at fixed knot(s) (27), was conducted in S-PLUS (MathSoft, Seattle) with a fixed knot at 15 map units.

Results

Genetic Distance vs. Geographic Distance. A plot of the average genetic distance within a binned distance class vs. the midgeographic distance for that binned distance class shows a very strong correlation (Fig. 2B, $R^2 = 0.90$, $df = 44$, $P < 0.001$). Such a correlation is a hallmark of isolation by distance and would not necessarily be expected under an ecological partitioning of viral variants according to habitat. Spatial association between genetic and geographic distance is confirmed in the Mantel statistical analysis. The observed Mantel statistic ($r = 0.394$, $P < 0.001$) is significantly outside the distribution of statistic values from the randomization test (Fig. 4).

Time Since Most Recent Common Ancestor (MRCA). RNA viruses often accumulate sequence variation in a clock-like fashion (28); therefore, sequence differences can be used to assess time to MRCA. We tested for a molecular clock (Table 1), which was only narrowly rejected for the SR ($P = 0.047$) as well as the SRDT model ($P = 0.043$). Failure to conform to a strict molecular clock, due to rate variation among sites or lineages, is common for RNA viruses (29). However, because it has been shown that even in the case of clock rejection, confidence limits may well include the true rate (29), and because rejection in our case was marginal, we felt that assuming clock-like behavior may still be a reasonable approximation of the molecular evolutionary process for Ontario fox rabies. Log likelihoods of the SR and

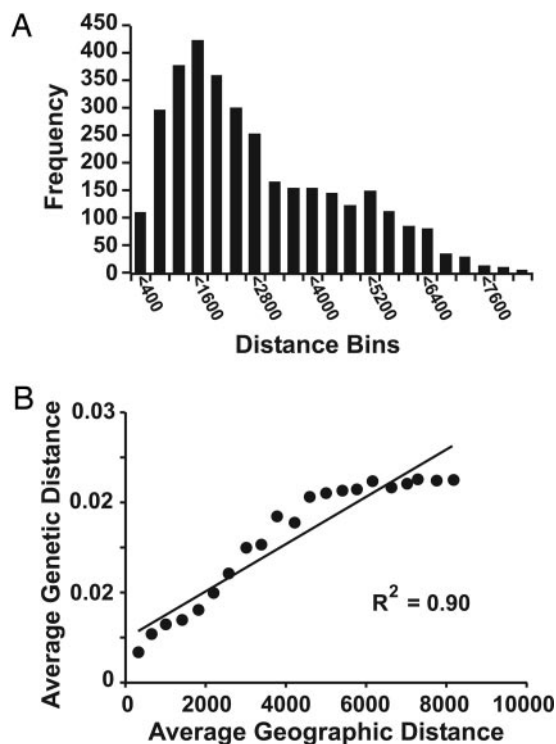


Fig. 2. Geographic distance versus genetic distance for fox rabies virus variants indicating isolation by distance. (A) The distribution of observations in each of the 45 different geographic distance bins is log-normally distributed. (B) Correlation between the average Euclidean distance of the geographic bin and the average genetic distance within each bin ($R^2 = 0.90$, $df = 44$, $P < 0.001$).

the SRDT model were virtually identical, indicating that differences in sampling dates had little effect on our data. Despite the short time span between samples, the SRDT model yielded an evolutionary rate estimate of $3.64 \pm 0.40 \times 10^{-4}$ per site per year. This rate conforms to the rate obtained by using a much larger data set ($3.46 \pm 0.12 \times 10^{-4}$ per site per year) encompassing G gene sequences from a variety of carnivore host species (R.B. and L.R., unpublished data). Based on our rate estimation from the Ontario fox data, the MRCA of fox rabies dates to 1960 (confidence interval, 1958–1963).

Geographic Gradients in Time Since MRCA. The Ontario fox rabies epidemic descended from northern Canada into the southern peninsula as an irregular wave, where one arm of the advancing wave moved through western Quebec, then crossed the Ottawa River and moved across eastern Ontario from the northeast. Another arm moved into the central portion of southern Ontario, then spread west and south through the remainder of the southern peninsula (30, 31).

The genetic structure of the propagating wave should reflect the combined ecological and evolutionary history of the viral population. If new viral variants emerge during wave advance, then the front of the wave should be characterized by the most recently evolved viral sequences. We measured the geographic distance from a fixed point at the southwestern extreme of the rabies distribution (point A, Fig. 1A) to each sample location. This point can be assumed to be the most recent one to have been reached by the epidemic wave because of being on the tip of a peninsula. A plot of distance from point A against the estimated number of substitutions since the MRCA for the viral sequence at the sample location shows a significant correlation (Fig. 3A, black line; $R^2 = 0.47$, $df = 82$, $P < 0.001$), indicating that the most

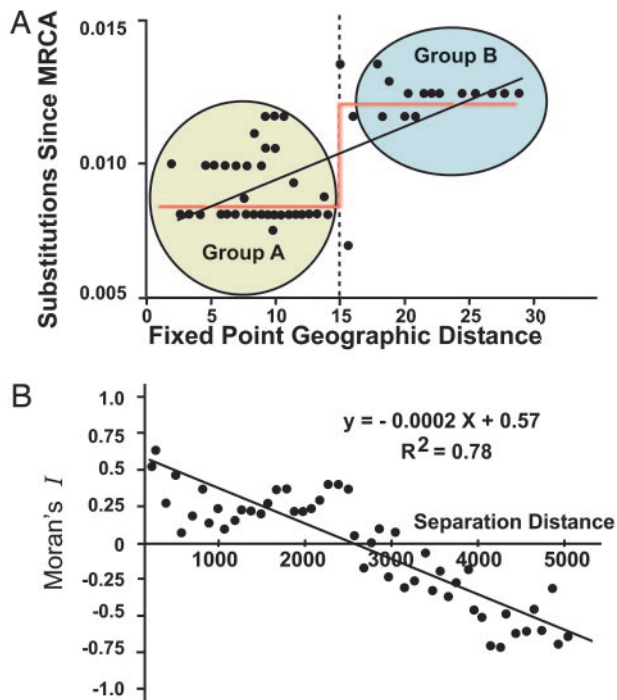


Fig. 3. Subpopulation structure and clinal variation in time since MRCA in fox rabies virus variants across southern Ontario. (A) Correlation between the estimated number of substitutions since MRCA and geographic distance from the sample to the “southern fixed point” (red arrow, Fig. 1A) for the 83 mapped Ontario red fox rabies virus samples ($R^2 = 0.47$, $df = 82$, $P < 0.001$). (B) Plot of Moran's I spatial autocorrelation coefficient among the substitutions since MRCA at different lagged distance classes over the southern Ontario peninsula.

divergent viral lineages relative to the MRCA are found near the geographic origin of epidemic expansion. Spatial autocorrelation (Moran's I) in divergence from the MRCA at different lagged geographic distances reveals a linear decline in spatial autocorrelation (Fig. 3B). Linear declines in spatial autocorrelation are often interpreted as indications of clinal variation (17). In our case, the apparent cline would be along the general direction of epidemic expansion.

However, the time since MRCA–distance relationship can be partitioned into two distinct halves (Group A and Group B, Fig. 3A, shaded colored ellipses). Group A corresponds to the western arm of the original invasion, and Group B corresponds to the eastern arm of the original invasion. Furthermore, the southern boundary of the Canadian Shield is a reasonable approximation of the boundary between the two groups, and the habitats and topography of these two regions are also quite distinct (Fig. 1A).

To explore this nonlinear relationship, we partitioned the data into samples less than or greater than 15 map units from point A (Fig. 3A) and constructed a spline regression in S-PLUS with a fixed knot at 15 map units. Spline regression fits the partitioned data by using the best linear regressions through each subset and joined at the fixed knot (27). The best-fit spline regression (Fig. 3A, red line; $R^2 = 0.61$) increases the R^2 by $\approx 14\%$, and the model essentially fits zero-slope lines at levels corresponding to the mean ages of viral sequences for each subset of the data. This pattern suggests there are two statistically distinguishable viral age groups: one older virus population circulating in the northeast and one younger population circulating in the southwest. This pattern is consistent with the pattern of initial invasion (30, 31).

How does isolation by distance affect the spatial pattern of

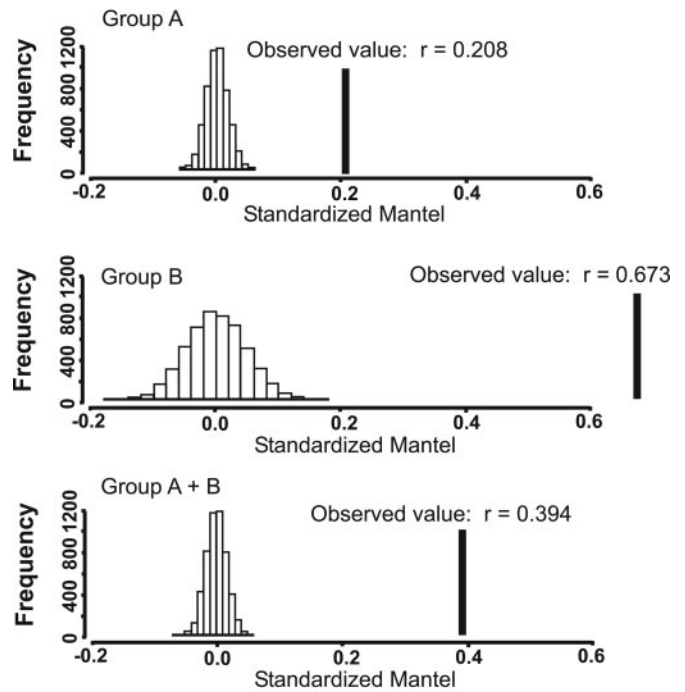


Fig. 4. Mantel statistic computed following Epperson (17) for the two viral groups (A and B) corresponding to the eastern and southern Ontario populations and combined groups (A + B). For each computation, we kept the geographic distance matrix constant and permuted the elements of the genetic distance matrix for each group. We randomly assigned genetic distances to geographic distances 5,000 times and created a distribution of values. For both Group A and Group B and combined groups (A + B), the observed statistic value is larger than the maximum value associated with permutations (P value < 0.001 for each analysis).

genetic variation within these subpopulations? The age–distance relationship was computed against a fixed geographic location to define subgroup distances. Within each subgroup, is there a genetic distance–geographic distance relationship? We performed Mantel tests for each subgroup based on the pair-wise genetic distance and geographic distance among all pairs of samples within a subgroup (Fig. 4). For both Group A (southern Ontario) and Group B (eastern Ontario), the observed statistic value is larger than the maximum value associated with the permutations (Group A, $r = 0.208$; Group B, $r = 0.673$; Groups A + B, $r = 0.394$; $P < 0.001$ for each test). Statistically significant Mantel statistics are indicative of isolation by distance within viral subpopulations. Moreover, the eastern subpopulation has a much greater degree of isolation by distance than does the more southern subpopulation A.

Discussion

The genetic differentiation of populations into distinct ecotypes as a consequence of natural selection is a common phenomenon in natural biological systems. We should expect, then, to see some evidence for local adaptation among pathogens associated with local host interactions or local environmental conditions. Indeed, there are excellent examples of local adaptation and differentiation within pathogen–host populations (32, 33).

However, the hypothesis of local adaptation generating spatial genetic organization of populations must be evaluated against an appropriate null model based on isolation by distance. In an earlier paper (16), we speculated that rabies virus genotypes in Ontario appear to be partitioned into distinct sequence variants, and these variants might be associated with specific habitats. Our current population genetic analysis of the Ontario epidemic

expansion suggests that the spatial organization of sequence variants is best explained largely as a consequence of neutral evolution coupled to limited local dispersal and mixing of viral variants. Although we did not specifically test for evidence (or absence) of adaptation here, comprehensive analyses of rabies G as well as the nucleoprotein and phosphoprotein gene sequences by us and others have provided no indication that positive selection plays an important role in the short-term evolution of these genes (34). Our current analysis suggests that $\approx 90\%$ of the geographic structure of the Ontario rabies virus population can be attributed to isolation by distance rather than to local adaptation.

Within this overall pattern, however, we have demonstrated that southern Ontario can be partitioned into an east and a southwest region linked to the initial pattern of invasion and the physical separation of the area by the Canadian shield. The overall geographic distribution of rabies virus genetic variants suggests the following scenario for the molecular evolution of the virus coupled to the spatial dynamics of the host. First, cross-species transmission of the arctic fox variant into red fox populations, which subsequently served as the principal reservoir host species, occurred before rabies decent into eastern and southern Ontario. Second, the virus population underwent a gradual spread over the eastern region where substantial restrictions to movement, imposed by limited local transmission by infected foxes, generated a pattern indicative of strong isolation by distance. This eastern Ontario population of viruses constitutes an early age group. Third, the eastern population of viruses invaded the southern section of Ontario in a second phase, giving rise to a later aged group of circulating viruses that shows reduced isolation by distance. This reduction can be largely ascribed to multiple waves of infection crossing the southern peninsula, but restricted transmission back into the eastern sections of Ontario. This would lead to increased mixing among

the southern isolates of the virus and the disruption of isolation by distance. The transgression of multiple waves of infection across the southern region (but absent in eastern Ontario) has been independently corroborated by temporal patterns of case reports of rabid animals (30, 31). In essence, then, two populations of virus of slightly different age exist on the Ontario Peninsula; each shows isolation by distance within their respective endemic region indicative of the historical dynamics of fox rabies spread during its decent into southern Ontario.

An understanding of the interaction between ecological dynamics, spatial spread, and evolutionary change in infectious diseases is of critical importance for understanding extant epidemiological patterns and prerequisite for constructing a predictive theory of disease emergence (8). The detection and analysis of molecular evolutionary change in RNA viral populations will be a critical tool in establishing these linkages, because evolutionary and ecological dynamics in these organisms occur on the same timescale. Statistical analysis of sequence evolution in space and time has been successfully used to detect wave-like patterns of spread in the Central African emergence of the Ebola Zaire virus variant (35) and the rapid spread of both European Bat Lyssaviruses 1 and 2 throughout Western Europe (E. Holmes, personal communication). We suspect that, as more gene sequence information becomes readily available, the use of integrated approaches such as the geographical analysis of sequence variants will become an indispensable aid in understanding patterns of disease emergence.

We thank Charles Rupprecht, Cathleen Hanlon, Andres Velasco-Villa, Richard Franka, Ann Schmitz, Serena Reeder, Tonya Mixson, Michelle Packard, and two anonymous reviewers for comments on the manuscript. This research was supported through National Institutes of Health Grant RO1-AI047498 and U.S. Department of Agriculture Grant 0371004129CA (to L.A.R.).

- Asokan, G., Vasanthan, A. & Prabhakaran, V. (2003) *Indian J. Anim. Sci.* **73**, 337–341.
- Daszak, P., Cunningham, A. A. & Hyatt, A. D. (2000) *Science* **287**, 443–449.
- Smolinski, M. S., Hamburg, A. A. & Lederberg, J. (2003) *Microbial Threats to Health: Emergence, Detection, and Response* (Natl. Acad. Press, Washington, DC).
- Stavrinos, J. & Guttman, D. S. (2004) *J. Virol.* **78**, 76–82.
- Domingo, E., Webster, R. & Holland, J. (1999) *Origin and Evolution of Viruses* (Academic, San Diego).
- Morse, S. (1993) *Emerging Viruses* (Oxford Univ. Press, Oxford, U.K.).
- Morse, S. (1994) *The Evolutionary Biology of Viruses* (Raven, New York).
- Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A. & Holmes, E. C. (2004) *Science* **303**, 327–332.
- Jenkins, S., Perry, B. & Winkler, W. (1998) *Rev. Infect. Dis.* **10**, Suppl., S620–S625.
- Prather, E. C., Bigler, W. J., Hoff, G. L. & Tomas, J. A. (1975) *Rabies in Florida* (Division of Health, Department of Health and Rehabilitative Services, State of Florida, Jacksonville, FL).
- Gordon, E. R., Curns, A. T., Krebs, J. W., Rupprecht, C. E., Real, L. A. & Childs, J. E. (2004) *Epidemiol. Infect.* **132**, 515–524.
- Rupprecht, C. E., Glickman, L. T., Spencer, P. A. & Wiktor, T. J. (1987) *Am. J. Epidemiol.* **126**, 298–309.
- Smith, J. S., Orciari, L. A., Yager, P. A., Seidel, H. D. & Warner, C. K. (1992) *J. Infect. Dis.* **166**, 296–307.
- Smith, J. S. & Seidel, H. D. (1993) *Prog. Med. Virol.* **40**, 82–106.
- Smith, J. S., Yager, P. A., Bigler, W. J. & Hartwig, E. C. J. (1990) *J. Wildlife Dis.* **26**, 473–485.
- Nadin-Davis, S. A., Sampath, M. I., Casey, G. A., Tinline, R. R. & Wandeler, A. I. (1999) *Epidemiol. Infect.* **123**, 325–336.
- Epperson, B. K. (2003) *Geographical Genetics* (Princeton Univ. Press, Princeton).
- Wright, S. (1943) *Genetics* **28**, 114–138.
- Higgins, D. G. & Sharp, P. M. (1988) *Gene* **73**, 237–244.
- Posada, D. & Buckley, T. (2004) *Syst. Biol.* **53**, 793–808.
- Posada, D. & Crandell, K. A. (1998) *Bioinformatics* **14**, 817–818.
- Yang, Z. (1997) *Comput. Appl. Biosci.* **13**, 555–556.
- Rambaut, A. (2000) *Bioinformatics* **16**, 395–399.
- Rousset, P. (2004) *Genetic Structure and Selection in Subdivided Populations* (Princeton Univ. Press, Princeton).
- Oden, N. L., Sokal, R. R., Fortin, M.-J. & Goebel, H. (1993) *Geogr. Anal.* **25**, 315–336.
- Jacquez, G. & Maruca, S. (2001) *Boundary Seer* (TerraSeer, Ann Arbor, MI).
- Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W. (1996) *Applied Linear Statistical Models* (McGraw-Hill, Boston).
- Moya, A., Elena, S., Bracho, A., Miralles, R. & Barrio, E. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 6967–6973.
- Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. (2002) *J. Mol. Evol.* **54**, 156–165.
- Tinline, R. R. (1988) in *Rabies*, eds. Cambell, J. B. & Carlton, K. M. (Kluwer, Boston), pp. 301–322.
- Tinline, R. R. & MacInnes, C. D. (2004) *J. Wildlife Dis.* **40**, 212–221.
- Kristt, A. C., Lively, C. M., Levri, E. P. & Jokela, J. (2000) *Parasitology* **121** (Pt 4), 395–401.
- Lively, C. M. & Dybdahl, M. F. (2000) *Nature* **405**, 679–681.
- Holmes, E. C., Woelk, C. H., Kassir, R. & Bourhy, H. (2002) *Virology* **292**, 247–257.
- Walsh, P., Biek, R. & Real, L. A. (2005) *PloS Biol.*, in press.