Correlation between geographic distance and genetic similarity in an international collection of bovine faecal *Escherichia coli* O157:H7 isolates

M. A. DAVIS^{1*}, D. D. HANCOCK¹, T. E. BESSER², D. H. RICE¹, C. J. HOVDE³, R. DIGIACOMO⁴, M. SAMADPOUR⁵ AND D. R. CALL²

(Accepted 28 March 2003)

SUMMARY

Evidence from epidemiological and molecular studies of bovine *Escherichia coli* O157:H7 suggests that strains are frequently transmitted across wide geographic distances. To test this hypothesis, we compared the geographic and genetic distance of a set of international bovine *Escherichia coli* O157:H7 isolates using the Mantel correlation. For a measure of genetic relatedness, pulsed-field gel electrophoresis of six different restriction enzyme digests was used to generate an average Dice similarity coefficient for each isolate pair. Geographic distance was calculated using latitude and longitude data for isolate source locations. The Mantel correlation between genetic similarity and the logarithm of geographic distance in kilometers was -0.21 (P < 0.001). The low magnitude of the Mantel correlation indicates that transmission over long distances is common. The occurrence of isolates from different continents on the same cluster of the dendrogram also supports the idea that *Escherichia coli* O157:H7 strains can be transferred with considerable frequency over global distances.

INTRODUCTION

The significance of *Escherichia coli* O157:H7 (O157) as a human pathogen has been known since 1982 [1]. While various food vehicles have been associated with human infection, foods of bovine origin are the most common vehicle [2–5]. There have also been an increasing number of outbreaks attributed to either direct human to animal contact, or to contact with a manure-contaminated environment [6–10]. Therefore,

research efforts aimed at reducing prevalence of O157 among cattle herds have great potential for reducing human exposure. In order to target critical control points for O157 on farms, it is necessary to understand how O157 is disseminated between cattle herds.

Results of molecular genetic subtyping of O157 from cattle herds suggest that new strains of O157 are introduced onto farms by some mechanism other than animal movement: (1) indistinguishable subtypes of O157 were isolated from herds located hundreds of kilometers apart, (2) the number of subtypes did not differ between open herds (where replacement animals were purchased from off the farm) and closed herds

¹ Field Disease Investigation Unit, Department of Veterinary Clinical Sciences, Washington State University, Pullman, WA 99164, USA

² Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA 99164, USA

³ Department of Microbiology, Molecular Biology and Biochemistry, University of Idaho, Moscow, ID 83844, USA

⁴ Division of Epidemiology, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195, USA

⁵ Department of Environmental Health, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195, USA

^{*} Author for correspondence: Department of Microbiology, Molecular Biology and Biochemistry, University of Idaho, Moscow, ID 83844, USA.

(where replacements were raised on the farm), and (3) in herds sampled longitudinally, new subtypes of O157 were observed to appear, or herds that were negative for more than a year became positive for O157, suggesting a source of new strains external to the farm [11, 12]. Although O157 has been found in wildlife [13], it seems unlikely that wildlife could account for transmission over hundreds of kilometers.

One plausible vehicle for widespread dissemination is cattle feed, which is frequently shipped over great geographic distances. For example, cottonseed is a common component of cattle diets in the northwestern United States, so must be frequently shipped from the southern states where it is produced [14]. Feeds are also shipped intercontinentally more frequently and in greater amounts by weight than are live cattle [15]. Generic E. coli are common in purchased feeds (30-50% of samples, [16]) and can replicate to high concentrations when feeds become wet [16]. O157 has been found at low prevalence in samples of component feeds prior to their exposure to cattle on the farm (personal observation). Because the typical dairy herd purchases hundreds of truckloads of feeds per year, even a low prevalence of O157 would result in several exposures per year. The importance of feed as a vehicle for transmission of Salmonella enterica has been well documented for pigs, poultry, and cattle [17-21].

We hypothesized that O157 strains are frequently transmitted across wide geographic distances, and that feed is a likely vehicle for transmission. Because of the difficulties inherent in testing this hypothesis directly, we chose to employ molecular subtyping to compare genetic and geographic distance between O157 isolates of bovine origin. If transmission by feed were common, one would expect to see very little localization of subtypes, and the correlation between genetic and geographic distance should be low. On the other hand, a high genetic-geographic distance correlation among isolates would not support the hypothesis that feeds play the major role in inter-herd dissemination. The Mantel test was chosen because it accommodates for the lack of independence between pairwise comparisons and has been used for geneticgeographic analyses of other species [22, 23].

METHODS

Isolates

Isolates of O157 from the United States were drawn from studies conducted by the Field Disease

Investigation Unit, Washington State University, Washington, USA [12, 13, 24, 25], and from the Food Animal Health Research Program, Ohio State University, Wooster, Ohio. Isolates from Japan, Australia, and Scotland were provided by the National Institute of Animal Health in Japan [26, 27], Food Safety and Quality Section, Food Science Australia, Brisbane Laboratory [28, 29], and the Regional Veterinary Centre, Inverness, Scotland [30], respectively. Confirmation of O157 was performed as in cited articles, and for the isolates from Ohio by multiplex PCR carried out at Washington State University [31, 32].

Pulsed-field gel electrophoresis (PFGE)

Isolates were grown overnight on sheep blood agar, suspended with sterile cotton swabs in 3 ml of TE buffer (100 mm Tris and 100 mm EDTA) and adjusted to an optical density of 1·3–1·4 at 610 nm. Aliquots of bacterial suspension (200 μ l) were transferred to 1.5-ml microcentrifuge tubes. Proteinase K (10 µl of 20 mg/ml) was added to each tube. InCert (FMC, Rockland, ME) agarose was added (1.6% with 1% sodium dodecyl sulfate in TE, 200 µl) to each tube, mixed, and immediately dispensed into plug molds (BioRad, Hercules, CA). After solidifying, plugs were transferred to round-bottomed tubes, and 1.5 ml of ES buffer (0.5 M EDTA, pH 9.0; 1 % sodium-lauroylsarcosine) and 40 μ l of proteinase K (20 mg/ml) were added to each tube. Plugs were then incubated with shaking at 54 °C for 1 h. Lysis buffer was then removed from each tube, plugs were washed in sterile distilled water (twice at 50 °C, 15 min), and then in TE buffer (four times, 50 °C, 15 min). Plug slices (2 mm) were digested with a restriction endonuclease (see Table 1 for times and temperatures). PFGE was performed on a CHEF-DRII PFGE apparatus (BioRad) using the following parameters: separation on an agarose (Seakem Gold agarose, FMC Bio Products, Rockland, ME) gel in 0.5× Tris-Borate-EDTA at 14 °C at 6 volts/cm. Agarose concentration, run time and pulse time differed according to the enzyme used (Table 1). Gels were stained with ethidium bromide and photographed with UV transillumination. The photographic image was captured digitally using a gel documentation system (AlphaImager 2000, Alpha Innotech Corporation, San Leandro, CA). Digital images were analysed using Bionumerics software (Applied Maths, Sint-Martens-Latem, Belgium). Both lambda DNA ladders (BioRad) and the standard

Enzyme	Digestion temperature (°C)	Digestion time (h)	Units of enzyme per plug	Agarose concentration (%)	Run time (h)	Initial switch time (s)	Final switch time (s)	Average number of bands
BlnI	37	16-20	10	1.0	22	2.2	54.2	12.8
VheI	37	16-20	15	1.0	24	1	15	12.4
PacI	37	16-20	24	1.2	24	0.1	11	16.7
SfiI	50	16-20	29	1.0	22	1	15	17:4
SpeI	37	16-20	25	1.0	22	5	25	10.5
YbaI	37	4–16	30	1.0	20	2.2	54.2	14.9

Table 1. Restriction enzyme digest conditions and electrophoresis conditions for each restriction enzyme used for PFGE of O157 isolates

isolate, G5244 [33], were used as size standards and for between-gel standardization in Bionumerics.

Latitude and longitude and United States trade data

Latitude and longitude data for locations outside the United States were obtained from the Gazetteer Search engine of the National Imagery and Mapping Agency web site at http://gnpswww.nima.mil/geonames/GNS/index.jsp. Latitude and longitude data for locations inside the United States were obtained from the US Gazetteer search engine of the US Census Bureau web site at http://www.census.gov/cgi-bin/gazetteer. Latitude and longitude in decimal degrees were used to calculate distance in kilometers using a published formula based on spherical trigonometry [34]. A limited number of distance calculations were checked using an atlas of the world [35].

Numbers of live cattle and metric tons of feed imported from and exported to the United States were available on the USDA Foreign Agricultural Service internet site, http://www.fas.usda.gov/ustrade/ [36]. Bulk, Intermediate, and Consumer-Oriented (BICO) commodity groups included all feeds and fodders except for poultry feeds, and all live bovine animals.

Data analysis

For each enzyme, Dice similarity coefficients were calculated in Bionumerics (Applied Maths). Percent tolerance settings varied according to the enzyme used (range: 1·0–3·0%), and were determined by minimizing the percent tolerance that resulted in 100% similarity between the standard isolate lanes. The Dice similarity coefficient is equal to the number of shared bands times two, divided by the total number of bands being compared [37]. Cluster analysis was performed in Bionumerics based on Dice coefficients and the

Table 2. Source of O157 isolates by year of sample collection and country of origin

	Australia	Japan	Scotland	United States	Total
1994	_		_	17	17
1995	1	_	_	5	6
1996	1	17	_	3	21
1997	3	_	_	_	3
1998	1	_	_	1	2
1999	1	_	11	4	16
2002	_		_	4	4
Total	7	17	11	34	69

unweighted pair method of arithmetic averages (UPGMA) to generate a composite dendrogram based on an unweighted average of six similarity coefficients from six enzymes. This average similarity for each isolate pair was used as a single estimate of relatedness, which was then used for the genetic–geographic comparison. The geographic distance matrix was computed using SAS (Cary, NC). The genetic similarity and geographic distance matrices were exported from SAS to Excel, in which the Mantel correlation [38] between genetic similarity and the logarithm (base 10) of the distance in kilometers was calculated using Poptools [39].

RESULTS

The total number of isolates was 69, with collection years ranging from 1994 through 2002 (Table 2). The Mantel correlation between genetic similarity (average Dice similarity coefficient from 6 enzymes) and the log base 10 of geographic distance in kilometers was -0.21 (P < 0.01). Figure 1 is the composite dendrogram from the average of six Dice similarities generated by six enzymes. There were several examples where isolates from separate continents occurred on

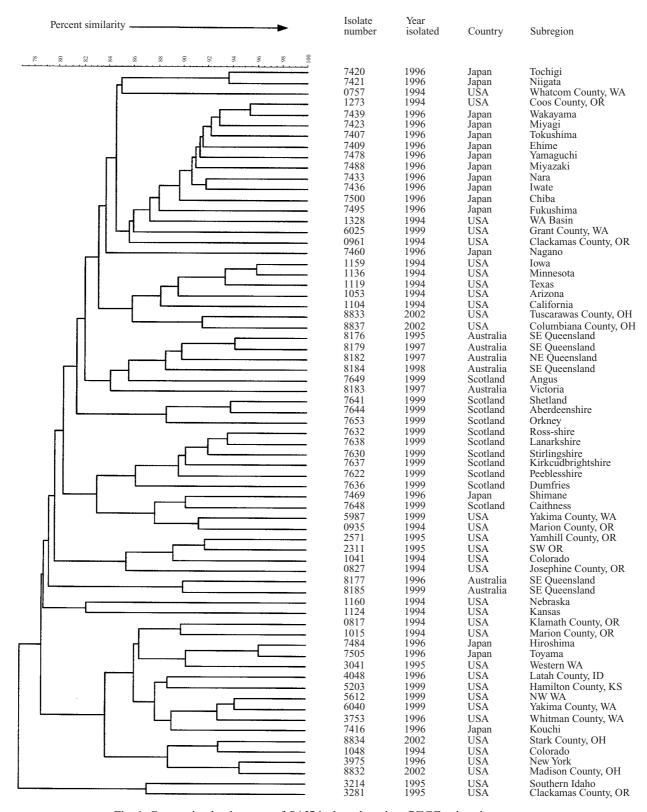


Fig. 1. Composite dendrogram of O157 isolates based on PFGE using six enzymes.

the same cluster, and so were likely to share a recent common ancestor, including strain 1273 from Oregon clustered with a six isolates from Japan, an isolate from Japan (7469) and from Scotland (7648) clustered together, and another isolate from Japan (7416) clustered with three isolates from Washington State

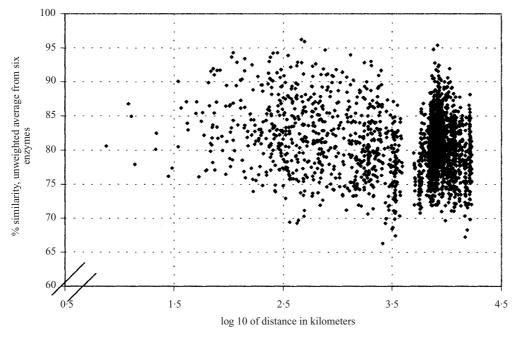


Fig. 2. Scatterplot of genetic similarity vs. log 10 geographic distance for 63 bovine faecal *Escherichia coli* O157:H7 isolates (2346 isolate pairs). The Mantel correlation was -0.21 (P < 0.01).

(Fig. 1). Figure 2 is a scatterplot of genetic similarity as measured by the average of six Dice coefficients plotted against the log base 10 of geographic distance in kilometers. Occurrences of isolate pairs with high similarity but from widely separated geographic locations occur in the upper-right of the scatterplot distribution.

DISCUSSION

The finding of a significant Mantel correlation confirms that transmission over a short distance is more likely to occur than over a long distance, however the low magnitude (-0.21) indicates that transmission over long distances is common. The occurrence of geographically distant isolates on the same clusters of the dendrogram also supports the frequent occurrence of such widespread transmission. In contrast, genetic–geographic analyses of infectious agents which are environmentally labile outside of a host, and therefore not subject to fomite transmission, have demonstrated strong geographic clustering of subtypes [40–42].

While the Mantel correlation and dendrogram support the idea that O157 is transferred with considerable frequency over global distances, how this occurs is unknown. Some reasonable hypotheses concerning vehicles of dissemination include the movements

of wild or domestic live animals, bird migration, human travel, and shipments of feeds. Most terrestrial wildlife species do not migrate over great distances, and none over oceans. Wild birds do migrate great distances, but the major migratory routes lie in a north-south direction and would not explain frequent east-west transoceanic transmission [43]. Human travel occurs frequently and transcontinentally, so cannot be ruled out as a possible means of dissemination of O157 strains. The introduction of O157 into a cattle herd from human faeces would most likely occur by faecal contamination of cattle feed at some point during the processing and transport of feed. In developed countries such as those represented in the present study, human faecal contamination of herd water supplies is probably rare, but faecal contamination of cattle feed does occur in parts of the United States as demonstrated by a relatively high incidence of Taenia saginata cysticercosis in feedlot cattle of the northwestern United States [44, 45].

Data on relative amounts of feed and live cattle that are shipped intercontinentally are consistent with the hypothesis that feed can be a vehicle for O157 dissemination. The isolates from Japan were distributed throughout the dendrogram in Figure 1 and frequently clustered with US isolates, whereas isolates from Australia and Scotland tended to cluster with each other.

During 1992 through 2001, United States exports of feeds and fodders to Japan averaged 2·4 million metric tons per year, while those to Australia and the United Kingdom averaged 0·8 million and 0·3 million, respectively [36]. Compared to feeds, animals are shipped in fewer amounts by weight and less frequently. Although intercontinental trade in live bovine animals is small and sporadic, a single live bovine animal could be colonized with high numbers of bacteria so it is difficult to rule out their role in dissemination of O157.

It is possible that our finding of similar isolates from geographically distant locations is a result of recurrent mutations (homoplasy); in other words, even though isolates have a high similarity based on restriction enzyme profiles, they may not truly share a recent common ancestor. This seems unlikely, because we used six enzymes rather than one or two to establish relatedness. As shown previously, the correlation between similarity coefficients generated by two single enzymes was poor (average 0.4), but that correlation improved when additional enzymes contributed to the similarity measure. This was determined by comparing the average similarity from independent sets of two and three enzymes. It was projected from those data that two independent sets of six enzymes would produce similarity measures that were highly correlated (r=0.8) (personal observation).

Unfortunately, the data available in the present study made it difficult to rule out time as a confounder. The isolates from Japan and Scotland were relatively clustered in time as well as space, which would bias our observations toward a higher correlation between genetic similarity and geographic distance. On the other hand, isolates from the US and Australia were obtained over several years, which would bias our observations toward a lower genetic—geographic correlation. The distribution of isolates in the dendrogram argues against a large effect of such biases, because isolates from Japan and the US occurred on clusters throughout the dendrogram, while most isolates from Australia clustered together.

Future opportunities for a more direct test of the hypothesis that feeds serve as a vehicle to introduce new strains of O157 into herds may be provided as the sensitivity of detection methods improves. A longitudinal study of small, closed herds fed heat-treated feeds compared to those fed untreated feeds may be feasible. This or other study design options should be encouraged in light of the results presented here.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge Dr Masato Akiba, Dr Rowland Cobbold, Dr Barti Synge, and Dr Jeff LeJeune for generously providing isolates. We would also like to acknowledge Yubei Zhang for technical assistance and Ravi Pallipamu for assistance with Bionumerics and all of the FDIU staff for essential contributions to this project.

Funding was provided in part by the National Alliance for Food Safety – USDA: Agricultural Research Service, in part by the Washington State University Agriculture Research Centre, and in part by the United States Department of Health and Human Services Food and Drug Administration (reference number FD-U-001608).

REFERENCES

- Wells JG, Davis BR, Wachsmuth IK, et al. Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype. J Clin Microbiol 1983; 18: 512–520.
- Wachsmuth IK, Sparling PH, Barrett T, Potter ME. Enterohemorrhagic *Escherichia coli* in the United States. FEMS Immunol Med Microbiol 1997; 18: 233–239.
- Mead PS, Finelli L, Lambert-Fair MA, et al. Risk factors for sporadic infection with *Escherichia coli* O157:H7. Arch Intern Med 1997; 157: 204–208.
- Armstrong GL, Hollingsworth J, Morris JGJ. Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. Epidemiol Rev 1996; 18: 29–51
- CDC. Summary of outbreaks of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli* reported to the CDC in 1999. http://www.cdc.gov/ncidod/dbmd/ diseaseinfo/files/ecoli_99summary.pdf (updated June, 2000). 2000.
- O'Brien SJ, Adak GK, Gilham C. Contact with farming environment as a major risk factor for Shiga toxin (Vero cytotoxin)-producing *Escherichia coli* O157 infection in humans. Emerg Infect Dis 2001; 7: 1049–1051.
- CDC. Outbreaks of *Escherichia coli* O157:H7 infections among children associated with farm visits Pennsylvania and Washington, 2000. MMWR 2001;
 50: 293–297.
- 8. Trevena WB, Willshaw GA, Cheasty T, Domingue G, Wray C. Transmission of Vero cytotoxin producing *Escherichia coli* O157 infection from farm animals to humans in Cornwall and west Devon. Commun Dis Public Health 1999; 2: 263–268.
- Trevena WB, Willshaw GA, Cheasty T, Wray C, Gallagher J. Vero cytotoxin-producing E. coli O157 infection associated with farms. Lancet 1996; 347: 60–61.

- Rice D, Hancock D, Vetter RL, Besser T. Escherichia coli O157 infection in a human linked to exposure to infected livestock. Vet Rec 1996; 138: 311.
- 11. Hancock D, Besser T, Lejeune J, Davis M, Rice D. The control of VTEC in the animal reservoir. Int J Food Microbiol 2001; **66**: 71–78.
- Rice DH, McMenamin KM, Pritchett LC, Hancock DD, Besser TE. Genetic subtyping of *Escherichia coli* O157 isolates from 41 Pacific Northwest USA cattle farms. Epidemiol Infect 1999; 122: 479–484.
- 13. Hancock DD, Besser TE, Rice DH, Ebel ED, Herriott DE, Carpenter LV. Multiple sources of *Escherichia coli* O157 in feedlots and dairy farms in the northwestern USA. Prev Vet Med 1998; **35**: 11–19.
- USDA. National Agricultural Statistics Service. http:// www.fas.usda.gov/nass/pubs/pubs.htm, 2002.
- USDA. Foreign Agricultural Service U.S. Trade Internet System. http://www.fas.usda.gov/ustrade/, 2002.
- Lynn TV, Hancock DD, Besser TE, et al. The occurrence and replication of *Escherichia coli* in cattle feeds.
 J Dairy Sci 1998; 81: 1102–1108.
- Glickman LT, McDonough PL, Shin SJ, Fairbrother JM, LaDue RL, King SE. Bovine salmonellosis attributed to *Salmonella anatum*-contaminated haylage and dietary stress. J Am Vet Med Assoc 1981; 178: 1268–1272.
- Anderson RJ, Walker RL, Hird DW, Blanchard PC. Case-control study of an outbreak of clinical disease attributable to *Salmonella menhaden* infection in eight dairy herds. J Am Vet Med Assoc 1997; 210: 528-530.
- Krytenburg DS, Hancock DD, Rice DH, Besser TE, Gay CC, Gay JM. A pilot survey of Salmonella enterica contamination of cattle feeds in the Pacific northwestern USA. Animal Feed Sci Technol 1998; 75: 75–79
- 20. Lindqvist N, Heinikainen S, Toivonen AM, Pelkonen S. Discrimination between endemic and feedborne *Salmonella infantis* infection in cattle by molecular typing. Epidemiol Infect 1999; **122**: 497–504.
- 21. Hinton MH. Infections and intoxications associated with animal feed and forage which may present a hazard to human health. Vet J 2000; **159**: 124–138.
- 22. Hutchinson DW, Templeton AR. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 1999; 53: 1898–1914.
- Gordon DM. The genetic structure of *Escherichia coli* populations in feral house mice. Microbiology 1997;
 143: 2039–2046.
- Hancock D, Rice D, Thomas LA, Dargatz D, Besser T. Epidemiology of *Escherichia coli* O157 in feedlot cattle. J Food Prot 1997; 60: 462–465.
- Hancock DD, Besser TE, Rice DH, Herriott DE, Tarr PI. A longitudinal study of *Escherichia coli* O157 in fourteen cattle herds. Epidemiol Infect 1997; 118: 193–195.
- Akiba M, Masuda T, Sameshima T, Katsuda K, Nakazawa M. Molecular typing of Escherichia coli

- O157:H7 (H-) isolates from cattle in Japan. Epidemiol Infect 1999; **122**: 337–341.
- Akiba M, Rice DH, Davis MA, et al. A comparison of *Escherichia coli* O157 isolates from cattle in Japan and the USA by molecular biological methods. Epidemiol Infect 2000; 125: 221–224.
- Cobbold R, Desmarchelier P. A longitudinal study of Shiga-toxigenic *Escherichia coli* (STEC) prevalence in three Australian dairy herds. Vet Microbiol 2000; 71: 125–137.
- Cobbold R, Desmarchelier P. Characterisation and clonal relationships of Shiga-toxigenic *Escherichia* coli (STEC) isolated from Australian dairy cattle. Vet Microbiol 2001; 79: 323–335.
- Synge B, Paiba G. Verocytotoxin-producing E coli O157. Vet Rec 2000; 147: 27.
- 31. Fagan PK, Hornitzky MA, Bettelheim KA, Djordjevic SP. Detection of Shiga-like toxin (stx1 and stx2), intimin (eaeA), and enterohemorrhagic Escherichia coli (EHEC) hemolysin (EHEC hlyA) genes in animal faeces by multiplex PCR. Appl Environ Microbiol 1999; 65: 868–872.
- Hu Y, Zhang Q, Meitzler JC. Rapid and sensitive detection of *Escherichia coli* O157:H7 in bovine faeces by a multiplex PCR. J Appl Microbiol 1999; 87: 867–876.
- 33. CDC. Standardized molecular subtyping of foodborne bacterial pathogens by pulsed-field gel electrophoresis. CDC Training Manual, Foodborne and Diarrheal Diseases Branch, Centers for Disease Control and Prevention, Ga, 1998.
- Ng E, Wilkins R, Perras A. How far is it to the nearest hospital? Calculating distances using the Statistics Canada Postal Code Conversion File. Health Rep 1993; 5: 179–188.
- Shupe JF, ed. National geographic atlas of the world, 6th edn. Washington, DC: National Geographic Society, 1992.
- 36. United States Department of Agriculture Foreign Agricultural Service. BICO Reports. http://www.fas.usda.gov/, 2002.
- 37. Dice LR. Measures of the amount of ecologic association between species. Ecology 1945; **26**: 297–302.
- 38. Manly BFJ. Randomization and Monte Carlo methods in biology. London: Chapman and Hall, 1991.
- Hood GM. PopTools CSIRO Sustainable Ecosystems, http://www.cse.csiro.au/CDG/poptools/index.htm, 2000.
- 40. Maruyama S, Kasten RW, Boulouis H, Gurfield NA, Katsube Y, Chomel BB. Genomic diversity of *Barton-ella henselae* isolates from domestic cats from Japan, the USA and France by pulsed-field gel electrophoresis. Vet Microbiol 2001; 79: 337–349.
- Meas S, Ohashi K, Sugimoto C, Onuma M. Phylogenetic relationships of bovine immunodeficiency virus in cattle and buffaloes based on surface envelope gene sequences. Brief report. Arch Virol 2001; 146: 1037–1045.
- 42. Forsberg R, Oleksiewicz MB, Petersen AK, Hein J, Bøtner A, Storgaard T. A molecular clock dates the common ancestor of European-type porcine

- reproductive and respiratory syndrome virus at more than 10 years before the emergence of disease. Virology 2001; **289**: 174–179.
- 43. Berthold P. Bird migration: a general survey. Oxford Ornithology Series. New York: Oxford University Press, 2001.
- 44. Hancock DD, Wikse SE, Lichtenwalner AB, Wescott RB, Gay CC. Distribution of bovine cysticercosis in Washington. Am J Vet Res 1989; **50**: 564–570.
- 45. Yoder DR, Ebel ED, Hancock DD, Combs BA. Investigation of an outbreak of bovine cysticercosis in an Idaho feedlot. J Am Vet Med Assoc 1994; 205: 45–50.