The Pennsylvania State University

The Graduate School

Department of Biology

ECOLOGY AND EVOLUTION OF THE *PARAMYXOVIRIDAE*

A Dissertation in

Biology

by

Laura W. Pomeroy

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2008

The dissertation of Laura W. Pomeroy was reviewed and approved* by the following:

Ottar N. Bjørnstad
Professor of Biology and Entomology and Adjunct Professor in Statistics
Dissertation Advisor
Chair of Committee

Réka Albert
Associate Professor of Physics and Biology

Bryan Grenfell
Alumni Professor of Biology

Edward C. Holmes
Professor of Biology

Peter J. Hudson
Willaman Professor of Biology

Douglas Cavener
Professor of Biology
Head of the Biology Department

*Signatures are on file in the Graduate School

ABSTRACT

Classical epidemiological theory uses compartmental susceptible-infected-removed (SIR) models to quantitatively explore epidemics caused by acute, immunizing viruses that sweep through local populations. While the basic theory explains viral-host dynamics in some systems, added model complexity creates a more accurate context for other systems. As a case study, I investigate the system of phocine distemper virus (PDV), a morbillivirus, in harbor seals (*Phoca vitulina*) in the North Sea. Using this dataset, I extend the classical theory by incorporating host age or stage heterogeneities, geographical heterogeneities, and host movement heterogeneities.

First, I investigated age or stage heterogeneities in transmission among the harbor seals in the Dutch 2002 PDV epidemic by creating three models to see which best fit the data. The model with the highest degree of heterogeneity best fit the host stage-structure ($p=0.0004$). I also estimated the "who acquires infection from whom" (WAIFW) matrix from detailed incidence data and confirmed with an R_0 calculation using next-generation formalism.

Next, I addressed geographic heterogeneity in the host population by addressing error inherent in the PDV incidence data in the entire North Sea with a Bayesian framework to estimate the initial population of susceptible individuals (S_0), the rate of pathogen transmission (β), and the time series of infected individuals with imperfect binomial reporting to the biweekly incidence time series and used this information to create distance and gravity based models to discriminate how different geographic locations are coupled by infected host movement. Results show that the distance model has a better fit to the seal stranding data, indicating that the distance between harbor seal haulouts drives the spatial spread of PDV.

Lastly, I looked at the epidemiological and evolutionary dynamics of the viral family *Paramyxoviridae* by investigating serially-sampled genomes of measles virus, mumps virus, and canine distemper virus. Using a Bayesian coalescent approach, we estimate viral substitution rates, the time to common ancestry and elements of their demographic history. Strikingly, the mean Time to the Most Recent Common Ancestor (TMRCA) was both similar and very recent among the viruses studied, ranging from only 58 to 91 years (1908 to 1943).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am greatly indebted to many people who supported my academic and personal endeavors throughout graduate school. Without the help and support of my advisor, committee, colleagues, friends, and family, this work would not have been possible.

Ottar Bjørnstad has been a wonderful mentor and advisor. Not only has he supported my academic projects, ideas, and growth, but he has also allowed me to pursue teaching and outreach experiences during my tenure in graduate school. He has shown a great deal of kindness, understanding, wisdom, and patience. He has made my experience in graduate school very positive, and I am very grateful for everything that he has taught me.

I am also appreciative to my committee, comprised of Réka Albert, Bryan Grenfell, Eddie Holmes, and Peter Hudson. Each member has been instrumental in my success, offering advice and encouragement. Special thanks to Peter Hudson, for forging my initial contact with collaborators at the University of St. Andrews and for the experiences during my rotation in his lab, and to Eddie Holmes, for allowing me to collaborate on a project which I found very exciting and for always offering clear and practical advice. In addition, thanks to Bryan Grenfell for constant support and encouragement in all aspects of my graduate career.

I thank the past and present members of the Bjørnstad and Grenfell Labs: Nita Bharti, Isabella Cattadori, Lawrence Chien, Matt Ferrari, Ashley Flannery, Mike Heinbach, Derek Johnson, Petra Klepac, Katia Koelle, Jennie Lavine, Jamie Lloyd-Smith, Angie Luis, Kim Peppin, Ginny Pitzer, Akiko Satake, Carrie Schwarz, Dietmar Schwarz, Conrad Stack, and Igor Volkov. Their ideas, comments, and discussions during lab meetings were very helpful. In particular, Matt Ferrari and Petra Klepac have been sources of advice, help, and humor during my entire time in graduate

school. Without them and their assistance, much of this work would not have been possible. Angie Luis has been a great office companion and advisor. In addition, the entire Center for Infectious Disease Dynamics (CIDD) community at Penn State has been extremely generous with their time and friendship. Special thanks are due to Rubing Chen, Kristle Krichbaum, Sandra Lass, Sarah Perkins, and Catherine Williams.

Other members of the Penn State community have been instrumental in my work. From biology, thanks are reserved for Kathryn McClintock, Kat Shea and the members of her lab. The entomology department has been very welcoming. Thank you to Roxie Smith and to all the main office staff.

This work has been a collaboration of many scientists, residing at three different research institutions. First, I would like to thank our collaborators at the University of St. Andrews. All of the researches at the Sea Mammal Research Unit (SMRU) and at the Centre for Research in Ecological and Enviromental Modelling (CREEM) at the University of St. Andrews in Scotland have been wonderfully supportive. Special thanks goes out to Catriona Stephenson and John Harwood for their willingness to share phocine distemper virus and seal haulout data, answer questions, offer theories and advice, and for their hospitality during my visit to Scotland. Second, Karin Harding from the University of Gothenburg and Michael Neubert and Petra Klepac from Woods' Hole Oceanographic Instiution have been very generous with both data and advice. Lastly, Thijs Kuiken and Jolianne Rijks at Erasmus Medical Center in the Netherlands, have been very generous with their data and knowledge. Special thanks are due to Petra Klepac and Jolianne Rijks. Working with them has been very enjoyable and knowledgeable. They have helped

me enormously, giving advice on all things seal related, long after academic commitments require them to do so.

Many friends have offered a great deal of support. Judy Racusin has supported me through life events, sickness, and many various endeavors. I cannot thank her enough. Also, thanks are due to Sarah Jen, who has supported my endeavors for the past ten years. Thank you to those who have made State College a wonderful place, including Judy Racusin, Theresa Foley, Petra Klepac, Loredana Verete, Zeynep Sezen, Danielle Garneau, Chris Lafty, Randa Jabbour, Emily Rauschert, Elena Angelo, Susan Easton, and Liz Liadis.

Ultimately, the greatest gratitude that I can express is for my family. Thank you to Marshall and Christopher Pomeroy, who possess a never-ending supply of love, support, and patience. My parents, Art and Mary Warlow, have supported and encouraged me in all aspects of my life – including my interest in both biology and mathematics – from childhood through my undergraduate and graduate studies. Thank you for all of your support over many years. Additional thanks are reserved for Evelyn Chabowski and Sherry, Dean, and Corey Pomeroy for their support and understanding.

Overview

Acute, immunizing viruses cause epidemics among social hosts, including both human and animal populations. Typically, the disease sweeps through local populations, causing illness and mortality that can affect human health and the status of many wildlife species. Classical epidemiological theory studies these dramatic dynamics and has been developed over multiple centuries to study the natural behavior of pathogens in human and wildlife hosts. Original investigations were case studies of important diseases in individuals or in closed populations, such as Panum's study of measles on the Faroe islands (Panum 1940). Early epidemiological studies involved systematic study of outbreaks, such as John Snow's famous study of a cholera outbreak in London in the mid 1800s. In the modern era, the dynamics between pathogens and host populations are studied quantitatively, using both the genetic data – readily available since the advent of high throughput molecular sequencing techniques – and mathematical and ecological models to describe and predict the temporal and spatial spread of disease.

Classical theory in the population-level dynamics of pathogens and host rely on the susceptible-infected-removed (SIR) model, which assumes that the host population is homogeneous in susceptibility, infectiousness, behavior, genetics, and all other characteristics. Although models that incorporate this assumption work well in many cases, this assumption may be too simplistic for many disease systems among human and wildlife hosts. Throughout this thesis, I work to extend classical epidemiological theory to incorporate four sources of heterogeneities: age differences in the host population, geographic differences in a host metapopulation, behavioral differences in a host metapopulation, and genetic differences in a viral population. First, I investigate host heterogeneities present in natural populations, including age or stage-related differences in transmission apparent in phocine distemper virus (PDV) in Holland

(chapter 2). Another source of heterogeneity can exist when local regions of a metapopulation exhibit unique characteristics that affect the spread of an epidemic throughout the metapopulation. I address unique regional characteristics, such as the size of susceptible population, transmission rate, and determine how they differ with geography in the spread of PDV throughout the North Sea (chapter 3) and how differences in host behavior relate to the spread of the virus (chapter 4). Lastly, I use heterogeneities in viral genomes to inform epidemic dynamics (chapter 5).

Morbilliviruses in Marine Mammals

In the past two decades, morbilliviruses have decimated marine mammal populations. In 1987 and 1988, half of the population of bottlenose dolphins (*Tursipos truncatus*) in the Atlantic coastal waters off of the United States where killed by a morbillivirus (Lipscomb *et al.* 1994b), while thousands of seals (*Phoca siberica*) in Lake Baikal died of canine distemper virus in the same year (Grachev *et al.* 1989). Closely related viruses affected striped dolphins (*Stenella coeruleoalba*) in the Mediterranean Sea in the early 1990s (Domingo *et al.* 1992; Domingo *et al.* 1995), bottlenose dolphins in the Gulf of Mexico in 1993 (Lipscomb *et al.* 1994a), and harbor porpoises (*Phocoena phocoena*) from the Black Sea in the late 1990s (Kennedy 1998; Muller *et al.* 2002). In 2000, thousands of Capsian Seals died from canine distemper virus in the Caspian Sea (Kennedy *et al.* 2000).

In addition, two outbreaks of PDV have affected seal populations throughout the entire North Sea: the first outbreak occurred in 1988, in which 18,000 to 23,000 harbor seals died (Hall *et al.* 2006; Harkonen *et al.* 2006). This mass mortality event caused by the viral epidemic began on the Danish island of Ånholt on April 12, 1988 and ended within the calendar year (Dietz *et al.* 1989; Hall *et al.* 2006; Harkonen *et al.* 2006). A second PDV outbreak occurred in 2002 with the same point of origin: initial cases of

harbor seal stranding and mortality occurred on May 4, 2002. In this epidemic, approximately 22,000 to 30,000 harbor seals died, resulting in the largest recorded mass mortality event in marine mammals (Jensen *et al.* 2002; Hall *et al.* 2006; Harkonen *et al.* 2006). Both epidemics spanned the North Sea coastline: areas affected included Kattegat-Skagerrak coasts, the Wadden Sea, and the coast of the United Kingdom (Hall *et al.* 2006; Harkonen *et al.* 2006).

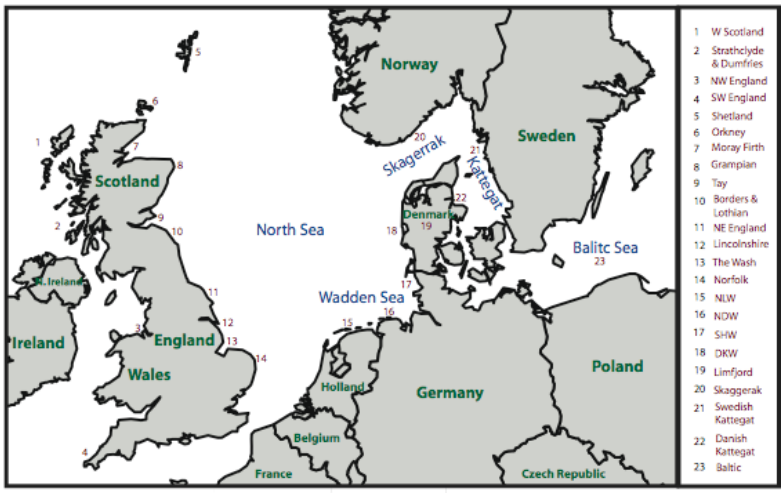

Figure 1-1. Locations of harbor seal strandings due to PDV in the North Sea. Complete stranding data that could be correlated with frequent harbor seal population surveys were available for 11 locations in 1988 and 21 locations in 2002. Study locations are highlighted and named on the map.

The causative agent in the harbor seal epidemic was a single-stranded, negative-sense RNA virus which is a member of the Morbillivirus genus, family *Paramyxoviridae* (Cosby *et al.* 1988; Mahy *et al.* 1988b; Osterhaus & Vedder 1988). Other members of the genus include measles virus, dolphin morbillivirius, porpoise morbillivirus, canine distemper virus, rindepest virus, and peste des petits ruminants virus (Barrett 1999).

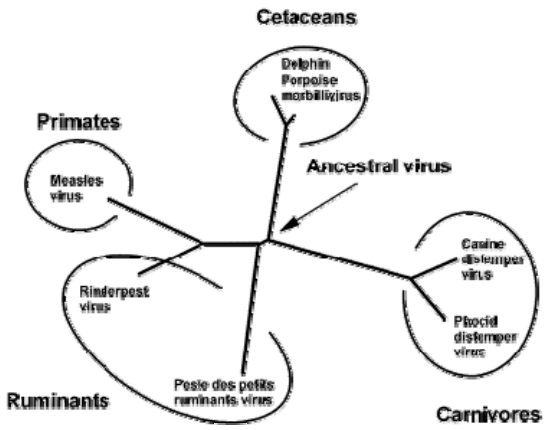

Figure 1-2. Phylogenetic relationships among the morbillivirus genus. The phylogenic relationships among the seven members of the morbillivirus genus are shown, among with the taxonomic order of their hosts (Barrett 1999, source).

In each individual, disease typically spans a two-week period, including both the latent and infectious disease stages (Osterhaus 1989; Harder 1990b; Baker 1992; Grenfell *et al.* 1992). Infected aerosol droplets transmit the disease, and the disease weakens the immune system (Baker & Ross 1992). Mortality is high among harbor seals (Rijks *et al.* 2005), partly due to co-infections (Baker & Ross 1992). The disease has had a decimating impact on the harbor seal population in the North Sea, affecting the population size and status since the late 1980s.

Harbor Seals

Harbor seals (*Phoca vitulina*), also known as common seals, populate the North Sea (Reijnders & Lankester 1990), with a minimum population estimate of 83,200 (Sea Mammal Research Unit 2007). The seals exhibit three stage classes: juveniles, subadults, and adults. Juveniles contain the pups of the year. Subadults are generally classified as seals aged 1 to 2 years for females and 1 to 3 years for males. Adults are female seals greater than 2 years old and male seals that are greater than 3 years old.

Adults can weigh between 80 to 100 kg and be greater than 120 cm in length (Sea Mammal Research Unit 2007) with lifespans that can range from 20 to greater than 35 years (Harkonen *et al.* 1999; Sea Mammal Research Unit 2007).

Harbor seals regularly "haulout" on sandy or rocky shores during periods of low tide to rest, birth, and molt (Sea Mammal Research Unit 2007). The pupping season ranges from May to July, with almost all females giving birth each year. Pups are nursed for two to four weeks; during this time they remain isolated from other seals with their mother. The molt occurs in August. The breeding season occurs in late summer, after the pups are weaned (Newby 1973; Godsell 1988a; Harkonen & Heide-Joergensen 1990; Traut 1999a).

These life history traits combine to form a wealth of data that can be used to modeling the harbor seal population dynamics with respect to PDV disease status. Models like these, in the framework of an SIR model, can lead to insights about transmission, heterogeneities, and spatial spread.

SIR Model

The basis for both phocine distemper models in chapters two though four is the susceptible-infected-recovered (SIR) model (Kermack & McKendrick 1927; Anderson & May 1991). In this model, the population is divided into three categories. Susceptible individuals (*S*) never experienced infection nor were exposed to the virus. Infected individuals (*I*) harbor the virus and convert susceptible individuals into infected individuals. Lastly, removed individuals (*R*) were previously infected and either recovered from the disease with conferred lifelong immunity or removed from the system due to mortality. A deterministic version of the model in continuous time would be:

$$\frac{dS}{dt} = -\phi S \qquad (1)$$

$$\frac{dI}{dt} = \phi S - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

where the rate of infection is determined by the force of infection, ϕ, and the rate of recovery or removal is given by γ. Mixing dynamics can either be modeled as density-dependent or frequency-dependent transmission through the force of infection (ϕ) term (McCallum *et al.* 2001). Density-dependent transmission assumes the rate of contact with infected hosts depends on host density, as follows

$$\phi = \beta I \quad (4)$$

where the force of infection (ϕ) is determined by the transmission rate (β), the number of infected individuals (I). Conversely, frequency-dependent transmission assumes the rate of contact with infected hosts is independent of host density, where the force of infection can be formulated as

$$\phi = \frac{\beta I}{N} \quad (5)$$

where the force of infection (ϕ) is determined by the transmission rate (β), the number of infected individuals (I) and scaled by the total number of individuals in the population (N). Haulouts are the units of transmission of PDV. When on the haulout, harbor seals rest on land in relatively immobile groupings. A seal will only pass infections to its nearest neighbors, no matter what the size of the seal group on the haulout. Because of these behaviors, PDV in harbor seals is best portrayed using frequency dependence, which is also consistent with previous studies (Swinton *et al.* 1999). Using this formalism, I investigate two scales of phocine distemper virus dynamics: the role of transmission between stages in a local PDV epidemic in the Dutch Wadden Sea (chapter 2) and the epidemic dynamics in the entire North Sea metapopulation (chapters 3 and 4). In this

way, I investigate different scales of viral-host dynamics, from the Dutch Wadden Sea to the entire North Sea metapopulation.

Viral Evolution

The largest scale of research in the dynamics of the *Paramyxoviridae* presented in this thesis investigates evolutionary and population wide demographics (chapter 5). RNA viruses, such as those that belong to the Morbillivirus genus and the family *Paramyxoviridae*, are known to have short generations and multiple errors introduced in each round of genomic replication, due to the lack of proofreading enzymes (Holland *et al.* 1982). Even though only a portion of the errors are fixed, RNA viruses still possess a staggering rate of substitution, with 10^{-3} substitutions per site, per year (Jenkins *et al.* 2002). By studying the history of mutation in a virus using serially-sampled sequences, we can use genetic data from canine distemper virus, mumps, and measles virus with a Bayesian coalescent approach (Drummond *et al.* 2002; Drummond *et al.* 2006), to understand and infer the demographic history of the *Paramyxoviridae*.

Coalescent Theory

Viral genetic sequences taken at serial times can inform genealogies of the sequences, the ancestral relationships or time to most recent common ancestor (TMRCA), and nucleotide substitution rates (Drummond *et al.* 2003). By following lineages back in time until the point where two lineages coalesce and possess a common ancestor, and using the fact that the probability of a coalescent event relates to the population size, one can estimate the past size of the infected population (Drummond *et al.* 2003). There are three main assumptions: that the virus does not recombine, that there is no positive selection, and that the population is panmixing.

Paramyxoviruses

Paramyxoviruses are negative-sense single-stranded RNA viruses that contain five or six structural genes at the 3' end and an L polymerase at the 5' end (Lamb & Kolakofsky 2001).

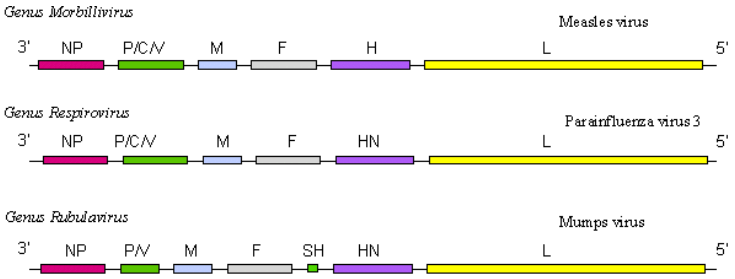

Figure 1-3. The genetic structure of the paramyxoviridae.Measles and canine distemper virus belong to the genus Morbillivirus. Their genome consists six genes that code for eight proteins. We used the nucleoprotein (NP) and hemagluttin (H) genes in our analysis. Mumps is a Rublavirus and contains seven genes that code for eight proteins. We used the hemagluttin (HN) gene in our analysis. Figure adapted from www.vadscorner.com/paramyxogenom.gif.

We investigated the nucleoprotein – which encapsulates the genome and associates with the polymerase and matrix protein (M) in replication and assembly, respectively (Lamb & Kolakofsky 2001) – in measles virus and the hemagluttinin in canine distemper virus, measles virus, and mumps virus. The hemagluttinin is an integral membrane glycoprotein that mediates host cell attachment (Lamb & Kolakofsky 2001).

Using this data in the Bayesian coalescent framework, we are able to estimate key demographic and evolutionary parameters, and therein gain important insights into the evolution and epidemiology of these viral pathogens (chapter 5).

Summary

Overall, classical epidemiological theory explains many aspects of viral-host dynamics; however, extensions to the theory can be made to fit other disease systems.. This thesis addresses heterogeneities in host populations in a local epidemic (chapter 2), geographic and regional heterogeneities that affect dynamics in a metapopulation (chapters 3 and 4), and molecular evolution with respect to host demography (chapter 5).

While this work addresses many quantitative aspects of epidemics among the *Paramyxoviridae*, there is a great deal of future work to be done, including the investigation of heterogeneities in other aspects of the North Sea seal metapopulation. To better understand and trace PDV spatial spread in the North Sea, temporally and spatially sampled sequences are needed from numerous marine mammal species in the region. Use of genetic data in this system could yield valuable insights into origins and patterns of disease spread. Ultimately, better understanding of uncertainties and heterogeneities in disease data will lead to clear pictures of viral-host dynamics in epidemics of acute, immunizing viruses.

References

Anderson R.M. & May R.M. (1991) *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.

Baker J.R. (1992) The Pathology of Phocine Distemper. *The Science of the Total Environment*, 115, 1-7

Baker J.R. & Ross H.M. (1992) The Role of Bacteria in Phocine Distemper. *The Science of the Total Environment*, 115, 9-14

Barrett T. (1999) Morbillivirus infections, with special emphasis on morbilliviruses of carnivores. *Veterinary Microbiology* 69, 3-13

Cosby S.L., McQuaid S., Duffy N., Lyons C., Rima B.K., Allan G.M., McCullough S.J., Kennedy S., Smyth J.A., McNeilly F., Craig C. & Orvell C. (1988) Characterization of a Seal Morbillivirus. *Nature*, 336, 115-116

Dietz R., Heide-Jorgensen M.-P. & Harkonen T. (1989) Mass Deaths of Harbor Seals *(Phoca Vitulina)* in Europe. *Ambio*, 18, 258-264

Domingo M., Vilafranca M., Visa J., Prats N., Trudgett A. & Visser I. (1995) Evidence for chronic morbillivirus infection in the Mediterranean striped dolphin (Stenella coeruleoalba). *Veterinary Microbiology*, 44, 229-239

Domingo M., Visa J., Pumarola M., Marco A.J., Ferrer L., Rabanal R. & Kennedy S. (1992) Pathological and immunocytochemical studies of morbillivirus infection of striped dolphins (Stenella coeruleoalba). *Veterinary Pathology* 29, 1-10

Drummond A.J., Ho S.Y.W., Phillips M.J. & Rambaut A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4, e88

Drummond A.J., Nicholls G.K., Rodrigo A.G. & Solomon W. (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, 161, 1307-1320

Drummond A.J., Pybus O.G., Rambaut A., Forsburg R. & Rodrigo A.G. (2003) Measurably evolving populations. *TRENDS in Ecology & Evolution*, 18, 481-488

Godsell J. (1988) Herd Formation and Haul-out Behavior in Harbor Seals *(Phoca vitulina)*. *Journal of Zoology, London*, 215, 83-98

Grachev M.A., Kumarev V.P., Mamaev L.V., Zorin V.L., Baranova L.V., Denikina N.N., Belikov S.I. & Petrov E.A. (1989) Distemper virus in Baikal seals. *Nature*, 338, 209

Grenfell B.T., Lonergan M.E. & Harwood J. (1992) Quantitative Investigations of the Epidemiology of Phocine Distemper Virus (PDV) in European Common Seal Populations. *Science of the Total Environment*, 115, 15-29

Hall A.J., Jepson P.D., Goodman S.J. & Harkonen T. (2006) Phocine distemper virus in the North and European Seas - Data and models, nature and nurture. *Biological Conservation*, 131, 221-229

Harder T., Willhaus, T.H., Frey, H.-R., and Liess, B. (1990) Morbillivirus Infections of Seals during the 1988 Epidemic in the Bay o Heligoland: III. Transmission Studies of Cell Culture-Propagated Phocine Distemper Virus in Harbor Seals (Phoca vitulina) and a Grey Seal (Halichoerus grypus): Clinical, Virological and Serological Results. *Journal of Veterinary Medicine*, 37, 641-650

Harkonen L., Dietz R., Reijnders P., Teilmann J., Harding K., Hall A., Brasseur S., Siebert U., Goodman S.J., Jepson P.D., Rasmussen T.D. & Thompson P. (2006) A Review of the 1988 and 2002 phocine distemper virus epidemics in European harbour seals. *Diseases of Aquatic Organisms*, 68, 115-130

Harkonen T., Harding K. & Lunneryd S.G. (1999) Age- and Sex-Specific Behaviour in Harbor Seals *Phoca vitulina* Leads to Biased Estimates of Vital Population Parameters. *Journal of Applied Ecology*, 36, 825-841

Harkonen T. & Heide-Joergensen M.P. (1990) Comparative life histories of East Atlantic and other harbour seal populations. *Ophelia*, 32, 211-235

Holland J., Spindler K., Horodyski F., Grabau E., Nichol S. & VandePol S. (1982) Rapid evolution of RNA genomes. *Science*, 215, 1577-1585

Jenkins G.M., Rambaut A., Pybus O.G. & Holmes E.C. (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *Journal of Molecular Evolution*, 54, 156-165

Jensen T., van de Bildt M., Dietz H.H., Andersen T.H., Hammer A.S., Kuiken T. & Osterhaus A. (2002) Another phocine distemper outbreak in Europe. *Science*, 297, 209-209

Kennedy S. (1998) Morbillivirus Infections in Aquatic Mammals *Journal of Comparative Pathology*, 119, 201-225

Kennedy S., Kuiken T., Jepson P.D., Deaville R., Forsyth M., Barrett T., van de Bildt M.W., Osterhaus A.D., Eybatov T., Duck C., Kydyrmanov A., Mitrofanov I. & Wilson S. (2000) Mass Die-Off of Caspian Seals Caused by Canine Distemper Virus. *Emerging Infectious Diseases*, 6, 637-639

Kermack W.O. & McKendrick A.G. (1927) Contributions to the mathematical theory of epidemics - I. *Proceedings of the Royal Society of Edinburgh A*, 115, 700-721

Lamb R.A. & Kolakofsky D. (2001) Paramyxoviradae: The Viruses and Their Replication. In: *Fields Virology* (eds. Knipe DM & Howley PM), pp. 1305-1340. Lippincott Williams & Wilkins, Philadelphia

Lipscomb T.P., Kennedy S., Moffett D. & Ford B.K. (1994a) Morbilliviral disease in an Atlantic bottlenose dolphin (Tursiops truncatus) from the Gulf of Mexico. *Journal of Wildlife Diseases*, 30, 572-576

Lipscomb T.P., Schulman F.Y., Moffett D. & Kennedy S. (1994b) Morbilliviral disease in Atlantic bottlenose dolphins (Tursiops truncatus) from the 1987-88 epizootic. *Journal of Wildlife Diseases*, 30, 567-571

Mahy B.W.J., Barrett T., Evans S., Anderson E.C. & Bostock C.J. (1988) Characterization of a Seal Morbillivirus. *Nature*, 336, 115

McCallum H., Barlow N. & Hone J. (2001) How should pathogen transmission be modelled? *TRENDS in Ecolgy & Evolution*, 16, 295-300

Muller G., Wunschmann A., Baumgartner W., Birkun A., Komakhidze A., Stanev T. & Joiris C.R. (2002) Immunohistological and serological investigations of morbillivirus infection in Black Sea harbour porpoises (Phocoena phocoena). *Veterinary Microbiology*, 87, 183-190

Newby T.C. (1973) Observations on the breeding behavior of the Harbor Seal in the State of Washington. *Journal of Mammology*, 54, 540-543

Osterhaus A.D., Uytdehaag, F.G., Visser, I.K., Vedder, E.J., Reijnders, P.J., Kuiper, J., and Brugge, H.N. (1989) Seal Vaccination Success. *Nature*, 337, 21

Osterhaus A.D.M.E. & Vedder E.J. (1988) Identification of Virus Causing Recent Seal Deaths. *Nature*, 355, 20

Panum P.L. (1940) Observations Made During the Epidemic of Measles on the Faroe Islands in the Year 1846. *Delta Omega*

Reijnders P.J.H. & Lankester K. (1990) Status of Marine Mammals in the North Sea. *Netherlands Journal of Sea Research*, 26, 427-435

Rijks J.M., Van de Bildt M.W.G., Jensen T., Philippa J.D.W., Osterhaus A. & Kuiken T. (2005) Phocine distemper outbreak, the Netherlands, 2002. *Emerging Infectious Diseases*, 11, 1945-1948

Sea Mammal Research Unit (2007) Special Committee on Seals Main Advice. In:

Swinton J., Harwood J., Gilligan C.A. & Hall A.J. (1999) Scaling of phocine distemper virus transmission with harbour seal community size. *Ecologie*, 30, 231-240

Traut I.M. (1999) Spacing among Harbor Seals *(Phoca vitulina viulina)* on Haul-out Sites in the Wadden Sea of Niedersachsen. *Z. Sugetierkunde*, 64, 51-53

Stage-Structured Transmission of Phocine Distemper Virus in the Dutch 2002 Outbreak

Laura W. Pomeroy[1,2], Petra Klepac[1,2], Ottar N. Bjørnstad[1,2,3], Thijs Kuiken[4], Albert DME Osterhaus[4], Jolianne M. Rijks[4,5]

[1] Department of Biology, the Pennsylvania State University, University Park, PA 16802 USA

[2] Center for Infectious Disease Dynamics, the Pennsylvania State University, University Park, PA 16802 USA

[3] Department of Entomology, the Pennsylvania State University, University Park, PA 16802 USA

[4] Department of Virology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, the Netherlands

[5] Dutch Wildlife Health Center, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584 CL, the Netherlands

Abstract

Heterogeneities in transmission are crucial, since these events can dictate epidemic dynamics. To elucidate transmission patterns of phocine distemper virus (PDV) between harbor seals (*Phoca vitulina*) during the 2002 Dutch outbreak, we created three models to distinguish transmission dynamics. A model in which the host population exhibited strong heterogeneous mixing best described PDV dynamics ($p=0.0004$), indicating that stage structured transmission occurs in the Dutch harbor seal population. To capture the stage transmission dynamics, we created a "who acquires infection from whom" (WAIFW) matrix solely from incidence data from seal strandings. Transmission between subadults and adults was very high, followed by transmission within the subadults. We confirmed the transmission estimates using the next-generation formalism to estimate R_0. The data produce quantitative transmission terms that can be used to describe roles of each stage class in the PDV outbreak; these findings can best be supported with harbor seal behavioral studies.

Introduction

Heterogeneities in transmission have long been recognized for shaping dynamics of infectious diseases. Heterogeneities can, for example, change invasion criteria (Woolhouse *et al.* 1997; Newman 2005; Ferrari *et al.* 2006a) and enhance spatial spread through superspreading (Lloyd-Smith *et al.* 2005). Complex biological mechanisms usually underlie heterogeneities such as spatial geography, gender, age class, individual immunological differences in susceptibility or infectiousness, superspreading events, behavioral factors, genetic variation and other individual variation (Anderson & May 1991; Altizer *et al.* 2003; Perkins *et al.* 2003; Lloyd-Smith *et al.* 2005). Several different approaches have been developed to account for such heterogeneities. For example, recognition of discrete classes permit the use of a more refined transmission term, $\beta_{i,j}$, which captures the rate at which an infectious individual of class *j* will infect a susceptible individual of class *i* (Anderson & May 1991). The separate classes can encompass gender, age, stage, social, immunological, physiological, or behavioral differences. These detailed transmission rates are usually modeled using the "who acquires infection from whom" (WAIFW) matrix (Anderson & May 1984; Schenzle 1984; Anderson & May 1985, 1991; Dobson 2004).

While the WAIFW matrix has proven to be of great theoretical utility, (Schenzle 1984; Anderson & May 1985; Dobson 2004; Kanaan & Farrington 2005) empirical approaches to estimation and characterization have often proven difficult because of lack of relevant data. Various efforts have employed contact tracing (Ferguson *et al.* 2001; Fraser *et al.* 2004) or inferred contacts (Edmunds *et al.* 1997; Huang & Rohani 2006; Wallinga *et al.* 2006) to determine how a virus is transmitted through different classes in a population. In this paper, we investigate whether it is possible to estimate elements in the WAIFW matrix from detailed age or stage incidence data. We ask three

nested questions: can we test the null hypothesis of homogenous mixing from such data? How well can we identify the WAIFW elements from such data in the absence of more detailed contact tracing? Finally, can we compare our estimation method with the theoretical next-generation formalism to estimate R_0 in structured populations?

To address these questions, we investigate a case study of phocine distemper virus (PDV) in harbor seals (*Phoca vitulina*) in the Netherlands during the 2002 epidemic (Rijks *et al.* 2005). PDV is a single stranded, negative sense RNA virus which is a member of the Morbillivirus genus, family *Paramyxoviridae* (Cosby *et al.* 1988; Mahy *et al.* 1988a; Osterhaus & Vedder 1988). In each individual, disease typically spans a two-week period, including both the latent and infectious disease stages (Osterhaus 1989; Harder 1990a; Baker 1992; Grenfell *et al.* 1992). Mortality is high (Rijks *et al.* 2005), partly due to co-infections.

Two outbreaks of PDV have affected seal populations throughout the entire North Sea region: the first outbreak occurred in 1988, in which 18,000 to 23,000 harbor seals died (Hall *et al.* 2006; Harkonen *et al.* 2006). This mass mortality event caused by the viral epidemic began on the Danish island of Ånholt on April 12, 1988 and ended within the calendar year (Dietz *et al.* 1989; Hall *et al.* 2006; Harkonen *et al.* 2006). A second PDV outbreak occurred in 2002 with the same point of origin: initial cases of harbor seal stranding and mortality occurred on May 4, 2002. In this epidemic, approximately 22,000 to 30,000 harbor seals died, resulting in the largest recorded mass mortality event in marine mammals (Jensen *et al.* 2002; Hall *et al.* 2006; Harkonen *et al.* 2006).

In the Netherlands, the first local case of PDV was found on June 16, 2002 on Vlieland and the local epidemic ceased at the end of November, as fully described in Rijks *et al.* (2005). In that time period, 2,284 seals were stranded along the Dutch coast, including 2,279 harbor seals and 5 gray seals. Interestingly, the timing of stranded seals

showed age-specificity. Not only was the index case a member of the subadult stage class, but the median stranding date of all subadults was significantly earlier than the median stranding date of both juvenile and adults (Rijks *et al.* 2005). Together, the stranding data implicate stage-structured disease transmission and heterogeneous host mixing in the 2002 Dutch epidemic. Previous models describing the spread of PDV throughout the North Sea have assumed homogeneous mixing among different harbor seal age or stage classes (Grenfell *et al.* 1992; De Koeijer *et al.* 1998; Swinton 1998; Swinton *et al.* 1998). Subsequently, we inquire if it is possible to estimate the WAIFW matrix from the detailed incidence data and investigate the evidence for non-homogenous mixing.

Methods

Seals were classified into stages based on body length of stranded carcasses, since only some of the stranded seals were precisely aged by counting the cementum layers a canine tooth. The juvenile class contained female seals less than 90 cm and male seals less than 95 cm. Subadults included females with body lengths between 90 cm and 120 cm and males with body lengths between 95 cm and 130 cm. Lastly, the adult category contained female seals with body lengths greater than 120 cm and males with body lengths greater than 130 cm. Using these body length classifications, the juvenile class contained most of the pups of the year, while the subadult class included most of the 1 and 2 year-old females and 1 to 3 year-old males. Finally, the adult class included most of the females older than 2 years and males older than 3 years.

The epidemic dynamics were captured with a susceptible-infected-removed (SIR) model, dividing the population into three categories based on their epidemiological state. Susceptible individuals never experienced infection nor were exposed to the virus. Infected individuals harbor the virus and are able to convert susceptible individuals into

infected individuals. Lastly, removed individuals were previously infected and either recovered from the disease with conferred lifelong immunity or are removed from the system due to mortality. Our model is defined in discrete time, with each time-step equal to one day.

We combined the three stage classes – juvenile, subadult, and adult – and the three epidemic classes – susceptible, infected, and removed – to capture both the population stage structure and the epidemic dynamics. This resulted in a model with nine total categories (equation 1).

$$N = \begin{array}{c} \begin{matrix} S & I & R \end{matrix} \\ \begin{pmatrix} n_{1,1} & n_{1,2} & n_{1,3} \\ n_{2,1} & n_{2,2} & n_{2,3} \\ n_{3,1} & n_{3,2} & n_{3,3} \end{pmatrix} \end{array} \begin{array}{l} \textit{juveniles} \\ \textit{subadults} \\ \textit{adults} \end{array} \qquad (1)$$

Since the lifespan of harbor seals is much longer than the duration of the PDV outbreak, we assumed that population size did not change during the outbreak – except for deaths due to infection – and fixed the initial population size (N) at 5400 based on population censuses (Ries *et al.* 1998; Trilaterial Seal Expert Group 2001). For each element $n_{i,j}$ in the matrix, the subscript i designates the stage-structure: the number 1 represents juveniles, the number 2 represents subadults, and the number 3 represents adults. Similarly, for each element $n_{i,j}$ in the matrix, the subscript j designates the epidemic category: the number 1 represents susceptible individuals, the number 2 represents infectious individuals, and the number 3 represents removed harbor seals.

The matrix N (equation 1) was reorganized into a population vector by stacking the rows of the matrix. The population vector then designates all juveniles, all subadults, and all adults. Within each stage class, epidemic categories are designated as susceptible individuals, infected individuals, and removed individuals as in equation 2, where ‘ specifies a vector transpose.

$$\mathbf{n} = \left(n_{1,1} \quad n_{1,2} \quad n_{1,3} \mid n_{2,1} \quad n_{2,2} \quad n_{2,3} \mid n_{3,1} \quad n_{3,2} \quad n_{3,3}\right)' \qquad (2)$$

Let $\beta_{i,j}$ be the probability per unit time of disease transmission between a susceptible individual in demographic class *i* and an infected individual in demographic class *j*. The transmission, or WAIFW matrix, is then a 3 by 3 matrix, $\beta = \left[\beta_{i,j}\right]$. We considered three different transmission scenarios and built three different models. In the first model, we considered homogeneous mixing among different stage classes and equal transmission rates ($\beta_{i,j} = \beta$ for all *i,j*). For the second model, we assumed weak heterogeneous mixing among the host population: within-stage transmission – the diagonal of the transmission matrix – was allowed to differ from between-stage transmission, which was designated by the off-diagonal elements. This difference was scaled by a coefficient, *k.* The transmission matrix is then:

$$\beta = \begin{pmatrix} k\beta_b & \beta_b & \beta_b \\ \beta_b & k\beta_b & \beta_b \\ \beta_b & \beta_b & k\beta_b \end{pmatrix} \qquad (3)$$

For the third model, we assumed strong heterogeneous mixing between stages by allowing the transmission term to vary with the interactions within and between stages (equation 4).

$$\beta = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix} \qquad (4)$$

The transmission matrix (equation 4) is symmetrical: $\beta_{i,j} = \beta_{j,i}$. This property captures an assumption of the model: each contact event between two hosts results in a bi-directional transmission process, where the probability of the two hosts infecting each other is equal. Moreover, this assumes that differences in susceptibility or infectiousness among stage classes do not significantly impact transmission dynamics (Anderson & May 1991).

Transmission probabilities were derived from the SIR model. The force of infection, ϕ, is the probability per unit time for a susceptible to become infected (Diekmann & Heesterbeek 2000). The recovery rate, γ, is the inverse of the average latent plus infectious periods (γ = 1/14). The transitions between epidemic categories within a stage class is then:

$$\mathbf{A}_i = \begin{pmatrix} 1-\phi(\mathbf{n}(t)) & 0 & 0 \\ \phi(\mathbf{n}(t)) & 1-\gamma & 0 \\ 0 & \gamma & 1 \end{pmatrix} \text{ where } \phi(\mathbf{n}(t)) = 1-\exp(-\sum_j \beta_{i,j} n_{j,2}(t)) \quad (5)$$

Epidemic transitions for all three stage classes were given by the transition matrix $\mathbf{A(n)}$, in which the block diagonal matrices $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ designate the epidemic transitions among juveniles, subadults and adults, respectively (equation 6).

$$\mathbf{A(n)} = \left(\begin{array}{c|c|c} \mathbf{A}_1 & 0 & 0 \\ \hline 0 & \mathbf{A}_2 & 0 \\ \hline 0 & 0 & \mathbf{A}_3 \end{array}\right) \quad (6)$$

Since we assumed that the epidemic dynamics are fast relative to demography, there are no transitions between stage classes, and the remaining elements of the block-

diagonal matrix $\mathbf{A(n)}$ are all zero. The epidemic trajectory is given by multiplying the population vector at time *t*, $\mathbf{n}(t)$, with the transition matrix $\mathbf{A(n)}$:

$$\mathbf{n}(t+1) = \mathbf{A}[\mathbf{n}(t)]\mathbf{n}(t) \tag{7}$$

This model (equation 7) is a discrete time approximation to the continuous time model:

$$\frac{d}{dt}\mathbf{n} = \left(\begin{array}{c|c|c} \mathbf{A}_1 & 0 & 0 \\ \hline 0 & \mathbf{A}_2 & 0 \\ \hline 0 & 0 & \mathbf{A}_3 \end{array}\right)\mathbf{n} \quad \text{where } \mathbf{A}_i = \begin{pmatrix} -\sum_j \beta_{i,j} n_{j,2} & 0 & 0 \\ \sum_j \beta_{i,j} n_{j,2} & -\gamma_i & 0 \\ 0 & \gamma_i & 0 \end{pmatrix} \tag{8}$$

Because of the underlying stage-structure there are three infected classes in this model. For models with multiple classes, R_0 can be derived using the next generation method (Diekmann *et al.* 1990; de Jong *et al.* 1994; Diekmann & Heesterbeek 2000; van den Driessche & Watmough 2002), where R_0 is given by the spectral radius, ρ, or the dominant eigenvalue of the next generation matrix, FV^{-1}:

$$R_0 = \rho\left[FV^{-1}\right] \tag{9}$$

To find the next generation matrix of a model with *s* compartments out of which *r* are infected, we let $n = n_i, \ldots, n_s$ be the number of individuals in each compartment; $F_i(n)$ be the rate at which newly infected individuals enter compartment *i*, let $\mathrm{V}_i^+(n)$ be the rate of entry of individuals into compartment *i* (including the transfer of infected individuals from one infective compartment to another), and $\mathrm{V}_i^-(n)$ be the rate at which individuals are

leaving compartment *i* . We defined V_i as $\mathrm{V}_i\ (n) = \mathrm{V}_i^{-}(n) - \mathrm{V}_i^{+}(n)$. The rate of change of compartment *i* is then $\frac{dn_i}{dt} = F_i - \mathrm{V}_i(n)$. We then formed the next generation matrix FV^{-1} by

$$F = \left\lfloor \frac{\partial F_i}{\partial n_j}(n_0) \right\rfloor \quad and \quad V = \left\lfloor \frac{\partial \mathrm{V}_i}{\partial n_j}(n_0) \right\rfloor \tag{10}$$

where *i,j = 1,…,r* and n_0 was the disease free equilibrium, at which the population remains in the absence of the disease (van den Driessche & Watmough 2002). The *(j,k)* entry of V^{-1} is the average amount of time an infective individual that was introduced into compartment *k* spends in compartment *j* during its lifetime. The *(i,j)* entry of *F* is the rate at which infected individuals in compartment *j* produce new infections in compartment *i*. Therefore, the entry *(i,k)* in the generation matrix FV^{-1} is the expected number of new infections in compartment *i* produced by an individual originally introduced into compartment *k*.

The matrix *F* shows the influx of new infections to the infectious compartments. Since we assumed that there are no transitions between the infectious classes due to growth during this acute PDV outbreak, matrix*V* reflects the rates at which individuals are leaving the infectious compartments due to recovery or death. At the disease free equilibrium the population consists wholly of susceptible individuals, so that

$$\mathbf{n_0} = [n_{11}(0) \quad 0 \quad 0 \quad n_{21}(0) \quad 0 \quad 0 \quad n_{31}(0) \quad 0 \quad 0]' \tag{11}$$

where ‘ designates a vector transpose.

F and *V* were constructed as follows:

$$F = \begin{pmatrix} \beta_{11}n_{11}(0) & \beta_{12}n_{11}(0) & \beta_{13}n_{11}(0) \\ \beta_{21}n_{21}(0) & \beta_{22}n_{21}(0) & \beta_{23}n_{21}(0) \\ \beta_{31}n_{31}(0) & \beta_{32}n_{21}(0) & \beta_{33}n_{31}(0) \end{pmatrix} \text{and } V = \begin{pmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \end{pmatrix} \tag{12}$$

The next generation matrix is thus

$$FV^{-1} = \begin{pmatrix} \frac{\beta_{11} n_{11}(0)}{\gamma} & \frac{\beta_{12} n_{11}(0)}{\gamma} & \frac{\beta_{13} n_{11}(0)}{\gamma} \\ \frac{\beta_{21} n_{21}(0)}{\gamma} & \frac{\beta_{22} n_{21}(0)}{\gamma} & \frac{\beta_{23} n_{21}(0)}{\gamma} \\ \frac{\beta_{31} n_{31}(0)}{\gamma} & \frac{\beta_{32} n_{21}(0)}{\gamma} & \frac{\beta_{33} n_{31}(0)}{\gamma} \end{pmatrix} \quad (13)$$

R_0 is given by the dominant eigenvalue of the next generation matrix (equation 13). We determined R_0 using equation 13 and our estimate of β.

Initial model conditions for the total population size (*N*), the infectious period, gamma – the inverse of the infectious period – population stage structure, and the length of the epidemic were derived from the literature (Table 2-1).

Table 2-1. Initial Model Conditions.

Parameter	**Value**	**Reference**
N (total population size)	5400	Ries *et al.* 1998; Trilateral Seal Expert Group 2001
Infectious Period	14 days	Swinton *et al.* 1998
Gamma (inverse of the infectious period)	1/14	Swinton *et al.* 1998
Population Stage Structure	15% juveniles 36% subadults 49% adults	(Abt 2002)
Length of Epidemic	180 days	Rijks *et al.* 2002

Initial conditions for the PDV SIR model, including population size (*N*), infectious period, gamma, population stage structure, and the length of the epidemic were obtained from the literature. Total population size (*N)* was estimated for the entire Dutch harbor seal population, and is approximately 40% greater than minimum numbers obtained in recent censuses.

The three models – homogeneous mixing, weak heterogeneous mixing, and strong heterogeneous mixing – were compared using the likelihood ratio test (LRT). Subsequently, p-values were calculated to determine which model best fits the data.

To obtain estimates for the WAIFW matrix, we used maximum likelihood techniques to find the values of the matrix elements which best fit the stage-specific incidence data (Rijks, 2005) and the probability of observing a stranded seal, *p* (equation 14). The probability of observation is a compound variable encompassing both the probability that a given seal, once infected by PDV, will strand and that the stranded seal will be encountered and observed. We assume Poisson likelihoods for disease incidence.

$$y \sim Po(pI_t) \tag{14}$$

Data for this model consisted of stranded seals from the Dutch islands of Vlieland, Terschelling, Ameland, Schiermonnikoog, and Texel, and from the mainland provinces of Friesland, Groningen, and Noord Holland. Point estimates were located by minimizing the negative log-likelihood of the data using simulated annealing (Belisle 1992) as implemented by the 'optim' function in R (R Development Core Team 2006). Strong colinearlity between elements in the WAIFW matrix led to a range of near equally likely results, since values for elements of the WAIFW matrix can compensate for each other to produce the same results in disease incidence. This colinearity was investigated by inverting the numerical Hessian matrix and converting variance-covariance matrix into the corresponding correlation matrix (McCullagh & Nelder 1989). If the pair-wise correlation coefficient between two elements was greater than 0.3, we used two-dimensional profile likelihoods to map plausible pair-wise parameter combinations including their two-dimensional 95% confidence intervals. We tested the significant deviation from homogeneous and weak heterogeneous mixing using likelihood ratio tests (LTRs).

All model building and parameter estimations were performed using R version 2.3.1 (R Development Core Team 2006). The next-generation estimates of R_0 were performed using Mathematica version 6 (Wolfram Research 2007).

Results

The three models created were compared using the likelihood ratio test (LRT). The first model, homogeneous mixing with uniform β, implies complete lack of stage-structure in the population. Results from the model selection tests (Table 2-2) show that the model with slight heterogeneous mixing has a better fit to the data than the model with homogeneous mixing (p=0.001). The strong heterogeneous mixing model fit the data better still (p=0.0004) (Table 2-2). When comparing the set of the three nested models, the best fit model overall was the model with the strong heterogeneous mixing which permitted unique within and between stage interactions (Table 2-2).

Table 2-2. Model Selection using the Likelihood Ratio Test (LRT)

Models	Likelihood Ratio	Degrees of Freedom	P-value	Better-fit Model
Homogeneous mixing Slight heterogeneous mixing	5.28	1	0.00116	Slight heterogeneous mixing
Slight heterogeneous mixing Strong heterogeneous mixing	10.14	4	0.0004385	Strong heterogeneous mixing

Three models, used to describe the spread of PDV in the Dutch Wadden Sea, were compared to see which best fit the data. These models include: homogeneous mixing with uniform β, slight heterogeneous mixing with β_b and $k\beta_b$, and the full stage-structure model with the symmetrical β matrix. The likelihood ratio, degrees of freedom, and p-values are listed for each pair-wise model comparison. Overall, the model incorporating the greatest degree of stage structure and symmetrical β matrix was the best-fit model to the data.

Using the model with strong heterogeneous mixing, chosen by the model selection test, we estimated point values for each of the elements in the WAIFW matrix according to the maximum likelihood estimates (Table 2-3).

Table 2-3. Parameter Point Estimates in the Strong Heterogeneous Mixing Model from One-Dimensional Likelihoods

Parameter	Demographic Classes	Value
$\beta_{1,1}$	Juvenile – Juvenile	5.56×10^{-5}
$\beta_{2,2}$	Subadult – Subadult	3.74×10^{-4}
$\beta_{3,3}$	Adult – Adult	8.08×10^{-5}
$\beta_{1,2} = \beta_{2,1}$	Juvenile – Subadult	9.09×10^{-5}
$\beta_{1,3} = \beta_{3,1}$	Juvenile – Adult	5.05×10^{-6}
$\beta_{2,3} = \beta_{3,2}$	Subadult – Adult	2.22×10^{-4}

Point estimates for each element in the symmetrical β matrix in the model incorporating full stage structure were obtained by maximum likelihood methods. The intra-stage transmission terms are designated by β_{11} for juveniles, β_{22} for subadults, and β_{33} for adults. Inter-stage transmission is symmetrical. Interactions between juveniles and subadults are designated by $\beta_{12} = \beta_{21}$, while transmission between juveniles and adults are designated by $\beta_{13} = \beta_{31}$. Finally, the transmission between subadults and adults is designated by $\beta_{23} = \beta_{32}$.

Among the juvenile stage, transmission with subadults comprised the greatest component of disease incidence ($\beta_{12} = 9.09 \times 10^{-5}$), closely followed by transmission within the juvenile stage ($\beta_{11} = 5.56 \times 10^{-5}$) and transmission between juveniles and adults ($\beta_{13} = 5.05 \times 10^{-6}$) (Table 2-3). Intra-stage transmission (β_{11}) and transmission between juveniles and subadults (β_{12}) provided clear maximum likelihoods with targeted values; the point estimate for transmission between juveniles and adults (β_{13}) was less obvious since it lies within a range of equally likely results. Subadults, in contrast, showed the greatest interaction with members of its own class ($\beta_{22} = 3.74 \times 10^{-4}$), followed by interactions with adults ($\beta_{23} = 2.22 \times 10^{-4}$) and then juveniles ($\beta_{21} = 5.56 \times 10^{-5}$) (Table 2-3). Again, intra-stage transmission (β_{22}) and transmission between subadults and juveniles (β_{21}) provided clear estimates; the point estimate for transmission between subadults and adults (β_{23}) was more obscure due to the fact that it lies within a range of results with similar likelihoods. Lastly, adult transmission was greatest with subadults

(β_{32} = 2.22 x 10^{-4}) and decreased with both adults (β_{33} = 8.08 x 10^{-5}) and juveniles (β_{31} = 5.05 x 10^{-6}) (Table 2-3). The transmission parameters for adults demonstrated the greatest degree of uncertainty among the three stages: each adult transmission term falls within a range of equally likely values.

The ambiguity in point estimates reflects the high degree of colinearity among the stage structured parameter point estimates in Table 2-3. In other words, the likelihood landscape is rugged: there are different combinations of parameter values that could result in equally fit likelihood values. Correlations between parameters, calculated from the unconstrained optimization of all parameters in the full stage-structure model, are shown in Table 2-4.

Table 2-4. Correlation Coefficients in the Full Stage Structure Model from the Unconstrained Optimization of All Parameters

	$\beta_{1,1}$	$\beta_{2,2}$	$\beta_{3,3}$	$\beta_{1,2} = \beta_{2,1}$	$\beta_{1,3} = \beta_{3,1}$	$\beta_{2,3} = \beta_{3,2}$
$\beta_{1,1}$	1.00	0.0201	0.0158	**- 0.842**	-0.0239	-0.00847
$\beta_{2,2}$	0.0201	1.00	-0.00818	**-0.506**	0.00953	-0.00649
$\beta_{3,3}$	0.0158	-0.00818	1.00	0.0484	**-0.853**	**-0.312**
$\beta_{1,2} = \beta_{2,1}$	**- 0.842**	**-0.506**	0.0484	1.00	-0.0827	-0.0337
$\beta_{1,3} = \beta_{3,1}$	-0.0239	0.00953	**-0.853**	-0.0827	1.00	0.0689
$\beta_{2,3} = \beta_{3,2}$	-0.00847	-0.00649	**-0.312**	-0.0337	0.0689	1.00

Hessian matrix and corresponding correlation matrix were calculated from the unconstrained optimization of all parameters in the model incorporating full stage structure. Both negative and positive correlations between the stage-specific transmission terms are identified. Correlations greater than 0.3 are highlighted. The intra-stage transmission terms are designated by β_{11} for juveniles, β_{22} for subadults, and β_{33} for adults. Inter-stage transmission is symmetrical. Interactions between juveniles and subadults are designated by $\beta_{12} = \beta_{21}$, while transmission between juveniles and adults are designated by $\beta_{13} = \beta_{31}$. Finally, the transmission between subadults and adults is designated by $\beta_{23} = \beta_{32}$.

Two variables are considered to be correlated if their correlation coefficient was greater than 0.3. Under this criteria, four sets of variables are highly correlated: intra-stage

juvenile transmission (β_{11}) and transmission between juveniles and subadults (β_{12} = β_{21}), intra-stage subadult transmission (β_{22}) and transmission between juveniles and subadults (β_{12} = β_{21}), intra-stage adult transmission (β_{33}) and transmission between juveniles and adults (β_{13} = β_{31}), and finally, intra-stage adult transmission (β_{33}) and transmission between subadults and adults (β_{23} = β_{32}) (Table 2-4).

Due to this high degree of colinearity in transmission parameters, as seen in the correlation coefficients, the maximum likelihood estimations result in ridges in likelihood space. Pair-wise estimates of the parameters permit likelihood contour maps to be drawn, where the contour lines show likelihood test statistics as in the sample contour map of the estimation of β_{13} = β_{31} and β_{33} in the full stage-structure model, where the maximum likelihood estimate is designated by a point and the bounds of the 95% confidence interval around the maximum likelihood estimate are shown in red (Figure 2-1).

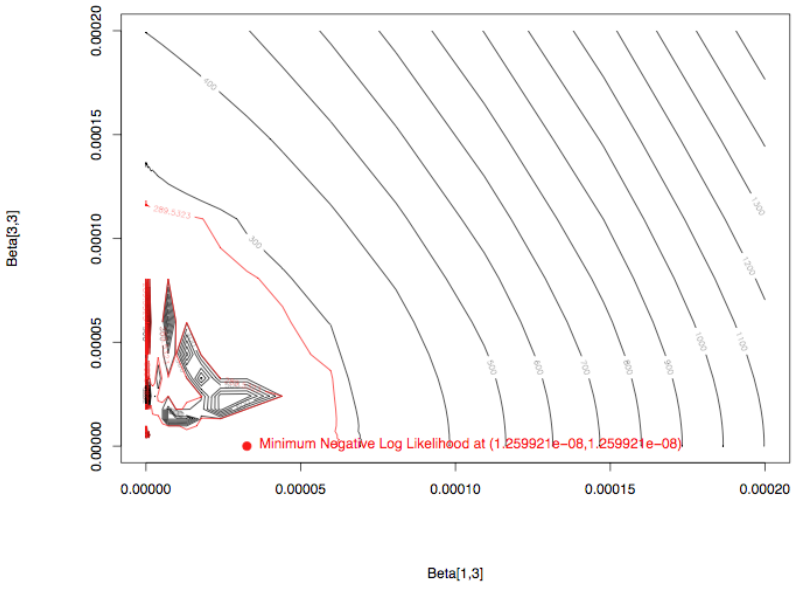

Figure 2-1. Likelihood Surface of the Pair-wise Estimation of $\beta_{1,3}$ = $\beta_{3,1}$ and $\beta_{3,3}$ in the Full Stage Structured Model. Pair-wise estimations for all elements in the symmetrical β matrix for the model incorporating full stage structure were obtained by maximum likelihood methods. The likelihoods for each pair-wise combination were plotted as a contour map, with the point estimates designated by a red point and 95% confidence intervals designated by red lines. A sample contour map is shown, with diagonal contours that highlight the unidentifiability of the two parameters graphed.

The contour maps (Figure 2-1) confirm the location of point estimates (Table 2-3) within valleys in likelihood space specifying nearly equally likely results.

The basic reproductive ratio, as given by the dominant eigenvalue of the next generation matrix (equation 12) using transmission estimates from the unconstrained simultaneous optimization of all parameters, gave an estimate of R_0 = 3.08, which falls directly in the range of other R_0 estimates for PDV (Swinton *et al.* 1998).

Discussion

Age or stage structured behavior dictates transmission of many diseases (Anderson & May 1982, 1984; Schenzle 1984; Anderson & May 1985) and evidence points to stage structured PDV transmission in harbor seals. For example, harbor seals show characteristics of discriminate interactions based on stage class (Wilson 1974; Sullivan 1982; Renouf & Lawson 1986, 1987; Godsell 1988b; Thompson *et al.* 1989; Traut 1999b) and PDV incidence has previously been shown to have signatures of stage dependence in the Dutch 2002 outbreak (Rijks *et al.* 2005). In this paper we develop a theoretical framework to incorporate stage structure in PDV epidemic SIR models. By creating and ranking three nested models ranging from a complete lack of stage structure with homogeneous mixing to strong heterogeneous mixing, we showed that added stage structure provided a better description of the data (Table 2-2). Overall, the model with strong heterogeneities was the best-fit model (Table 2-2), indicating that the

harbor seal population in the Dutch Wadden Sea transmitted PDV in a stage dependent manner during the 2002 epidemic.

Using the full stage structure model with the symmetrical β matrix, we were able to determine elements of the WAIFW matrix from incidence data alone (Table 2-3), illuminating both mechanisms of epidemic spread and harbor seal contact structure. The highest transmission occurs among the subadult class and between subadults and adults, while the lowest transmission occurs between juveniles and adults. Other combinations of stage classes show intermediate levels of transmission. We were able to verify our estimates by calculating R_0 using the next-generation matrix, which resulted in a value that is acceptable for PDV ($R_0 = 3.08$).

Among all of the classes for which elements of the WAIFW matrix have been defined, subadults emerge as an interesting and vital category. Since the index case and most early cases were in the subadult category (Rijks *et al.* 2005), it logically follows that subadult-subadult transmission would be high since there are many behavioral contacts that may be epidemiologically relevant.

However, the elements in the WAIFW matrix are highly correlated, so it is difficult to unambiguously estimate certain pairs of parameters (Table 2-4). For example, transmission between juveniles and subadults are highly correlated with both intra-stage juvenile transmission and intra-stage subadult transmission. This makes it difficult to distinguish between an intra-stage transmission chain and a “pathogen rain” that infects the juvenile class from many other stages. Another reason for the correlation may stem from the fact that the stage classes were based on a proxy for age, namely sex and body length. Exact age determination may have lessened the correlations. Nevertheless, correlation between elements of the WAIFW matrix set limits to the indentifiability of the matrix. Encouragingly, though, likelihood ratio tests appear to have an ability to distinguish heterogeneous from homogeneous mixing in stage-structured data.

The WAIFW matrix is a catalog of "who acquires infection from whom;" therefore, it is a reflection of transmission events and not a direct reflection of the contact process among healthy or immune seals. For instance, the stage class which introduces the disease into the population appears to influence the strength of the transmission coefficients. Many early cases in the 2002 Dutch PDV epidemic were subadults (Rijks *et al.* 2005), which are reflected in the high transmission between juveniles and subadults, among subadults, and between adults and subadults (Table 2-3). In addition, seals may have altered their behavior due to disease. Nevertheless, the resulting transmission rates also seem to a mirror harbor seal behavior under normal conditions. For example, from point estimates, juvenile transmission is greatest with subadults and decreases with other juveniles and adults (Table 2-3). The juveniles stranded in September, and therefore generally more than two weeks after weaning. At this time they are no longer in relative isolation with their mothers (Newby 1973; Godsell 1988b; Traut 1999b).

Although many insights on PDV transmission between harbor seal stage classes in the Dutch Wadden Sea have been gained by the incidence data and resulting WAIFW matrix, the ambiguity in the estimates highlights the need for additional behavioral data. To make the theoretical results relevant, information about the number and type of intra-stage and inter-stage interactions must be determined. For example, specific contact measurements between and among stage-classes could be determined based on the distance between individuals from observations or aerial photographs. In particular, focusing on the behavior of the subadult stage, since the index case fell into this category, would be the most helpful. These additional behavioral data will constrain the degree of uncertainty in the parameter estimates to biologically reasonable and relevant ranges.

In conclusion, elements in the WAIFW matrix that provide information about transmission dynamics within and between stage classes of harbor seals can be

estimated based on stage incidence data alone. Combining our statistical methodology with the next-generation formalism further allows us to estimate R_0. However, identifiably and uncertainty issues still exist within these parameter estimates, highlighting the need for additional behavioral data to restrict the ranges of the theoretical parameter estimations to biologically plausible and realistic values. Stage-structure clearly plays an important role in the dynamics of this epidemic.

Acknowledgements

The authors would like to thank Bryan Grenfell for his assistance with both theory and analysis, Michael Neubert and Hal Caswell for help with the model formulation, and Angela Luis for helpful discussions. In addition, the authors wish to thank the volunteers who covered the Dutch coast daily in search of stranded seals; the people working at the Seal Research and Rehabilitation Center in Pieterburen for logistical support; V.O.P. Containers for providing the location to necropsy the seals; the staff of the Dutch Ministry of Agriculture, Nature and Food Quality (LNV-Noord) for providing access to and help with the centralized seal registration data; the Common Wadden Sea Secretariat and the Trilateral Seal Expert Group for international coordination of the outbreak. Laura Pomeroy was supported by the National Science Foundation, under the NSF Graduate Teaching Fellowship in K-12 Education (DGE-0338240).

References

Abt K.F. (2002) Phanologie und Populationskynamik des Seehundes (*Phoca vitulina*) im Wattenmeer: Grundlangen zur Messung von Statusparametern In. Christian-Albrechts-Universitat zu Kiel, Kiel, Germany

Altizer S., Nunn C.L., Thrall P.H., Gittleman J.L., Antonovics J., Cunningham A.A., Dobson A.P., Ezenwa V., Jones K.E., Pedersen A.B., Poss M. & Pulliam J.R.C. (2003) Social organization and parasite risk in mammals: Integrating theory and empirical studies. *Annual Review of Ecology Evolution and Systematics*, 34, 517-547

Anderson R.M. & May R.M. (1984) Spatial, Temporal, and Genetic Heterogeneity in Host Populations and the Design of Immunization Programs. *Mathematical Medicine and Biology*, 1, 233-266

Anderson R.M. & May R.M. (1985) Age-Related Changes in the Rate of Disease Transmission: Implications for the Design of Vaccine Programmes. *Journal of Hygiene (Cambridge)*, 94, 365-436

Anderson R.M. & May R.M. (1991) *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.

Baker J.R. (1992) The Pathology of Phocine Distemper. *The Science of the Total Environment*, 115, 1-7

Belisle C.J.P. (1992) Convergence Theorems for a Class of Simulated Annealing Algorithms. *Journal of Applied Probability*, 29, 855-895

Cosby S.L., McQuaid S., Duffy N., Lyons C., Rima B.K., Allan G.M., McCullough S.J., Kennedy S., Smyth J.A., McNeilly F., Craig C. & Orvell C. (1988) Characterization of a Seal Morbillivirus. *Nature*, 336, 115-116

de Jong M., Diekmann O. & Heesterbeek J. (1994) The Computation of R0 for Discrete-Time Epidemic Models with Dynamic Heterogeneity. *Mathematical Biosciences*, 97-114

De Koeijer A., Diekmann O. & Reijnders P. (1998) Modelling the spread of phocine distemper virus among harbour seals. *Bulletin of Mathematical Biology*, 60, 585-596

Diekmann O. & Heesterbeek J.A.P. (2000) *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation*. Wiley, New York, NY.

Diekmann O., Heesterbeek J.A.P. & Metz J.A.J. (1990) On the Definition and the Computation of the Basic Reproduction Ratio R0 in Models for Infectious Diseases in Heterogeneous Populations. *Journal of Mathematical Biology*, 28, 365-382

Dietz R., Heide-Jorgensen M.-P. & Harkonen T. (1989) Mass Deaths of Harbor Seals *(Phoca Vitulina)* in Europe. *Ambio*, 18, 258-264

Dobson A. (2004) Population dynamics of pathogens with multiple host species. *American Naturalist*, 164, S64-S78

Edmunds W.J., O'Callaghan C.J. & Nokes D.J. (1997) Who Mixes with Whom? A Method to Determine the Contact Patterns of Adults that May Lead to the Spread of Airborne Infections. *Proceedings of the Royal Society B*, 264, 949-957

Ferguson N.M., Donnelly C.A. & Anderson R.M. (2001) The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science*, 292, 1155-1160

Ferrari M.J., Bansal S., Meyers L.A. & Bjornstad O.N. (2006) Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society B*, 273, 2743-2748

Fraser C., Riley S., Anderson R.M. & Ferguson N.M. (2004) Factors That Make a Disease Outbreak Controllable. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 6146-6151

Godsell J. (1988) Herd Formation and Haul-out Behavior in Harbor Seals *(Phoca vitulna)*. *Journal of Zoology, London*, 215, 83-98

Grenfell B.T., Lonergan M.E. & Harwood J. (1992) Quantitative Investigations of the Epidemiology of Phocine Distemper Virus (PDV) in European Common Seal Populations. *Science of the Total Environment*, 115, 15-29

Hall A.J., Jepson P.D., Goodman S.J. & Harkonen T. (2006) Phocine distemper virus in the North and European Seas - Data and models, nature and nurture. *Biological Conservation*, 131, 221-229

Harder T., Willhaus, T.H., Frey, H.-R., and Liess, B. (1990) Morbillivirus Infections of Seals during the 1988 Epidemic in the Bay o Heligoland: III. Transmission Studies of Cell Culture-Propagated Phocine Distemper Virus in Harbor Seals (Phoca vitulina) and a Grey Seal (Halichoerus grypus): Clinical, Virological and Serological Results. *Journal of Veterinary Medicine*, 37, 641-650

Harkonen L., Dietz R., Reijnders P., Teilmann J., Harding K., Hall A., Brasseur S., Siebert U., Goodman S.J., Jepson P.D., Rasmussen T.D. & Thompson P. (2006) A Review of the 1988 and 2002 phocine distemper virus epidemics in European harbour seals. *Diseases of Aquatic Organisms*, 68, 115-130

Huang Y. & Rohani P. (2006) Age-Structured Effects and Disease Interference in Childhood Infections. *Proceedings of the Royal Society B*, 273, 1229-1237

Jensen T., van de Bildt M., Dietz H.H., Andersen T.H., Hammer A.S., Kuiken T. & Osterhaus A. (2002) Another phocine distemper outbreak in Europe. *Science*, 297, 209-209

Kanaan M.N. & Farrington C.P. (2005) Matrix Models for Childhood Infections: A Bayesian Approach with Applications to Rubella and Mumps. *Epidemiology and Infection*, 133, 1009-1021

Lloyd-Smith J.O., Schreiber S.J., Kopp P.E. & Getz W.M. (2005) Superspreading and the effect of individual variation on disease emergence. *Nature*, 438, 355-359

Mahy B.W.J., Barrett T., Evans S., Anderson E.C. & Bostock C.J. (1988) Characterization of a Seal Morbillivirus. *Nature*, 336, 115

McCullagh P. & Nelder J.A. (1989) *Generalized linear models*. 511 edn. Chapman and Hall, New York, New York.

Newby T.C. (1973) Observations on the breeding behavior of the Harbor Seal in the State of Washington. *Journal of Mammology*, 54, 540-543

Newman M.E.J. (2005) Threshold effects for two pathogens spreading on a network. *Phys. Rev. Lett.*, 95, 108701

Osterhaus A.D., Uytdehaag, F.G., Visser, I.K., Vedder, E.J., Reijnders, P.J., Kuiper, J., and Brugge, H.N. (1989) Seal Vaccination Success. *Nature*, 337, 21

Osterhaus A.D.M.E. & Vedder E.J. (1988) Identification of Virus Causing Recent Seal Deaths. *Nature*, 355, 20

Perkins S.E., Cattadori I.M., Tagliapietra V., Rizzoli A.P. & Hudson P.J. (2003) Empirical evidence for key hosts in persistence of a tick-borne disease. *International Journal for Parasitology*, 33, 909-917

R Development Core Team (2006) R: A Language and Environment for Statistical Computing. In: (ed. Computing RFfS), Vienna, Austria

Renouf D. & Lawson J.W. (1987) Quantitative Aspects of Harbour Seals *(Phoca vitulina)* Play. *Journal of Zoology, London*, 212

Ries E.H., Hiby L.R. & Reijnders P.J.H. (1998) Maximum Likelihood Population Size Estimation of Harbour Seals in the Dutch Wadden Sea Based on a Mark Recapture Experiment. *Journal of Applied Ecology*, 35, 332-339

Rijks J.M., Van de Bildt M.W.G., Jensen T., Philippa J.D.W., Osterhaus A. & Kuiken T. (2005) Phocine distemper outbreak, the Netherlands, 2002. *Emerging Infectious Diseases*, 11, 1945-1948

Schenzle D. (1984) An Age-Structured Model of Pre- and Post-Vaccination Measles Transmission. *Mathematical Medicine and Biology*, 1, 169-191

Sullivan R.M. (1982) Antagonistic Behavior and Dominance Relationships in the Harbor Seals. *Journal of Mammology*, 63, 544-569

Swinton J. (1998) Extinction times and phase transitions for spatially structured closed epidemics. *Bulletin of Mathematical Biology*, 60, 215-230

Swinton J., Harwood J., Grenfell B.T. & Gilligan C.A. (1998) Persistence thresholds for phocine distemper virus infection in harbour seal Phoca vitulina metapopulations. *Journal of Animal Ecology*, 67, 54-68

Thompson P.M., Fedak M.A., McConnell B.J. & Nicholas K.S. (1989) Seasonal and Sex-Related Variation in the Activity Patterns of Common Seals *(Phoca vitulina)*. *Journal of Applied Ecology*, 26, 521-535

Traut I.M. (1999) Spacing among Harbor Seals *(Phoca vitulina vitulina)* on Haul-out Sites in the Wadden Sea of Niedersachsen. *Z. Sugetierkunde*, 64, 51-53

Trilateral Seal Expert Group (2001) Common Seals in the Wadden Sea in 2001. In: *Wadden Sea Newsletter 2001*, p. 3. Common Wadden Sea Secretariat

van den Driessche P. & Watmough J. (2002) Reproduction Numbers and Subthreshold Endemic Equilibria for Compartmental Models of Disease Transmission. *Mathematical Biosciences*, 180, 32-34

Wallinga J., Teunis P. & Kretzschmar M. (2006) Using Data on Social Contacts to Estimate Age-Specific Transmission Parameters for Respiratory-Spread Infectious Agents. *American Journal of Epidemiology*, 164, 936-944

Wilson S. (1974) Juvenile Play of the Common Seal, *Phoca vitulina vitulina*, with Comparative Notes on the Gray Seal *Halichoerus grupus*. *Behaviour*, 48, 37-60

Wolfram Research I. (2007) Mathematica. In, Champaign, IL

Woolhouse M.E.J., Dye C., Etard J.-F., Smith T., Charlwood J.D., Garnett G.P., Hagan P., Hii J.L.K., Ndhlovu P.D., Quinnel R.J., Watts C.H., Chandiwana S.K. & Anderson R.M. (1997) Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 338-342

A Bayesian Approach to Reconstructing Wildlife Epidemics From Imperfect Data: Phocine Distemper Virus as an Example

Laura W. Pomeroy[1,2], Matthew J. Ferrari[1,2], Catriona Stephenson[3,4], John Harwood[3,4], Karin Harding[5], Ottar N. Bjørnstad[1,2,6]

[1]Department of Biology, Pennsylvania State University, University Park, PA 16802 USA

[2]Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA 16802 USA

[3]Sea Mammal Research Unit, University of St. Andrews, St Andrews, Fife, KY16 9LZ UK

[4]Centre for Research into Ecological and Environmental Modelling, University of St. Andrews, St Andrews, Fife, KY16 9LZ UK

[5]Department of Marine Ecology, University of Gothenburg, SE-405 30 Gothenburg, Sweden

[6]Department of Entomology, Pennsylvania State University, University Park, PA 16802 USA

Abstract

Wildlife disease data possess bias, error, and uncertainty that pose an obstacle to analysis and further research. We propose a framework to address error in both process and observation through a state-space model. Using the 1988 and 2002 phocine distemper virus (PDV) outbreaks among harbor seals in the North Sea, we use a Bayesian framework to estimate the initial population of susceptible individuals (S_0), the rate of pathogen transmission (β), and the time series of infected individuals with imperfect binomial reporting to the biweekly incidence time series. The overall disease burden calculated in this framework is comparable with theorecial estimates. The Bayesian framework presented offers a method to address uncertainties inherent in many wildlife disease datasets.

Introduction

Both observation and process error contribute to uncertainty in ecological research (Harwood & Stokes 2003; Clark & Bjornstad 2004). One major source of uncertainty is errors in the data itself (Legendre 1992). In many cases, the data can be biased, indirect, or inconsistently sampled. Another source of uncertainty is the stochasticity inherent in the ecological process: variability is an intrinsic component of ecological population and disease dynamics. Oftentimes, reconstructing actual processes on the basis of observational uncertainty is a critical challenge to research progress. Here, we explore this issue using a statistical model that accounts for both aspects of uncertainty. As a case study, we investigate two phocine distemper virus (PDV) outbreaks that have recently occurred in the North Sea. Phocine distemper virus is a negative-sense, single stranded, newly emerged member of the genus Morbillivirus, family *Paramyxoviridae* (Cosby *et al.* 1988; Mahy *et al.* 1988b; Osterhaus & Vedder 1988). In the host, disease typically spans a two-week period, which includes both the latent and infectious stages (Osterhaus 1989; Harder 1990a; Baker 1992; Grenfell *et al.* 1992). Mortality is high (Rijks *et al.* 2005), partly due to co-infections and secondary complications (Baker & Ross 1992). Two outbreaks of PDV have affected seal populations throughout the entire North Sea: the first outbreak occurred in 1988, in which 18,000 to 23,000 harbor seals died (Hall *et al.* 2006; Harkonen *et al.* 2006). This mass mortality event caused by the viral epidemic began on the Danish island of Ånholt on April 12, 1988 and ended within the calendar year (Dietz *et al.* 1989; Hall *et al.* 2006; Harkonen *et al.* 2006). A second PDV outbreak occurred in 2002 with the same point of origin: initial cases of harbor seal stranding and mortality occurred on May 4, 2002. In this epidemic, approximately 22,000 to 30,000 harbor seals died, resulting in the largest recorded mass mortality event in marine mammals (Jensen *et al.* 2002; Hall *et al.* 2006; Harkonen *et al.* 2006). Both epidemics spanned the North Sea coastline: areas affected included Kattegat-Skagerrak coasts, the Wadden Sea, and the coast of the United Kingdom (Hall *et al.* 2006; Harkonen *et al.* 2006).

Wildlife epidemic data often possess a high degree of uncertainty, and the data from the PDV epidemics are no exception. Since the virgin epidemic in 1988 was caused by a newly emerging pathogen and caught the susceptible population of harbor seals – as well as the biologists, ecologists and naturalists that monitor the seals – by surprise, there was little infrastructure in place to monitor the mass stranding of harbor seals that occurred throughout Europe. Therefore, the data are likely to possess some degree of bias, inconsistency, and uncertainty.

As a precursor to a full model for this epidemic metapopulation (see Chapter 3), we need to quantify three crucial parameters: the size of the initial population of susceptible individuals (S_0), the rate of pathogen transmission (β), and the probability of observing an infected seal (P_{obs}). Ultimately, to understand the spatial spread we need to make further statistical inference on the most likely true epidemic trajectories at each location, in the presence of the observational uncertainty. Observational error is inevitable in wildlife disease data for many reasons, including underreporting, spatial and temporal error. Underreporting occurs for at least two reasons. First, although case fatality rates are high, a fraction of seals recover and will not be counted among the stranded seals. Second, some dead seals may be carried by currents and lost to sea (Rijks *et al.* 2005). Seals that strand and are recorded may contribute to spatial errors – if currents carried them away from their haulout of origin – and temporal errors – if they drifted at sea long enough so that their count is attributed to the wrong epidemic generation. Using a model that accounts for both data and process error, we estimate the three basic parameters and provide posterior densities for the time course of infection for each affected local population.

Methods

The epidemiological data consist of time-series counts of stranded harbor seals by geographical area for both the 1988 and 2002 epidemics. Although data are available from

throughout the North Sea, sampling regime varied greatly by country and by year. In the Kattegat-Skagerrak haulout sites, beaches were surveyed daily, twice weekly, or thrice weekly depending on the intensity and status of the epidemic. Stranded seals were classified by age and either tagged or removed from the beaches after recording (Heide-Jorgensen & Harkonen 1992). Wadden Sea strandings were reported to a telephone service and examined and categorized by veterinary staff (Rijks *et al.* 2005). In the United Kingdom, reports of stranded seals were made by the public and collated by one of a number of professional organizations (Thompson & Miller 1992).

Population sizes and trends are also monitored in all locations, but again, methods vary by location. In the Kattegat-Skagerrak sites, harbor seal populations have been monitored since the 1970s by annual aerial surveys (Heide-Jorgensen & Harkonen 1988; Harding *et al.* 2002; Harkonen *et al.* 2002). Similarly, the Wadden Sea harbor seal populations have been monitored by aerial surveys since the early 1970s (Ries *et al.* 1999). Harbor seals on the coast of the United Kingdom are monitored in regular aerial surveys or counts made by boat (Duck *et al.* 2007). In Northeast Scotland, counts are made weekly or bimonthly in June and July at low tide (Thompson & Miller 1992).

We estimated the geographical variability in the size of the initial population of susceptible individuals (S_0), the rate of pathogen transmission (β), and the time series of infected individuals with imperfect binomial reporting to the biweekly incidence time series for both the 1988 and 2002 epidemics using Bayesian Markov chain Monte Carlo methods (Morton & Finkenstädt 2005; Ferrari *et al.* 2008).

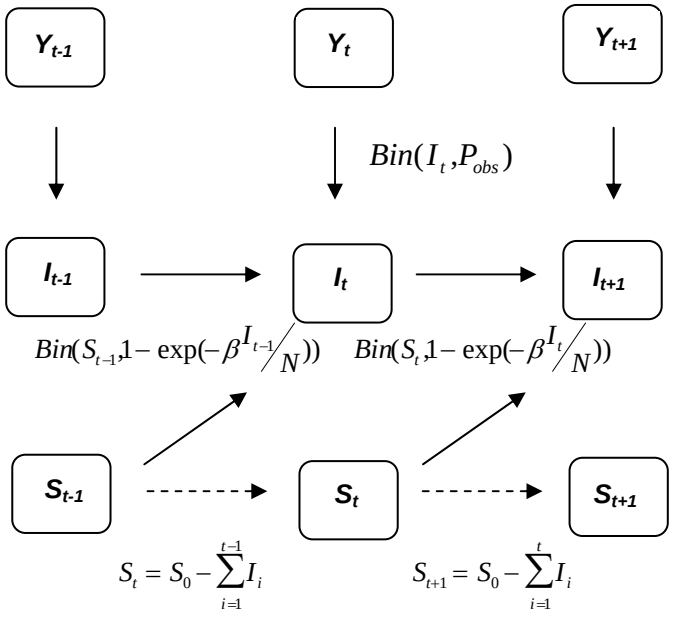

Figure 3-1. The relationships between the data model, process model, and estimated parameters. In the model, *Y* designates the data and is represented in the top row of the graph. This is the only observed state, or the seal stranding data. The bottom rows represent the unobserved time series of infected harbor seals, *I*, and susceptible harbor seals, *S*, which must be estimated. Solid lines reflect probabilistic (binomial) transitions, and dotted lines indicate deterministic transitions.

We model the time series of reported cases, *Y*, using a state-space model with unobserved epidemic states, *S* and *I*. Let Y_t be the reported number of cases summed over a two-week period. We assume this to be related to the true number of unobserved cases, I_t, through a binomial observation process with reporting rate P_{obs} according to

$$Y_t \sim Bin(I_t, P_{obs}) \quad (1)$$

where *Bin (a,b)* designates a binomial mass function where *a* is the number of trials and *b* is the success probability.

A susceptible seal – that has had no previous exposure to the virus – is assumed to have β potentially transmission-relevant infectious contacts per two-week time-step. Of these an

expected fraction I_t/N will be with infectious seals, where N is the total population size. Thus, the number of new infectious cases in time interval *t*, can assumed to be a time-varying binomial draw from the number of susceptible seals, S_t, according to the chain-binomial model (Bailey 1957; Ferrari *et al.* 2005):

$$I_{t+1} \sim Bin(S_t, 1 - e^{-\frac{\beta I_t}{N}}) \quad (2)$$

The transmission term β is a compound parameter encompassing the rate of contact among seals and the probability that a contact is sufficient for infection. As the contact rate may vary geographically with both haulout size and density, we estimated separate transmission terms for each haulout.

We assume that the epidemic process is fast relative to demography; therefore, the only change in the susceptible class occurs when a susceptible individual becomes infected. Thus the number of susceptible seals at time *t* is the number initially susceptible, S_0, less the cumulative number infected,

$$S_t = S_0 - \sum_{i=1}^{t-1} I_i . \quad (3)$$

We fit the state space model (equations 1-3) to the reported incidence for each haulout for the 1988 and 2002 epidemics using Bayesian Markov chain Monte Carlo (MCMC) methods (Morton & Finkenstädt 2005; Ferrari *et al.* 2008) (Appendix A). For each haulout, we estimated the initial population of susceptible individuals (S_0), the transmission term (β), the reporting probability (P_{obs}), and the unobserved time series of infected and susceptible seals (*I* and *S*) for each haulout.

The literature dictated initial values for the total population size (table 3-1, table 3-2). The epidemic generation time, which is the length of the latent plus infectious periods, was assumed to be 14 days (Swinton *et al.* 1998).

Table 3-1. Initial estimates of total population size (*N*) population size for the 1988 PDV epidemic.

Location	Minimum Estimate of Population Size (N_{min})	Reference
Danish Kattegat	2500	(Heide-Joergensen & Harkonen 1988)
Swedish Kattegat	2500	(Heide-Joergensen & Harkonen 1988)
Limfjord	2000	(Trilateral Seal Expert Group 2000)
Danish Wadden Sea	5200	(Trilateral Seal Expert Group 2000)
Schleswig-Holstein Wadden Sea	3400	(Trilateral Seal Expert Group 2000)
Niedersachsen Wadden Sea	1300	(Trilateral Seal Expert Group 2000)
Dutch Wadden Sea	3300	(Hall *et al.* 1992)
Norfolk	467	(Hall *et al.* 1992)
Tay	900	(Hall *et al.* 1992)

Initial estimates of total population size (*N*) used in the model.
Values were taken directly from the literature.

Table 3-2. Initial estimates of total population size (*N*) population size for the 2002 PDV epidemic.

Location	Minimum Estimate of Population Size (N_{min})	Reference
Kattegat	5,814	(Advisory Committee on Ecosystem 2001)
Skagerrak	3,658	(Advisory Committee on Ecosystem 2001)
Baltic	585	(Duck & Thompson 2002)
Limfjord	800	(Trilateral Seal Expert Group 2000)
Schleswig-Holstein Wadden Sea	9,570	(Reinekinig 2002)
Niedersachsen Wadden Sea	3,220	(Reinekinig 2002)
Dutch Wadden Sea	3,600	(Reinekinig 2002)
NE England	200	Adjusted from (Duck & Thompson 2002)
Lincolnshire	233	(Duck & Thompson 2002)
Norfolk	4,041	(Duck & Thompson 2002)
SW England	100	Adjusted from (Duck & Thompson 2002)
NW England	50	Adjusted from (Duck & Thompson 2002)
Shetland	4,883	(Duck & Thompson 2002)
Orkney	7,752	(Duck & Thompson 2002)
Grampian	126	(Duck & Thompson 2002)
Borders & Lothian	40	(Duck & Thompson 2002)
Strathclyde &	7,915	(Duck & Thompson 2002)

Dumfries		
W Scotland	6,955	(Duck *et al.* 2007)
Wales	200	Adjusted from (Duck & Thompson 2002)
N Ireland	400	(Duck & Thompson 2002)

Initial estimates of total population size (N) used in the model. Values were either taken directly from the literature or adjusted to be greater than the total number of stranded seals in the area. Adjusted counts were based on population surveys from the reference noted.

Many studies have argued that the counted populations (N_{min}, table 3-1, table 3-2) of harbor seals are a downward-biased estimate of the true population size by as much as one third (Thompson *et al.* 1992; Ries *et al.* 1998; Harkonen *et al.* 1999). Therefore, initial values of S_0 were set to to N_{min} / 0.66. We used a Poisson prior for S_0, an uninformative gamma prior for β, and a beta prior with for P_{obs} with a mean of 0.75 and variance of 0.0375. Candidate values in the MCMC were proposed using independent random walks for S_0, β, and P_{obs} according to the standard recipe for the Gibbs sampler (Casella & George 1992). This, unfortunately, does not work for the time series of infected harbor seals: the sum of the infected cases over the course of the epidemic is restricted to be less than or equal to the initial number of susceptibles, S_0. Thus, if the I_t's were to be proposed sequentially (according to standard Gibbs), the last time steps will tend to mix poorly and explore a biased part of the state-space because of the restriction on this sum. Therefore, we propose the entire time series of infected cases simulataneously as a multivariate random walk.

The default burn-in period were set at 500,000 iterations, but could be greater if visual inspection of the chains suggested transients that lasted longer. Chains were run for 1,000,000 iterations, or longer if mixing was slow. For inference, the chains were thinned to every tenth step to reduce undue influence of autocorrelation. Ten chains were run for each location, with different starting values; point estimates for S_0, P_{obs}, β and the infected time-series were found by taking the mean of each of the 10 posterior distributions generated by the 10 chains. To

check for robustness, these results were compared to the mean of the posterior distribution of one chain of 20 million iterations for any given location.

Results

We estimated S_0, P_{obs}, β and the infected time-series and calculated the 95% credible interval (CI) around each estimate for both the 1988 PDV epidemic (Table 3-3) and the 2002 PDV epidemic (Table 3-4). Given our model formulation (equations 1-3), the rate of transmission, β, corresponds to R_0 if the whole population was susceptible at the onset of the epidemic. The probability of observation, P_{obs}, is a compound variable encompassing the probability that a given seal, once infected by PDV, will die, will strand and that the stranded seal will be counted. Finally, the initial number of susceptible individuals, S_0, is an estimate of the total susceptible population of harbor seals at each location before the epidemic began.

Data from ten geographic areas permitted us to estimate the three parameters and reconstruct the epidemic time-series and estimate the final epidemic size for the 1988 epidemic and the 95% credible interval (table 3-3). The endpoints of the 95% CI are the 2.5th and 97.5th quantiles.

Table 3-3. Parameter estimates for the 1988 PDV Epidemic.

Region	β (95% CI)	P_{obs} (95% CI)	S_0 (95% CI)	Final Epidemic Size (95% CI)
Danish Kattegat	3.58 (3.42 – 3.76)	0.0645 (0.0670 – 0.0732)	4610 (4326 – 4775)	4587 (4083 – 5044)
Swedish Kattegat	3.04 (2.93 – 3.14)	0.482 (0.460 – 0.506)	4834 (4707 – 4943)	4816 (4623 – 5029)
Limfjord	3.62 (1.78 – 4.55)	0.344 (0.294 – 0.480)	1230 (1186 – 1346)	1130 (467 – 2137)
Danish Wadden Sea	2.19 (2.12 – 2.28)	0.312 (0.294 – 0.330)	4010 (3922 – 4082)	3997 (3721 – 4290)
Schleswig-Holstein Wadden	2.10 (2.05 –	0.537 (0.521 – 0.554)	10684 (10498 –	10664 (10308 –

Sea	2.15)		10860)	11043)
Niedersachsen Wadden Sea	2.76 (2.67 – 2.85)	0.145 (0.135 –0.155)	6871 (6704 – 7004)	6856 (6522 – 7191)
Dutch Wadden Sea	3.01 (2.84 – 3.18)	0.191 (0.173 – 0.213)	2300 (2171 – 2374)	2280 (2048 – 2512)
Norfolk	4.78 (3.43 – 5.05)	0.0783 (0.0701 – 0.0893)	5340 (5019 – 5538)	5330 (3858 – 6499)
Tay	1.31 (1.02 – 2.04)	0.602 (0.174 – 0.914)	707 (688 – 727)	78 (42 – 277)
Moray Firth	1.10 (1.00 – 1.31)	0.737 (0.560 – 0.937)	1363 (1341 –1385)	96 (63 – 161)

Estimates of β, P_{obs}, and S_0 for the 1988 epidemic. Values are noted in each box, followed by the 95% CI in parentheses.

Estimates for all three parameters varied with location: β ranged from 1.10 in Moray Firth to 4.78 in Norfolk (table 3-3), which are similar to other estimates for R_0 (Swinton *et al.* 1998; Klepac 2007). Scotland appears to have the lowest transmission rates, while the Wadden Sea has mid range rates (table 3-3). The highest estimated transmission rates were for Norfolk, England at a large haulout of harbor seals commonly known as “the Wash” (table 3-3). These values for transmission parallel the severity of the epidemic as illustrated by summary statistics of incidence and total epidemic mortality (Harkonen *et al.* 2006). The probability of observation, P_{obs}, ranges from a low of 0.0783 in Norfolk, England (the Wash) to a high of 0.737 in the Moray Firth, with mid-range values in the Wadden Sea. We also estimated the initial number of susceptible individuals, designated by S_0, which was larger than the latest census in all cases. Lastly, we reconstructed the time-series of infected individuals. By summing the estimated number of infected individuals over the entire epidemic, we also estimated the final epidemic size. Theoretically, the final size of of a closed epidemic in a homogenously mixing population is expected to be $1-e^{-R_0}$ (e.g., (Swinton 1998). We compared our estimated epidemics with this theoretical expectation for the 1988 (table 3-4).

Table 3-4. Final Epidemic Size for the 1988 PDV Epidemic.

Region	Reconstructed Final Epidemic	Theoretical Final Epidemic
Danish Kattegat	0.995	0.972
Swedish Kattegat	0.996	0.952
Limfjord	0.919	0.973
Danish Wadden Sea	0.997	0.888
Schleswig-Holstein Wadden Sea	0.998	0.877
Niedersachsen Wadden Sea	0.998	0.937
Dutch Wadden Sea	0.991	0.951
Norfolk	0.998	0.992
Tay	0.110	0.730
Moray Firth	0.070	0.667

The reconstructed final epidemic was calculated by dividing the final epidemic size (table 3-3) by S_0 (table 3-3). The theoretical final epidemic size was calculated by $1-e^{-R_0}$, where β (from the epidemic reconstruction, table 3-3) is equal to R_0.

The comparable estimates for the three basic parameters and final epidemic size for the 2002 epidemic are shown in table 3-5 and 3-6, respectively.

Table 3-5. Parameter estimates for the 2002 PDV Epidemic.

Region	β (95% CI)	P_{obs} (95% CI)	S_0 (95% CI)	Final Size of Epidemic (95% CI)
Danish Kattegat	3.40 (3.26 – 3.54)	0.249 (0.233 – 0.267)	4488 (4309 – 4606)	4470 (4158 – 4757)
Swedish Kattegat	3.19 (3.07 – 3.31)	0.498 (0.475 – 0.524)	4453 (4335 – 4548)	4435 (4210 – 4671)
Skagerrak	4.25 (3.26 – 4.60)	0.124 (0.113 – 0.149)	6378 (5564 – 6705)	6349 (4718 – 7384)
Baltic	1.93 (1.36 – 3.48)	0.261 (0.182 – 0.456)	889 (872 – 942)	558 (195 – 1583)
Limfjord	3.71 (1.14 – 4.56)	0.306 (0.265 – 0.520)	1248 (1187 – 1335)	1189 (536 – 1833)
Schleswig-Holstein Wadden	4.31 (1.01 –	0.203 (0.155 – 0.543)	16,621 (14,402 –	16,278 (4355 –

Sea	4.54)		21,672)	29,797)
Niedersachsen Wadden Sea	3.29 (3.19 – 3.40)	0.432 (0.414 – 0.452)	9594 (9262 – 9876)	9558 (9079 – 10,046)
Dutch Wadden Sea	3.02 (2.90 – 3.16)	0.329 (0.312 – 0.349)	6752 (6464 – 6946)	6725 (6259 – 7173)
NE England	2.45 (2.14 – 2.78)	0.798 (0.703 – 0.903)	351 (334 – 368)	341 (307 – 398)
Lincolnshire	2.03 (1.85 – 2.22)	0.951 (0.874 – 0.997)	530 (520 – 540)	527 (518 – 546)
Norfolk	3.37 (1.89 – 3.48)	0.218 (0.207 – 0.353)	7869 (6132 – 8071)	7844 (7021 – 8341)
SW England	1.66 (1.15 – 2.65)	0.619 (0.367 – 0.880)	152 (143 – 161)	82 (41 – 181)
NW England	1.16 (1.01 – 1.65)	0.775 (0.519 – 0.982)	74 (66 – 83)	31 (30 – 40)
Shetland	1.32 (1.03 – 1.81)	0.590 (0.300 – 0.830)	7398 (7339 – 7458)	80 (54 – 171)
Orkney	2.18 (1.60 – 2.75)	0.152 (0.0663 – 0.424)	11746 (11665 – 11826)	1291 (564 – 2381)
Grampian	2.34 (1.90 – 2.78)	0.505 (0.406 – 0.637)	210 (190 – 225)	196 (129 – 278)
Borders & Lothian	2.11 (1.61 – 2.74)	0.921 (0.754 – 0.996)	72 (67 – 78)	68 (59 – 89)
Strathclyde & Dumfries	1.04 (1.00 – 1.19)	0.745 (0.547 – 0.960)	11990 (11909 – 12071)	73 (47 – 131)
W Scotland	1.23 (1.01 – 1.76)	0.0732 (0.0177 – 0.492)	10537 (10461 – 10613)	1042 (129 – 3486)
Wales	2.20 (1.95 – 2.48)	0.477 (0.405 – 0.560)	362 (338 – 381)	353 (292 – 425)
N Ireland	3.10 (1.33 –3.55)	0.173 (0.137 – 0.406)	645 (596 – 691)	610 (183 – 981)

Estimates of β, P_{obs}, and S_0 for the 1988 epidemic. Values are noted in each box, followed by the 95% CI in parentheses.

Again, estimates for transmission rate, the probability of observation, and the initial number of susceptible individuals varied with location. Estimates of the transmission rate (β) ranges from a low of 1.23 in Western Scotland to 4.25 in the Skagerrak. Observation probabilities had a much larger range than in the previous epidemic, spanning from 0.0732 in Western Scotland to 0.950 in Lincolnshire. We also estimated the initial number of susceptible individuals, designated by S_0, which was larger than the last census in all cases. Lastly, we reconstructed the time-series of infected individuals; by summing the time-series over the entire epidemic, we can estimate the final size of the epidemic (table 3-6).

Table 3-6. Final Epidemic Size for the 2002 PDV Epidemic.

Region	Reconstructed Final Epidemic	Theoretical Final Epidemic
Danish Kattegat	0.967	0.996
Swedish Kattegat	0.959	0.996
Skagerrak	0.986	0.996
Baltic	0.855	0.628
Limfjord	0.976	0.953
Schleswig-Holstein Wadden Sea	0.987	0.980
Niedersachsen Wadden Sea	0.963	0.996
Dutch Wadden Sea	0.951	0.996
NE England	0.914	0.972
Lincolnshire	0.869	0.994
Norfolk	0.966	0.997
SW England	0.810	0.539
NW England	0.687	0.419
Shetland	0.733	0.011
Orkney	0.887	0.110
Grampian	0.904	0.933
Borders & Lothian	0.879	0.944
Strathclyde & Dumfries	0.647	0.006
W Scotland	0.708	0.099
Wales	0.889	0.975
N Ireland	0.955	0.946

The reconstructed final epidemic was calculated by dividing the final epidemic size (table 5) by S_0 (table 5). The theoretical final epidemic size was calculated by $1-e^{-R_0}$, where β (from the epidemic reconstruction, table 5) is equal to R_0.

Estimates for β (or R_0 assuming all seals were susceptible prior to the epidemics) shows some consitent geographic structures: in 1988, the largest values were found near Denmark, Sweden, and Norfolk (the Wash), England while the smallest values were found in Scotland (table 3-3, figure 3-2).

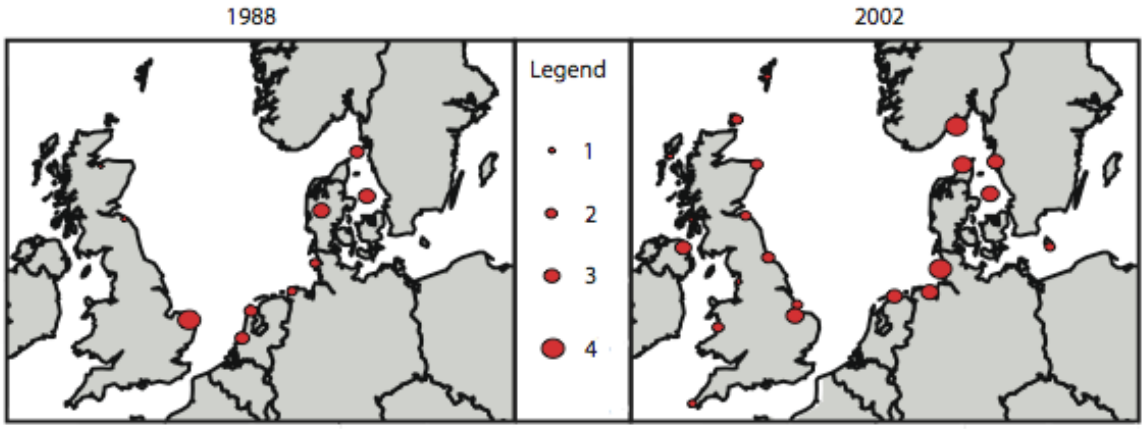

Figure 3-2. Point estimates of β by location in the North Sea for 1988 (left panel) and 2002 (right panel). The radius of the red circle represents the value of the parameter.

In 2002, large values were found in Denmark and Sweden, mid-range values were found in the Wadden Sea, and smaller values were found in the northern United Kingdom (table 3-4, figure 3-2).

Finally, we estimated the time-series of infected individuals for each location and for each epidemic. The marginal posterior for the unobserved number of infected cases is given by an epidemic curve (Appendix A). The reconstructed epidemics always had greater incidence than the data alone; furthermore, the reconstructed time-series would tend to follow smoother epidemic curves and tended to resolve the irregularities in the incidence curves from the data. Four comparisons of the data and the reconstructed time-series are shown below (figure 3-3).

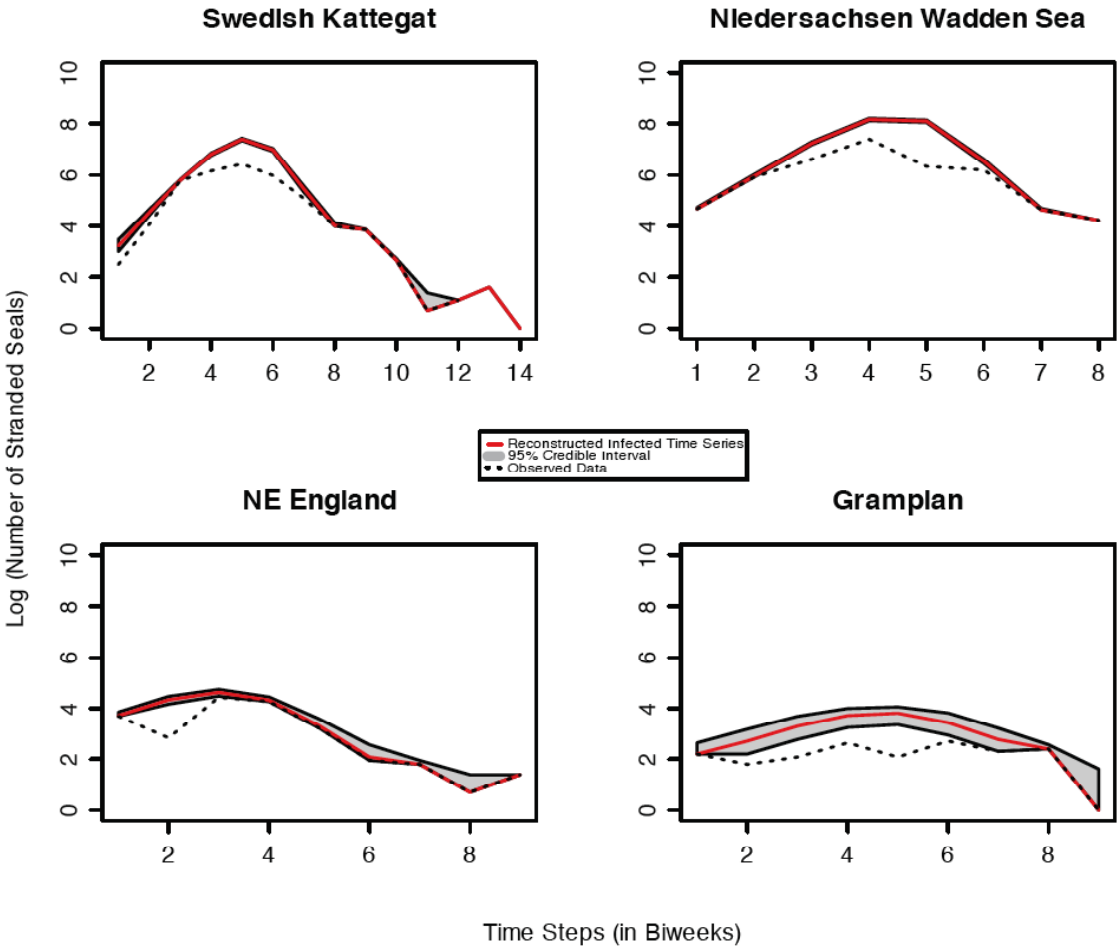

Figure 3-3. Epidemic reconstruction for four locations in the 2002 PDV epidemic on a log-scale.

Although the two epidemics were very similar, parameter estimates differed between the 1988 and 2002 epidemics (figure 3-4). In 1988, β was highest in Norfolk and lowest in the Schleswig-Holstein Wadden Sea, among the seven locations where comparative estimates are available.

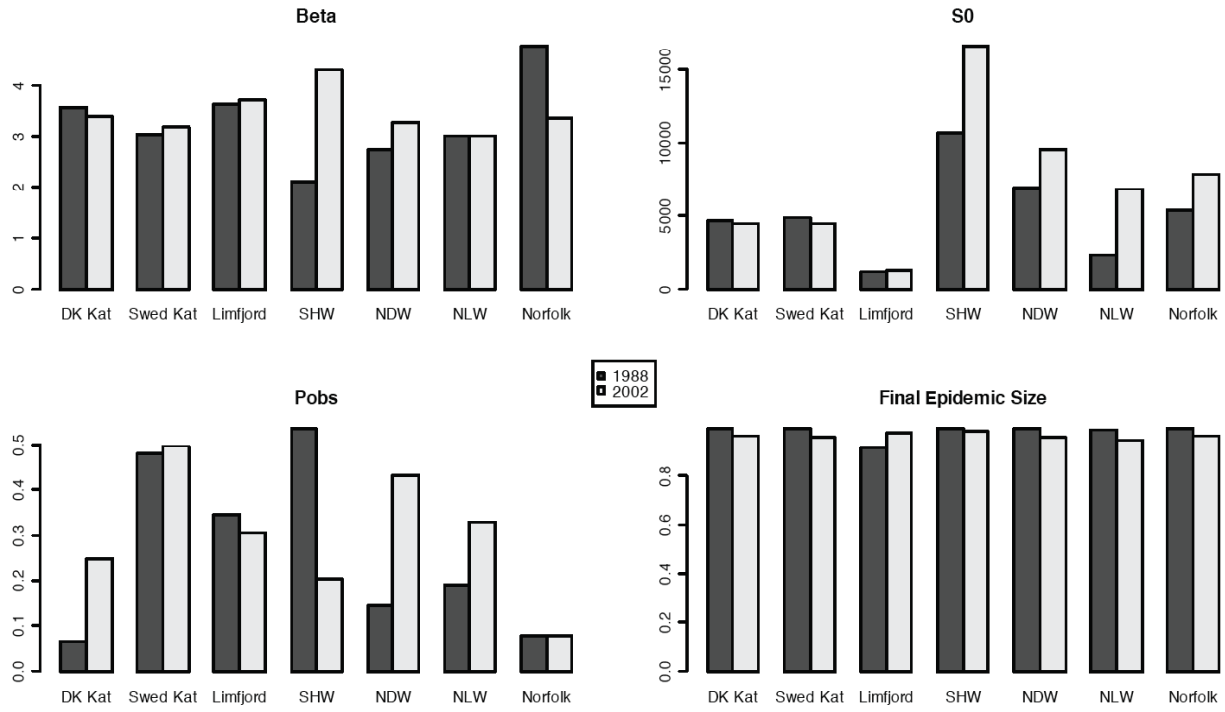

Figure 3-4. Comparison of Parameter Estimates for 1988 (dark bars) and 2002 (light bars). Estimates of β, P_{obs}, S_0, and the final epidemic size for the 1988 and 2002 epidemics.

This pattern was partially reversed in 2002: the Schleswig-Holstein Wadden Sea showed the largest value of β. Initial population sizes (S_0) for the two PDV outbreaks were comparable for the Kattegat and Limfjord, while later populations were greater for the Wadden Sea regions and Norfolk, or "the Wash." The probability of observation varied among location and year; the final epidemic size was over 90% for all locations in both years (figure 3-4).

Discussion

The epidemic reconstruction within the Bayesian framework has two main contributions to understanding the dynamics during the PDV epidemics: first, in clarifying true values from the uncertainties inherent in the data, such as the initial number of susceptible individuals (S_0),

transmission rate (β) and the probability of observation (P_{obs}) and second, it allows us to reconstruct the epidemic curves from imperfect time-series of incidence. It must be noted, however, that the overall disease burden was calculated from the marginal posterior (Appendix A), which constricted the posterior to resemble an epidemic curve, despite the data quality or resolution.

Estimates for the initial number of susceptible individuals (S_0) and the transmission rate (β), the latter which represents R_0 if susceptibility was 100% at the beginning of the epidemic, fall within previously published values. Uncorrected population censuses count seals on haulouts – considered to be a minimum estimate of the population size – have been thought to underreport regional population estimates by up to 60% (Thompson *et al.* 1992; Ries *et al.* 1998; Harkonen *et al.* 1999). Our estimates of the size of the initial susceptible population (table 3-3, table 3-4) fall within this range when compared to the most recent population census (table 3-1, table 3-2). In addition, previous R_0 estimates range from 1.5 to 4.0 for different regions throughout the North Sea (Swinton *et al.* 1998; Klepac 2007), which are similar in magnitude to our estimates (table 3-3, table 3-4, figure 3-2).

Additionally, the parameter estimates illuminate and support published observations of harbor seal behavior. For instance, estimates for S_0 support a well cited argument that even if population census are performed at a time to maximize the number of seals on the haulout, many seals will not be counted since they will not be on the haulout (Thompson *et al.* 1992; Ries *et al.* 1998; Harkonen *et al.* 1999). In addition, our estimates for β illustrate that different epidemic dynamics occurred in different regions: populations had different incidence, prevalence, mortality, and transmission rates (table 3-3, table 3-4, figure 3-2), (Hall *et al.* 2006; Harkonen *et al.* 2006). Finally, the compound probability of observation estimate (P_{obs}), differs widely based on location (table 3-3, table 3-4). Differences in local dynamics – stage structured strandings in the Dutch Wadden Sea (Rijks *et al.* 2005) in contrast with the lack of structure

found at the Wash in Norfolk, England (Thompson *et al.* 2005) – may affect pre and post epidemic population counts, since these may also show a sex and age bias (Harkonen *et al.* 1999). In addition, the chance of observing a stranded seal relates to the proximity to human observers and to sea tidal behavior, since strandings have occurred where there is no evidence of resident seal haulouts.

Final epidemic sizes from reconstructed parameters and from theoretical predictions offered similar results for both 1988 and 2002, with some exceptions. In both 1988 and 2002, the final size of the epidemic from the reconstruction was much smaller than the theoretical predictions for the Scottish haulouts, including the Tay, Moray Firth, Shetland, Orkney, etc (table 3-4, table 3-6). The Scottish haulouts, in general, have shown disease burdens less than those predicted and less than those experienced by other haulouts, for reasons which are not yet understood (Hall *et al.* 2006; Harkonen *et al.* 2006). Other discrepancies, such as some regions with slightly higher disease burden calculated from the reconstruction than that predicted by theory may be attributable to the geographical misplacement of seal bodies due to tidal currents (Rijks *et al.* 2005).

In wildlife disease systems, such as PDV in harbor seals in the North Sea, many different sources of error contribute to known and unknown biases in the data, which leads to problems with analysis, interpretation, model building, etc. To understand the disease dynamics, it is important to reconstruct the true epidemic trajectories for use in models that can be used to estimate spatial spread of the virus and understand its behavior in the harbor seal metapopulation. Bayesian Markov Chain Monte Carlo methods offer a flexible framework to deal with inference in the face of observational errors and unobserved states.

Acknowledgements

The authors would like to thank the volunteers and researchers that performed the population censuses and seal stranding counts. In addition, we would like to thank Petra Klepac

and Jolianne Rijks for assistance with North Sea data sets, sources, and interpretations as well as general advice on the PDV epidemics that occurred in the North Sea. We are also grateful to Angela Luis for helpful discussions. Laura Pomeroy was supported by the National Science Foundation, under the NSF Graduate Teaching Fellowship in K-12 Education (DGE-0338240).

References

Advisory Committee on Ecosystem (2001) Report of the Working Group on Marine Mammal Population Dynamics and Habitats. In. International Council for the Exploration of the Sea, Copenhagen, Denmark

Bailey N.T.J. (1957) *The Mathematical Theory of Epidemics*. Griffin, New York.

Baker J.R. (1992) The Pathology of Phocine Distemper. *The Science of the Total Environment*, 115, 1-7

Baker J.R. & Ross H.M. (1992) The Role of Bacteria in Phocine Distemper. *The Science of the Total Environment*, 115, 9-14

Casella G. & George E.I. (1992) Explaining the Gibbs sampler. *American Statistician*, 46, 167-174

Cosby S.L., McQuaid S., Duffy N., Lyons C., Rima B.K., Allan G.M., McCullough S.J., Kennedy S., Smyth J.A., McNeilly F., Craig C. & Orvell C. (1988) Characterization of a Seal Morbillivirus. *Nature*, 336, 115-116

Dietz R., Heide-Jorgensen M.-P. & Harkonen T. (1989) Mass Deaths of Harbor Seals *(Phoca Vitulina)* in Europe. *Ambio*, 18, 258-264

Duck C.D. & Thompson D. (2002) The status of the British common seal populatons. In: *Special Committee on Seals (SCOS) 2002*, pp. 35-38. Sea Mammal Research Unit, St. Andrews, Fife, Scotland

Duck C.D., Thompson D. & B.L. M. (2007) The status of the British common seal populations. In: *Special Committee on Seals (SCOS) 2007*, pp. 50-61. Sea Mammal Research Unit, St. Andrews, Fife, Scotland

Ferrari M.J., Bjørnstad O.N. & Dobson A.P. (2005) Estimation and inference for R0 of an infectious disease using a removal method. *Mathematical Biosciences*, 198, 14-26

Ferrari M.J., Grais R.F., Bhart i.N., Conlan A.J., Bjørnstad O.N., Wolfson L.J., Guerin P.J., Djibo A. & Grenfell B.T. (2008) The dynamics of measles in sub-Saharan Africa. *Nature*, 451, 679-684

Grenfell B.T., Lonergan M.E. & Harwood J. (1992) Quantitative Investigations of the Epidemiology of Phocine Distemper Virus (PDV) in European Common Seal Populations. *Science of the Total Environment*, 115, 15-29

Hall A.J., Jepson P.D., Goodman S.J. & Harkonen T. (2006) Phocine distemper virus in the North and European Seas - Data and models, nature and nurture. *Biological Conservation*, 131, 221-229

Hall A.J., Pomeroy P.P. & Harwood J. (1992) The Descriptive Epizootiology of Phocine Distemper in the Uk During 1988/89. *Science of the Total Environment*, 115, 31-44

Harder T., Willhaus, T.H., Frey, H.-R., and Liess, B. (1990) Mobillivirus Infections of Seals during the 1988 Epidemic in the Bay o Heligoland: III. Transmission Studies of Cell Culture-Propagated Phocine Distemper Virus in Harbor Seals (Phoca vitulina) and a Grey Seal (Halichoerus grypus): Clinical, Virological and Serological Results. *Journal of Veterinary Medicine*, 37, 641-650

Harkonen L., Dietz R., Reijnders P., Teilmann J., Harding K., Hall A., Brasseur S., Siebert U., Goodman S.J., Jepson P.D., Rasmussen T.D. & Thompson P. (2006) A Review of the 1988 and 2002 phocine distemper virus epidemics in European harbour seals. *Diseases of Aquatic Organisms*, 68, 115-130

Harkonen T., Harding K. & Lunneryd S.G. (1999) Age- and Sex-Specific Behaviour in Harbor Seals *Phoca vitulina* Leads to Biased Estimates of Vital Population Parameters. *Journal of Applied Ecology*, 36, 825-841

Heide-Joergensen M.-P. & Harkonen T. (1988) Rebuilding seal stocks in the Kattegat-Skagerrak. *Marine Mammal Science*, 4, 231-246

Heide-Jorgensen M.-P. & Harkonen T. (1992) Epizootiology of the Seal Disease in the Eastern North Sea. *Journal of Applied Ecology*, 29, 99-107

Jensen T., van de Bildt M., Dietz H.H., Andersen T.H., Hammer A.S., Kuiken T. & Osterhaus A. (2002) Another phocine distemper outbreak in Europe. *Science*, 297, 209-209

Klepac P. (2007) Mathematical Epidemiology of Phocine Distemper Virus in Harbor Seals. In: *Biological Oceanography*. MIT/WHOI, Woods Hole, MA

Legendre P. (1992) Real Data are Messy. *Statistics and Computing*, 3, 197-199

Mahy B.W.J., Barrett T., Evans S., Anderson E.C. & Bostock C.J. (1988) Characterization of a Seal Morbillivirus. *Nature*, 336, 115

Morton A.M. & Finkenstädt B.F. (2005) Discrete-time modelling of disease incidence time series by using Markov Chain Monte Carlo methods. *Applied Statistics*, 54, 575-594

Osterhaus A.D., Uytdehaag, F.G., Visser, I.K., Vedder, E.J., Reijnders, P.J., Kuiper, J., and Brugge, H.N. (1989) Seal Vaccination Success. *Nature*, 337, 21

Osterhaus A.D.M.E. & Vedder E.J. (1988) Identification of Virus Causing Recent Seal Deaths. *Nature*, 355, 20

Reinekinig B. (2002) Phocine Distemper Virus amongst Seals in 2002. In: *Wadden Sea Newsletter* Common Wadden Sea Secretariat, Wilhelmshaven, Germany

Ries E.H., Hiby L.R. & Reijnders P.J.H. (1998) Maximum Likelihood Population Size Estimation of Harbour Seals in the Dutch Wadden Sea Based on a Mark Recapture Experiment. *Journal of Applied Ecology*, 35, 332-339

Ries E.H., Traut I.M., Brinkman A.G. & Reijnders P.J.H. (1999) Net dispersal of harbour seals within the Wadden Sea before and after the 1988 epizootic. *Journal of Sea Reasearch*, 41, 233-244

Rijks J.M., Van de Bildt M.W.G., Jensen T., Philippa J.D.W., Osterhaus A. & Kuiken T. (2005) Phocine distemper outbreak, the Netherlands, 2002. *Emerging Infectious Diseases*, 11, 1945-1948

Swinton J. (1998) Extinction times and phase transitions for spatially structured closed epidemics. *Bulletin of Mathematical Biology*, 60, 215-230

Swinton J., Harwood J., Grenfell B.T. & Gilligan C.A. (1998) Persistence thresholds for phocine distemper virus infection in harbour seal Phoca vitulina metapopulations. *Journal of Animal Ecology*, 67, 54-68

Thompson D., Lonergan M. & Duck C. (2005) Population dynamics of harbour seals *Phoca vitulina* in England: monitoring growth and catastrophic declines. *Journal of Applied Ecology*, 42, 638-648

Thompson P.M. & Miller D. (1992) Phocine distemper virus outbreak in the Moray Firth common seal population: an estimate of mortality. *The Science of the Total Environment*, 115, 57-65

Thompson P.M., Tollit D.J., Wood D., Corpe H.M., Hammond P.S. & Mackay A. (1992) Estimating Harbour Seal Abundance and Status in an Estuarine Habitat in North-East Scotland. *Journal of Applied Ecology*, 34, 43-52

Trilateral Seal Expert Group (2000) Common Seals in the Wadden Sea in 1999. In: *Wadden Sea Newsletter*, Wilhelmshaven, Germany

Appendix A: Marginal posteriors for PDV MCMC chain.

The likelihoods for both the unobserved state process, and the observed data in the PDV state-space model are all binomial, where

$$L(X \mid Y, p) = \binom{Y}{X} p^X (1-p)^{Y-X} \ .$$

The marginal posterior for the unobserved number of infected cases at time t is given by

$$P(I_t \mid I_{t-1}, I_{t+1}) \propto L(I_t \mid S_{t-1}, 1-e^{-\beta \frac{I_{t-1}}{N}}) L(I_t + 1 \mid S_t, 1-e^{-\beta \frac{I_t}{N}}) L(Y_t \mid I_t, P_{obs})$$

where S_t is given by equation 3 (main text).

The marginal posteriors for β, S_0, and P_{obs} are given by

$$P(\beta \mid I, S, S_0) \propto \prod_{i=2}^{T} L(I_i \mid S_{i-1}, 1-e^{-\beta \frac{I_{i-1}}{N}}),$$

$$P(S_0 \mid I, S, \beta) \propto \prod_{i=2}^{T} L(I_i \mid S_{i-1}, 1-e^{-\beta \frac{I_{i-1}}{N}}),$$

$$P(P_{obs} \mid I, Y) \propto \prod_{i=1}^{T} L(Y_i \mid I_i, P_{obs}).$$

Modeling the Spatial Spread of Phocine Distmper Virus in the North Sea

Laura W. Pomeroy[1,2], Catriona Stephenson[3,4], John Harwood[3,4], Karin Harding[5], Ottar N. Bjørnstad[1,2,6]

[1]Department of Biology, Pennsylvania State University, University Park, PA 16802 USA

[2]Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA 16802 USA

[3]Sea Mammal Research Unit, University of St. Andrews, St Andrews, Fife, KY16 9LZ UK

[4]Centre for Research into Ecological and Environmental Modelling, University of St. Andrews, St Andrews, Fife, KY16 9LZ UK

[5]Department of Marine Ecology, University of Gothenburg, SE-405 30 Gothenburg, Sweden

[6]Department of Entomology, Pennsylvania State University, University Park, PA 16802 USA

Abstract

Acute infections cause epidemics in many wildlife populations, affecting their abundance and status. We aim to determine the principles that dictate spatial spread in social animals by using a gravity model and a distance or diffusion model to fit the spread of the phocine distemper virus (PDV) epidemic in the North Sea in 2002. Initial model conditions are based on previous epidemic reconstructions within a Bayesian framework. Results show that the distance model has a better fit to the seal stranding data, indicating that the distance between harbor seal haulouts drives the spatial spread of PDV. Further models should include the population and dynamics of the gray seal population in the North Sea, which may be responsible for vectoring PDV spread among harbor seals.

Introduction

Acute, virulent infections pose important threats to wildlife and are considerable conservation concerns for a number of species. More specifically, viruses with high case fatality rates that induce strong immunity following recovery experience episodic outbreaks followed by the local extinction of the pathogen and sometimes the host, as well. Many members within the Morbillivirus genus belong to this category of pathogens. Continued circulation requires repeated reintroduction in all but extremely large host communities (Grenfell & Harwood 1997). Understanding the laws governing the spatial dissemination of infection, therefore, is of great importance. Simple distance-based diffusion models appear to adequately represent the spread of certain infections, such as rabies (Murray *et al.* 1986), albeit spatial heterogeneities sometimes result in geographically-varying diffusion rates (Smith *et al.* 2002). Recent research has focused on whether more complex models are required to understand the spread among social and group-dwelling hosts. The gravity model of spatial spread proposes that host dispersal among local communities depends not only on the distance between them but is also a bi-linear function of their sizes (Haynes & Fotheringham 1984). The gravity formulation has received strong empirical support for a number of human infections such as measles (Xia *et al.* 2004) and influenza (Viboud *et al.* 2006). It is, however, unclear the extent to which it is an important idea for infections in social wildlife. It is important to determine if these models are needed for the system of phocine distemper virus (PDV) in harbor seals in the North Sea, since PDV is a Morbillivirus that shows patterns of periodic outbreaks that quickly sweep through a local population and pose a threat to the conservation of the species (Harding *et al.* 2002).

Over the past two decades, phocine distemper virus (PDV), a recently emerged member of the Morbillivirus genus (Cosby *et al.* 1988; Mahy *et al.* 1988b; Osterhaus &

Vedder 1988), caused two mass mortality events among harbor seals in the North Sea (Cosby *et al.* 1988; Mahy *et al.* 1988b; Osterhaus & Vedder 1988; Dietz *et al.* 1989; Rima *et al.* 1992; Jensen *et al.* 2002; Hall *et al.* 2006; Harkonen *et al.* 2006). The first epidemic occurred in 1988, in which 18,000 to 23,000 harbor seals died (Hall *et al.* 2006; Harkonen *et al.* 2006). This mass mortality event caused by the viral epidemic began on the Danish island of Ånholt on April 12, 1988 and ended within the calendar year (Dietz *et al.* 1989; Hall *et al.* 2006; Harkonen *et al.* 2006). A second PDV outbreak occurred in 2002 with the same point of origin: initial cases of harbor seal stranding and mortality occurred on May 4, 2002. In this epidemic, approximately 22,000 to 30,000 harbor seals died, resulting in the largest recorded mass mortality event in marine mammals (Jensen *et al.* 2002; Hall *et al.* 2006; Harkonen *et al.* 2006).

In this parasite-host system, the causative agent is a single stranded, negative sense RNA virus which is a member of the family *Paramyxoviridae (Cosby et al. 1988; Mahy et al. 1988b; Osterhaus & Vedder 1988)* which affects the host with disease that spans a two-week period, including both the latent and infectious disease stages (Osterhaus 1989; Harder 1990a; Baker 1992; Grenfell *et al.* 1992). Mortality is high (Rijks *et al.* 2005), partly due to co-infections (Baker & Ross 1992).

The disease spread throughout the North Sea in both 1988 and 2002 and affected most haulouts along the coast (Hall *et al.* 2006; Harkonen *et al.* 2006). While both epidemics began on an island in the Danish Kattegat, slightly different spatial spread patterns were observed, as fully described in Harkonen *et al.* (2006). Notably, the disease spread regionally in 1988 in accordance with diffusion models except for three surprising jumps where more distant haulouts were affected earlier than expected: once to the more distant Dutch Wadden Sea before other haulouts in the Wadden Sea, once from a large harbor seal haulout in southwest England, commonly known as the Wash, to Ireland and lastly, from the Wash to Scotland (Harkonen *et al.* 2006). Similar patterns

of spread were observed in 2002, including the unexpected jump to the Dutch Wadden Sea; in contrast with the 1988 epidemic, certain areas, including the Danish Wadden Sea, German Wadden Sea, and Scotland, showed little or no evidence of a phocine distemper virus epidemic (Harkonen *et al.* 2006).

The detailed pattern of PDV spread illustrates a challenge in modeling disease spread with the susceptible-infected-removed (SIR) model framework (Kermack & McKendrick 1927; Anderson & May 1991): to accurately model the disease dynamics, we must choose a model that accurately describes host population interactions (McCallum *et al.* 2001). For social hosts, such as common seals, two aspects of transmission must be considered. These include both the local transmission among individuals within a "patch," or haulout, and the transmission between patches, which mirrors host or vector movement (Xia *et al.* 2004). In this study, we contrast the relative merits of two nested models – a distance-based model that assumes that the spatial spread of disease is strictly dependent on the size of the donor community and the distance between two communities and the gravity model – to determine the best-fit to seal stranding data from the 2002 epidemic.

Methods

To study the spatial spread, we use the susceptible-infected-removed (SIR) framework for local epidemic dynamics, dividing the local population into three categories based on their epidemiological state (Grenfell & Dobson 1995). Susceptible individuals have never been exposed to the virus nor experienced infection. We assume that local epidemics, which sweep through local communities in two to five months, are fast relative to harbor seal demography, since harbor seals have an annual summer pupping season. Therefore, we assume closed epidemics: the only change in the susceptible category is when susceptible individuals are converted into infected and

infectious individuals. We use a discrete-time formulation where the time-step is the epidemic generation time (latent and infectious periods) of approximately 14 days. The balance equation for the susceptible population in the local community *k* is then:

$$S_{k,t+1} = S_{k,t} - I_{k,t+1} \quad (1)$$

where *S* represents the number of susceptible individuals and *I* represents the number of infected individuals. The time-series of PDV cases can be modeled using a TSIR model, with the infected individuals calculated as the expectation for the chain-binomial model [e.g. Bailey 1957; Chapter 3] according to

$$I_{k,t+1} = S_{k,t} \cdot (1 - e^{-\phi_{k,t}}) \quad (2)$$

where ϕ is the force of infection and $e^{-\phi_{k,t}}$ represents the probability that a susceptible individual will escape infection in epidemic generation *t*. Within the epidemic metapopulation setting, this force of infection will depend both on the number of resident infected individuals and on migrant or transient infected individuals, here designated as *m*. The local force of infection is thus:

$$\phi_{k,t} = \frac{\beta_k (I_{k,t} + m_{k,t})}{N_k} \quad (3)$$

where β is the transmission rate and *N* is the total number of individuals on each haulout, as determined from an epidemic reconstruction (Chapter 3). For simplicity, we will assume that the number of transients is so small that they don't affect the local population size. If we let y_j represent the fraction of infected individuals at location *j*, then according to the gravity formulation, the transient term in the force-of-infection is:

$$m_{k,t} = \theta N_k^{\tau_1} \sum_{j \neq k} \frac{N_j^{\tau_2} y_j}{d_{jk}^{\rho}} \quad (4)$$

where θ represents the spatial coupling constant, N_k and N_j represents the recipient and donor populations, respectively. The parameter τ_1 scales how immigration rates depends

on recipient population size and τ_2 scales how emmigration depends on donor population size. Lastly, d_{jk} represents the distance between haulouts *j* and *k*. We measure this as the shortest seaway routes using ArcView. The distance decay is scaled by ρ. If τ_1 = 0, then this gravity model collapses to the distance model:

$$m_{k,t} = \theta \sum_{j \neq k} \frac{N_j^{\tau_2} y_j}{d^{\rho}} \qquad (5)$$

Removed individuals are previously infected and either recovered from the disease with conferred lifelong immunity or suffer mortality from the disease. Since the length of each time-step equals the latent and infectious periods, all infected individuals (I_t) will enter the removed compartment (R_{t+1}) after one time-step. Thus

$$R_{k,t+1} = R_{k,t} + I_{k,t}. \qquad (6)$$

Our focus, here, is on making inference regarding the most appropriate model for the topology of spatial spread – i.e., estimating θ, τ_1, τ_2 and ρ. This requires detailed data on local and regional counts of susceptibles and infecteds through time, as well as the local susceptible population size ($S_{0,k}$) and local transmission rates (β_k). For these quantities, we use the local estimates and reconstructed epidemic trajectories from previous Bayesian Markov chain Monte Carlo (MCMC) state space analysis of haulout-specific time-series of incidence (Chapter 3: equations 1-3). Conditional on these previous estimates regarding the non-spatial aspects of spread, we use likelihood theory to estimate and make inference on the four spatial parameters. For this we also need to consider the reporting rate, $\rho_{k,obs}$: a nuisance parameter for which our previous MCMC analysis (Chapter 3) also provides estimate. For the likelihood inference we assume that the biweek-specific, stranded seals at haul-out *k*, $x_{k,t}$, reflect “counting errors” according to:

$$x_{k,t} \sim Po\left(\hat{p}_{k,obs}\hat{S}_{k,t}(1-e^{\phi_{k,t}})\right), \quad (7)$$

where $\hat{p}_{k,obs}$ is the estimated site-specific reporting rate (Chapter 3), and $\hat{S}_{k,t}$ is the MCMC-reconstructed susceptible time-series. We model the time-specific force-of-infection, $\phi_{k,t}$, according to:

$$\phi_{k,t} = \frac{\hat{\beta}_k(\hat{I}_{k,t} + \theta\sum_{j\neq k}\frac{N_j^{\tau_2}\hat{y}_j}{d^{\rho}})}{N_k}. \quad (8)$$

Again $\hat{\beta}_k$ and $\hat{I}_{k,t}$ represent previous MCMC estimates for the site-specific transmission rate and the reconstructed infected time-series (chapter 3). The reporting rate is a compound variable encompassing both the probability that a given seal, once infected by PDV, will strand and that the stranded seal will be encountered and observed. Equations (7) and (8) allow maximum likelihood estimation on θ, ρ, τ_1, and τ_2. Because of possible co-linearity among these estimates, we base our inference on profile likelihoods. We first profile on τ_1 and τ_2 marginal on θ, ρ. Then, we calculate the likelihood profile for ρ conditional on the maximum likelihood estimates for τ_1 and τ_2 and marginal on θ.

Results

To determine the model that best fits the PDV regional data, we must determine the parameter estimates for θ, ρ, τ_1, and τ_2. Initial model fitting exercises provided estimates of both τ_1, and τ_2, as shown on the grid (figure 4-1).

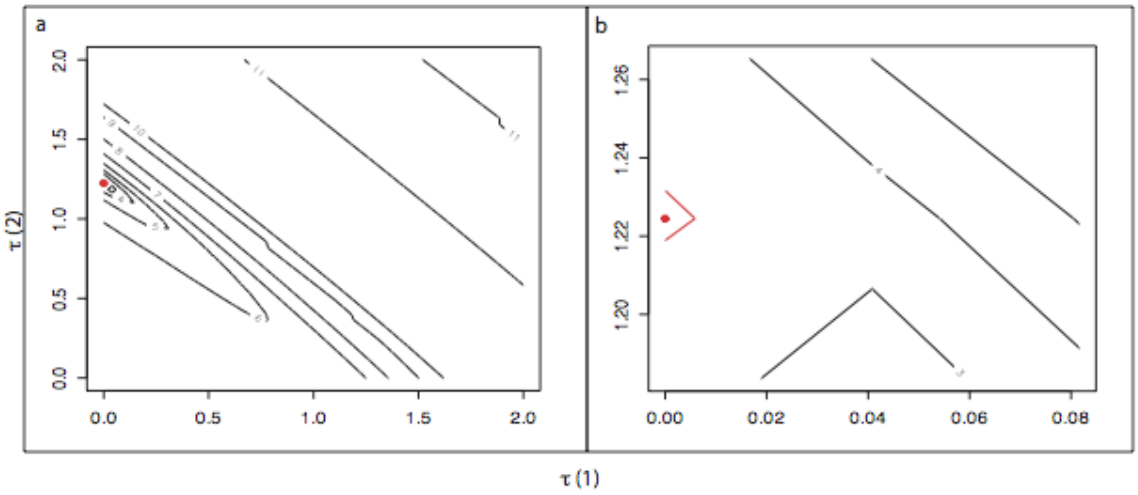

Figure 4-1. Maximum likelihood estimate of τ_1 and τ_2, which are the paratmeres that tune the affect of the donor and recipient population sizes. (a) Results from the simultaneous maximum likelihood estimation of τ_1 and τ_2 over a grid from 0 to 2. The maximum likelihood estimate found was $\tau_1 = 0$, $\tau_2 = 1.2$. (b) Zoom view on the maximum likelihood estimate at the 95% confidence interval, depicted in red.

Maximum likelihood estimates of τ_1, and τ_2 indicate that since $\tau_1 = 0$, the gravity model formulation (equation 4) collapses to the nested distance model (equation 5), where $\tau_2 = 1.2$. Therefore, the number of transient migrants in the force of infection depends on the size of the donor population, scaled by the estimate of τ_2, but not on the size of the recipient population.

The next step in fitting the most parsimonious model to 2002 PDV seal stranding data was to find the maximum likelihood estimate of ρ, which corresponds to the minimum value of the negative log likelihood in practice (figure 4-2).

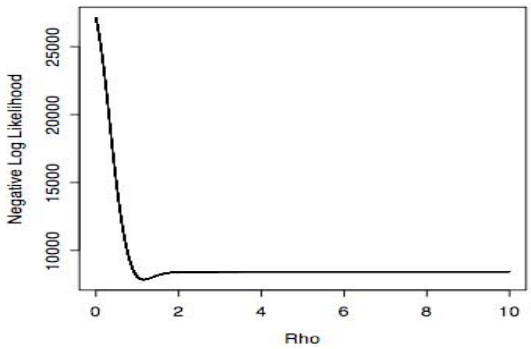

Figure 4-2. Maximum likelihood estimate of ρ, the exponential tuning parameter on the pair-wise distance between haulouts. Results from the maximum likelihood estimation of ρ over a vector from 0 to 10. The maximum likelihood estimate, which can also be calculated as the minimum negative log likelihood, is ρ=1.2.

The estimate of ρ determines how the transient migrants behave with distance between two haulouts. As the distance between locations increases, the number of migrants decreases as the recprical of the distance raised to the power of ρ. Final estimation of the number of transient migrants relates scales with θ, which was estimated as 1.1. The final model, which dictates the topology of the spatial spread of PDV in the North Sea, can be summarized with the four parameter estimates (table 4-1).

Table 4-1. Parameter estimates.

Parameter	Estimate (95% CI)
Theta (θ)	1.13 (1.13 – 1.14)
Rho (ρ)	1.2 (0.653 -)
Tau 1 (τ_1)	0 (0 - 0.004)
Tau 2 (τ_2)	1.2 (1.16-1.24)

Maximum likelihood estimates of θ, ρ, τ_1 and τ_2, which represent a scaling term for the number of infectious migrants, a term that determines how the transient migrants behave with distance between two haulouts, and the two parameters that scale how the infectious migrants scale with the donor and recipient population sizes. The estimates and their 95% confidence intervals are noted.

Rho shows only a lower bound in the 95% confidence interval since the right shoulder of the graph (figure 4-2) lies below the upper bound of the 95% confidence interval. From these results, it is clear that the distance model is the best fit to the PDV dynamics in the North Sea.

Based on the parameter estimates, we can calculate the number of migrant harbor seals designated $m_{k,t}$, to various recipient haulout populations in the North Sea using the distance formulation (equation 5). The Dutch and German areas of the Wadden Sea gain the most immigrants while the smallest number of migrant seals flock to Scotland (figure 4-3).

Figure 4-3. Infectious Immigrants throughout the North Sea. Number of infectious immigrants ($m_{k,t}$) at major harbor seal haulout sites throughout the North Sea. The number of immigrants to a location are represented by the radius of the red circle.

Infectious immigrant numbers reflect both the infection history of the epidemic and the dispersal patterns of hosts and vectors. Intuitively, locations that had an early first date of local seal strandings – indicating that they local PDV epidemic occurred relatively early with respect to the regional North Sea epidemic – had small numbers of infectious immigrants (figure 4-3). These locations include harbor seal haulouts in Denmark, Kattegat and Skagerrak (figure 4-3). Locations with later dates of first local seal strandings, such as haulouts in the Wadden Sea, had larger numbers of infectious

immigrants (figure 4-3). However, immigration numbers also reflect the movement of hosts or vectors throughout the North Sea. Our findings indicate that areas of the Wadden Sea have high numbers of infectious immigrants while Scotland has the lowest numbers of infectious immigrants.

Discussion

Gravity models, and similar models that incorporate not only distance but also the attractiveness characteristics of competing destinations, have had a high degree of success describing disease dynamics and spatial spread of disease in human populations and systems with insect vectors or highly social animals (Xia *et al.* 2004; Ferrari *et al.* 2006b; Viboud *et al.* 2006). Since harbor seals are social animals that congregate on haulouts, and since PDV displayed patchy spatial spread in both epidemics, the gravity model formulation seemed to be a likely candidate to describe the topology of spatial spread in the North Sea. In addition, numerous models have been used to describe the spread of PDV in the North Sea in both the 1988 and 2002 epidemics (Grenfell *et al.* 1992; De Koeijer *et al.* 1998; Swinton 1998; Swinton *et al.* 1998; Harding *et al.* 2002; Hall *et al.* 2006; Harkonen *et al.* 2006; Klepac 2007), with the idea that more complex models were needed to adequately describe the disease dynamics (Hall *et al.* 2006). However, when comparing two models – the distance model and the gravity model – it is apparent that the distance model is sufficient to explain the disease dynamics and to lend insights into the seal host and/or vector dynamics.

The fact that the distance model is a better fit than the gravity model lends insight into PDV spatial spread, but it may also lend insight into harbor seal behavior. Perhaps the gravity model is needed to describe the host behavior of extremely social animals – such as humans – or the behavior of animals that visit many locations – such as foraging insects. But traditionally, harbor seals fall into neither of these categories. Harbor seals

have been known to interact with others of their species: mother and pup form a tight bond for weeks after birth, subadults display periods of social play, adults males aggressively interact, and adult males and females will mate (Renouf & Lawson 1986; Lawson & Renouf 1987; Renouf & Lawson 1987; Godsell 1988b). However, these seals spend much of their time in isolation, during periods of foraging, haul compared to highly social animals. It may be true that the lack of the high degree of sociality lacking in harbor seal populations permits the spatial spread of disease to follow a distance-based model of spread. In addition, other aspects of the harbor seal population may make the gravity model less applicable. For instance, the population sizes used in this study are more homogeneous than the ones used in previous studies, which show greater variation (Xia *et al.* 2004; Viboud *et al.* 2006). These factors may combine to explain some reasons behind the greater support for the distance model.

Alternatively, the result may have been confounded by the properties and uncertainties in the population size data. First, the haulouts were defined in a two-step process that includes both the location and status of harbor seal haulout sites and the sites where seal strandings occurred. This process may have led to true haulout sites to be missed or for haulout sites to be artificially created. Second, the population sizes are based on the reconstructed parameter (S_0) from the Bayesian state-space model (chapter 3), which may include some degree of error, since the true population size for the haulouts is unknown. Third, the model formulation assumes that there is no change in the social structure over the course of the epidemic, since the model uses the total number of seals (N) instead of the most recent population size (S_{t-1}). Since the 2002 epidemic is the second outbreak to occur in 14 years, it is possible that the entire population may not be susceptible; some may have conferred immunity from the earlier outbreak in 1988. Our estimates of immunity, extrapolated from life expectancy tables

(Harkonen *et al.* 2002), show that as much as 15% of the population could be immune. These uncertainties could combine to create sources of error in the model.

Lastly, this model is somewhat simplistic in the fact that it accounts for harbor seals, but not gray seals, which are also found throughout the North Sea (Sea Mammal Research Unit 2007). It is thought that gray seals are a vector species that carry the virus, remain relatively unaffected, and pass the virus to the common seals (Harkonen *et al.* 2006). Other support for vector gray seals include antibody data: they harbor antibodies in the interepidemic period and pups born after epidemic test positive for PDV antibodies (Barrett *et al.* 1995; Thompson *et al.* 2002). If gray seals are, in fact, vectors of the virus, the spatial dynamics may be driven by gray seal haulout population sizes and the distances between gray seal haulouts.

In conclusion, the distance model provides a better fit to the 2002 PDV epidemic stranding data in the North Sea than the gravity model, implying that the disease was more likely to spread to closer harbor seal haulouts. Factors that play a role in the spread of the disease include spatial coupling (θ), donor population sizes (N_j), and the seaway distance between haulouts. Together, these factors can quantify the number of infectious immigrants in the system, which was highest for the Dutch and German areas of the Wadden Sea. Future models to describe the system should include dynamics of the gray seal population, which may play a role in the spatial spread of PDV among harbor seals.

Acknowledgements

The authors would like to thank the volunteers and researchers that performed the population censuses and seal stranding counts. In addition, we would like to thank Petra Klepac and Jolianne Rijks for assistance with North Sea data sets, sources, and interpretations as well as general advice on the PDV epidemics that occurred in the

North Sea. We are also grateful to Angela Luis and Matt Ferrari for helpful discussions. Laura Pomeroy was supported by the National Science Foundation, under the NSF Graduate Teaching Fellowship in K-12 Education (DGE-0338240).

References

Abt K.F. (2002) Phanologie und Populationskynamik des Seehundes (*Phoca vitulina*) im Wattenmeer: Grundlangen zur Messung von Statusparametern In. Christian-Albrechts-Universitat zu Kiel, Kiel, Germany

Advisory Committee on Ecosystem (2001) Report of the Working Group on Marine Mammal Population Dynamics and Habitats. In. International Council for the Exploration of the Sea, Copenhagen, Denmark

Altizer S., Nunn C.L., Thrall P.H., Gittleman J.L., Antonovics J., Cunningham A.A., Dobson A.P., Ezenwa V., Jones K.E., Pedersen A.B., Poss M. & Pulliam J.R.C. (2003) Social organization and parasite risk in mammals: Integrating theory and empirical studies. *Annual Review of Ecology Evolution and Systematics*, 34, 517-547

Anderson R.M. & May R.M. (1982) The Control of Communicalbe Diseases by Age-Specific Immunization Schedules. *Lancet*, 1, 160

Anderson R.M. & May R.M. (1984) Spatial, Temporal, and Genetic Heterogeneity in Host Populations adn the Design of Immunization Programs. *Mathematical Medicine and Biology*, 1, 233-266

Anderson R.M. & May R.M. (1985) Age-Related Changes in the Rate of Disease Transmission: Implications for the Design of Vaccine Programmes. *Journal of Hygine (Cambridge)*, 94, 365-436

Anderson R.M. & May R.M. (1991) *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.

Bailey N.T.J. (1957) *The Mathematical Theory of Epidemics*. Griffin, New York.

Baker J.R. (1992) The Pathology of Phocine Distemper. *The Science of the Total Environment*, 115, 1-7

Baker J.R. & Ross H.M. (1992) The Role of Bacteria in Phocine Distemper. *The Science of the Total Environment*, 115, 9-14

Barrett T. (1999) Morbillivirus infections, with special emphasis on morbilliviruses of carnivores. *Veterinary Microbiology* 69, 3-13

Barrett T., Blixenkrone-Moller M., Di Guardo G., Domingo M., Duignan P., Hall A., Mamaev L. & Osterhaus A.D.M.E. (1995) Morbilliviruses in aquatic mammals - report on round table discussion. *Vet Microbiol*, 44, 261-265

Belisle C.J.P. (1992) Convergence Theorems for a Class of Simulated Annealing Algorithms. *Journal of Applied Probability*, 29, 855-895

Casella G. & George E.I. (1992) Explaining the Gibbs sampler. *American Statistician*, 46, 167-174

Cosby S.L., McQuaid S., Duffy N., Lyons C., Rima B.K., Allan G.M., McCullough S.J., Kennedy S., Smyth J.A., McNeilly F., Craig C. & Orvell C. (1988) Characterization of a Seal Morbillivirus. *Nature*, 336, 115-116

de Jong M., Diekmann O. & Heesterbeek J. (1994) The Compuation of R0 for Discrete-Time Epidemic Models with Dyanmic Heterogeneity. *Mathematical Biosciences*, 97-114

De Koeijer A., Diekmann O. & Reijnders P. (1998) Modelling the spread of phocine distemper virus among harbour seals. *Bulletin of Mathematical Biology*, 60, 585-596

Diekmann O. & Heesterbeek J.A.P. (2000) *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation*. Wiley, New York, NY.

Diekmann O., Heesterbeek J.A.P. & Metz J.A.J. (1990) On the Definition and the Computation of the Basic Reproduction Ratio R0 in Models for Infectious Diseases in Heterogenous Populations. *Journal of Mathematical Biology*, 28, 365-382

Dietz R., Heide-Jorgensen M.-P. & Harkonen T. (1989) Mass Deaths of Harbor Seals *(Phoca Vitulina)* in Europe. *Ambio*, 18, 258-264

Dobson A. (2004) Population dynamics of pathogens with multiple host species. *American Naturalist*, 164, S64-S78

Domingo M., Vilafranca M., Visa J., Prats N., Trudgett A. & Visser I. (1995) Evidence for chronic morbillivirus infection in the Mediterranean striped dolphin (Stenella coeruleoalba). *Veterinary Microbiology*, 44, 229-239

Domingo M., Visa J., Pumarola M., Marco A.J., Ferrer L., Rabanal R. & Kennedy S. (1992) Pathological and immunocytochemical studies of morbillivirus infection of striped dolphins (Stenella coeruleoalba). *Veterinary Pathology* 29, 1-10

Donnelly P.J. T.S. (1995) Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.*, 29, 401 - 442

Drummond A.J. R.A., Shapiro B., and Pybus O.G. (2005) Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Mol Biol Evol*, 22, 1185-1192

Drummond A.J., Ho S.Y.W., Phillips M.J. & Rambaut A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4, e88

Drummond A.J., Nicholls G.K., Rodrigo A.G. & Solomon W. (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, 161, 1307-1320

Drummond A.J., Pybus O.G., Rambaut A., Forsburg R. & Rodrigo A.G. (2003) Measurably evolving populations. *TRENDS in Ecolgy & Evolution*, 18, 481-488

Duck C.D. & Thompson D. (2002) The status of the British common seal populatons. In: *Special Committee on Seals (SCOS) 2002*, pp. 35-38. Sea Mammal Research Unit, St. Andrews, Fife, Scotland

Duck C.D., Thompson D. & B.L. M. (2007) The status of the British common seal populations. In: *Special Committee on Seals (SCOS) 2007*, pp. 50-61. Sea Mammal Research Unit, St. Andrews, Fife, Scotland

Edmunds W.J., O'Callaghan C.J. & Nokes D.J. (1997) Who Mixes with Whom? A Method to Determine the Contact Patterns of Adults that May Lead to the Spread of Airborne Infections. *Proceedings of the Royal Society B*, 264, 949-957

Ferguson N.M., Donnelly C.A. & Anderson R.M. (2001) The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science*, 292, 1155-1160

Ferrari M.J., Bansal S., Meyers L.A. & Bjornstad O.N. (2006a) Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society B*, 273, 2743-2748

Ferrari M.J., Bjornstad O.N., Partain J.L. & Anonovics J. (2006b) A Gravity Model for the Spread of a Pollinator-Borne Plant Pathogen. *The American Naturalist*, 168, 294-303

Ferrari M.J., Bjørnstad O.N. & Dobson A.P. (2005) Estimation and inference for R0 of an infectious disease using a removal method. *Mathematical Biosciences*, 198, 14-26

Ferrari M.J., Grais R.F., Bhart i.N., Conlan A.J., Bjørnstad O.N., Wolfson L.J., Guerin P.J., Djibo A. & Grenfell B.T. (2008) The dynamics of measles in sub-Saharan Africa. *Nature*, 451, 679-684

Fraser C., Riley S., Anderson R.M. & Ferguson N.M. (2004) Factors That Make a Disease Outbreak Controllable. *Proceedings of the National Academey of Sciences of the United States of America*, 101, 6146-6151

Godsell J. (1988a) Herd Formation and Haul-out Behavior in Harbor Seals *(Phoca vitulina)*. *Journal of Zoology, London*, 215, 83-98

Godsell J. (1988b) Herd Formation and Haul-out Behavior in Harbor Seals *(Phoca vitulna)*. *Journal of Zoology, London*, 215, 83-98

Grachev M.A., Kumarev V.P., Mamaev L.V., Zorin V.L., Baranova L.V., Denikina N.N., Belikov S.I. & Petrov E.A. (1989) Distemper virus in Baikal seals. *Nature*, 338, 209

Grenfell B. & Harwood J. (1997) (Meta)population dynamics of infectious diseases. . *Trends in Ecology and Evolution* 12, 395-399

Grenfell B.T. & Dobson A.P. (1995) *Ecology of infectious diseases in natural populations*. Cambridge University Press, Cambridge.

Grenfell B.T., Lonergan M.E. & Harwood J. (1992) Quantitative Investigations of the Epidemiology of Phocine Distemper Virus (PDV) in European Common Seal Populations. *Science of the Total Environment*, 115, 15-29

Hall A.J., Jepson P.D., Goodman S.J. & Harkonen T. (2006) Phocine distemper virus in the North and European Seas - Data and models, nature and nurture. *Biological Conservation*, 131, 221-229

Hall A.J., Pomeroy P.P. & Harwood J. (1992) The Descriptive Epizootiology of Phocine Distemper in the Uk During 1988/89. *Science of the Total Environment*, 115, 31-44

Harder T., Willhaus, T.H., Frey, H.-R., and Liess, B. (1990a) Mobillivirus Infections of Seals during the 1988 Epidemic in the Bay o Heligoland: III. Transmission Studies of Cell Culture-Propagated Phocine Distemper Virus in Harbor Seals (Phoca vitulina) and a Grey Seal (Halichoerus grypus): Clinical, Virological and Serological Results. *Journal of Veterinary Medicine*, 37, 641-650

Harder T., Willhaus, T.H., Frey, H.-R., and Liess, B. (1990b) Morbillivirus Infections of Seals during the 1988 Epidemic in the Bay o Heligoland: III. Transmission Studies of Cell Culture-Propagated Phocine Distemper Virus in Harbor Seals

(Phoca vitulina) and a Grey Seal (Halichoerus grypus): Clinical, Virological and Serological Results. *Journal of Veterinary Medicine*, 37, 641-650

Harding K.C., Harkonen T. & Caswell H. (2002) The 2002 European seal plague: epidemiology and population consequences. *Ecology Letters*, 5, 727-732

Harkonen L., Dietz R., Reijnders P., Teilmann J., Harding K., Hall A., Brasseur S., Siebert U., Goodman S.J., Jepson P.D., Rasmussen T.D. & Thompson P. (2006) A Review of the 1988 and 2002 phocine distemper virus epidemics in European harbour seals. *Diseases of Aquatic Organisms*, 68, 115-130

Harkonen T., Harding K. & Heide-Jorgensen M.-P. (2002) Rates of Increase in Age-Structured Populatons: A Lesson from the European Harbour Seals. *Canadian Journal of Zoology*, 80, 1498-1510

Harkonen T., Harding K. & Lunneryd S.G. (1999) Age- and Sex-Specific Behaviour in Harbor Seals *Phoca vitulina* Leads to Biased Estimates of Vital Population Parameters. *Journal of Applied Ecology*, 36, 825-841

Harkonen T. & Heide-Joergensen M.P. (1990) Comparative life histories of East Atlantic and other harbour seal populations. *Ophelia*, 32, 211-235

Haynes K.E. & Fotheringham A.S. (1984) *Gravity and Spatial Interaction Models*. Sage Publications, Beverly Hills, California.

Heide-Joergensen M.-P. & Harkonen T. (1988) Rebuilding seal stocks in the Kattegat-Skagerrak. *Marine Mammal Science*, 4, 231-246

Heide-Jorgensen M.-P. & Harkonen T. (1992) Epizootiology of the Seal Disease in the Eastern North Sea. *Journal of Applied Ecology*, 29, 99-107

Holland J., Spindler K., Horodyski F., Grabau E., Nichol S. & VandePol S. (1982) Rapid evolution of RNA genomes. *Science*, 215, 1577-1585

Huang Y. & Rohani P. (2006) Age-Structured Effects and Disease Interference in Childhood Infections. *Proceedings of the Royal Society B*, 273, 1229-1237

Hudson R.R. (1990) Gene Genealogies and the Coalescent Process. *Oxf. Surv. Evol. Biol.*, 7, 1-44

Jenkins G.M., Rambaut A., Pybus O.G. & Holmes E.C. (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *Journal of Molecular Evolution*, 54, 156-165

Jensen T., van de Bildt M., Dietz H.H., Andersen T.H., Hammer A.S., Kuiken T. & Osterhaus A. (2002) Another phocine distemper outbreak in Europe. *Science*, 297, 209-209

Kanaan M.N. & Farrington C.P. (2005) Matrix Models for Childhood Infections: A Bayesian Approach with Applications to Rubella and Mumps. *Epidemiology and Infection*, 133, 1009-1021

Kennedy S. (1998) Morbillivirus Infections in Aquatic Mammals *Journal of Comparative Pathology*, 119, 201-225

Kennedy S., Kuiken T., Jepson P.D., Deaville R., Forsyth M., Barrett T., van de Bildt M.W., Osterhaus A.D., Eybatov T., Duck C., Kydyrmanov A., Mitrofanov I. & Wilson S. (2000) Mass Die-Off of Caspian Seals Caused by Canine Distemper Virus. *Emerging Infectious Diseases*, 6, 637-639

Kermack W.O. & McKendrick A.G. (1927) Contributions to the mathematical theory of epidemics - I. *Proceedings of the Royal Society of Edinburgh A*, 115, 700-721

Kingman J.F.C. (1982) The coalescent. *Stochastic Process. Appl*, 13, 235 - 248

Klepac P. (2007) Mathematical Epidemiology of Phocine Distemper Virus in Harbor Seals. In: *Biological Oceanography*. MIT/WHOI, Woods Hole, MA

Lamb R.A. & Kolakofsky D. (2001) Paramyxoviradae: The Viruses and Their Replication. In: *Fields Virology* (eds. Knipe DM & Howley PM), pp. 1305-1340. Lippincott Williams & Wilkins, Philadelphia

Lawson J.W. & Renouf D. (1987) Bonding and Weaning in Harbor Seals, *Phoca vitulina*. *Journal of Mammology*, 68, 445-449

Legendre P. (1992) Real Data are Messy. *Statistics and Computing*, 3, 197-199

Lipscomb T.P., Kennedy S., Moffett D. & Ford B.K. (1994a) Morbilliviral disease in an Atlantic bottlenose dolphin (Tursiops truncatus) from the Gulf of Mexico. *Journal of Wildlife Diseases*, 30, 572-576

Lipscomb T.P., Schulman F.Y., Moffett D. & Kennedy S. (1994b) Morbilliviral disease in Atlantic bottlenose dolphins (Tursiops truncatus) from the 1987-88 epizootic. *Journal of Wildlife Diseases*, 30, 567-571

Lloyd-Smith J.O., Schreiber S.J., Kopp P.E. & Getz W.M. (2005) Superspreading and the effect of individual variation on disease emergence. *Nature*, 438, 355-359

Mahy B.W.J., Barrett T., Evans S., Anderson E.C. & Bostock C.J. (1988a) Characterization of a Seal Morbiliivirus. *Nature*, 336, 115

Mahy B.W.J., Barrett T., Evans S., Anderson E.C. & Bostock C.J. (1988b) Characterization of a Seal Morbillivirus. *Nature*, 336, 115

McCallum H., Barlow N. & Hone J. (2001) How should pathogen transmission be modelled? *TRENDS in Ecolgy & Evolution*, 16, 295-300

McCullagh P. & Nelder J.A. (1989) *Generalized linear models*. 511 edn. Chapman and Hall, New York, New York.

Morton A.M. & Finkenstädt B.F. (2005) Discrete-time modelling of disease incidence time series by using Markov Chain Monte Carlo methods. *Applied Statistics*, 54, 575-594

Muller G., Wunschmann A., Baumgartner W., Birkun A., Komakhidze A., Stanev T. & Joiris C.R. (2002) Immunohistological and serological investigations of morbillivirus infection in Black Sea harbour porpoises (Phocoena phocoena). *Veterinary Microbiology*, 87, 183-190

Murray J.D., Stanley E.A. & D.L. B. (1986) On the Spatial Spread of Rabies among Foxes. *Proceedings of the Royal Society of London Series B-Biological Sciences* 229, 111-150

Newby T.C. (1973) Observations on the breeding behavior of the Harbor Seal in the State of Washington. *Journal of Mammology*, 54, 540-543

Newman M.E.J. (2005) Threshold effects for two pathogens spreading on a network. *Phys. Rev. Lett.*, 95, 108701

Osterhaus A.D., Uytdehaag, F.G., Visser, I.K., Vedder, E.J., Reijnders, P.J., Kuiper, J., and Brugge, H.N. (1989) Seal Vaccination Success. *Nature*, 337, 21

Osterhaus A.D.M.E. & Vedder E.J. (1988) Identification of Virus Causing Recent Seal Deaths. *Nature*, 355, 20

Panum P.L. (1940) Observations Made During the Epidemic of Measles on the Faroe Islands in the Year 1846. *Delta Omega*

Perkins S.E., Cattadori I.M., Tagliapietra V., Rizzoli A.P. & Hudson P.J. (2003) Empirical evidence for key hosts in persistence of a tick-borne disease. *International Journal for Parasitology*, 33, 909-917

R Development Core Team (2006) R: A Language and Environment for Statistical Computing. In: (ed. Computing RFfS), Vienna, Austria

Reijnders P.J.H. & Lankester K. (1990) Status of Marine Mammals in the North Sea. *Netherlands Journal of Sea Research*, 26, 427-435

Reinekinig B. (2002) Phocine Distemper Virus amongst Seals in 2002. In: *Wadden Sea Newsletter* Common Wadden Sea Secretariat, Wilhelmshaven, Germany

Renouf D. & Lawson J.W. (1986) Play in Harbor Seals *(Phoca vitulina)*. *Journal of Zoology, London*, 208, 73-86

Renouf D. & Lawson J.W. (1987) Quantative Aspects of Harbour Seals *(Phoca vitulina)* Play. *Journal of Zoology, London*, 212

Ries E.H., Hiby L.R. & Reijnders P.J.H. (1998) Maximum Likelihood Population Size Estimation of Harbour Seals in the Dutch Wadden Sea Based on a Mark Recapture Experiment. *Journal of Applied Ecology*, 35, 332-339

Ries E.H., Traut I.M., Brinkman A.G. & Reijnders P.J.H. (1999) Net dispersal of harbour seals within the Wadden Sea before and after the 1988 epizootic. *Journal of Sea Reasearch*, 41, 233-244

Rijks J.M., Van de Bildt M.W.G., Jensen T., Philippa J.D.W., Osterhaus A. & Kuiken T. (2005) Phocine distemper outbreak, the Netherlands, 2002. *Emerging Infectious Diseases*, 11, 1945-1948

Rima B.K., Curran M.D. & Kennedy S. (1992) Phocine Distemper Virus, the Agent Responsible for the 1988 Mass Mortality of Seals. *Science of the Total Environment*, 115, 45-55

Schenzle D. (1984) An Age-Structured Model of Pre- and Post-Vaccination Measles Transmission. *Mathematical Medicine and Biology*, 1, 169-191

Sea Mammal Research Unit (2007) Special Committee on Seals Main Advice. In:

Smith D.L., Lucey B., Waller L.A., Childs J.E. & Real L.A. (2002) Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3668-3672

Sullivan R.M. (1982) Antagonistic Behavior and Dominance Relationships in the Harbor Seals. *Journal of Mammology*, 63, 544-569

Swinton J. (1998) Extinction times and phase transitions for spatially structured closed epidemics. *Bulletin of Mathematical Biology*, 60, 215-230

Swinton J., Harwood J., Gilligan C.A. & Hall A.J. (1999) Scaling of phocine distemper virus transmission with harbour seal community size. *Ecologie*, 30, 231-240

Swinton J., Harwood J., Grenfell B.T. & Gilligan C.A. (1998) Persistence thresholds for phocine distemper virus infection in harbour seal Phoca vitulina metapopulations. *Journal of Animal Ecology*, 67, 54-68

Thompson D., Lonergan M. & Duck C. (2005) Population dynamics of harbour seals *Phoca vitulina* in England: monitoring growth and catastrophic declines. *Journal of Applied Ecology*, 42, 638-648

Thompson P.M., Fedak M.A., McConnell B.J. & Nicholas K.S. (1989) Seasonal and Sex-Related Variation in the Activity Patterns of Common Seals *(Phoca vitulina)*. *Journal of Applied Ecology*, 26, 521-535

Thompson P.M. & Miller D. (1992) Phocine distemper virus outbreak in the Moray Firth common seal population: an estimate of mortality. *The Science of the Total Environment*, 115, 57-65

Thompson P.M., Thompson H. & Hall A. (2002) Prevalence of morbilliviurs antibodies in Scottish seals. *Vet Rec*, 151, 609-610

Thompson P.M., Tollit D.J., Wood D., Corpe H.M., Hammond P.S. & Mackay A. (1992) Estimating Harbour Seal Abundance and Status in an Estuarine Habitat in North-East Scotland. *Journal of Applied Ecology*, 34, 43-52

Traut I.M. (1999a) Spacing among Harbor Seals *(Phoca vitulina viulina)* on Haul-out Sites in the Wadden Sea of Niedersachsen. *Z. Sugetierkunde*, 64, 51-53

Traut I.M. (1999b) Spacing among Harbor Seals *(Phoca vitulina viulina)* on Haul-out Sites in the Wadden Seaw of Niedersachsen. *Z. Sugetierkunde*, 64, 51-53

Trilateral Seal Expert Group (2000) Common Seals in the Wadden Sea in 1999. In: *Wadden Sea Newsletter*, Wilhelmshaven, Germany

Trilaterial Seal Expert Group (2001) Common Seals in the Wadden Sea in 2001. In: *Wadden Sea Newsletter 2001*, p. 3. Common Wadden Sea Secretariat

van den Driessche P. & Watmough J. (2002) Reproduction Numbers and Subthreshold Endemic Equilibria for Compartmental Models of Disease Transmisison. *Mathematical Biosciences*, 180, 32-34

Viboud C., Bjørnstad O.N., Smith D.L., Simonsen L., Miller M.A. & Grenfell B.T. (2006) Synchrony, waves and spatial hierarchies in the spread of influenza. *Science*, 312, 447-451

Wallinga J., Teunis P. & Kretzschmar M. (2006) Using Data on Social Contacts to Estimate Age-Specific Transmission Parameters for Respirtatory-Spread Infectious Agents. *American Journal of Epidemiology*, 164, 936-944

Wilson S. (1974) Juvenile Play of the Common Seal, *Phoca vitulina vitulina*, with Comparative Notes on the Gray Seal *Halichoerus grupus*. *Behaviour*, 48, 37-60

Wolfram Research I. (2007) Mathematica. In, Champaign, IL

Woolhouse M.E.J., Dye C., Etard J.-F., Smith T., Charlwood J.D., Garnett G.P., Hagan P., Hii J.L.K., Ndhlovu P.D., Quinnel R.J., Watts C.H., Chandiwana S.K. & Anderson R.M. (1997) Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academey of Sciences of the United States of America*, 94, 338-342

Xia Y.C., Bjornstad O.N. & Grenfell B.T. (2004) Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *American Naturalist*, 164, 267-281

The Evolutionary and Epidemiological Dynamics of the *Paramyxoviridae*

Laura W. Pomeroy[1], Ottar N. Bjørnstad[1,2,3] & Edward C. Holmes[1,2]

[1]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA.

[2]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA.

[3]Department of Entomology, The Pennsylvania State University, University Park, PA 16802, USA.

Abstract. Paramyxoviruses are responsible for considerable disease burden in human and wildlife populations: measles and mumps continue to affect the health of children worldwide, while canine distemper virus causes serious morbidity and mortality in a wide range of mammalian species. Although these viruses have been studied extensively at both the epidemiological and phylogenetic scales, little has been done to integrate these two types of data. Using a Bayesian coalescent approach, we infer the evolutionary and epidemiological dynamics of measles, mumps and canine distemper viruses. Our analysis yielded data on viral substitution rates, the time to common ancestry and elements of their demographic history. Estimates of rates of evolutionary change were similar to those observed in other RNA viruses, ranging from 6.585 to 11.350 x 10^{-4} nucleotide substitutions per site, per year. Strikingly, the mean Time to the Most Recent Common Ancestor (TMRCA) was both similar and very recent among the viruses studied, ranging from only 58 to 91 years (1908 to 1943). Worldwide, the paramyxoviruses viruses studied here have maintained a relatively constant level of genetic diversity. However, detailed heterchronous samples illustrate more complex dynamics in some epidemic populations, and the relatively low levels of genetic diversity (population size) in all three viruses is likely to reflect the population bottlenecks that follow recurrent outbreaks.

Introduction

Viral epidemics caused by negative-sense viruses of the family *Paramyxoviridae* have plagued animal populations for millennia. Three of these viruses – measles, mumps, and canine distemper virus – affect animals ranging from livestock, canines, and humans (Lamb and Kolakofsky 2001; Fauquet et al. 2005), and have had a major economic and demographic impact in these species. A major disease burden is also associated with rinderpest in cattle, Newcastle disease in poultry, and phocine distemper in seals (Lamb and Kolakofsky 2001; Barrett 1994; Griffin 2001).

The first documented outbreaks of measles date to the 2^{nd} – 4^{th} centuries in China and the Roman Empire (Orvell 1994), while the closest relative of measles virus – rinderpest virus – affected cattle even earlier (Carbone and Wolinsky 2001). Measles virus (genus *Morbillivirus*) affects nearly 30 million people annually, with a mortality of 454,000 (WHO 2007), largely in developing countries (WHO 2007). In populations with recurring multi-annual epidemics measles cases are usually clustered among children and spread through aerosol droplets (Anderson and May 1991; Grenfell and Harwood 1997; Griffin 2001). The virus causes acute systematic disease (Orvell 1994) and confers life-long immunity on those who recover (Griffin 2001; Panum 1939).

At the epidemiological level, measles epidemics display violent cyclic dynamics with a 1 – 5 year periodicity of outbreaks in a given location (Orvell 1994). Two critical epidemiological parameters have been estimated for this important human pathogen: the reproductive number (R_0) and the critical community size (CCS). R_0 is the number of secondary cases that arise from each primary case in a totally susceptible population and estimates of this parameter for measles range from 15 to 20 (Griffin 2001). CCS is the minimum number of individuals in a population required for the disease to persist in a population, and has been estimated as 250,000 – 500,000 for measles (Griffin 2001; Bartlett 1957). At the phylogenetic level, variability in the hemagglutinin (H) glycoprotein identifies eight major groups with 15 component

genotypes, and which exhibit temporal and spatial segregation (Griffin 2001; Bellini and Rota 1998).

Mumps, the viral agent of which is classified within the genus *Rublavirus*, was first described by Hippocrates in the 5th century BC (Carbone and Wolinsky 2001; Rima 1994) and spreads through aerosol droplets (Carbone and Wolinsky 2001). Mumps virus, like measles virus, has a very narrow host range with humans the only species that can actively transmit the virus, and other species acting as dead-end hosts (Rima 1994). Epidemics of mumps occur in both developed and developing countries due to a lack of vaccination campaigns or vaccine failure (Carbone and Wolinsky 2001). These epidemics also display cyclical dynamics with annual outbreaks. However, the periodicity of these outbreaks can vary by as much as 2 to 7 years (Rima 1994). At the population level, R_0 has been estimated at 5 (Rima 1994) and CCS = 200,000 with school terms forcing epidemic patterns (Carbone and Wolinsky 2001). Analyses of genetic variability in the hemagluttinin-neuraminidase (HN) proteins have designated 11 circulating genotypes (Orvell et al. 2002).

While canine distemper virus (CDV) shares some similarities with measles virus, including its taxonomic identification within the genus *Morbillivirus*, it seems to have emerged more recently, with the first case described in 1905 (Griffin 2001). CDV also has a broader host range and can infect many mammalian species, including members of the *Felidae* (cats), *Canidae* (dogs, foxes, etc.), *Mustelidae* (weasles), and *Proconidae* (raccoons) (Barrett 1994; Griffin 2001). Although a vaccine was developed for domestic dogs in 1950, limited use means that CDV remains prevalent in many populations (Barrett 1994). CDV also displays a high degree of genetic diversity that rivals, or exceeds, that of measles virus (Bolt et al. 1997); again, genetic diversity is greatest in the hemagluttinin (H) protein and displays spatial differentiation (Mochizuki et al. 1999).

Together, the measles, mumps, and canine distemper viruses are among some of the most thoroughly studied viruses at the epidemiological scale, and all three viruses exhibit major

population bottlenecks due to their cyclic epidemiological dynamics. However, little work has been done to determine how these epidemiological dynamics affect their phylogenetic patterns as inferred from gene sequence data. By performing a Bayesian coalescent analysis using serially-sampled gene sequence data (Drummond et al. 2002; Drummond et al. 2006)(Kingman 1982; Hudson 1990; Donnelly P.J. 1995), we are able to estimate key demographic and evolutionary parameters, and gain important insights into the evolution and epidemiology of these viral pathogens.

Methods

Data Sets

All viral sequences were downloaded from GenBank and manually aligned using the Se-Al program (Rambaut 1996). Alignment files for all data sets are available from the authors on request.

For the analysis of measles virus, both the H and N (nucleoprotein) genes were utilized. H gene sequences were collected from localities worldwide, notably Africa. Once vaccine strains and those associated with the persistent disease manifestation subacute sclerosing panencephalitis (SSPE) cases were removed, as these are expected to exhibit different evolutionary dynamics (Woelk et al. 2002), 215 taxa remained with an alignment length of 1851 bp. Due to computational constraints associated with the coalescent analysis of large numbers of sequences, 120 taxa were randomly selected from this larger data set. To ensure that this sub-sampling introduced no bias into the analysis, it was performed an additional five times independently and all coalescent analyses were run on these five sub-sampled data sets. Similarly, we compiled a global sample of measles virus N gene sequences. Again, vaccine strains and those associated with SSPE were removed, which resulted in a final data set of 107 taxa, 1575 bp. The mumps data set comprised HN sequences collected globally, although with a particular emphasis on sequences from Asia. Once the vaccine strains were removed, 27

taxa remain for analysis, 1746 bp in length. Finally, in the case of CDV, H gene sequences were sampled from Japan, the USA, and Western Europe. Vaccine strains and passaged strains were removed from the data set, resulting in 35 taxa of 1949 bp.

Evolutionary Analysis

Rates of nucleotide substitution per site, the Time to the Most Recent Common Ancestor (TMRCA), and key aspects of demographic history, particularly changes in genetic diversity (an indicator of population size under strictly neutral evolution) quantified as $N_e\tau$ where N_e is the effective number of infections and τ is the infection-to-infection generation time, were estimated using a Bayesian Markov Chain Monte Carlo (MCMC) method available in the BEAST package (Drummond et al. 2002; Drummond and Rambaut 2003). This method analyzes the distribution of branch lengths among viruses isolated at different times (year of collection) among millions of sampled trees. A variety of models of demographic history were investigated – constant population size, exponential population growth, logistic population growth, expansion population growth – and incorporating both relaxed (uncorrelated exponential) and constant molecular clocks (Drummond et al. 2006). For each data set, the best-fit model of nucleotide substitution was determined using MODELTEST (Posada and Crandall 1998). In the case of the two measles virus data sets and the mumps HN gene the favored model was a close relative of the most general GTR+I+G_4 model. For the CDV H gene a simpler GTR+I model was preferred. All models were compared using Akaike's Information Criterion (AIC). To infer demographic histories we depicted the changing profile of N_et through time in a Bayesian skyline plot (Drummond et al. 2005)(Drummond A.J. 2005). In all cases statistical uncertainty in parameter values across the sampled trees is given by the 95% Highest Probability Density (HPD) values. For all models, chains were run until convergence was achieved (as assessed using the TRACER program; http://evolve.zoo.ox.ac.uk/software.html?id=tracer). The BEAST analysis

was also used to infer a maximum *a posteriori* (MAP) tree for each data set, in which tip times correspond to the year of sampling.

Gene and site-specific selection pressures for all four data sets were measured as the ratio of nonsynonymous (d_N) to synonymous substitutions (d_S) per site (d_N/d_S), estimated using the Single Likelihood Ancestor Counting (SLAC) and Random Effects Likelihood (REL) maximum likelihood methods available at the Datamonkey facility (Kosakovsky Pond and Frost 2005). Because of the large number of sequences in the measles virus data sets, our analysis in this case were restricted to the SLAC method. In all cases we utilized the general reversible substitution (GTR) model with input neighbor-joining trees.

Results

Evolutionary dynamics of the paramyxoviruses

Maximum *a posteriori* (MAP) trees were inferred for each paramyxovirus data set and are shown in Figures 5-1a – 5-4a (equivalent maximum likelihood trees are available from the authors on request). In each case the temporal structure in the data is reflected in the diversity of tip times, which provide a means of estimating evolutionary dynamics.

Rates of evolutionary change, measured as the number of nucleotide substitutions per site, per year (subs/site/year) were estimated using a Bayesian coalescent method (Table 5-1). In all cases a relaxed (uncorrelated exponential) molecular clock was a better fit to the data than a strict molecular clock, and the best-fit demographic model was either constant population size or exponential population growth (see below), although parameter estimates were consistent among models. Notably, similar parameter values were estimated for all three viruses and with overlapping HPD values in all cases (Table 5-1).

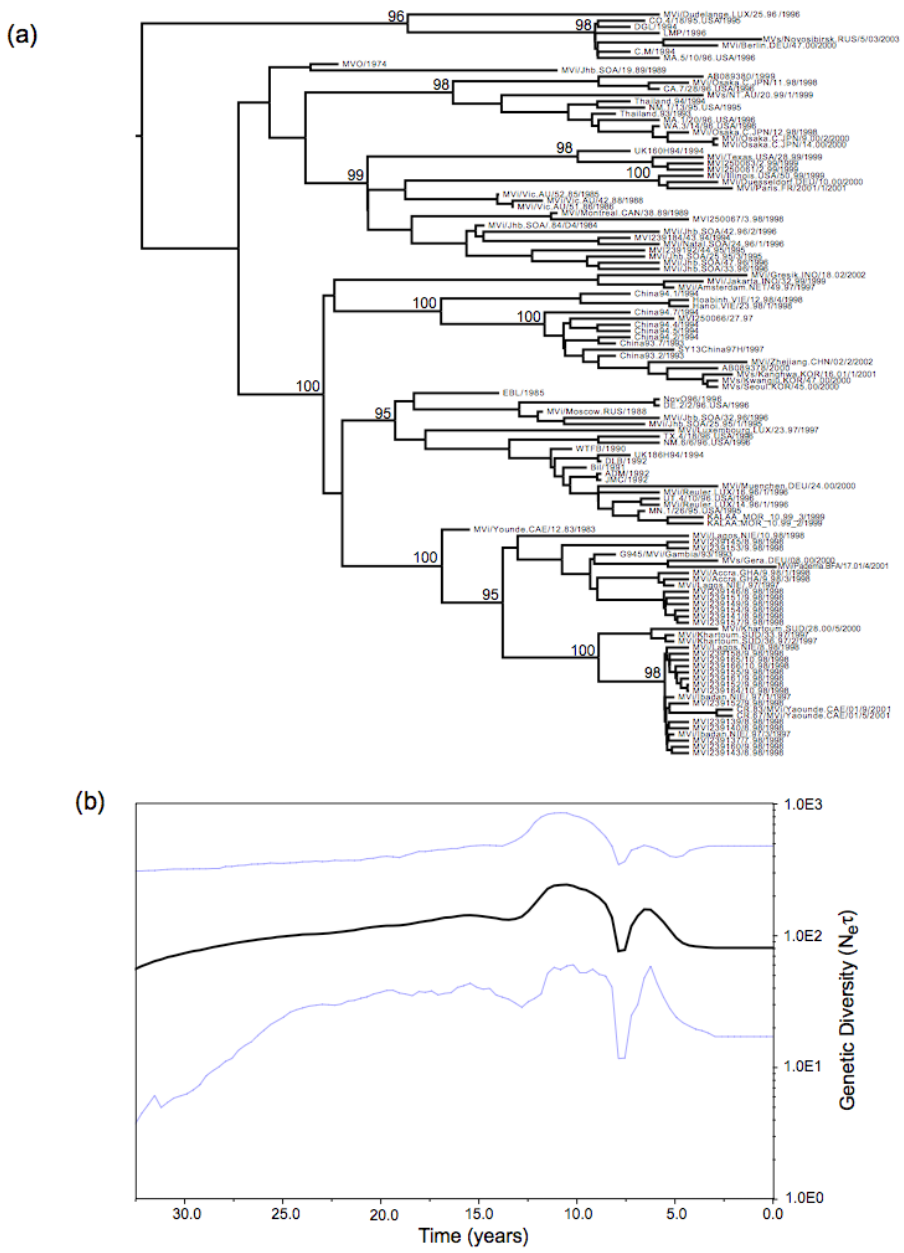

Figure 5-1. (a) Maximum *a posteriori* (MAP) tree of the H gene of measles virus. Tip times reflect the year of sampling (which is provided by the final four digits in the isolate name) and correspond to the time-scale provided in part (b) of the figure. (b) Bayesian skyline plot of the changing levels of genetic diversity ($N_e t$) of the measles virus H gene sampled between 1965-2003. The bold line represents the median estimate while the light lines depict the 95% HPD values.

Measles virus was found to have a mean substitution rate of 6.585 x 10^{-4} for the H gene and 8.693 x 10^{-4} for the N gene; the HN gene mumps virus exhibited a substitution rate of 9.168 x

10^{-4} subs/site/year, while the H gene of CDV had the highest mean substitution rate at 11.350 x 10^{-4} subs/site/year. In the case of the measles H gene, similar rates were found in all sub-sampled data sets (Table 5-1), indicating that sampling has not introduced major biases into these analyses. These rates are slightly higher than those previously estimated for the paramyxoviruses (Jenkins et al. 2002), although with overlapping sampling errors, which most likely reflects the use of strict molecular clocks in earlier analyses, and relaxed molecular clocks in the current study.

Using the same coalescent approach we also estimated the TMRCA (i.e. the age of the sampled genetic diversity) of each virus (Table 5-1). For the measles virus H gene, the mean TMRCA was ~60 years before the most recent sequence, collected in 2003, which equates to approximately 1943. The measles N gene analysis yielded a similar result: the mean TMRCA was ~77 years before the most recent sequence, isolated in 2003. The mean TMRCA in the HN gene of mumps virus was ~91 years before 1999, dating the MRCA to ~1908, although with wide HPD values reflecting the small sample size. Finally, for CDV, the mean TMRCA was ~58 years before the most recent sequence, sampled in 2001, suggesting that the ancestor of these sequences existed close to 1943. Despite differences in sample size, all analyses point to a similar time-frame for the age of the current genetic diversity in the measles, mumps and canine distemper viruses, equating to the first half of the 20th century.

Table 5-1**.** Bayesian estimates of substitution and demographic parameters in the paramyxoviruses studied here.

Virus	Gene	No. of Sequences	Date Range of Sequences	Molecular Clock	Demographic Model	Substitution Rate, 10^{-4} subs/site/year (95% HPD)	TMRCA, years (95% HPD)	TMRCA date (95% HPD)
Measles	H	120	1974 - 2003	Relaxed	Exponential growth	6.585	60	1943
						(4.792 – 8.306)	(36 - 90)	(1913 – 1967)
						[6.341 – 8.154][a]	[57 – 63][a]	[1940 – 1946][a]
Measles	N	107	1977 - 2003	Relaxed	Constant	8.693	77	1926
						(5.886 – 11.270)	(37 - 140)	(1863 – 1966)
Mumps	HN	27	1945 - 1999	Relaxed	Constant	9.168	91	1908
						(4.832 – 14.170)	(58 – 149)	(1850 – 1941)
Canine Distemper	H	35	1982 - 2001	Relaxed	Constant	11.650	58	1943
						(5.438 – 18.050)	(27 – 107)	(1894 – 1974)

[a]Range of mean values in 5 additional sub-sampled data sets of 120 sequences.

Population Demography

The serially-sampled coalescent method available in the BEAST program also allows the inference of the modes and rates for population growth. However, because all three viruses undergo cyclic epidemiological dynamics and coalescent methods that deal, quantitatively, with such dynamics are currently unavailable, it is not possible accurately estimate population growth rates in these circumstances. We therefore chose to depict population dynamics using a Bayesian skyline plot which provides a piece-wise graphical depiction of changes in genetic diversity (population size) through time (Drummond et al. 2005). Ecologically, measles epidemics occur in multi-annual cycles; hints of this periodicity can be seen in the Bayesian skyline plots for this virus. First, in the case of the H gene (Figure 5-1b) a peak in population size occurs in ~1991 (although note the HPD values in each case), followed by a population bottleneck in ~1995. This peak in population size is also apparent in the 5 additional sub-sampled H gene data sets, although the size of the bottleneck varies in magnitude (not shown, available from the authors on request). An identical demographic signal is apparent in the Bayesian skyline plot of the N gene of measles virus (Figure 5-2b), with a sharp population decline clearly depicted. Notably, the timing of the population decline depicted in both genes corresponds with the advent of WHO vaccination campaigns in six southern African countries – Botswana, Malawi, Namibia, South Africa, Swaziland, and Zimbabwe (Uzicanin et al. 2002; WHO 1999) – and African sequences constitute a major component of our measles data sets. Although vaccination campaigns may cause a decrease in cases due to depletion of susceptible individuals, additional epidemics can follow when susceptible individuals are replenished through births or immigration. Given a more intensive sampling regime, in which multiple viruses are collected each year from individual localities, it should therefore be possible to recover the complex cyclic behavior of measles virus.

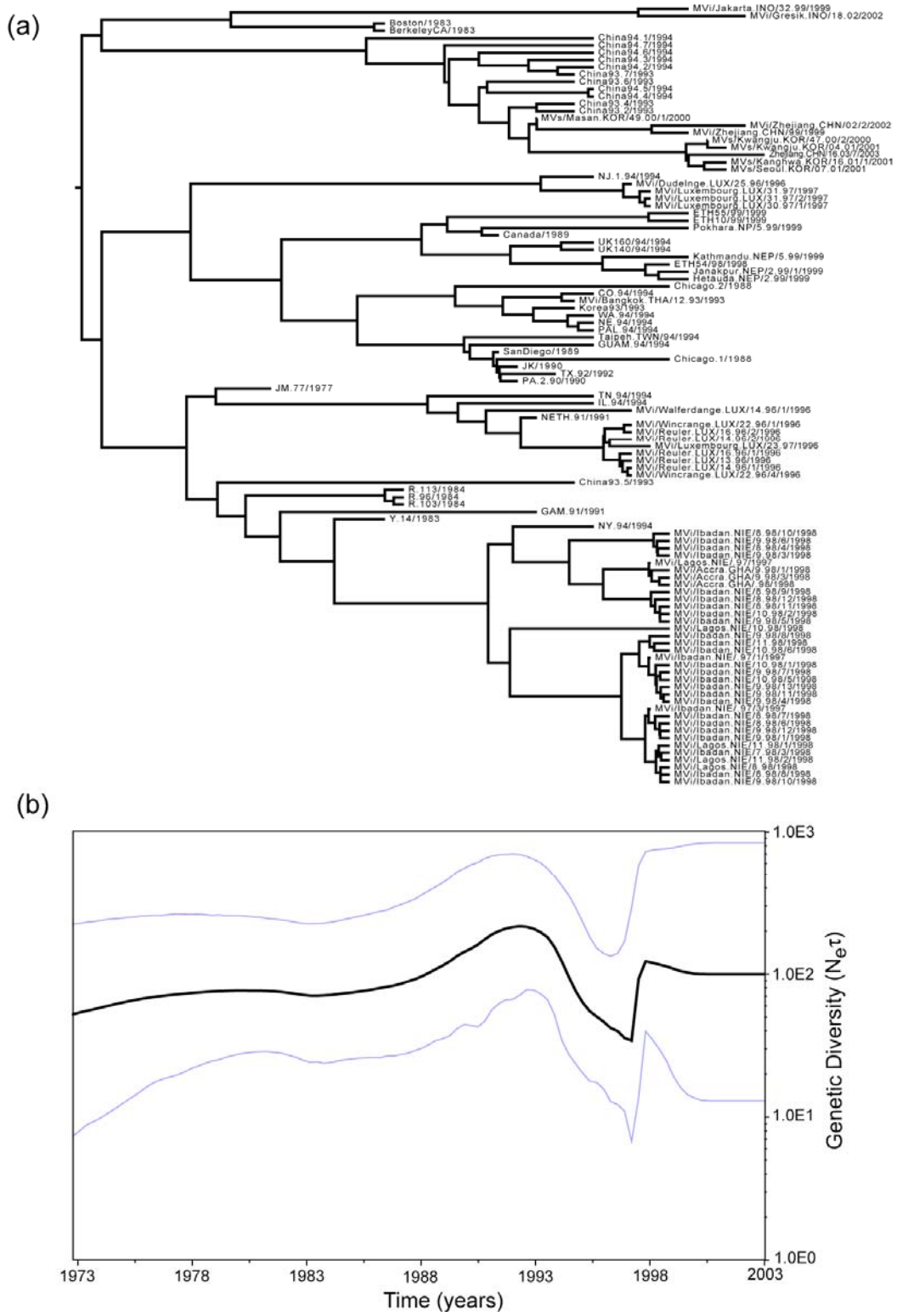

Figure5-2. (a) Maximum *a posteriori* (MAP) tree of the N gene of measles virus. (b) Bayesian skyline plot of the measles virus N gene sampled between 1962-2003.

In the case of mumps virus, the small sample size precludes a detailed analysis of population dynamics, although the Bayesian skyline plot reveals that N_et is consistently low (see below) (Figure 5-3b).

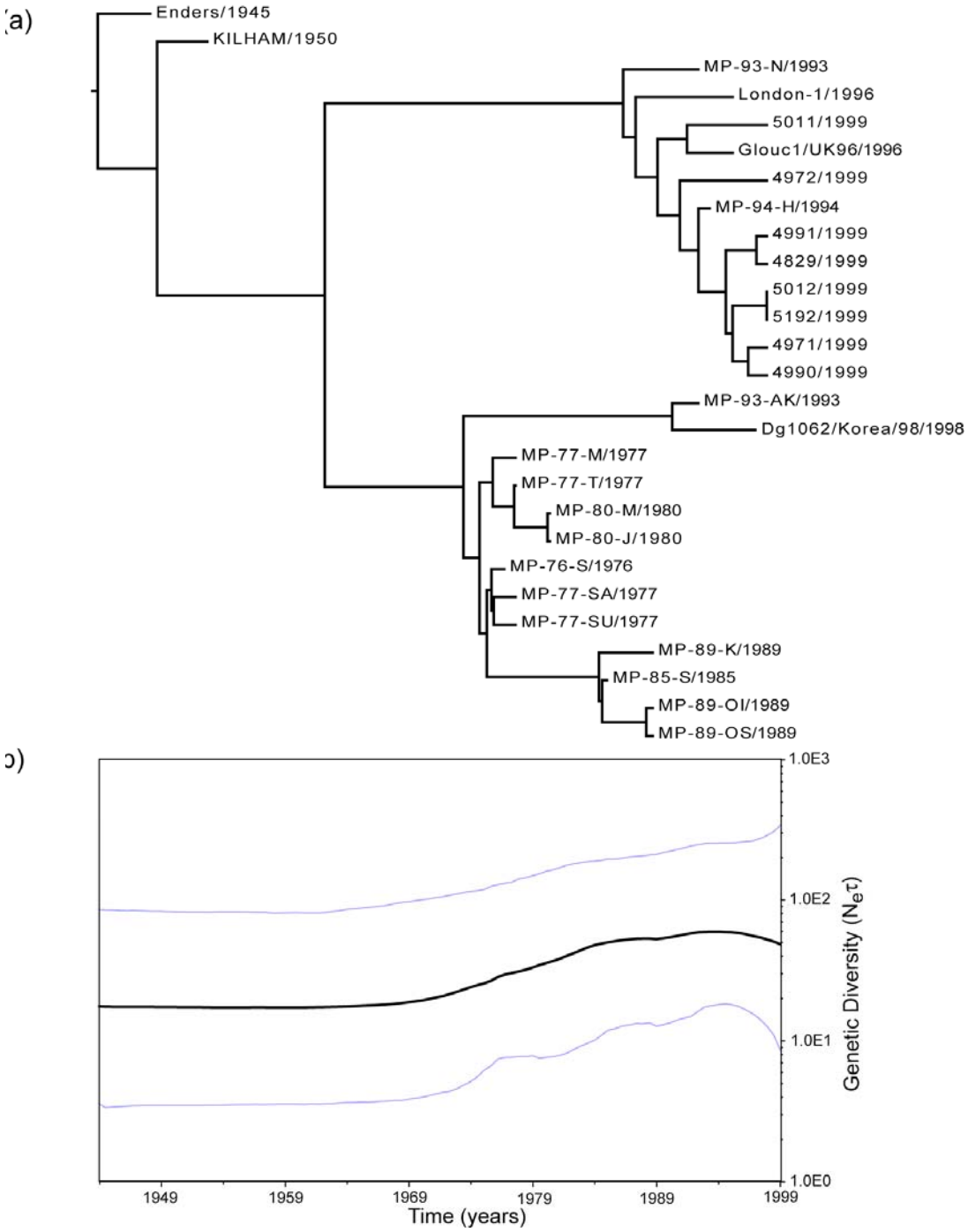

Figure 5-3. (a) Maximum *a posteriori* (MAP) tree of the HN gene of mumps virus. (b) Bayesian skyline plot of the mumps virus HN gene sampled between 1950-2000.

Similar constraints imposed by sample size also apply to CDV, the population size of which appears to be relatively constant in the Bayesian skyline plot and with a low $N_e t$ (Figure 5-4b).

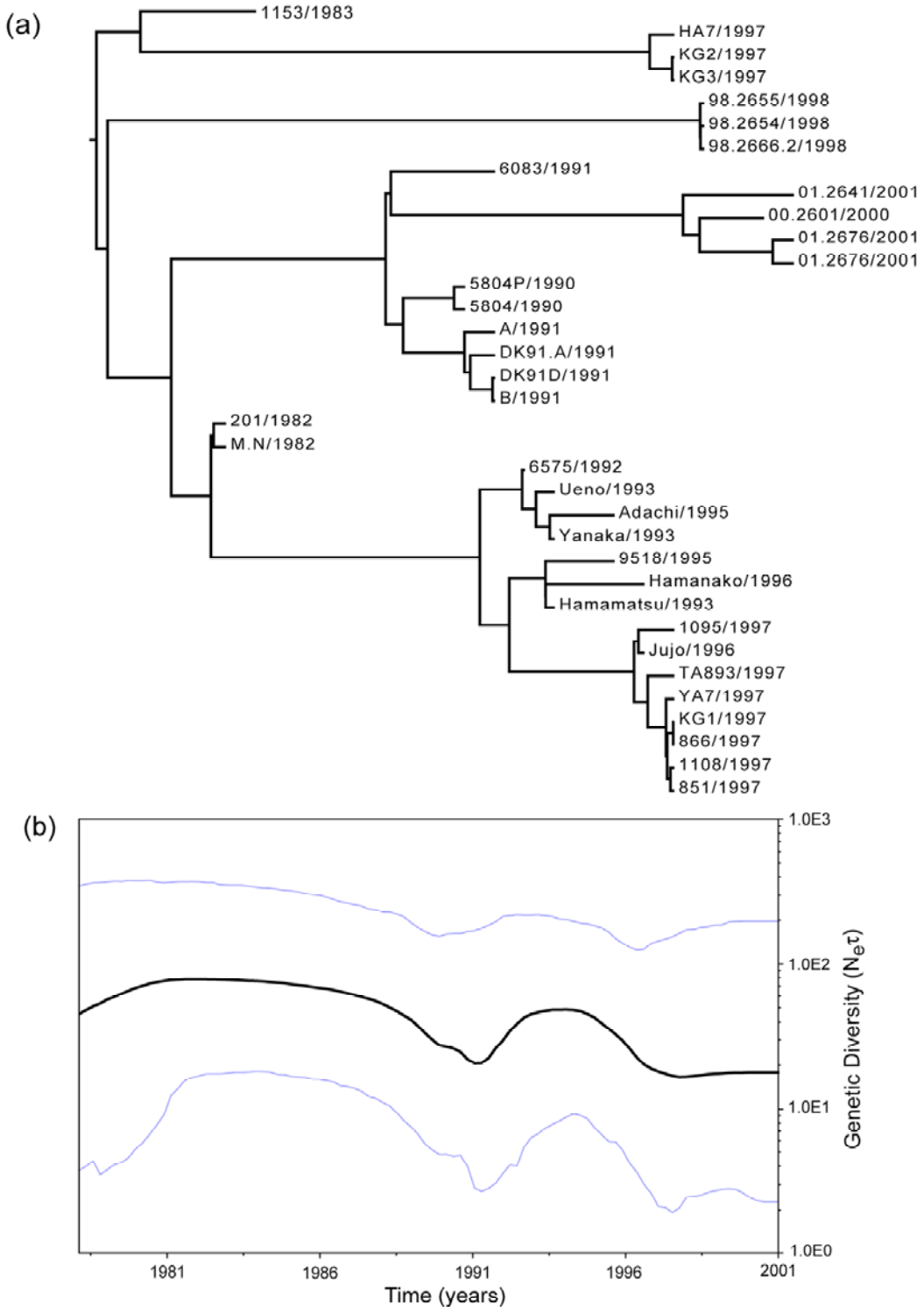

Figure 5-4. (a) Maximum *a posteriori* (MAP) tree of the H gene of canine distemper virus. (b) Bayesian skyline plot of the CDV H gene sampled between 1982-2001.

Strikingly, in all the viruses studied here our estimates of genetic diversity, measured as $N_e t$, are between 10-100 are hence considerably lower than those previously recorded in chronic infections such as HIV (Robbins et al. 2003) and hepatitis C (Pybus et al. 2001). Such consistently low levels of genetic diversity suggest that recurrent population bottlenecks, a characteristic of paramyxovirus population dynamics, are regularly purging, and hence limiting, genetic variation.

Selection Pressures

To obtain a measure of the selection pressures acting on the three paramyxoviruses studied here we computed the relative numbers of nonsynonymous (d_N) to synonymous (d_S) substitutions per site (ratio d_N/d_S) using two likelihood-based methods (SLAC and REL; Kosakovsky Pond and Frost 2005). This analysis revealed little evidence for positive selection, although an abundance of negatively selected sites (Table 5-2).

Table 5-2. Summary of selection pressures acting in the paramyxoviruses studied.

Data Set	Mean d_N/d_S	SLAC	SLAC	REL
		No. positively selected sites (codon position)[a]	No. negatively selected sites[a]	No. positively selected sites (codon position)[a]
Measles H	0.232	1 (476)	82	NA
Measles N	0.155	1 (329)	95	NA
Mumps HN	0.143	0	49	2 (353, 402)
Canine Distemper N	0.274	0	21	0

[a] $p < 0.1$ (no positively selected sites were detected at $p < 0.05$).

Overall, across all the sequences analyzed, only four sites were found to be positively selected at a significance value of $p < 0.1$; one site each for the measles H and N genes using SLAC, and two in the mumps HN gene under REL. No sites were found be selected in the CDV H gene. No positively selected sites in any virus were detected using a more stringent significance value of $p < 0.05$.

Discussion

Our results indicate that the nucleotide substitution rates of measles, mumps and canine distemper viruses resemble those of many other RNA viruses (Hanada et al. 2004; Jenkins et al. 2002). Measles virus has traditionally been thought to be a relatively conserved RNA virus, since it confers life-long immunity upon its hosts or those in the vaccinated class (Panum 1939), and because some earlier studies revealed low levels of both antigenic and genetic variation (Rota et al. 1992). These observations led to suggestions that measles – and its relatives within in the *Paramyxoviridae* – evolve anomalously slowly. However, our analyses indicate that all three paramyxoviruses studied here have substitution rates in the range 10^{-3} to 10^{-4} subs/site/year, and so typical of the rapid mutation and replication dynamics that define RNA viruses (Domingo and Holland 1997). It is therefore important to note that high rates of molecular evolutionary change do not necessarily translate into high levels of antigenic variation.

More striking was that the TMRCA of all three viruses was surprisingly recent and always within the last century. In principle, such shallow genetic diversity can be attributed to one of four causes: the erroneous estimation of substitution rates and divergence times, recent cross-species transmission, a large-scale and global population bottleneck which purged pre-existing genetic diversity, or neutral evolution within a small effective population. The first of these four theories – estimation error – seems unlikely

as substitution rates of between 10^{-3} and 10^{-4} subs/site/year have been estimated previously for the paramyxoviruses (Hanada et al. 2004; Jenkins et al. 2002; Woelk et al. 2001; Woelk et al. 2002) and for RNA viruses in general. Similarly, we regard recent cross-species transmission as unlikely since measles-like disease has been documented in human populations for millennia, associated with the start of human urbanization (Orvell 1994; Carbone and Wolinsky 2001; Rima 2004).

A more reasonable explanation for the recent dates of common ancestry is a large-scale and geographically extensive population bottleneck (as opposed to local cyclical dynamics), either neutral or selectively determined, which would have purged most standing genetic variation. Although possible, it seems unlikely that such bottlenecks would occur in all the paramyxoviruses studied here on roughly the same time-scale. Further, host immune selection, either mediated by antibody or T-cell responses, is only sporadically observed in measles virus and is not associated with global selective sweeps (Woelk et al. 2001). Indeed, we found only weak evidence for positive selection in the sequence data analyzed here.

Finally, and perhaps most likely of all, the recent age of the three viruses may simply be a fundamental property of neutral evolutionary dynamics. Under a haploid Wright-Fisher model, the expected mean TMRCA is $2N_e$ generations (Ewens 2004). Although there are 30 million measles cases per year (WHO 2007), the number of cases in the troughs between epidemic peaks is evidently much lower. Indeed, estimates for case numbers during epidemic troughs, which provide a more realistic measure of effective population size, can be as low as 5000 individuals in specific localities (Grenfell, pers. comm.). Under these parameters, and assuming a generation time (mean waiting time between host infections) of 9-13 days (Grenfell, pers. comm.), the mean TMRCA of measles virus is 90,000 – 130,000 days, or ~250-300 years – and with a large variance. It is therefore possible that a history of neutral genetic drift alone is

sufficient to cause the shallow genetic diversity observed in the paramxyoviruses studied here, particularly given their complex cyclical dynamics which will reduce effective population sizes in the long-term.

Acknowledgements

We thank Rubing Chen for assistance with the BEAST analyses, Bryan Grenfell for advice on measles demography and statistics and two anonymous reviewers for useful comments. Laura Pomeroy was supported by the National Science Foundation, under the NSF Graduate Teaching Fellowship in K-12 Education (DGE-0338240). This work was also supported by NIH grant number GM080533-01.

References

Anderson RM, May RM (1991) Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford.

Barrett T (1994) Rinderpest and Distemper Viruses. In: Webster RG, Granoff A (eds) Encyclopedia of Virology. Academic Press, New York, pp. 1260-1269.

Bartlett MS (1957) Measles periodicity and community size. J R Stat Soc A 120:48-70.

Bellini WJ, Rota PA (1998) Genetic diversity of wild-type measles viruses: implications for global measles elimination programs. Emerg Infect Dis 4:29-35.

Bolt G, Jensen TD, Gottschalck E, Arctander P, Appel MJ, Buckland R, Blixenkrone-Moller M (1997) Genetic diversity of the attachment (H) protein gene of current field isolates of canine distemper virus. J Gen Virol 78:367-372.

Carbone KM, Wolinsky JS (2001) Mumps Virus. In: Fields BN, Knipe DM, Howley PM (eds) Virology, Lippincott Williams & Wilkins: Philadelphia, PA, pp. 1381-1400.

Domingo E, Holland JJ (1997) RNA virus mutations for fitness and survival. Ann Rev Microbiol 51:151-178.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol 4(5):e88.

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161:1307-1320.

Drummond AJ, Rambaut A (2003) BEAST v1.0, Available from http://evolve.zoo.ox.ac.uk/beast/.

Drummond AJ, Rambaut A, Shapiro B, and Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22:1185-1192.

Ewens WJ (2004) Mathematical Population Genetics. Second Edition. Springer-Verlag: New York.

Fauquet, CM, Mayo MA, Maniloff J, Desselberger U, Ball LA, ed. (2005) Virus Taxomomy: Classification and Nomenclature of Viruses. Elsevier Academic Press: New York.

Grenfell B, Harwood J (1997) (Meta)population dynamics of infectious diseases. Trend Ecol Evol 12:395-399.

Griffin, DE (2001) Measles Virus. In: Fields BN, Knipe DM, Howley PM (eds) Virology. Lippincott Williams & Wilkins: Philadelphia, pp. 1401-1441.

Hanada K, Suzuki Y, Gojobori T (2004) A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. Mol Biol Evol 21:1074-1080.

Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J Mol Evol 54:156-165.

Kosakovsky Pond SL, Frost SDW (2005) Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21: 2531-2533.

Lamb RA, Kolakofsky D (2001) Paramyxoviridae: The Viruses and Their Replication. In: Fields BN, Knipe DM, Howley PM (eds) Virology. Lippincott Williams & Wilkins: Philadelphia, p. 1305-1340.

Mochizuki M, Hashimoto M, Hagiwara S, Yoshida Y, Ishiguro S (1999) Genotypes of canine distemper virus determined by analysis of the hemagglutinin genes of recent isolates from dogs in Japan. J Clin Microbiol 37:2936-2942.

Orvell C (1994) Measles Virus. In: Webster RG, Granoff A (eds) Encyclopedia of Virology. Academic Press: New York, pp. 838-847.

Orvell C, Tecle T, Johansson B, Saito H, Samuelson A (2002) Antigenic relationships between six genotypes of the small hydrophobic protein gene of mumps virus. J Gen Virol 83:2489-2496.

Panum PL (1939) Observations made during the epidemic of measles on Faroe Islands in the year 1846. Med Class 3:829-886.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH (2001) The epidemic behaviour of the hepatitis C virus. Science 292:2323-2325.

Rambaut A (1996) Se-Al: Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk/

Rima BK (1994) Mumps Virus. In: Webster RG, Granoff A (eds) Encyclopedia of Virology. Academic Press: New York, pp. 876 – 883.

Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairoj AS, Brown TM, Salemi M, Vandamme A-M, Kalish ML (2003) US human immunodeficiency virus type 1 epidemic: Date of origin, population history, and characterization of early strains. J. Virol. 77:6359-6366.

Rota JS, Hummel KB, Rota PA, Bellini WJ (1992) Genetic variability of the glycoprotein genes of current wild-type measles isolates. Virology 188:135-142

Uzicanin A, Eggers R, Webb E, Harris B, Durrheim D, Ogunbanjo G, Isaacs V, Hawkridge A, Biellik R, Strebel P (2002) Impact of the 1996-1997 Supplementary Measles Vaccination Campaigns in South Africa. Int J Epidemiol 31: 968-976.

WHO (2007) Measles Fact Sheet. http://www.who.int/mediacentre/factsheets/fs286/en/.

WHO (1999) Progress toward measles elimination – Southern Africa, 1996-1998. MMWR 48:585-589.

Woelk CH, Jin L, Holmes EC, Brown DWG (2001) Immune and artificial selection in the hemagglutin (H) glycoprotein of measles virus. J Gen Virol 82:2463-2474.

Woelk CH, Pybus OG, Jin L, Brown DWG, Holmes EC (2002) Increased positive selection pressure in persistent (SSPE) versus acute measles virus infections. J Gen Virol 83:1419-1430.

AUTHOR CONTRIBUTIONS AND NOTES

Chapter 2

Laura Pomeroy implemented the stage-structured matrix model model, estimated parameters, performed likelihood ratio tests, and wrote the manuscript. Petra Klepac, Ottar Bjørnstad, Thijs Kuiken, Albert Osterhaus, and Jolianne Rijks were co-authors on this paper and assisted in the preparation and editing of the manuscript. Thijs Kuiken, Albert Osterhaus, and Jolianne Rijks coordinated and assisted with seal stranding data collection, seal stage classification, and necropsies. Petra Klepac developed the stage-structured matrix model and R_0 estimator; she also completed the R_0 estimation. Ottar Bjørnstad helped design and assisted with analyses and parameter estimation.

Chapter 3

Laura Pomeroy helped implement the model, estimated parameters, completed analyses, and wrote the manuscript. Matt Ferrari and Ottar Bjørnstad were co-authors on this paper and assisted in the preparation and editing of the manuscript; they also designed and helped implement the model. Catriona Stephenson, John Harwood, and Karin Harding contributed harbor seal epidemic data.

Chapter 4

Laura Pomeroy implemented the model, estimated parameters, and wrote the manuscript. Ottar Bjørnstad was a co-author on this paper and assisted in the preparation and editing of the manuscript; in addition, he designed the model. Catriona Stephenson, John Harwood, and Karin Harding contributed harbor seal epidemic data.

Chapter 5

This chapter was previously published as Pomeroy, L.W, Bjørnstad, O.N., and Holmes, E.C. 2008. The Evolutionary and Epidemiological Dynamics of the *Paramyxoviridae*. Journal of Molecular Evolution, vol. 66, no. 2, pp. 98-106. Laura Pomeroy collected data from GenBank, implemented the model, estimated parameters, and wrote the manuscript. Eddie Holmes and Ottar Bjørnstad were co-authors on this paper and assisted in the preparation and editing of the manuscript. Eddie Holmes designed the analysis.

VITA: LAURA WARLOW POMEROY

EDUCATION

The Pennsylvania State University — September 2003 to August 2008
Ph.D. Biology — University Park, PA

Purdue University — December 2002
B.S. Molecular Biology with Honors in Research — Minors: Chemistry and Spanish

PUBLICATION

Pomeroy, L.W., Bjørnstad, O.N., and E.C. Holmes. (2008) "A Molecular Perspective on the Demographic History of the *Paramyxovirdae*." Journal of Molecular Evolution, 66(2): 98-106 (doi: 10.1007/s00239-007-9040-x)

RESEARCH EXPERIENCE

Graduate Research Assistant — Sept 2003 – June 2008
The Pennsylvania State University — University Park, PA

Laboratory Technician — January 2003 – July 2003
Purdue University — West Lafayette, IN

Undergraduate Research Assistant — January 1999 – August 2002
Purdue University & Roswell Park Cancer Institute — W Lafayette, IN & Buffalo, NY

TEACHING EXPERIENCE

Science Methods for Elementary Education Majors — Nov 2005 to May 2007
Penn State, SUNY Oneonta, JMU — PA, NY, and VA

Teaching Assistant for Biology 110 — Fall Semester, 2004
The Pennsylvania State University — University Park, PA

Teaching Intern for Biology 195B — Fall Semester, 2002
Purdue University — West Lafayette, IN

OUTREACH EXPERIENCE

GREATT, NSF GK-12 Fellow — Aug 2005 – Aug 2007
The Pennsylvania State University — University Park, PA

Action Potential Summer Science Camp — June 2004 – Aug 2006
The Pennsylvania State University — University Park, PA

AWARDS

- Braddock Fellowship, The Pennsylvania State University (2007-2008)
- National Science Foundation GK-12 Fellowship (2005-2007)
- Biology Department Travel Grant, The Pennsylvania State University (2004, 2005, 2006)
- EEID Travel Grant, Cornell University (2006)
- Kananaskis Workshop Travel Grant, University of Calgary (2006)
- College of Agricultural Sciences Travel Grant – Gentry Endowment, The Pennsylvania State University (2005)
- Women in Science and Engineering (WISE) Institute Travel Grant, The Pennsylvania State University (2004)
- National Science Foundation Graduate Fellowship Honorable Mention (2004)
- University Graduate Fellowship, The Pennsylvania State University (2003-2004)
- Braddock Fellowship, The Pennsylvania State University (2003-2004)