THE UNIVERSITY OF CHICAGO

STRAIN DIVERSITY AND THE POPULATION DYNAMICS OF ROTAVIRUS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY AND EVOLUTION

BY
PAMELA MARTINEZ VARGAS

CHICAGO, ILLINOIS
DECEMBER 2017

A mis padres, Ines Vargas y Manuel Martinez.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am tremendously grateful to Mercedes, my advisor and mentor. Without her support, this work would not have been possible. In the last 6 years I have learned many skills from her. Passion, creativity, and perseverance are some of her qualities that I truly admire. Mercedes has been a source of inspiration for my career, not just for her outstanding curriculum, but also for the way that she leads her laboratory, which is more like a community than a place to work. As other alumni of the lab have stated before, I hope that her legacy is reflected in my own scientific path.

I would also like to thank the members of my committee. Aaron King, Sarah Cobey, Stefano Allesina, and Greg Dwyer for their feedback and insightful comments on this dissertation. Aaron King has also been one of my collaborators, and has been involved directly in my development as a scientist. He has always been open to discuss even simple questions, keeping a close connection in spite of the distance, and being incredibly patient. I want to thank Robert Woods, former post-doc and collaborator, who taught me how to connect ideas between theory and practice, and that having a life outside of academia is essential for being successful. Thanks to Sarah Cobey and Annette Ostling (University of Michigan) for allowing me to be their teaching assistant and being model educators and models for women in science.

I want to thank the Pascual lab community. Thanks to Ed, Andres, Shai, Xiangjun, Qixin, Victoria, Ruby, and many others that have been members of the lab for the last several years for their support and thoughtful discussions. Especially thanks to Mauricio, who has been like a brother to me, always there and remarkably tolerant. I also need to thank people in the administrative office, Audrey, Bonnie, Connie, and Mary for always being willing to help.

I am grateful to have met wonderful people in Ann Arbor and Chicago. Thanks to Dea, Jordan, Marian, Doug, Mariela, Hine, Liz, Dani, Shane, Darli, Sole, Pablo, and Brian for making me be a better person.

Finally, I want to thank my family, and in particular my parents, to whom this dissertation is dedicated. My curiosity and excitement towards science would not have initiated without them. They have always been supportive, in bad and good times. I feel proud to have them as my parents.

ABSTRACT

The ecological and evolutionary mechanisms that influence the population dynamics of infectious diseases remain an active area of research of relevance to fundamental biology and applied public health. Here, I have considered rotavirus as a study system to better understand the effect of environmental drivers on transmission dynamics, and strain competition in the context of evolution on pathogen diversity. I first studied the effect of climate variability on rotavirus population dynamics in a heterogeneous urban landscape, and compared to rural ones in a developing country, Bangladesh. To carry out this project, I implemented a process-based model of pathogen spread in the core and periphery of Dhaka, and relied on surveillance data to address the role of the monsoons and how this varied in different parts of the city. These results show that the inclusion of spatial information is essential to better understand the seasonality and inter-annual variation of the disease transmission and its response to climate anomalies. I then extended the epidemiological model to consider different serotypes and to interrogate serotype-specific surveillance data to understand the role of host immunity on rotavirus diversity. The main results of this work reveal a strong effect of generalized immunity on shaping this diversity. Lastly, I developed a theoretical approach to understand the effect of host-pathogen heterogeneity on pathogen diversity in the context of different immune responses. These results emphasize the importance of considering two kinds of phenotypic differences underlying also distinct kinds of competition when studying pathogen diversity, namely absolute differences that allow viruses to exploit different hosts (as resources), and frequency-dependent differences related to antigenic traits and resulting immunity. In particular, I emphasized the recognition that these two trait axes may not be independent. Improvements of current vaccines of limited efficiency in the developing world should result from an increased understanding of ecological and evolutionary mechanisms that shape rotavirus diversity.

CHAPTER 1

INTRODUCTION

While the field of ecology focuses on how organisms interact with each other and with the environment, evolution attempts to understand the processes by which individuals change over time, due to heritable modifications. Both ecology and evolution generate patterns at different organizational levels; importantly, their respective processes can occur on similar time scales for pathogens due to large population sizes and short generation times. Thus, the interplay of ecology and evolution is particularly relevant and amenable to study in these systems. The mechanisms that shape pathogen diversity and their consequences for epidemiological processes are today active fields of study. Here I use rotavirus as a study system to address open questions on population dynamics at the intersection of these two fields.

Rotavirus as a study system

For several reasons, the segmented RNA virus rotavirus is an ideal study system for seeking a deeper understanding of mechanisms driving pathogen population dynamics and strain diversity. It is the major cause of diarrheal disease in infants and young children worldwide, responsible for 40% of childhood gastroenteritis hospitalizations and 37% of diarrhea-related deaths in children younger than 5 years old (Parashar et al. 2003, Tate et al. 2012). It exhibits substantial diversity of strains or 'serotypes' globally, with the dominant serotype varying across space and time, for reasons that are poorly understood. Moreover, host heterogeneity exists and creates a variable resource for rotavirus spread, as serotypes bind host types based on histo-blood antigens in a differential manner. Finally, the disease remains a major public health problem in developing countries, where current vaccines are less efficient than in the USA and Europe, for reasons that are also poorly understood despite their importance to reducing transmission.

The diversity that interests me here is the antigenic diversity that influences how the virus is recognized by the immune system, and determines patterns of cross-protection be-

tween different strains. This is the diversity that underlies competition for hosts at the population level. To understand its basis, it is useful to consider that the virus is composed of 11 segments of double-stranded RNA which encode for 12 proteins in total. Only two proteins, VP7 and VP4, are targets of the immune system and define the serotype-genotype classification based on G and P types respectively. At least 12 G types and 15 P types have been identified in humans, with over 70 G-P combinations observed in different regions of the world (Bányai et al. 2012).

In terms of rotavirus evolution, mutation rate is of the order of $5 \cdot 10^{-5}$ per nucleotide (Blackhall et al. 1996), while the rate of evolutionary change is between $1.1 \cdot 10^{-3}$ and $8.7 \cdot 10^{-4}$ nucleotide substitutions per site per year (Donker and Kirkwood 2012, Matthijnssens et al. 2010). Reassortment of the RNA segments is a common mechanism to generate genetic novelty; however, fitness differences from distinct combinations may arise, limiting the number of dominant serotypes globally observed (Bányai et al. 2009, 2012, Dóró et al. 2014). Recombination seems to be rare and does not play an important role in the evolution of rotavirus (Woods 2015).

Additionally, it has been shown that rotavirus strains bind to human histo-blood group antigens differently (Shirato et al. 2008, Van Trang et al. 2014, Lui et al. 2016). This variation in the host affinity could explain patterns that suggest specific host susceptibility to particular strains, as well as changes in the susceptibility to specific strains at different ages, due to glycan modifications during neonatal development (Nordgren et al. 2014, Ramani et al. 2013, Van Trang et al. 2014).

As a diarrheal disease that transmits primarily through the fecal-oral route, rotavirus mainly affects developing countries. My research focused on incidence data from Bangladesh, a country where almost 40 million people still live in poverty (Ferdous et al. 2015). Specifically, I worked with data from Dhaka, the capital of Bangladesh, the home of more than 14 million people and the fastest growing megacity in the world.

To address rotavirus' diversity in Dhaka, the first step is to address the population dynamics of the disease in such an urban landscape. Both intrinsic factors, such as population-level 'herd' immunity, and extrinsic factors, such as climate forcing, can modulate disease dynamics, including seasonality and inter-annual variability. In particular, Dhaka is strongly affected by the annual monsoons, the major environmental disturbance in this region. The city further exhibits considerable spatial heterogeneity, including environments that range from highly urban in its core to more rural in the periphery. The population dynamics of infectious diseases in such heterogeneous urban landscapes remain poorly understood. More generally, climate variability is viewed as a regional phenomenon and therefore, its effects on the population dynamics of infectious diseases tend to be studied in time over aggregated space, with no consideration of spatial heterogeneity.

Chapter 2 analyzes how climate variables and spatial heterogeneity impact rotavirus dynamics to gain a better understanding of transmission and risk of infection at a more local scale than the whole city. Specifically, I studied the role of the monsoons in modulating the transmission of rotavirus in Dhaka and whether such an effect may vary spatially in different regions of the city. To carry out this study, I combined a process-based model with an extensive disease surveillance record spanning 22 years. This research provides a basis for the study of the population dynamics of serotypes.

In chapter 3, I incorporated serotype data to analyze the role of immunity on the strain diversity observed in this region. The distribution and abundance of rotavirus serotypes are highly variable with dominant strains fluctuating in both time and space, yet the reasons for this variation are not well understood. I incorporated knowledge on the population dynamics from the previous chapter to further investigate the effect of immunological processes and specifically quantify the relative importance of specific and generalized immunity in determining rotavirus diversity patterns in Dhaka. These processes represent two different ways in which pathogens compete from hosts as infection confers protection, either to specific types or regardless of such specificity and just as a function of the number of previous infections.

In Chapter 4, I implemented a theoretical framework to address whether and how host variability influences the coexistence of multiple strains. I specifically evaluated the impact of differential host-pathogen affinity on the evolution of strain diversity in the presence of a variable immune responses. These results highlight the importance of considering both host heterogeneity and immunity in understanding pathogen diversity below the species level. These two kinds of variation map respectively onto two different forces recognized in ecology for competitive interactions (Chesson 2000). Population-level immunity corresponds to frequency-dependent competition for hosts, conferring an advantage to the rare and a disadvantage to the common. It is otherwise neutral in the sense of not conferring any absolute advantage. Thus, antigenic traits underlying the acquisition of specific immunity are associated with phenotypic differences that ecologists have come to call 'niche' differences. In contrast, host heterogeneity gives absolute advantage to the strains that can bind a specific type. Thus, variation in binding sites is associated with what has been called 'fitness' differences in traits. Most research on pathogen diversity has focused on antigenic traits and not on the combined effect of both kinds of traits. Here, I go further and ask what happens if these axes are not independent.

Conclusions and future directions are included in chapter 5, where I discuss the contributions and limitations of this dissertation.

CHAPTER 2

DIFFERENTIAL AND ENHANCED RESPONSE TO CLIMATE FORCING IN DIARRHEAL DISEASE DUE TO ROTAVIRUS ACROSS A MEGACITY OF THE DEVELOPING WORLD

[Originally published as: Martinez, P.P., A.A. King, M. Yunus, A.S.G. Faruque, M. Pascual. 2016. Differential and enhanced response to climate forcing in diarrheal disease due to rotavirus across a megacity of the developing world. Proceedings of the National Academy of Sciences 113(15): 4092–4097]

2.1 Introduction

Many infectious diseases, especially those that are water-borne and vector-borne, have been shown to exhibit significant interannual variability in the size of seasonal outbreaks (e.g. Cazelles et al. 2005, Koelle et al. 2005, Laneri et al. 2010, Nagao and Koelle 2008, Pascual et al. 2000, Zhou et al. 2004). Identification of climate factors shaping interannual and seasonal variability is prerequisite to an understanding of the basic transmission biology of these environmentally-driven diseases, and of their response to climate change. The impact of climate factors on the population dynamics of infectious diseases has typically been addressed at large spatial scales by aggregating surveillance data over whole countries, regions, and cities (e.g., Cazelles et al. 2005, Hoshen and Morse 2004, Koelle et al. 2005, Laneri et al. 2010, Sultan et al. 2005). Global climate drivers, such as the El Niño Southern Oscillation (ENSO), are expected to operate over large spatial scales, synchronizing fluctuations of disease incidence across space (i.e., the Moran effect in population dynamics; Hudson and Cattadori 1999, Moran 1953). A recent study has shown however that the spatio-temporal dynamics of cholera in Dhaka, Bangladesh, are not homogeneous at intraurban scales Reiner et al. 2016. Two regions or clusters were identified, corresponding respectively to the highly populated core and the more rural periphery. The urban core was shown to be more climate-

sensitive, acting to propagate climate perturbations in cholera infection risk to the rest of the city. Our study addresses whether such spatial heterogeneity is also visible in the response to climate forcing at seasonal and interannual time scales in another major diarrheal infection, rotavirus. Because, unlike cholera, rotavirus does not appear to possess an environmental transmission pathway, it has not been viewed as a climate-sensitive infection. Here, we examine the role of the monsoons, and in particular flooding, in modulating the transmission of rotavirus. We inquire into whether such an effect might vary across local scales within a large urban environment.

Rotavirus is the most common cause of diarrhea among infants and young children worldwide, responsible for 40% of childhood gastroenteritis hospitalizations and 37% of diarrhea-related deaths in children younger than 5 years Parashar et al. 2003, Tate et al. 2012. This disease was recently reported to be the second most commonly isolated pathogen after *Vibrio cholerae* among adults attending urban and rural treatment facilities in Bangladesh Ferdous et al. 2015. Rotavirus is transmitted primarily via the fecal-oral route and two main seasonal patterns have been described. In temperate regions, incidence tends to peak during winter months Bresee et al. 2004, Cook et al. 1990, DSouza et al. 2008, Konno et al. 1983, Nakagomi et al. 2005, Purohit et al. 1998. In the tropics, incidence exhibit less pronounced seasonal variation Cook et al. 1990, Levy et al. 2009, with year-round incidence and peaks during summer or fall following monsoon rains, as in Bangladesh Hieber et al. 1978, Stoll et al. 1982. Despite these contrasting patterns, few correlational studies have statistically associated environmental variables with rotavirus incidence Hashizume et al. 2008, Rahman et al. 2007. We focus here on the monsoon season and on flooding in particular, which is both a major environmental disturbance in Bangladesh and one of the more prominent local manifestations of ENSO in this regionCash et al. 2008. Specifically, we implement stochastic transmission models (figure 2.2) to explore the effects of climate forcing within and between years, and across regions of the city. More specifically, they allow us to estimate and compare seasonal transmission intensity in the core and periphery of the city.

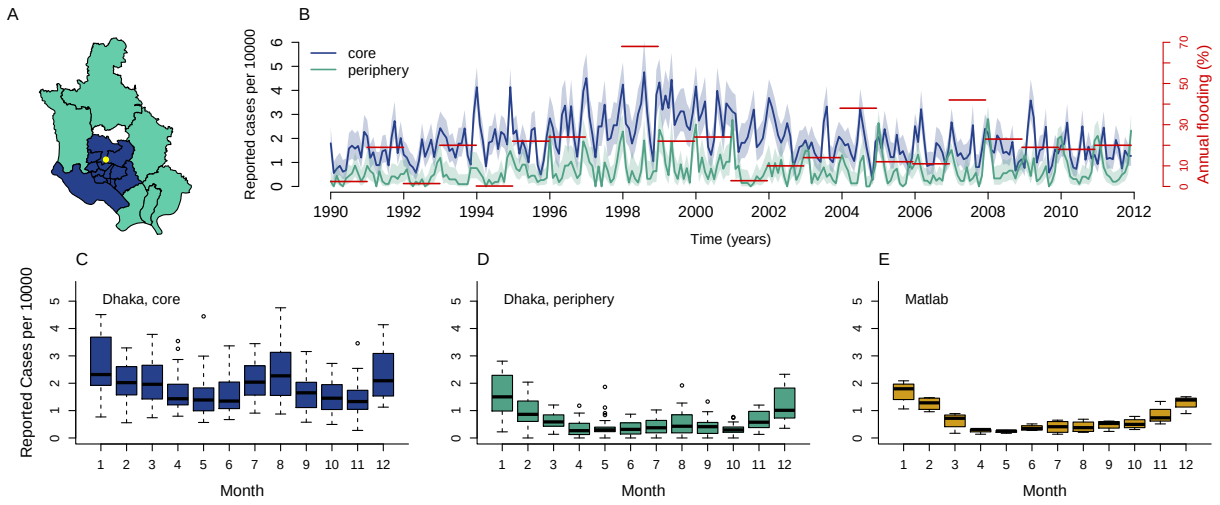

Figure 2.1: Data. (A) The administrative subdivisions or thanas of Dhaka are divided into two groups: the core (blue) and the periphery region (cyan), according to the grouping proposed by Reiner et al. 2012 Reiner et al. 2016. The thana in white was excluded from the analyses because of a lack of data before 2001, when it was created. (B) Monthly cases of rotavirus from 1990 to 2011 aggregated by region and normalized by population size and annual flooding index (in red), the 95% confidence intervals of a binomial distribution test are shaded in light color. (C-E) Boxplot of the normalized rotavirus cases by month per region. Matlab is a rural area of Bangladesh, located 55km south-east of Dhaka.

Our analyses show that the risk of rotavirus is far higher within the urban core and aggregation of cases reveals distinct seasonal patterns in core and periphery, with two peaks per year in the former (in the winter and monsoon seasons), and a single winter peak with the second one usually absent in the latter. Pronounced differences are further identified in the force of infection throughout the year and in its response to the monsoons, with distinct effects at the seasonal and interannual time scales for the core and periphery of the city respectively. These findings underscore the importance of spatial heterogeneity when addressing the role of climate forcing in urban environments. An infection that is not typically considered climate sensitive can be seen to be so within the highly populated core of the city. We discuss implications for the sensitivity of infectious diseases to changes in climate, in the context of the accelerating growth of cities in decades to come.

2.2 Methods

2.2.1 Data

Records of rotavirus cases confirmed by enzyme-linked immunosorbent assay (ELISA) were obtained from the Dhaka Hospital, through the ongoing and long-term surveillance program of the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B). A sample was taken for every 25th patient who visited the hospital from 1990 to 1995, and for every 50th patient from 1996 to 2011. These samples were extrapolated to correct for the frequency of sampling and to produce a consistent time series across time. The cases were aggregated by month per administrative subdivision or thana. The population for each thana was exponentially interpolated from the decadal censuses (1981, 1991, 2001) to generate the monthly values. The cases were then aggregated for the two different parts of the city proposed by Reiner *et al.*, 2012 (core and periphery, figure 2.1A). The monthly data for Matlab of rotavirus cases for the period 2010 to 2013 were obtained from the Matlab hospital. The flooding data consist of the annual percentage of country area flooded, provided by the Annual Flood Report of the Flood Forecasting and Warning Centre, Bangladesh. We drive the model with the flooding anomaly F (the flooding index with its mean value subtracted).

2.2.2 Transmission model

To describe the population dynamics of the disease within a region, we adapted the model proposed by Pitzer *et al.* in 2009 Pitzer et al. 2009 and coupled transmission between regions (figure 2.2). Newborns enter the M_i class where they are protected from infection by maternal antibodies. This maternally acquired immunity wanes at rate ω as individuals are transferred to the first susceptible class S_{1i}. Because individuals may be infected multiple times during their lifetime Velazquez et al. 1996, gaining immunity by repeated exposure, we considered a structure with multiple classes of infected (I_1, I_2, I_3) and susceptible individuals (S_1, S_2, S_3). Although the recovery class is not included, the model takes into account partial immunity

from previous exposure to the pathogen through a reduction in susceptibility following the first infection (σ). Other epidemiological parameters are described in table 2.2. The set of stochastic differential equations is given by:

$$\begin{aligned}
\frac{dM_i}{dt} &= (\mu P_i + \frac{dP_i}{dt}) - \omega M_i - \mu M_i \\
\frac{dS_{1i}}{dt} &= \omega M_i - \lambda_i S_{1i} - \mu S_{1i} \\
\frac{dI_{1i}}{dt} &= \lambda_i S_{1i} - \gamma I_{1i} - \mu I_{1i} \\
\frac{dS_{2i}}{dt} &= \gamma I_{1i} - \lambda_i \sigma S_{2i} - \mu S_{2i} \\
\frac{dI_{2i}}{dt} &= \lambda_i \sigma S_{2i} - \gamma I_{2i} - \mu I_{2i} \\
\frac{dS_{3i}}{dt} &= \gamma I_{2i} + \eta\gamma I_{3i} - \lambda_i \sigma^2 S_{3i} - \mu S_{3i} \\
\frac{dI_{3i}}{dt} &= \lambda_i \sigma^2 S_{3i} - \eta\gamma I_{3i} - \mu I_{3i}
\end{aligned} \tag{2.1}$$

P_i represents population size, and the flow of newborns combined with the death rate of each class results in population numbers equal to those observed for the growth of the city. Furthermore, the force of infection (or rate of transmission per susceptible individual) of each population is given by the following expression:

$$\lambda_i = \beta_i \left[\alpha_i \frac{(I_{1i} + I_{2i} + I_{3i})}{P_i} + (1 - \alpha_j) \frac{(I_{1j} + I_{2j} + I_{3j})}{P_i} \right] \tag{2.2}$$

where β_i is the transmission rate for populations i and (1 - α_i) refers to the movement rate of Infected individuals from region i to j and vice versa. The transmission rate is in turn given by

$$\beta_i = \exp \left[\sum_{k=1}^{6} b_{ki} s_k \; + \; b_{fi} s_4 F \right] \left[\frac{d\,\Gamma}{dt} \right] \tag{2.3}$$

and includes three components: 1) periodic functions of time to incorporate the seasonality through 6 splines s_k (Figure 2.11) and their respective coefficients b_k; 2) the interannual effect of flooding F and 3) environmental noise through a Gamma distribution Γ, which represents stochastic variability absent in the climate covariate (see Laneri et al. 2010 and the Supplementary Information for details). We note that in this expression the effect of climate forcing by the monsoons enters at two different time scales, seasonal and interannual. The seasonal effect is represented implicitly by the coefficients b_{ki}, specifically by b_{4i} since the fourth spline (s_4) peaks during the monsoon months (Figure 2.11); we therefore interpret the seasonal component quantified by this term as the average influence of the monsoon, which might include effects of humidity and temperature, and not necessarily or uniquely those of flooding. Additional variability is introduced across years through explicit consideration of an interannual effect of flooding (F), which as a major manifestation of the monsoons is also localized during those same months by its dependency on s_4 and multiplicatively modulates this seasonal component (equation 2.3).

We assume that reported cases are sampled from a negative binomial distribution, allowing for measurement noise: $\text{cases}_t \sim \text{NegBin}(\rho C_i, k_i)$ with mean ρC_i and overdispersion k_i. ρ is the reporting rate and C_i are the symptomatic infected for population i coming from the sum of the individuals entering in class I_1 and I_2 at time t (figure 2.2). Two additional models are described in the Supplement. The first one adds an additional parameter to the above model to allow for differential infectiousness of the infected classes. The second one allows us to investigate the hypothesis that the differences between core and periphery arise from biases in hospital attendance for core and periphery.

2.2.3 Parameter estimation and model selection

The estimation of both parameters and initial conditions for all state variables was carried out with an iterated filtering algorithm (MIF) implemented in the R package 'pomp' for Partially Observed Markov Processes (King et al. 2016). This algorithm maximizes the

likelihood and allows for the inclusion of both measurement and process noise, in addition to hidden variables, a typical limitation of surveillance records which provide a time series for a single observed variable per region. The initial search of parameter space was performed with a grid of 10000 random parameter combinations, and the output of this search was used as the initial conditions of a more local search. We repeated this process until the maximum likelihood value was stationary. All parameters were estimated, except for μ which was based on the average lifespan of an individual.

Likelihood-based criteria were used for model selection, including a likelihood ratio test (LR Test) since the models are nested, and the Akaike Information Criterion (AIC), which penalizes the likelihood based on the number of parameters, thus taking into account model complexity.

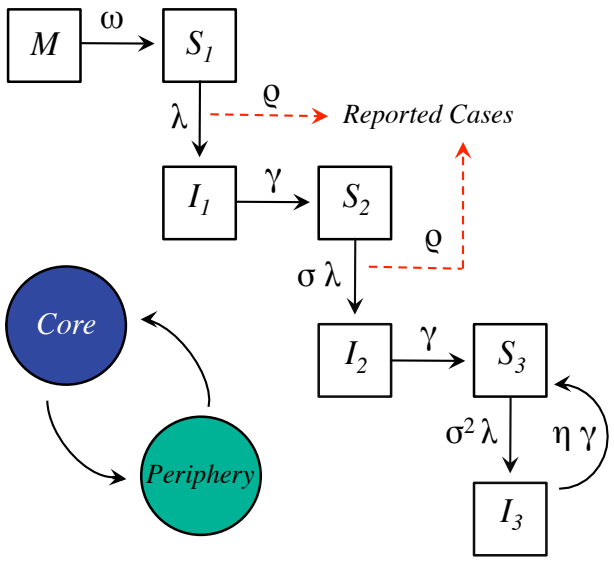

Figure 2.2: Diagram for the transmission model of rotavirus. The arrows indicate rates of flow among compartments. Each population, for core and periphery respectively, is divided into the following classes: Newborn (M), Susceptible (S_1, S_2, S_3) and Infected (I_1, I_2, I_3). The three levels of susceptible and infected individuals are meant to represent the recurrent exposure of individuals to the pathogen, as they acquire increasing protection and eventually become asymptomatic (after the second infection). The effect of movement between populations is incorporated within the force of infection (equation 4.2).

2.3 Results

The aggregation of the 22 years of data by core and periphery (figure 2.1A) reveals that the incidence rate in the core is almost 3 times that in the periphery (figure 2.1B). The two zones display distinct, hitherto unremarked, patterns of seasonality, which are also representative of variation at the thana level (Figure 2.5). Specifically, rotavirus cases in the core exhibit a temporal pattern similar to that described previously for tropical countries, with one peak during winter months and another during the monsoon (figure 2.1C). The average number of reported cases cumulated from June to September for the monsoon season is comparable to that obtained from November to February for the winter season (6% smaller). However, the second peak in cases is less pronounced in the periphery, and less than half the size of the winter peak (57% smaller, figure 2.1D). A similar seasonal pattern is found in Matlab, a rural area 55km south-east of Dhaka (figure 2.1E). These seasonal patterns suggest important differences in disease transmission within the city. In terms of climate fluctuations, severity of flooding shows year-to-year variation, with the most severe floods recorded for the years of 1998, 2004, and 2007 (figure 2.1B).

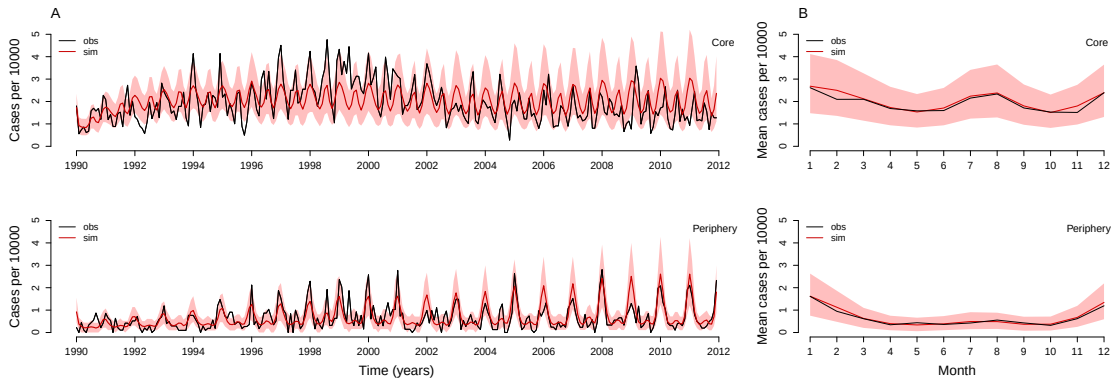

Figure 2.3: Comparison of simulated cases with those reported for the core and periphery of Dhaka. (A) Time series and (B) seasonal pattern for the observed cases (black) and the mean of 1000 model simulations (red). The 10% and 90% percentiles of the simulated data are shaded in light red. The model simulations are not next step predictions but numerical simulations of the model forward for the whole time period of the study starting with estimated initial conditions.

Table 2.1: A likelihood-based comparison of the different models

Model	log-likelihood	SE	no. param	AIC	LR Test
With flooding effect	-1577.6	0.33	25	3205.2	
Without flooding effect	-1582.2	0.35	23	3210.4	p-value = 0.01

For the two different models tested, the best model includes flooding as a covariate and performs significantly better than the model without this interannual effect based on a likelihood ratio test (p-value = 0.01, Table 2.1). Figure 2.3 illustrates simulations of the monthly cases from estimated initial conditions from 1990 to 2012 for the best model compared to the observed cases. The best-fitting model captures the interannual variation (figure 2.3A) and the main seasonal pattern of the reported cases in both regions (figure 2.3B). Moreover, this model reveals a striking difference in the force of infection (the instantaneous infection risk to each susceptible individual) between both regions, with larger values in the core than in the periphery (figure 2.4A-B). This difference is of an order of magnitude for the monsoon season, suggesting a higher intensity of transmission during this period. Interestingly, the estimated seasonality of the force of infection for the core of the city is similar to that of cases in tropical countries, with one peak during winter and another during the monsoon, and sustained transmission throughout the year. In contrast, the force of infection in the periphery shows one dominant peak during winter and a typically much weaker peak during the monsoons, with a more epidemic pattern with deeper troughs in between those seasons. The second peak is apparent here only in years with large flooding events (Figure 2.6): in the periphery, the transmission rate during the monsoon season has a much lower average (Figure 2.8), yet a more pronounced response to large floods (figure 2.4C).

The maximum likelihood estimates (MLE) of the parameters reveal additional features of the dynamics (Table 2.2, figure 2.7 and 2.8): the coupling between core and periphery appears weak ($\alpha_c = 0.96$ and $\alpha_p = 0.87$); maternal immunity of newborn individuals ($1/\omega$) wanes rapidly; the duration of infection ($1/\gamma$) is between 7 and 15 days; and individuals in

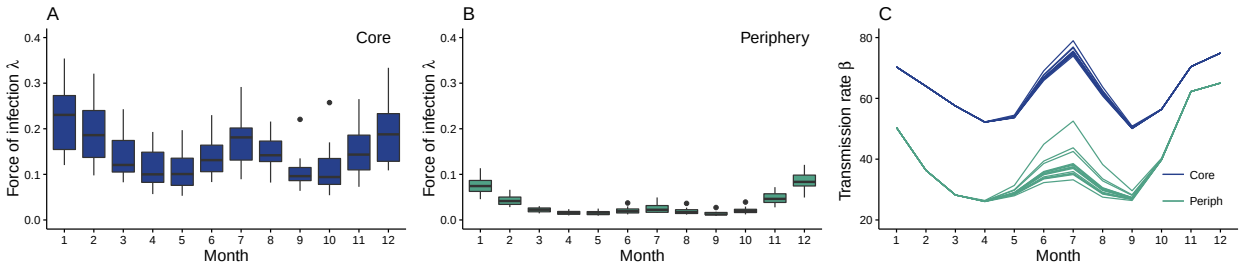

Figure 2.4: Force of infection and transmission rate from one simulation, for the period 1997 - 2011. (A-B) Boxplot of the force of infection by month per region. The force of infection is defined as the per capita rate at which susceptible individuals become infected. (C) Transmission rate by month per region. The maximum peak value relative to the mean for the monsoon season equals 1.05 and 1.36 for the core and periphery respectively. For the corresponding estimates of R_0, see Supporting Information.

the I_3 class have an infection that is long lasting ($1/\eta$, about 16 times longer). As we discuss below, we interpret the value of this duration as indicating the presence of a transmission reservoir.

Finally, an alternative explanation for the higher overall number of cases in the core than in the periphery, and for their differential seasonality, might be that access to the hospital which is located in the former region (Figure 1) is limited and especially impaired during the monsoons. Differences in hospital attendance rather than in the force of infection would explain the empirical patterns. To test this alternative hypothesis, an additional model was considered in which the transmission rate is the same for both regions but their reporting rate differs. Four reporting rates are estimated to allow for different values outside and inside the monsoon season (Jun-Sept), for core and periphery respectively. The log-likelihood of this model was 29 units lower than our best model (log-likelihood = -1606.9 and AIC = 3256 versus log-likelihood = -1577.6 and AIC = 3205). Additional arguments against this alternative explanation are presented below.

2.4 Discussion

Consideration of two different parts of the city, a densely populated core and a more rural periphery, appears essential to understand seasonal and interannual variation of this major diarrheal infection in response to climate forcing by the monsoons. In particular, our analysis reveals pronounced spatial heterogeneity in both the overall magnitude and the temporal pattern of disease risk within the city. Our results provide evidence that these two regions respond differentially to climate forcing, consistent with previous findings for cholera in the same region Reiner et al. 2016. The contrasting seasonal patterns described here within the city are consistent with the findings of a recent meta-analysis on rotavirus seasonality proposing that climatic conditions and the degree of country development are better predictors than latitude or geographic location per se Levy et al. 2009, Patel et al. 2013. The cholera study Reiner et al. 2016 also suggested that this level of aggregation (core and periphery) is congruent with spatial variation in socioeconomic conditions including population density (figure 2.9). These factors may act to modulate the effect of climate variables at local spatial scales by enhancing contact and population susceptibility. It is interesting to note that the climate sensitivity of cholera has been traditionally explained with reference to the residence of its etiological agent, the bacterium *Vibrio cholerae*, in aquatic environments independent of the human host Colwell 1996. The fact that rotavirus displays climate sensitivity too, in the absence of such an environmental reservoir, suggests that the causal links in the climate-disease connection are not strictly dependent on specific ecological mechanisms related to transmission pathways Lipp et al. 2002. A similar association with Sea Surface Temperatures in the Pacific for cholera and shigellosis in Bangladesh suggested that ENSO acts mainly via the modulation of secondary transmission in these diarrheal diseases (e.g. by increasing exposure to contaminated water as well as person-to-person contact) Cash et al. 2014.

In addition to socioeconomic and demographic factors, the two regions of the city might more directly differ in susceptibility to flooding itself. Flooding risk and extent appear

Table 2.2: Parameter estimates and confidence intervals

1-3	Description	Value	C.I.
$1/\mu$	Average lifespan (yr)	50	Fixed
$1/\gamma$	Duration of infection (dy)	10	7 - 15
$1/\omega$	Duration of maternal immunity (dy)	1	0 - 110
ρ	Reporting rate (10^{-3})	2.3	1.8 - 3.6
α_c	Coupling core	0.96	0.90 - 0.98
α_p	Coupling periphery	0.87	0.31 - 1.00
σ	Susceptibility reduction	0.19	0.12 - 0.24
η	Additional duration of infection	0.06	0.04 - 0.09

heterogeneous throughout the city (Dewan and Corner 2014), but in a way that is not consistent with this explanation. Moreover, during extreme floods, such as the floods of 1998, more than 50% of the city's area can be inundated (Faisal et al. 2003).

Our results suggest the presence of a transmission reservoir that maintains transmission between seasons and primes the seasonal response of the system to the monsoons. The existence of such a reservoir is implied by the long duration of infection in the most immune class (I_3), whose individuals are asymptomatic in our model. The more endemic epidemiological pattern of the core region is generated in the model via this reservoir conjoined with high transmission rates. In the core, the monsoon-season peak is highly regular, and the relatively small interannual variability in this peak suggests rapid saturation of the transmission system during the season. In the periphery, by contrast, low transmission rates lead to a more epidemic pattern of dynamics (deeper troughs) where the magnitude of the second peak is limited by the rate of transmission during the monsoon season, which varies substantially from year to year. Thus the degree of interannual and seasonal sensitivity to climate would vary across a transmission gradient, as documented for climate-sensitive diseases such as malaria via comparisons of disease dynamics in endemic vs fringe regions Laneri et al. 2015, Pascual 2015, and here within an urban landscape.

The validity of our conclusions vis-à-vis transmission presupposes that the patterns we see in the incidences are not merely due to differential hospital-seeking behavior in the

two parts of the city. Three pieces of evidence suggest that this is not the case. First, the similarity in the seasonal pattern and overall incidence rate between the periphery of Dhaka and the rural area of Matlab, 40 km South East and the site of an intensive and well-established surveillance system, support that the observed differences are not merely an artifact of biases in hospital attendance rates. Second, rotavirus rates do not track seasonal rates in hospital attendance rates (for example, the former decline in the periphery in March-April-May when values of the latter increase, Fig. 2.10). Hospital seeking behavior is driven by a variety of diarrheal infections, in particular by cholera during those months. Finally, we formulated a model specifically allowing for differential reporting rates between the regions and seasons; this model is rejected by the model selection criteria (p-value < 0.0001).

Relationships among force of infection, re-infection frequency, and severity of disease shape the age distribution of cases. Under similar case age distributions across the city, which we find for Dhaka (Fig. 2.12), the estimated higher force of infection in the core implies that children there also experience more frequent bouts of infection. Pitzer et al. Pitzer et al. 2011 fitted a model similar to ours to age-distribution data, under the assumption of stationarity. We have focused here instead on the dynamic behavior of the system and fitted the model to time series of incidence, ignoring age structure. The fact that these contrasting approaches lead to models that differ in several particulars, including the basic reproduction number, R_0, emphasizes the remaining uncertainties in model structure and parameter values. To resolve these issues, it will be useful to fit rotavirus transmission models to time series of age-specific incidence. Such an effort should further elucidate rotavirus epidemiology and in particular, the relationship between age structure and force of infection in models with more complex structure than the well-understood SIR formulation.

The coupling of the population dynamics of the virus between the two regions of the city is weak in our best model. This result could be further investigated in the future with explicit spatiotemporal information on human movement within the city based on mobile phone data and census data, as these methods continue to develop and have already contributed

to a better spatial understanding of transmission in other infectious diseases Buckee et al. 2013, Dalziel et al. 2013, Richardson et al. 2013, Wesolowski et al. 2012, 2013. Further research of relevance to early warnings would also benefit from better data on flooding with higher spatio-temporal resolution, more congruent with the spatial scale of the city and the temporal scale of the seasonal transmission dynamics. A finer resolution would allow the explicit consideration of flooding as a seasonal driver. In addition, investigation of seasonal forecasts from hydrological models is also warranted. Finally, other model structures should be investigated to better examine the nature of the reservoir, and to take into account existing serotype variation within the virus population.

To conclude, our results underscore the importance of considering the spatial heterogeneity of large urban environments when analyzing climate-driven transmission dynamics. Urban heterogeneity can enhance sensitivity of transmission dynamics to climate factors via demographic and socio-economic conditions, even in infectious diseases that are not necessarily recognized as climate-sensitive to begin with. These conditions can facilitate the persistence of reservoirs of infection and in so doing facilitate responses to anomalous climate events and seasonal environmental variation, especially in the populated cities of the developing world and in the future under climate change.

2.5 Acknowledgments

We thank two anonymous referees for their insightful comments. This research was partially supported by the National Oceanic and Atmospheric Administration (Grant # F020704). The case data used in this paper were collected with the support of ICDDR,B and its donors who provide unrestricted support to ICDDR,B for its operation and research. Current donors providing unrestricted support include the Government of the People's Republic of Bangladesh; the Department of Foreign Affairs, Trade and Development Canada (DFATD); the Swedish International Development Cooperation Agency (Sida); and the Department for International Development (DFID, UK Aid). We gratefully acknowledge these donors for

their support and commitment to ICDDR,B's research efforts. A.A.K. was supported by the Research and Policy in Infectious Disease Dynamics program of the Science and Technology Directorate, U.S. Department of Homeland Security and the Fogarty International Center, U.S. National Institutes of Health (NIH), and also by research grants from the National Institutes of Health (grant #1R01AI101155) and MIDAS, National Institute of General Medical Sciences (grant #U54-GM111274). For data access, co-authors from the ICDDR,B should be contacted and data access would be guided by the data sharing policy of ICDDR,B.

2.6 Supporting information

2.6.1 Alternative transmission models

Differential infectiousness of the infected classes

Variations in infectiousness across the infected classes (I_1, I_2, I_3) are included in the parameter ξ, which acts to scale this quantity in the force of infection, given by:

$$\lambda_i = \beta_i \left[\alpha_i \frac{(I_{1i} + \xi I_{2i} + \xi^2 I_{3i})}{P_i} + (1 - \alpha_j) \frac{(I_{1j} + \xi I_{2j} + \xi^2 I_{3j})}{P_i} \right] \tag{2.4}$$

The log-likelihood for the MLE of this model equals -1577.5 with the estimated $\xi = 0.97$ (C.I. = 0.59 - 1.00). Because this model did not perform significantly better than the model without variation in infectiousness and because the confidence intervals of ξ include 1, we decided to fix ξ at this value to decrease the number of parameters.

Differential reporting rate

We implemented an alternative model that assumes the same transmission rate across the city, but includes different reporting rates across the seasons (monsoon and non-monsoon months, from June to September) and between the core and periphery. As a consequence, the number of parameters of this model is smaller (by 4) than that of our main model in the

Methods. The measurement model is given by: $\text{cases}_t \sim \text{NegBin}(\rho_{ij}C_i, k_i)$, where ρ_{ij} is the reporting rate for region i and season j. As described in the Methods section, k_i denotes the overdispersion and C_i, the new symptomatic infections for population i, consisting of the accumulated individuals entering the classes I_1 and I_2 during the time interval of sampling (δt). The expression for the transmission rate simplifies to:

$$\beta = \exp\left[\sum_{k=1}^{6} b_k s_k \; + \; b_f s_4 F\right]\left[\frac{d\,\Gamma}{dt}\right] \tag{2.5}$$

The maximum log-likelihood for this model is equal to -1606.9, with an AIC of 3256.

2.6.2 Basic reproductive number R_0

The basic reproductive number, or R_0, refers to the number of secondary infections in a completely susceptible population (i.e. for the whole susceptible population in class S_1). This was estimated by using the next generation method, resulting in the following expression: $R_{0i} = \frac{\bar{\beta}_i}{\gamma+\mu}$, where $\bar{\beta}_i$ is the average transmission rate for region i. The value for the core is $R_{0c} = 1.80$ and for the periphery, $R_{0p} = 1.11$.

2.6.3 Supplementary figures

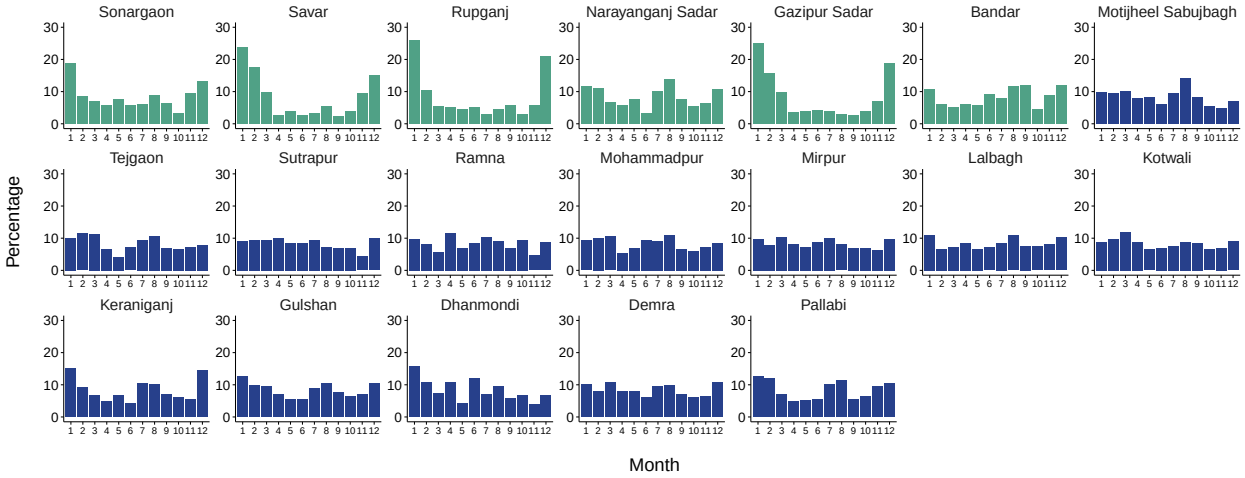

Figure 2.5: Percentage of reported cases per month for each thana. Each plot corresponds to a given thana. Thanas that are part of the core are depicted in dark blue and those that belong to the periphery, in cyan. Cases were first added for all years (1990-2012) and then the percentage corresponding to each month was computed. Thanas in the core exhibit a more endemic behavior with less pronounced troughs between seasons; those in the periphery typically show a higher peak in the winter and deeper troughs following this outbreak, with low levels of cases during the monsoons.

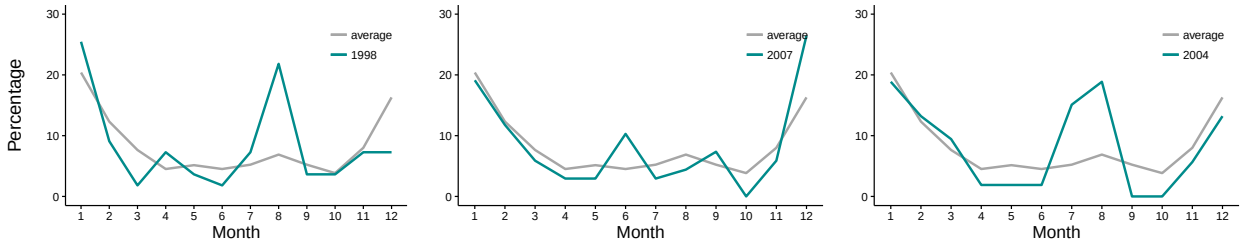

Figure 2.6: Percentage of the total cases falling in a given month on average (in gray) and for years with anomalous flooding (in cyan) for the periphery. Values of the flooding index for 1998, 2007 and 2004 are 68%, 42% and 38% respectively.

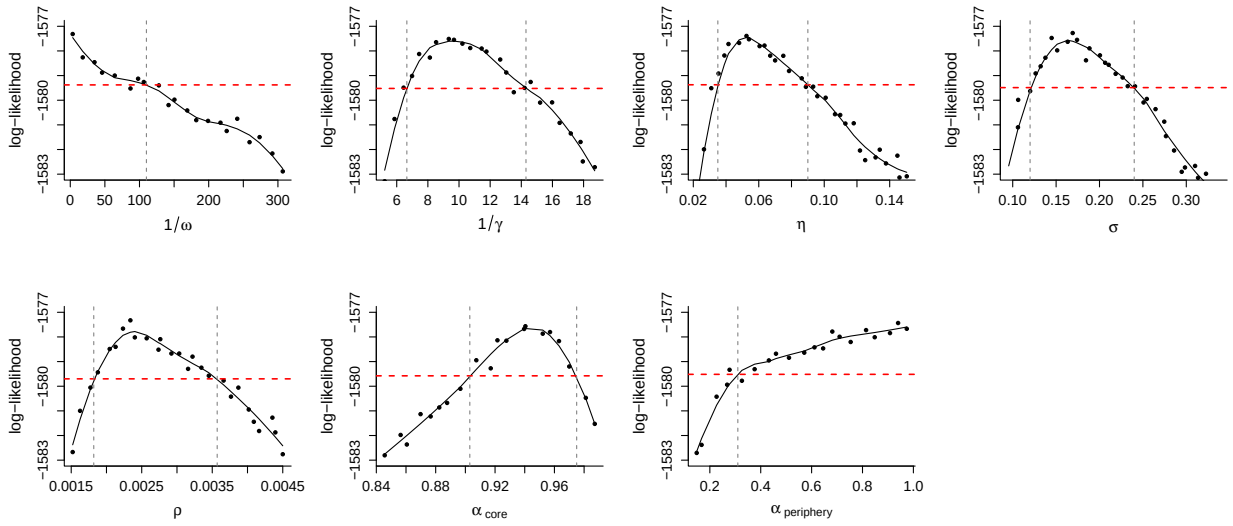

Figure 2.7: Log-likelihood profiles for epidemiological parameters. The intersection between the red dashed line and the vertical dashed lines indicates the 95% confidence interval of each parameter.

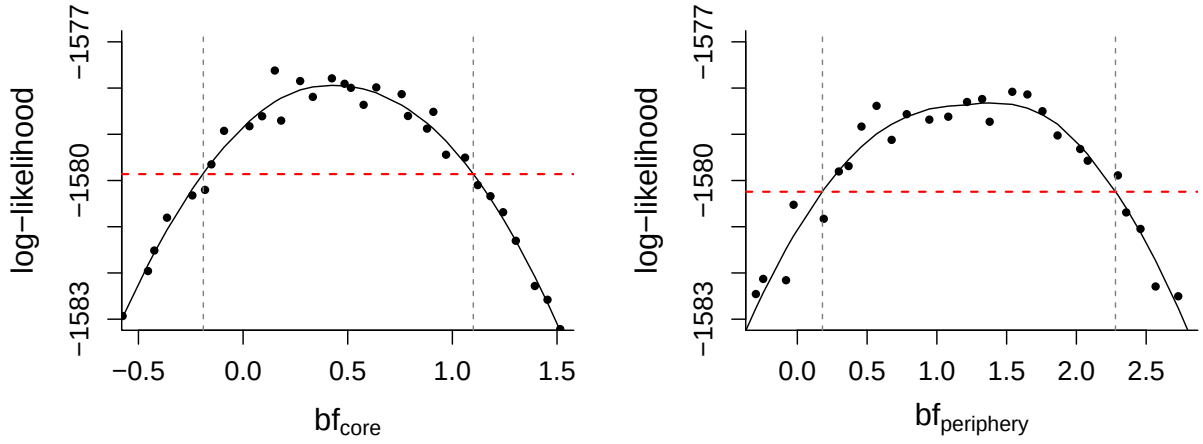

Figure 2.8: Coefficient bf representing the interannual effect of flooding for the core and periphery. The likelihood profile curves are shown, with the intersection between the solid and dashed lines indicating the 95% confidence interval for each region.

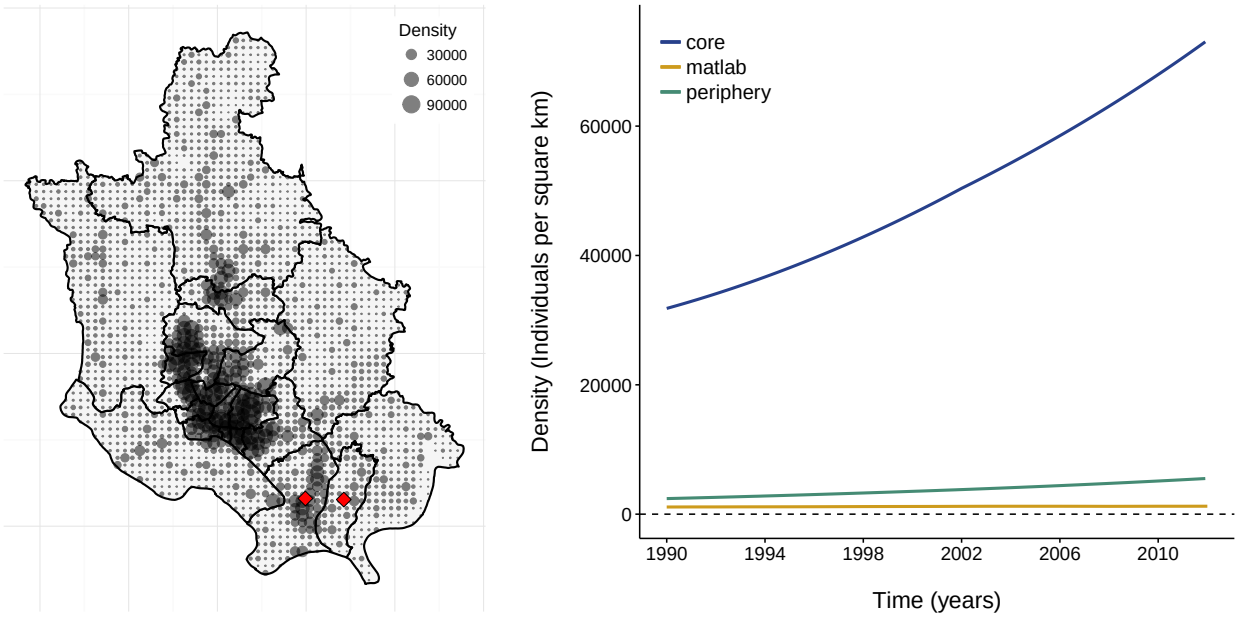

Figure 2.9: Population density. (Left) Map of the population density for 2010. Each grey dot refers to the density at a 1 km^2 resolution. Diamonds in red label the peripheral thanas of Narayanganj Sadar and Bandar. Note the concentration of density values larger than is typical the case in other areas of the periphery, especially for Nrayaganj Sadar, although not as high as in the core of the city. Population data were downloaded from http://web.ornl.gov/sci/landscan. (Right) Average population density across thanas within each region, and for the rural area South of Dhaka known as Matlab.

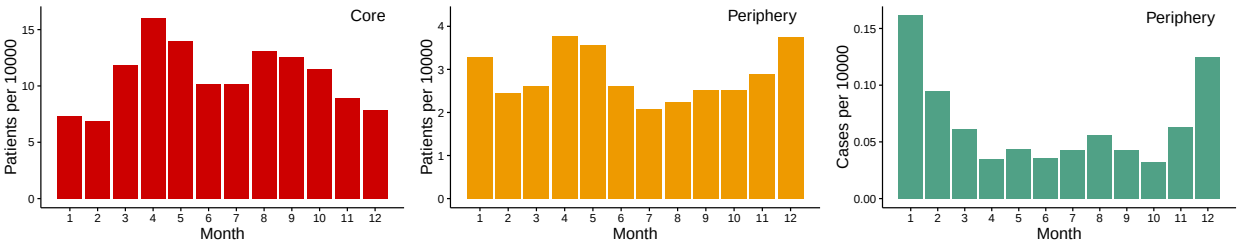

Figure 2.10: Seasonality of the number of patients visiting the hospital in core (left) and periphery (center). The seasonality in the number of rotavirus cases in the periphery (right) does not track that of the number of patients. Rotavirus cases decline in March-April-May when hospital attendance increases (concurrent with the first peak of cholera).

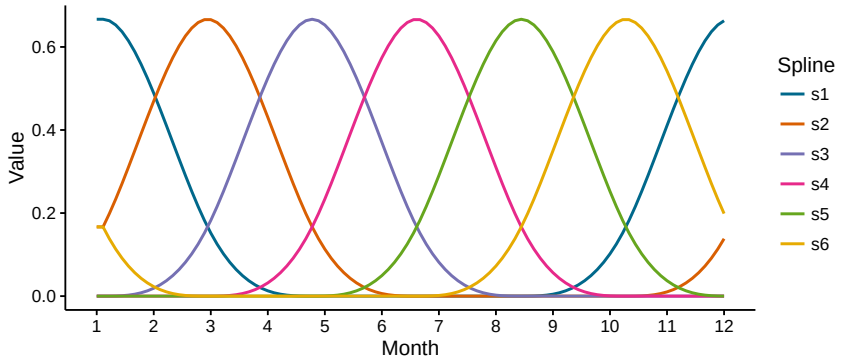

Figure 2.11: Periodic splines. The six beta splines ($s_1 ... s_6$) used in the expression and estimation of the transmission rate. Note the localization of the fourth spline (magenta) during the monsoon months.

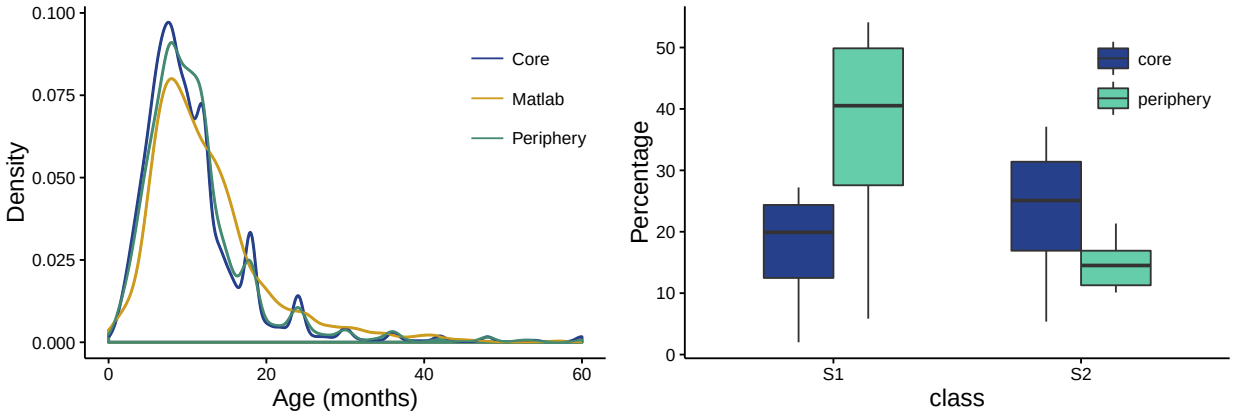

Figure 2.12: Age distribution of cases per region (left) and boxplot of the proportion of individuals in classes S_1 and S_2 from simulations of the model (right).

CHAPTER 3

IMPACT OF SPECIFIC AND GENERALIZED IMMUNITY ON ROTAVIRUS ANTIGENIC DIVERSITY

3.1 Introduction

As the most common cause of diarrheal disease worldwide, rotavirus contributes 40% of the total hospitalizations of young children due to gastroenteritis and accounts for more than 450,000 deaths every year globally (Parashar et al. 2003, Tate et al. 2012). The antigenic diversity of the virus is defined by variation in its coat proteins exposed to recognition by the immune system. In particular, two outer layer proteins, VP4 and VP7, are targets of the immune system and define the serotype-genotype classification based on the designated P and G types respectively. Around 12 G types and 15 P types have been found in humans (Bányai et al. 2012), and the most predominant serotype worldwide is currently G1P8. At local scales, however, the dominant strain can vary across years, even in the absence of vaccination, for reasons that are not yet well understood (Hungerford et al. 2016, Pitzer et al. 2011, 2015).

Two kinds of immunity underlie competition for hosts as a function of their history of exposure, depending on whether strain specificity matters or not, versus simply the number of previous infections regardless of their antigenic identity. The serotype-specific immune response is known as 'homotypic', whereas generalized immunity generated by antibodies that can neutralize more than one rotavirus serotype is designated as 'heterotypic' (Chiba et al. 1986). With a deterministic model for rotavirus transmission, Pitzer et al. 2015 showed that the homotypic immunity must be stronger than the heterotypic one in order to qualitatively explain patterns in the distribution of rotavirus serotypes in a developed country, Belgium, after the introduction of the vaccine. Here, I extend and complement this previous study by considering the temporal variation of serotypes in a developing country, Bangladesh, where the efficacy of the rotavirus vaccine (Rotarix) is only 41% (Zaman et al. 2017). With a

process-based model that allows for both demographic and measurement noise, I analyzed 11 years of rotavirus incidence data at the serotype level, for a total of 6 strains, to examine the role of specific and generalized immunity. I specifically quantified the reduction in susceptibility following a first infection due to the generalized immune response and whether this reduction differs significantly from that generated by specific immunity acquired by exposure either to a strain that shares the same G-type, the same P-type, or to the exact same strain. The main results show that a strong generalized immunity is needed to recover the serotype dynamics of rotavirus in the city of Dhaka, an effect that cannot be distinguished from that of exposure to the same P-type. It follows that this P-type would be mainly responsible for the generalized response. Interestingly, the results further indicate a role of specific immunity via the G-type, which although weak is still meaningful for the epidemiological dynamics and antigenic diversity patterns of the virus. I discuss the consequences of the intensity of the immune response in the context of disease control and vaccine implementation for developing countries.

3.2 Methods

3.2.1 Serotype data

Retrospective records for the rotavirus cases were obtained from the Dhaka Hospital, through the ongoing surveillance program of the ICDDR,B, which has been conducted over the years in a systematic manner. Specifically, stool samples were collected to determine the presence of enteric pathogens in 2% of the patients attending the hospital for treatment of diarrhea by an enzyme-linked immunosorbent assay (ELISA). Around 10% of the positive rotavirus cases between 2001 and 2011 were serotyped using multiplexed RT-PCR. These cases were aggregated here into two-month bins by serotype. The population for the city of Dhaka was exponentially interpolated from the decadal censuses (2001 and 2011) to generate the bi-monthly values.

3.2.2 *Transmission model*

A stochastic system of ordinary differentials was formulated to model the transmission dynamics of the virus. This model was parameterized based on a time series spanning 11 years. The model is similar to the one described in chapter 1, with additional features to incorporate the serotype-specific dynamics. Based on the dominant strains present in the data, I defined the set of serotypes ai, where $a \in \{G_1, G_2, G_9, G_N\}$ and $i \in \{P_4, P_8, P_N\}$.

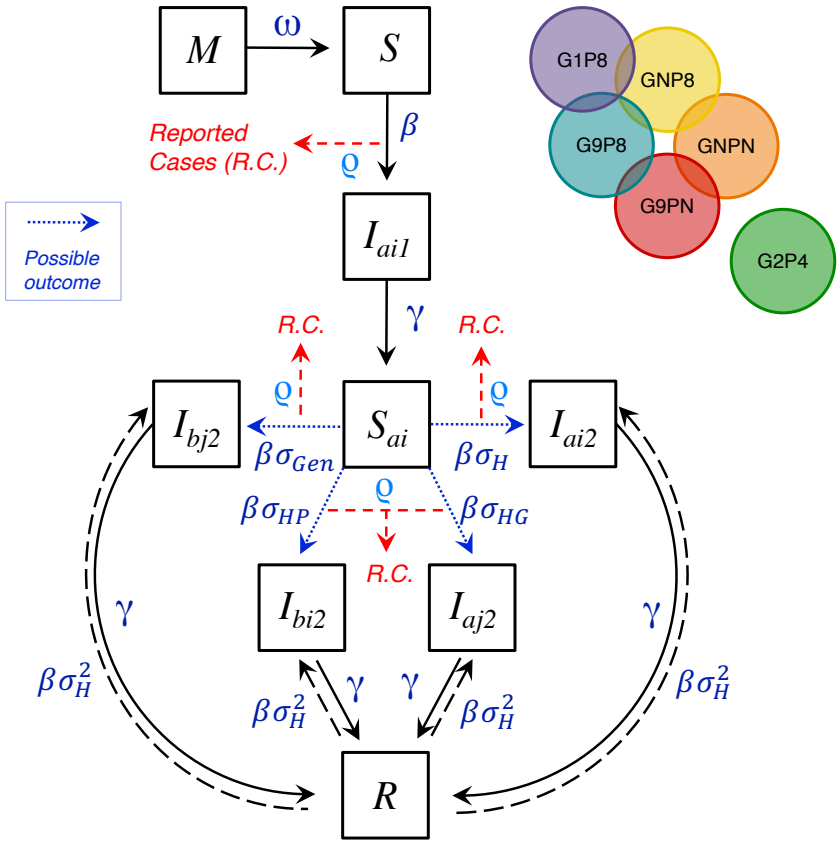

Figure 3.1: Diagram of the transmission model. The population is divided into the following classes: Newborn M, first Susceptible S, first Infected I_{ai1}, second Susceptible S_{ai}, second Infected I_{ai2}, and Recovered R, with $a \in \{G_1, G_2, G_9, G_N\}$ and $j \in \{P_4, P_8, P_N\}$. The diagram on the right illustrates the serotypes and their overlap partners.

The population dynamics of these serotypes is then specified by the following set of equations:

$$
\begin{aligned}
\frac{dM}{dt} &= \left(\mu P + \frac{dP}{dt}\right) - \omega M \frac{d\Gamma}{dt} - \mu M \\
\frac{dS}{dt} &= \omega M \frac{d\Gamma}{dt} - \beta S \left(\frac{\sum I_{ai}}{P}\right) - \mu S \\
\frac{dI_{ai_1}}{dt} &= \beta S \left(\frac{I_{ai}}{P}\right) - \gamma I_{ai_1} - \mu I_{ai_1} \\
\frac{dS_{ai}}{dt} &= \gamma I_{ai_1} - \beta S_{ai} \left[\sigma_{Gen} \left(\sum_{\substack{b \neq a \\ j \neq i}} I_{bj} + \sigma_{HG} \sum_{j \neq i} I_{aj} + \sigma_{HP} \sum_{b \neq a} I_{bi} + \sigma_H I_{ai}\right)\right] \frac{1}{P} - \mu S_{ai} \\
\frac{dI_{ai_2}}{dt} &= \beta \left[\sigma_{Gen} \left(\sum_{\substack{b \neq a \\ j \neq i}} S_{bj} + \sigma_{HG} \sum_{j \neq i} S_{aj} + \sigma_{HP} \sum_{b \neq a} S_{bi} + \sigma_H S_{ai} + \sigma_H^2 R\right)\right] \frac{I_{ai}}{P} - (\gamma + \mu) I_{ai_2} \\
\frac{dR}{dt} &= \gamma \sum I_{ai_2} - \beta R \sigma_{Gen} \sigma_H^2 \left(\frac{\sum I_{ai}}{P}\right) - \mu R
\end{aligned}
\tag{3.1}
$$

where P refers to population size and β is the transmission rate, which incorporates a seasonal effect through 6 splines s_y and their coefficients b_y: $\beta = \exp \sum b_y s_y$. The process noise is included through a Gamma distribution Γ, which represents stochastic variability in the rate at which the maternal antibodies wane. The total reported cases are sampled with a negative binomial distribution: $\text{cases}_t \sim \text{NegBin}(\rho C, k)$ with mean ρC and overdispersion k, being ρ the reporting rate and C the infected individuals that attend the hospital. The reported cases arise only from those individuals leaving the classes S and S_{ai} at time t, since individuals who are reinfected from the recovery class R are considered asymptomatic (figure 3.1). The serotype cases are sampled from a multinomial distribution $\sim \text{Multinomial}(F, (f_1, ..., f_n))$, where F is the number of reported cases that are serotyped, and f_i is the incidence of serotype i. A total of 6 serotypes (G1P8, G2P4, G1P8, G9PN, GNP8, GNPN) were considered when fitting the model. I considered serotypes denoted by GN and PN to aggregated all G and P types other than G1, G2, and G9, and P4 and P8 respectively. This aggregation is justified by the low incidence of these P and G types in the area for the 11 years of the data analyzed.

3.2.3 Parameter estimation

Parameter estimation was conducted via a sequential Monte Carlo method for likelihood maximization with an iterated filtering algorithm (MIF2) available in the R package 'pomp' (King et al. 2016). This method allows for the inclusion of two types of noise, dynamical (or process) and measurement noise, as well as for hidden variables, whose values have not been measured (here all M, S, I, and R variables). A grid of 50000 randomly generated parameter combinations was considered for the initial search. Parameter estimates from this first search were then used as the initial values for the following iteration. This process was repeated until reaching the maximum likelihood estimate. The average lifespan $1/\mu$ and the duration of infection $1/\gamma$ were fixed to 50 years and 10 days respectively, based on the epidemiology of rotavirus.

3.3 Results

About 10% of the reported cases from the Dhaka Hospital have been analyzed to detect the serotype (figure 3.2A). The overall distribution of the serotyped cases for the period 2001–2011 shows that about half of the cases belong to serotypes G1P8 and G9P8 (figure 3.2B), consistent with the pattern observed in global strain diversity (Bányai et al. 2012). Interestingly, 20% of the cases correspond to G2P4, a completely non-overlapping serotype with respect to the two most dominant ones, which has been shown to be more prevalent in this region since 2005 (Rahman et al. 2007). The remaining 30% belongs to GNP8, GNPN, and G9PN. The distribution of the G and P types over time shows distinct patterns of inter-annual variation (figure 3.2C). The predominant G type changes over time, consistent with what has been reported for other locations (Pitzer et al. 2015), suggesting that frequency-dependent processes are acting at this level of organization. In contrast, the P types are relatively stable, with P8 as dominant for most years (figure 3.2C). In terms of the serotype incidence over time (figure 3.2D), the last two years (2009-2011) show an increase in the

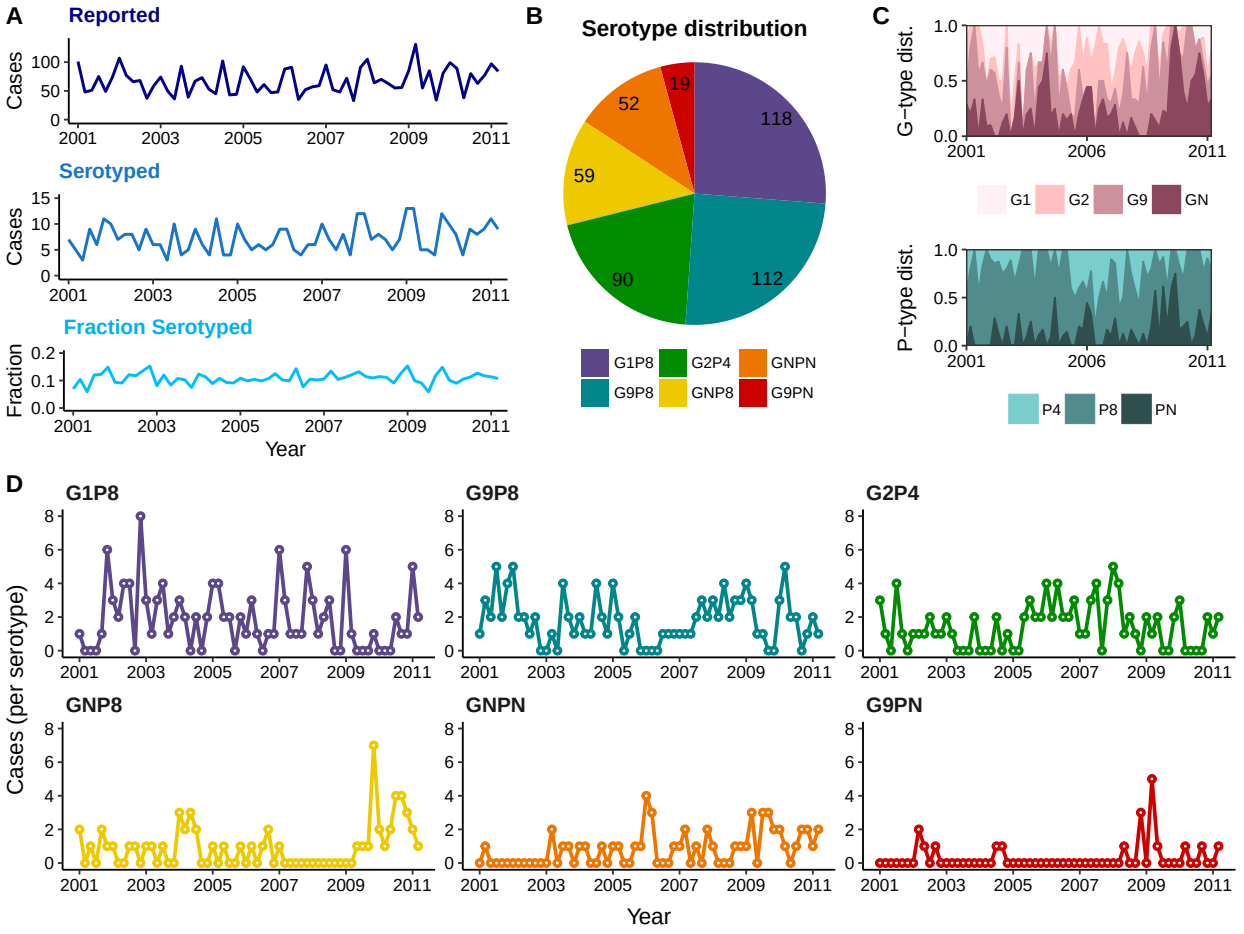

Figure 3.2: Data. A) Total positive rotavirus cases, number of cases serotyped, and the fraction serotyped over time. B) Serotype distribution of the cases aggregated over the whole period of time (2001 - 2011). C) G-type and P-type distributions over time. D) Bi-monthly serotype cases over time.

reported cases for G9PN, GNP8 and GNPN, consistent with the changes observed in the G type and P type distributions for the same time period (figure 3.2C), suggesting a shift in the immunity of the host population that promotes higher diversity.

Figure 3.3 shows the comparison of the observed data to the simulations from the model of figure 3.1 for parameters from maximum likelihood estimates. The best model is able to reproduce the inter-annual variation of the total reported cases as well as the main fluctuations in the incidence of each serotype, with most of the observed cases enclosed within the 95% confidence intervals (fig. 3.3 bottom).

Importantly, the model reveals distinctive features of the immune response that underlie the population dynamics of rotavirus. The reduction in susceptibility due to generalized immunity σ_{Gen} is about 98.8%; in other words, 11 of each 1000 susceptible individuals who have been previously infected and have subsequent contact with the pathogen would be infected for a second time. In contrast, a reduction in susceptibility due a previous contact with the same P-type σ_{HP} appears almost negligible. I have assumed that a generalized immune response is always attributable to those individuals who have been previously exposed to the virus. This means that the effect of the specific immunity occurs in addition to that of the generalized response. Taking this into account, a value of σ_{HP} close to 1 implies that exposure to exactly the same P-type does not provide additional protection than just being exposed to a completely non-overlapping serotype. The situation is different when contact occurs with a virus that shares the G-type with past infections. The estimate of σ_{HG} for the best model is ~ 0.36, and thus, an additional 64% reduction in the risk of infection is

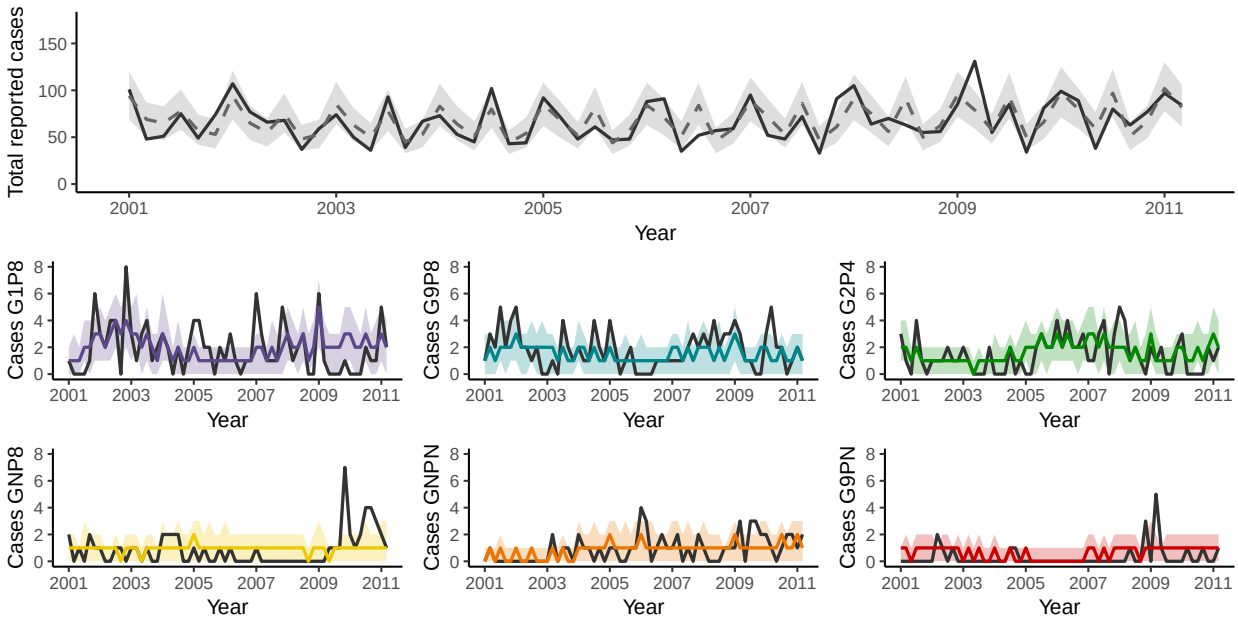

Figure 3.3: Fitting. Comparison of the observed cases, shown in black, with the mean of 1000 model simulations for the total number of cases (gray) and the reported cases at the serotype level (color). The 10-90% percentiles of the simulations are shaded in lighter color.

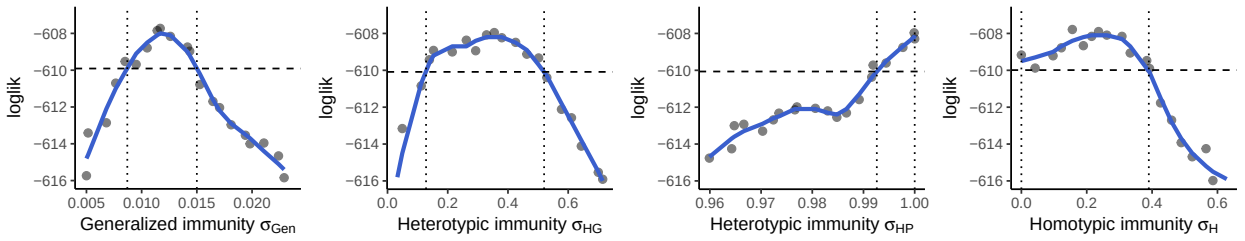

Figure 3.4: Profile likelihood. The likelihood profile curves for the immunological parameters are shown in blue, with the 95% confidence interval cut-off indicated by dashed lines. Parameters $\sigma_{Gen}, \sigma_{HG}, \sigma_{HP}, \sigma_{H}$ refer to fully heterotypic response, heterotypic response against G, heterotypic response against P, and homotypic response, respectively.

expressed as $\sigma_{Gen}\sigma_{HG}$. I have also explored the effect of being exposed to exactly the same serotype in the past. This effect is reflected in the parameter σ_{H}, which has a value of 0.25, although it not differ significantly from 0. Profile likelihood and confidence intervals for these immunological parameters are shown in Figure 3.4.

Finally, I analyzed the population dynamics of some of the hidden variables to examine features of the serotype fluctuations over time. Figure 3.5 illustrates the time series for the susceptible and infected individuals, aggregated at the annual level. I find a long-period cyclical behavior of the individuals infected with different serotypes, with pronounced anti-phase oscillations of G1P8 and G2P4, two nonoverlapping strains, that do not exclude other, partially overlapping, strains. Although long-period cycles are also observed in the time series of the susceptible population, these dynamics are transient and later dominated by the rapid growth of the population in Dhaka (fig. 3.5B). To examine the nature of the oscillations present in the dynamics of infected individuals, I predicted the incidence for 20 years (from 2011 to 2030) based on the model. Simulations show that the amplitude of the oscillations tends to decrease over time, with G1P8 and G2P4 remaining the dominant serotypes, and all remaining strains coexisting at lower incidence. This analysis excludes however the invasion of new types, due to migration or reassortment events, which can disturb the selection imposed by the immune system, as I discuss in the next section.

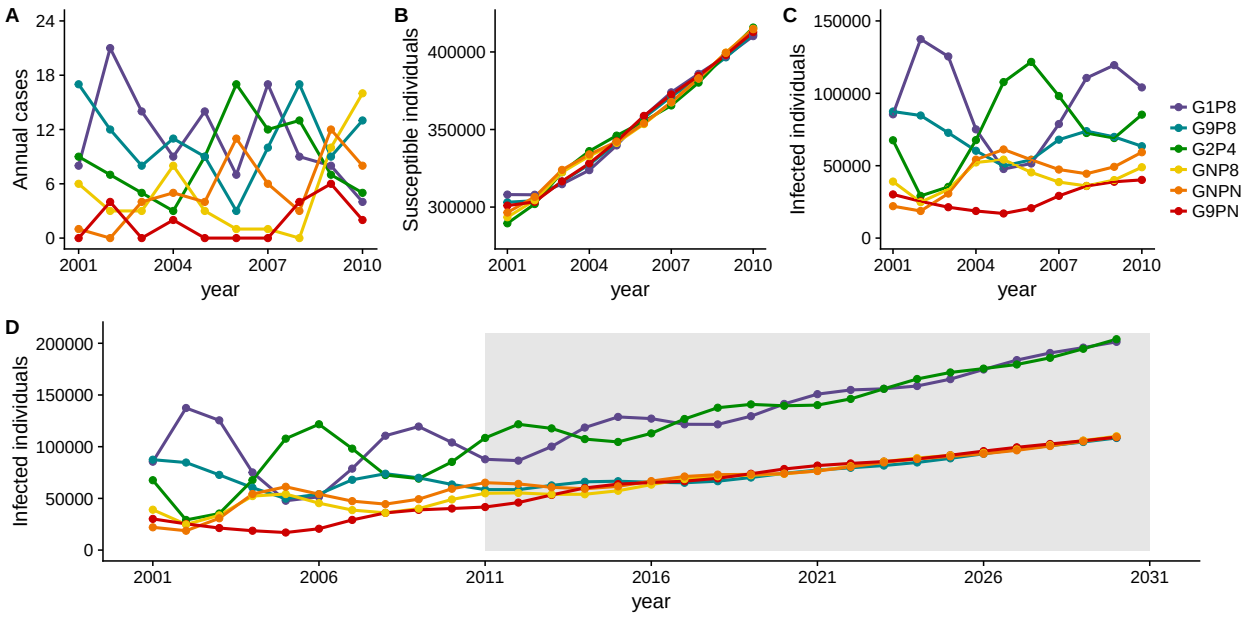

Figure 3.5: Epidemiological variables. A) Annual observed cases. B) Mean susceptible individuals to the particular strain C) Mean infected individuals. D) Mean and predicted infected individuals. The prediction from 2012 to 2031 is shown in grey background.

3.4 Conclusions

Different levels of selection imposed by the immune system on antigenically diverse pathogens can lead to distinct regimes of coexistence, including single strain dominance, cyclical or chaotic behavior and coexistence of non-overlapping strains (Gupta et al. 1998). Here, I investigate the role of specific and generalized immunity in rotavirus antigenic diversity in Dhaka by fitting a mechanistic model to 6 inter-dependent time series for different serotypes. The results from this study reveal a strong immune response for secondary infections, even when the serotypes of the first and second exposure do not share the same G or P type, in contrast with previous findings for rotavirus in Belgium, where the vaccine has been successful at reducing the incidence of this pathogen (Pitzer et al. 2015).

Based on the estimates of partially and fully heterotypic responses, the epitopes located in the VP7 protein seem to trigger a stronger specific response, while the ones located in the VP4 protein cause a more general, heterotypic immune response. These results are con-

sistent with recent empirical evidence revealing broad cross-protection caused by antibodies binding VP4 of serotypically distinct strains as the dominant immune response against rotavirus, and a strong partially heterotypic response only present for VP7 (Nair et al. 2017). The strong heterotypic response produced against VP4 is also consistent with the highly conserved features and functional constraints of this protein, which mediates cell attachment and penetration (Dormitzer et al. 2004), as well as a higher number of G than P types observed in common G–P combination of human rotavirus serotypes (Bányai et al. 2012).

The results presented here also underscore the importance of high heterotypic immunity in the maintenance of strain diversity. In the absence of such generalized immune response, theoretical work has shown that the most likely scenario is one in which only non-overlapping serotypes can coexist (Gupta et al. 1998). My analyses show that two non-overlapping serotypes, G1P8 and G2P4, are dominant, limiting the incidence of other strains that share the same G-type. This niche partitioning can be broken however by repeated invasions of new types and by reassortment events, which constantly disrupt the selective pressure imposed by the immune system. The global distribution of G-P associations indicates the presence of 12 G types, 15 P types, and more than 70 G-P combinations, with only 6 serotypes common worldwide (Bányai et al. 2012). Particular G-P associations appear to exhibit higher fitness and therefore,, reassortment events may lead to less fit serotypes that may not persist in the long term (McDonald et al. 2009).

In summary, these findings emphasize the differences in the homotypic (strain specific) and heterotypic (cross-protective) immunity against rotavirus. The identification of the epitopes in VP4 triggering the generalized response would enable a more effective vaccine, especially for developing countries such as Bangladesh, where the virus is responsible for more than 6000 deaths each year in children younger than 5 years of age (Tate et al. 2012).

CHAPTER 4

ENRICHED PATHOGEN DIVERSITY UNDER HOST-TYPE HETEROGENEITY AND IMMUNE-MEDIATED COMPETITION

[Originally published as: Martinez, P.P., R.J. Woods, M. Pascual. 2017. Enriched pathogen diversity under host-type heterogeneity and immune-mediated competition. bioRxiv 195065]

4.1 Introduction

In recent decades, the increased availability of molecular data has enhanced theoretical understanding of the role of epidemiological, immunological, and evolutionary processes in shaping pathogen diversity (Grenfell et al. 2004, Lipsitch and O'Hagan 2007, Volz et al. 2013). The mechanisms that underlie the phenotypic diversity of pathogens remain however an active area of research of relevance to epidemiology and responses to intervention.

Questions on strain diversity within pathogen populations exhibit clear similarity to those posed for species diversity within ecological communities. For the latter, two kinds of mechanisms for long-term species coexistence are distinguished: those known as 'stabilizing' which operate in a frequency-dependent manner as they confer an advantage to the rare; and those known as 'equalizing' because they minimize absolute fitness differences between species which would otherwise lead to dominance and exclusion (Chesson 2000). Equalizing effects reduce differences in fitness components such as growth rate, reproduction and survival. Given absolute fitness differences that promote exclusion, coexistence can also result from the existence of trade-offs as in the classical example provided by Tilman's R^* theory (Tilman 1982): multiple species that compete for the same resources can coexist if they have significantly different resource requirements and rely on the presence of a trade-off in the ability to consume these resources.

In pathogens, frequency-dependent selection imposed by the immune system can result in different strains coexisting within a population (Gupta et al. 1996, Gupta et al. 1998, Gupta and Maiden 2001, Gog and Grenfell 2002, Gomes et al. 2002). The antigenic determinants that trigger the specific immune response are the phenotypic traits that underlie competition for hosts at the population level: once a host is protected against a given antigenic variant, it is no longer a resource for the growth of this specific type. Changes in antigenic sites allow the pathogen to escape immunity in the host population, so that strains that are sufficiently distant in antigenic space can coexist by limiting similarity (Macarthur and Levins 1967). Coexistence arises here from purely stabilizing differences due to the frequency of given antigenic types, with no fixed fitness differences between individual pathogens.

Another aspect of strain diversity in viruses is associated with fixed fitness differences and follows from phenotypic variation in the host population itself (Shirato et al. 2008, Van Trang et al. 2014). In particular, host variation has the potential to promote strain diversity if strains differ in their ability to bind host receptors. This is the case for norovirus and rotavirus which recognize human histo-blood group antigens as receptors, with the ABO group and Lewis b antigen structures determining the strain's specificity for the host (Shirato et al. 2008, Van Trang et al. 2014).

Using an adaptive dynamics approach, here we address how differences in host-pathogen affinity (equalizing effect) interact with selection imposed by specific immunity (stabilizing effect) to influence strain diversity. Our results highlight the importance of considering the constraint imposed by a link between these two mechanisms when considering the stable coexistence of pathogen strains. We discuss the relevance of these findings in the context of epidemic control implementations.

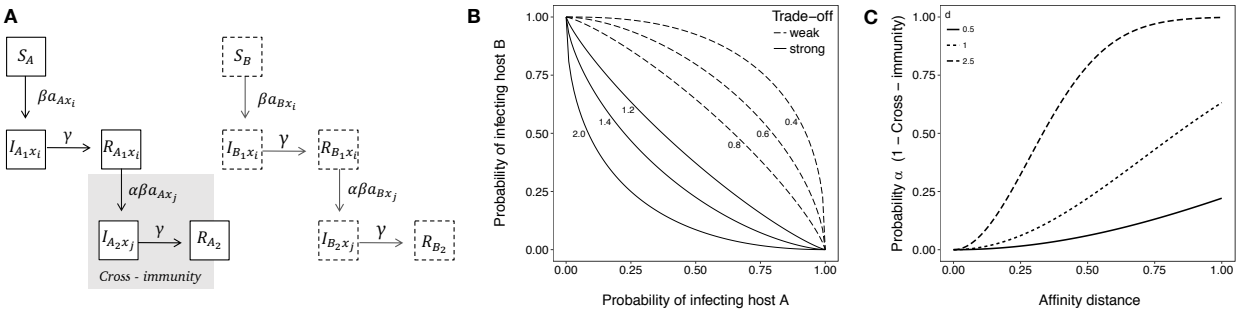

Figure 4.1: A) Transmission model. S, I, R refer respectively to the classes of Susceptible, Infected and Recovered individuals. The diagram on the left corresponds to host A, and the one on the right, to host B. B) Affinity trade-off. The trade-off can be strong ($s > 1$) or weak ($s < 1$). The different curves correspond to different values of s as indicated by the numbers next to each of them. C) Cross-immunity function. The probability α (1-cross-immunity) is plotted as a function of the affinity distance between current and previous infections, for different values of d.

4.2 Methods

4.2.1 Transmission model

We implemented an extended SIR transmission model that considers two types of hosts, A and B, which can differ in their binding affinity to two or more strains (fig 4.1A). Within each host population, new borns enter the susceptible class S. When hosts become infected, they move first to the infected (and infectious) class I_1, and then, to the first recovery class R_1 according to the recovery rate γ. Individuals within the R_1 class can be infected by a different strain with probability α denoting the degree of cross-immunity conferred by the first infection. Individuals can then transition to a second infected class I_2, followed by a second recovery class R_2, with the same recovery rate γ. Individuals in the R_2 compartment are assumed to have acquired permanent immunity (the notation 1, 2 refers to the order of the events and not to the infection with a particular strain). Finally, P refers to the population size and μ to the birth/death rate. The complete set of differential equations is described in the Supporting Information.

The force of infection consists of the contact rate β and the probability of being infected after contact. Here, this probability specifically refers to host affinity, which defines the identity of each strain x_i and is characterized by the trait a_{x_i}. An increase in the affinity for one host is assumed to occur at the expense of a decreasing affinity for the other so that $a_{Ax_i}^{1/s} + a_{Bx_i}^{1/s} = 1$ (Egas et al. 2004). This trade-off means that each strain can be highly specialized for only one host (A or B), with shape determined by parameter s. Values of $s > 1$ result in a strong trade-off, while $s < 1$ generates a weak trade-off (fig 4.1B).

To compare the biological scenario in which the degree of differences in host-binding affinity is independent from the intensity of cross-immunity, to one in which these two phenotypes are linked, we considered first a model with fixed values of α and then one in which α depends on the distance between the affinity values of the current and past infection (a_{x_i} and a_{x_j} respectively): $\alpha = 1 - e^{-(d(a_{x_i} - a_{x_j}))^2}$, where the parameter d determines how fast the immune protection decays as a function of the distance (Gog and Grenfell 2002, fig 4.1C). This assumption implies that changes in the protein that binds the host receptor have an effect in both the degree of cross-immunity (*i.e.* antigenic distance) and the affinity for the host.

4.2.2 Evolutionary dynamics

An adaptive dynamics framework was used to analyze the different strategies under which the system may evolve (Metz et al. 1996, Dieckmann and Law 1996, Geritz et al. 1997). This framework assumes that mutants have a small difference in the phenotype of interest with respect to the resident's individuals. Mutants can invade the system if the growth rate or 'invasion fitness' is positive. Changes in the invasion fitness associated with changes in the mutant phenotype determine the direction in which the population will evolve. For each combination of the pathogen's trait, we estimated the endemic equilibrium numerically, which allowed us to obtain the selection gradient as well as the mutual invasibility plots. See Supporting Information for further details.

4.3 Results

We first analyzed the evolutionary dynamics of the system under a high degree of cross-immunity (for fixed $\alpha = 0.08$). Figure 4.2A-E shows the mutual invasibility plots for weak and strong trade-offs in the ability to bind the receptor of different hosts. When there is a weak trade-off in host utilization (values of $s < 1$), the system evolves towards a singular 'generalist' strategy. In contrast, for values of s greater than 1, the strains evolve to be 'specialists', either on different hosts (top left and bottom right), or on the same host (bottom left and top right of the mutual invasibility plot). The evolutionary outcome is dependent upon the initial values of the affinity trait a_{x_i}, which can lead the system in the direction of an evolutionary branching point (attractors that then allow branching toward specialization on different hosts), and/or away from an unstable repellor, which can trap the strains in regions where they both specialize on the same host. As we vary the strength of the affinity trade-off, the evolutionary outcomes are summarized in the bifurcation diagram (fig 4.4A), which shows that the system is dominated by a Continuously Stable Strategy (CSS, generalist strategy) for any value of $s < 1$. Likewise, for $s \geq 2$ the only possible outcome is a repellor, and thus a completely specialized trait. From high to low values of s, a bifurcation occurs at $s \approx 2$ where two repellors emerge together with a saddle node until s approaches 1. As we show next, $s < 1$ is a universal condition for the generalist strategy, while the value of s at which the bifurcation occurs depends on α and epidemiological parameters.

We then considered a model in which the strength of cross-immunity (α) depends on the host affinity distance between current and past infections (see Methods for detail). This can be the case when the receptor's binding site also functions as an antigen (*e.g.* Taylor et al. 1987, Ruggeri and Greenberg 1991, Taylor and Dimmock 1994). This assumption implies that changes in binding affinity have a direct effect on antigenic distance, and that therefore, these changes affect the degree of specific immunity acquired from past infections. Figure 4.2F-J summarizes the findings from this analysis for a broad strain space ($d = 0.5$), for which the average probability α is 0.08. This result shows that the coexistence of semi-

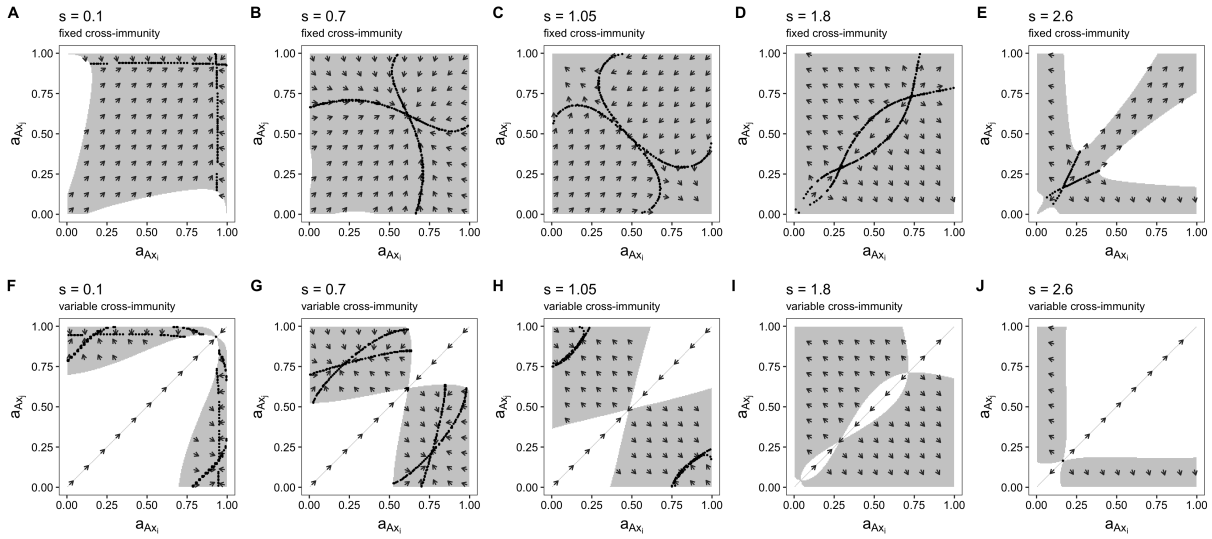

Figure 4.2: Trait evolution plots for the affinity for host A, for different values of trade-off strength s. Comparison of the evolutionary outcomes for the scenario in which α is fixed ($\alpha = 0.08$, A-E) and the one in which cross-immunity α depends on the binding affinity distance (between the traits a_{x_i} and a_{x_j}; $d = 0.5$, F-J). The arrows show the directions of evolution. The grey area indicates the region for which the pair of traits a_{x_i} and a_{x_j} can mutually invade and coexist (i.e. the invasion fitness is positive for both). The arrows indicate the direction of evolution as the result of successive invasion events. Dotted black lines refer to the adaptive isoclines, where the selection gradient vanishes. The intersection of these isoclines denotes singular strategies, which can be stable or unstable. If the strategy is stable, then the two traits that are given by this intersection stably coexist (*i.e.* they result in an evolutionary stable strategy).

specialized strains is likely for a wide range of values of s, with the potential evolutionary outcomes shown in fig 4.4B. The area in the mutual invasibility plot under which two residents would coexist varies with the value of s in a nonlinear way (fig 4.5); this area refers to the percentage of possible trait combinations under which the invasion fitness (*i.e.* growth rate) is positive for both strains, showing that intermediate values of the affinity trade-off allow for greater coexistence. Interestingly, a model with fixed α exhibits more transient ecological coexistence than the one with a variable cross-immunity, but this coexistence does not last when considering evolutionary time scales.

For the above results, cross immunity was strong (as indicated by the low values of α and d). We now examine the effect of both an even higher and lower cross immunity. Not surprising based on our first analysis, the impact of cross-immunity is dependent upon the

value of the host affinity trade-off when α is fixed and independent from host affinity. Under weak host affinity trade-offs ($s < 1$), the system evolves to a unique Evolutionarily Stable Strategy (ESS), which is also a CSS, over the full range of cross-immunity (fig 4.3A). In this situation, the strains have the same ability to infect both host types (*i.e.* same trait value),

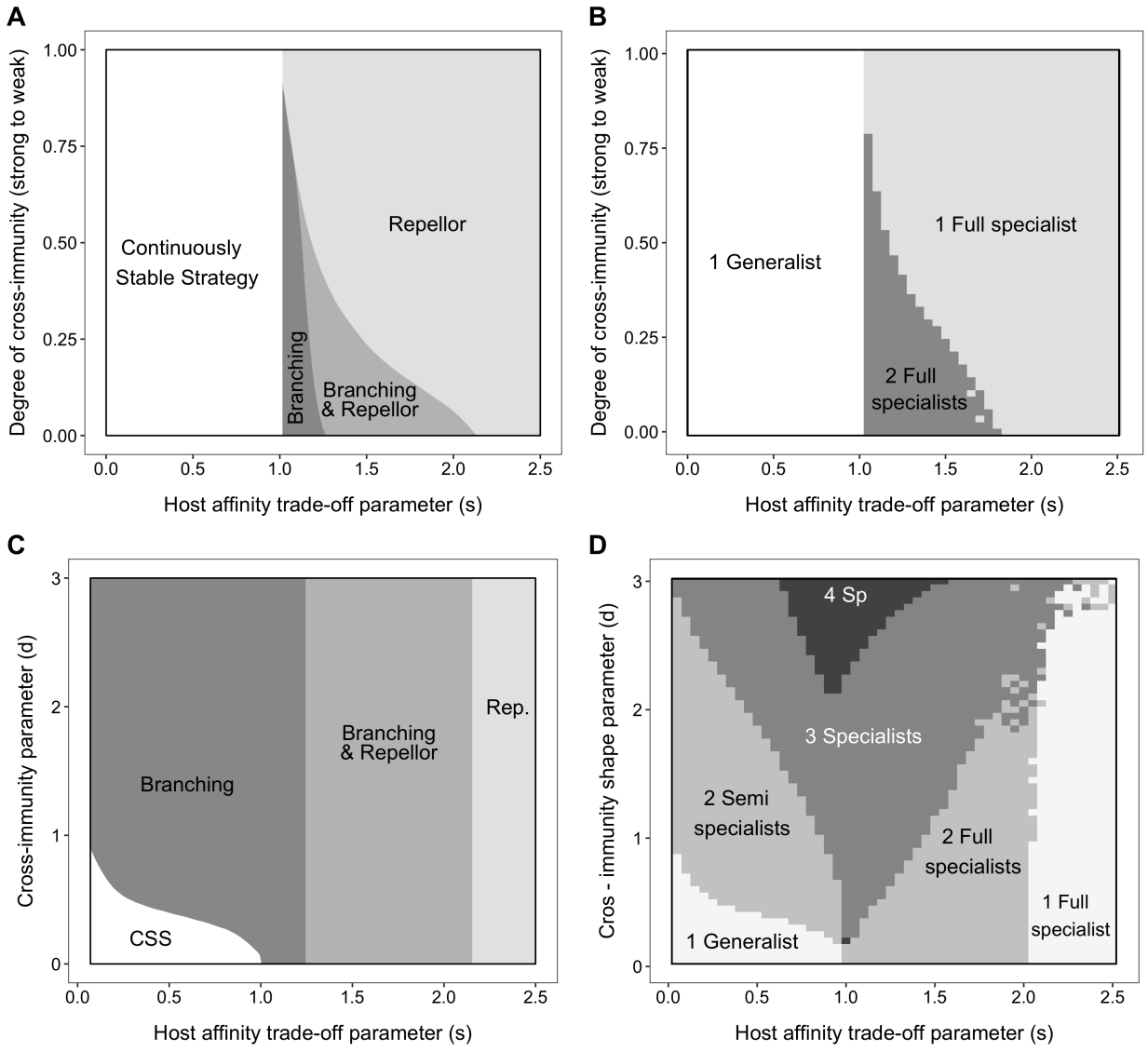

Figure 4.3: Long term behavior as a function of the strength of cross-immunity. Parameter space under which the system evolves to a CSS, Repellor and/or Evolutionary Branching strategy with their respective conversion into specialist or generalist. Complete cross-immunity is indicated by $\alpha = 0$ and $d = 0$. Figures A-B and C-D correspond to a fixed and variable α, respectively. Figures B and D were calculated from the output of numerical simulations (with an initial affinity value a_{Ax_i} of 0.5).

becoming neutral with respect to the binding affinity. When the trade-off is stronger ($s > 1$), evolutionary branching occurs under strong cross-immunity; that is, the population evolves to a dimorphic population and each strain becomes completely specialized for different hosts, avoiding competition. Finally, under a very strong trade-off ($s > 2$), the most likely scenario is one in which both strains become specialized for the same host (evolutionary repellors), this is because even very strong cross-immunity cannot push one of the populations to utilize the other host due to the steep trade-off. Stable coexistence of two specialized strains would be possible, however, if the evolutionary steps were larger, as is the case for zoonotic introductions. In summary, a weak trade-off favors the evolution of generalist strains that are neutral with respect to their ability to bind the host receptor, while the evolutionary outcome for a strong trade-off is determined by initial conditions, where the two strains can become completely specialized for different hosts or for the same one (fig 4.3B).

In contrast to the fixed cross-immunity, the parameter space for the model with variable α in which the two phenotypes (differences in binding and immunity) are linked, is dominated by evolutionary branching points, followed by repellors and CSS (fig 4.3C). After the branching occurs, the population becomes polymorphic with completely or partially specialized pathogens (fig 4.3D, fig 4.6), where the coexistence of multiple strains is now evolutionarily stable. This result also indicates that higher values of d, and thus narrower cross-immunity width (fig 4.1C), amplify the coexistence of semi-specialized strains, which is consistent with observations in nature for pathogens like rotavirus, as we discuss next.

4.4 Discussion

We have presented a transmission model that ties together host-pathogen affinity with the frequency dependence selection caused by specific immune responses. Previous studies of pathogen diversity have considered these two aspects of competitive dynamics independently. Concordant with previous results, we show that a weak trade-off leads to generalists, while a strong trade-off leads to full specialists, when the degree of specific immunity is fixed

(Egas et al. 2004). By contrast, if the strength of cross-immunity is functionally related to the difference in host affinity, the dynamics of the system do not follow from the simpler models. Our results show that coexistence of multiple semi-specialized strains able to infect both types of hosts is a common evolutionary scenario, favoring diversity. These results also suggest that populations need a weaker immune response to overcome the affinity trade-off (indicated by higher values of d, fig 4.3D) to maintain 2 or more phenotypes coexisting.

Two of the main assumptions of the adaptive dynamics framework are that the system initially consists of a monomorphic population (*i.e.* all pathogens have the same trait), and that most mutations have a small effect. These assumptions highly constrain the outcome by repellors pushing the two types to specialize on the same host (trajectories along the diagonal in the Trait evolution plots). However, viruses like flu and rotavirus can experience zoonotic introduction of somewhat different types, pushing the system into the region where two strains have very different affinity values. The system under these conditions could lead to the utilization of multiple hosts, especially when cross immunity is dependent upon the affinity difference, with the evolution of semi-specialized strains, as commonly seen in rotavirus (*i.e.* many of the P types can bind multiple HBGAs) (Van Trang et al. 2014, Nordgren et al. 2014, Ayouni et al. 2015, Ma et al. 2015, Lui et al. 2016, Ramani et al. 2016, Sun et al. 2016a, Sun et al. 2016b, Sun et al. 2016c, Zhang et al. 2016).

Another important assumption of this work is the consideration of a trait that has an effect in both immune recognition as well as host entry. Several studies in viral pathogens have shown that the major antigenic sites or epitopes are located in the same subunits of the proteins responsible for binding the host-receptor (Taylor et al. 1987, Ruggeri and Greenberg 1991, Taylor and Dimmock 1994). For the case of rotavirus, the major antigenic sites of the VP4 protein are in the VP8 subunit (Larralde et al. 1991, Kirkwood et al. 1996), the same subunit responsible of cell attachment inside the host (Ruggeri and Greenberg 1991). Additionally, strong association between infection by the predominant rotavirus genotypes and histo-blood group antigens have been found in children (Van Trang et al. 2014), implying

affinity differentiation and the presence of a trade-off in the way that the serotypes interact with the host receptor. Rotavirus' serotypes are defined based on the G and P genotypes that encode for the outer capsid protein (VP7) and the cell attachment protein (VP4) respectively. For human infections, around 11 P-types have been described (Matthijnssens et al. 2011), suggesting that strain diversity is higher than expected based solely on the number of blood types present in the population, and that therefore strains may not be completely specialized for a certain host type. The number of studies showing the association between histo-blood group types in the host and strain specificity has increased substantially in the last decade (*e.g.* Shirato et al. 2008, Van Trang et al. 2014, others). These studies have only addressed however the affinity for the host, leaving open questions about the role of immunity in the observed diversity, especially when the distribution of rotavirus serotypes differs among different countries (Gentsch et al. 2005). Previous theoretical work has explored the relationship between the binding site and antigenic response (Ndifon et al. 2009). Inspired by the population dynamics of influenza, the authors showed that the presence of epitopes located far from the receptor-binding site may disrupt viral neutralization causing steric interference between antibodies. This consideration is of relevance as changes in the receptor binding site can also modify the recognition by the immune system, with epidemiological consequences that are still poorly understood. It would also be valuable to further investigate whether viruses with significant diversity tend to use a range of similar receptors, or if less diverse viruses have become fixed on a single receptor with no similar receptors, as a consequence of different degrees of cross-immunity and trade-off. These empirical studies could make possible the evaluation of predictions made by theoretical models like the one presented in this article, with potential contributions to more effective vaccine development, particularly in developing countries where vaccine efficacy has been limited (Zaman et al. 2017).

4.5 Acknowledgments

We are grateful for the support of the Fogarty International Center at NIH [Program on the Ecology and Evolution of Infectious Diseases, EEID, R01-TW009670] to M.P., and the support provided by NIH K08 award 9(5K08AI119182-03) to R.J.W.

4.6 Supporting information

4.6.1 Transmission model

The population dynamics of the disease was represented with an SIR (Susceptive–Infected–Recovered) model that considers two populations of hosts (A and B) and multiple pathogen strains that differ in their affinity to bind the host receptor. The set of equations that describe the dynamics of the population A is given by:

$$
\begin{aligned}
\frac{dS_A}{dt} &= \mu P_A - \beta \frac{S_A}{P_A} \sum_{i=1}^{n} a_{Ax_i}(I_{A_1x_i} + I_{B_1x_i} + I_{A_2x_i} + I_{B_2x_i}) - \mu S_A \\
\frac{dI_{A_1x_i}}{dt} &= \beta \frac{S_A}{P_A} a_{Ax_i}(I_{A_1x_i} + I_{B_1x_i} + I_{A_2x_i} + I_{B_2x_i}) - \gamma I_{A_1x_i} - \mu I_{A_1x_i} \\
\frac{dR_{A_1x_i}}{dt} &= \gamma I_{A_1x_i} - \beta \frac{R_{A_1x_i}}{P_A} \sum_{j \neq i} \alpha_{ij}\ a_{Ax_j}(I_{A_1x_j} + I_{B_1x_j} + I_{A_2x_j} + I_{B_2x_j}) - \mu R_{A_1x_i} \\
\frac{dI_{A_2x_j}}{dt} &= \beta \sum_{i \neq j} \alpha_{ij} \frac{R_{A_1x_i}}{P_A} - \gamma I_{A_2x_j} - \mu I_{A_2x_j} \\
\frac{dR_A}{dt} &= \gamma \sum_{i=1}^{n} I_{A_2x_i} - \mu R_A
\end{aligned}
\tag{4.1}
$$

Similarly, the equations that characterize the population of host type B is given by:

$$\begin{aligned}
\frac{dS_B}{dt} &= \mu P_B - \beta \frac{S_B}{P_B} \sum_{i=1}^{n} a_{Bx_i}(I_{A_1x_i} + I_{B_1x_i} + I_{A_2x_i} + I_{B_2x_i}) - \mu S_B \\
\frac{dI_{B_1x_i}}{dt} &= \beta \frac{S_B}{P_B} a_{Bx_i}(I_{A_1x_i} + I_{B_1x_i} + I_{A_2x_i} + I_{B_2x_i}) - \gamma I_{B_1x_i} - \mu I_{B_1x_i} \\
\frac{dR_{B_1x_i}}{dt} &= \gamma I_{B_1x_i} - \beta \frac{R_{B_1x_i}}{P_B} \sum_{j \neq i} \alpha_{ij}\ a_{Bx_j}(I_{A_1x_j} + I_{B_1x_j} + I_{A_2x_j} + I_{B_2x_j}) - \mu R_{B_1x_i} \\
\frac{dI_{B_2x_j}}{dt} &= \beta \sum_{i \neq j} \alpha_{ij} \frac{R_{B_1x_i}}{P_B} - \gamma I_{B_2x_j} - \mu I_{B_2x_j} \\
\frac{dR_B}{dt} &= \gamma \sum_{i=1}^{n} I_{B_2x_i} - \mu R_B
\end{aligned} \tag{4.2}$$

The first infection is represented by the pathogen x_i and its respective affinity for hosts A (a_{Ax_i}) and B (a_{Bx_i}). The second infection is then characterized by pathogen x_j. The identity of each strain is defined by its ability to bind the host, with affinity values between 0 and 1. After the second infection (by a different strain), the host becomes immune to all future infections. For the first part of our analyzes we have fixed the value of α (between 0 and 1), and therefore α_{ij} is independent from the identity of the strains. For the second part, α_{ij} is a function of the distance between a_{Ax_i} and a_{Bx_i}, as described in the Methods section and fig 4.1C. The epidemiological parameters that we have used are described in table 4.1.

4.6.2 Evolutionary dynamics

In order to determine the ability of a new mutant to invade the system, we first estimated the conditions under which the system goes to a stable endemic equilibrium. The equilibrium will be stable if the real part of the leading eigenvalue of the corresponding Jacobian matrix is negative. We then included the equations that involve the behavior of the mutant. The new Jacobian matrix has a block-triangular form, where the diagonal blocks correspond

respectively to the Jacobian of the residents J_{res} (*i.e.* the system in the absence of the mutant) and the Jacobian of the mutant J_{mut}, as follows:

$$\begin{bmatrix} J_{res} & J \\ 0 & J_{mut} \end{bmatrix} \tag{4.3}$$

Therefore, the eigenvalues of the new system are those of these block-triangular matrix. Since the real part of the eigenvalue of J_{res} is negative, the stability of the new system is determined by the leading eigenvalue of J_{mut} (λ). If the real part of λ is positive, the equilibrium becomes unstable and thus the population of the mutant will increase over time, invading the system; conversely, if $\lambda < 0$, the mutant will not be able to invade and will go extinct. The invasion fitness of the mutant (x_k) is then defined by λ, and has the following form:

$$\begin{aligned} \lambda(x_i, x_j, x_k) = & \frac{a_{Ax_k}\beta(\alpha_{ik}R^*_{A_1x_i} + \alpha_{jk}R^*_{A_1x_j} + S^*_A)}{P^*_A} + \\ & \frac{a_{Bx_k}\beta(\alpha_{ik}R^*_{B_1x_i} + \alpha_{jk}R^*_{B_1x_j} + S^*_B)}{P^*_B} - \gamma - \mu \end{aligned} \tag{4.4}$$

where $a_{Bx_k} = (1 - a_{Ax_k}^{1/s})^s$ and α_{ik} can be a fixed value, or a function of the distance between a_{x_i} and a_{x_k}: $\alpha_{ik} = 1 - e^{-(d(a_{x_i} - a_{x_k}))^2}$. The direction of evolution as the result of successive invasions is determined by the selection gradient g, given by the first derivative of λ with respect to the trait of the mutant k and evaluated at the resident traits values i and j:

$$g_i(x_i, x_j) = \frac{\partial\lambda(x_i, x_j, x_k)}{\partial a_{Ax_k}}|_{k=i} \tag{4.5}$$

$$g_j(x_i, x_j) = \frac{\partial\lambda(x_i, x_j, x_k)}{\partial a_{Ax_k}}|_{k=j} \tag{4.6}$$

If the sign of the selection gradient is positive, mutants with traits higher than those of the residents will be successful, whereas if $g < 0$, mutants with traits lower than those of the residents will invade. The evolutionary singular strategy is characterized by the pair of traits a_{Ax_i} and a_{Ax_j} for which both values of selection gradients are zero.

4.6.3 Supplementary figures

Table 4.1: Epidemiological parameters.

Description (parameter)	Value
Birth/death rate (μ)	0.067 (1/year)
Recovery rate (γ)	4 (1/year)
Contact rate (β)	20

Parameter values where chosen based on the feasibility of steady states of the system for ecological time.

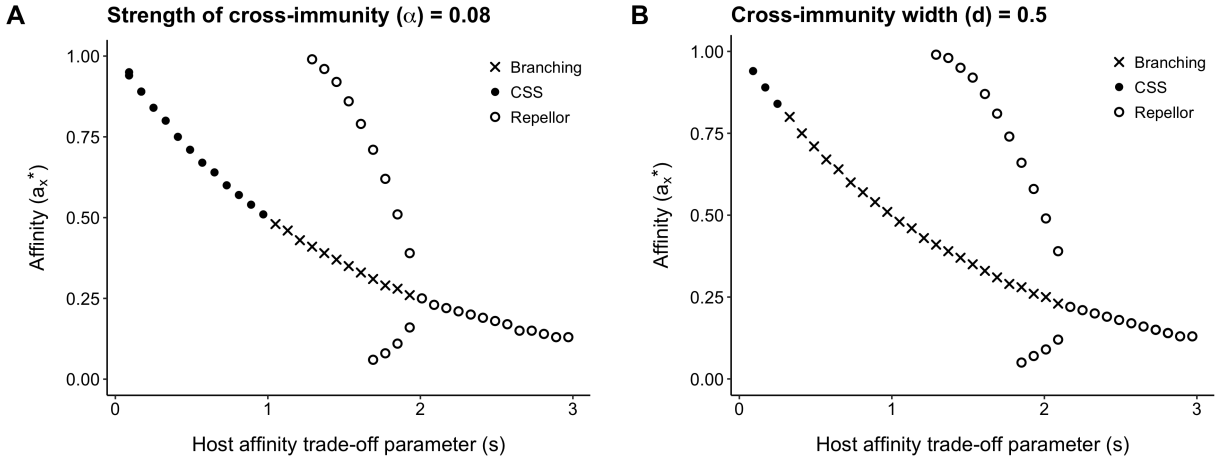

Figure 4.4: Bifurcation diagram. Evolutionary strategies as a function of the trade-off parameter s. The strategies under which the system may evolve are identified as: a Continuously Stable Strategy (CSS), a local evolutionarily stable strategy whose affinity evolves until it reaches the CSS and cannot be invaded by a nearby mutant; an Evolutionary branching: the affinity trait evolves towards the branching point and once this point is reached, the traits diverge and the population becomes dimorphic; a Repellor: an evolutionary unstable strategy. The filled dots, open dots, and x symbols indicate CSS, repellors, and branching strategies, respectively, and the grey arrows denote the direction of monomorphic evolution.

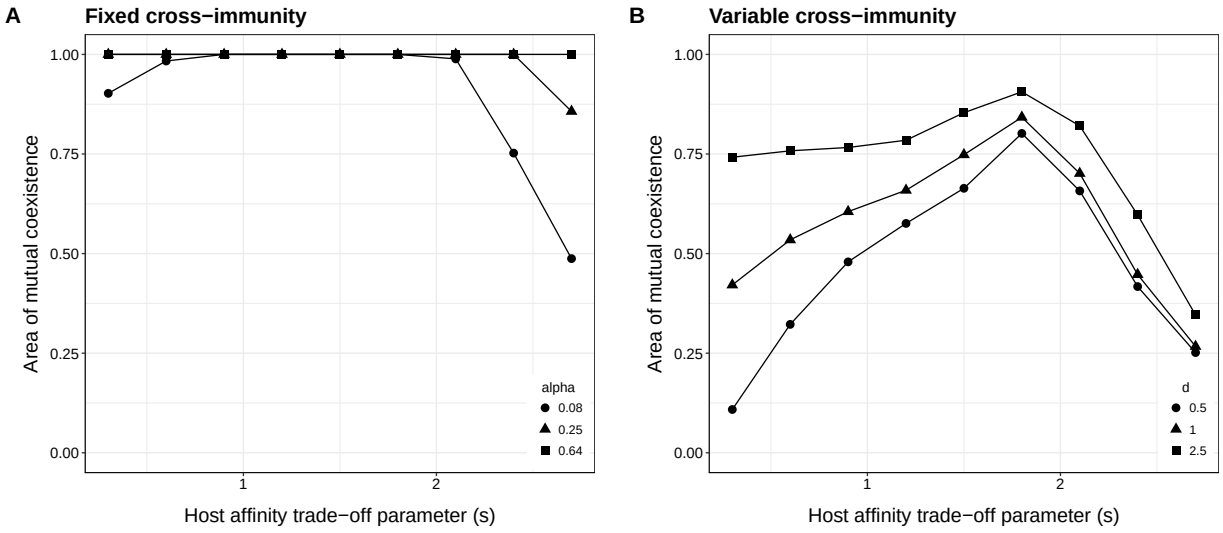

Figure 4.5: Percentage of all possible combinations of traits that allow for mutual invasion and coexistence as a function of s, for different values of fixed cross-immunity (A), and different values of d (variable cross-immunity, B).

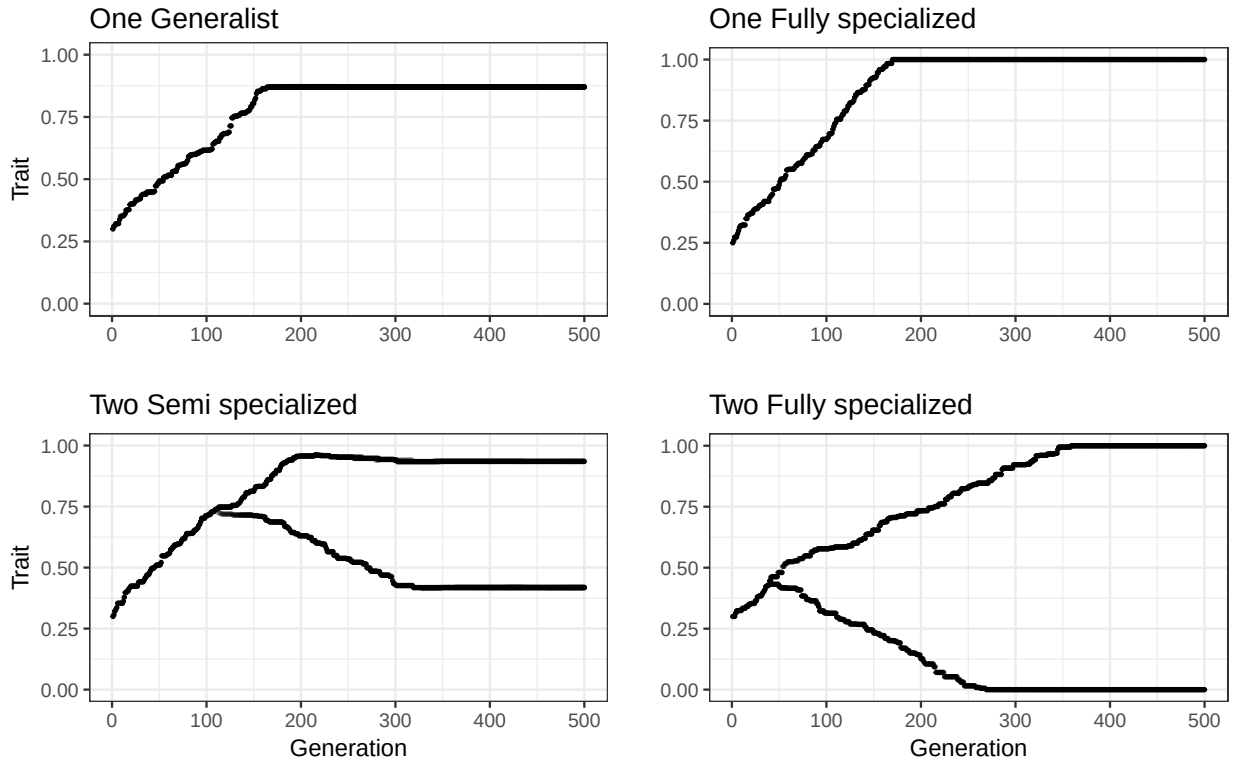

Figure 4.6: Representative evolutionary outcomes of affinity traits over time.

CHAPTER 5

CONCLUSION

5.1 Concluding remarks

Disease ecology involves the integration of mechanisms at different organizational scales, from interactions between pathogens and their hosts (e.g. immunology) to the population dynamics of transmission. Phylodynamics seeks to explain both epidemiological and phylogenetic patterns, including temporal fluctuations in disease incidence and the genetic diversity and evolutionary trajectories of the pathogen (Grenfell et al. 2004). The main goal of this dissertation was to improve our understanding of the epidemiological, ecological, and evolutionary mechanisms responsible for the population dynamics of rotavirus and its diversity.

In terms of epidemiological patterns, I focused on the interface between the population dynamics of rotavirus and environmental variability. Several infectious diseases exhibit significant variation in the size of seasonal outbreaks (e.g. Koelle et al. 2005, Laneri et al. 2010, Pascual et al. 2000). Since environmental variables may act at different spatial and temporal scales, identification of climate factors influencing disease transmission can be challenging, especially in the context of a large urban environment. Chapter 2 addressed this goal by quantifying the effect of flooding on the incidence of rotavirus in Dhaka, Bangladesh. I showed that the distinction of different parts of the city is essential to reveal the seasonal component of the disease transmission and its response to the monsoons. Specifically, the core and the periphery of Dhaka respond differentially to climate forcing, facilitating the persistence of reservoirs and thus the response to climate anomalies. This suggests that infectious diseases that have not been seen as climate-sensitive can become so under the environmental conditions found in megacities of the developing world.

For the components affecting the antigenic patterns of rotavirus, previous analyses in Dhaka have shown fluctuations in the strain distribution over time (Rahman et al. 2007); however, the mechanisms underlying this variability are still unknown. It is likely that

genetic changes, spatial structure, host heterogeneity, and environmental fluctuations all play a role. I took advantage of the strong conceptual connections with competitive systems in ecology in general. For infectious diseases, the hosts are seen as the resource and the pathogen population is subdivided into phenotypic strains that compete for those resources. Therefore, once a strain becomes frequent it depletes its resource and a rarer strain acquires a competitive advantage. This frequency-dependent interaction with selection mediated through immunity and cross-immunity sets the stage for a competitive system whose strain diversity emerges from the interplay of the population dynamics of the disease and the evolution of the pathogen. Here, I have looked at two particular aspects: the effect of specific and generalized immune response, and the effect of host heterogeneity in shaping rotavirus diversity.

In chapter 3, I evaluated the epidemiological parameters associated with the immune response against rotavirus. Strong generalized immunity together with a high specific immunity against VP7 explain the diversity patterns observed in Dhaka. Improvements of the current vaccines are possible if information about the specific epitopes in VP4 generating the generalized response is included.

In chapter 4, I considered the effects of host variation. I explored the consequences of host heterogeneity in the diversity of rotavirus by implementing a theoretical work that examines the long term evolution of host-pathogen in the presence of the immune response. These results showed that an association between the strength of cross-immunity and the difference in host affinity is needed to recover the long-term coexistence of several strains. Thus, the interplay of equalizing and stabilizing mechanisms is critical to better understand diversity patterns in pathogens.

5.2 Future directions

I highlight here specific areas still under research that can contribute to a deeper knowledge of both rotavirus and the population dynamics of RNA viruses in general.

5.2.1 *High dimensional systems*

High dimensionality is a common feature of biological systems, especially infectious diseases. Dynamical systems involving pathogens can exhibit not only rapid temporal variation, but also spatial variation and changes in host susceptibility at different ages, all of which complicate eradication strategies. Nowadays, technological developments such as cellular telephone data provide highly accurate spatio-temporal information on human movement (Richardson et al. 2013, Wesolowski et al. 2013), and have already contributed to a better understanding of transmission and disease risk at a local scale (Buckee et al. 2013, Dalziel et al. 2013, Wesolowski et al. 2012). However, the development of statistical inference methods for stochastic dynamical systems at high spatio-temporal resolution is still lacking. Extensions of existing methods based on likelihood and particle filters are good candidates to address this challenge. Future work to decrease the high computational cost and improve mathematical optimization would allow the inclusion of multiple spatial locations to fit models and have a meaningful impact on future epidemic control and public health outcomes.

5.2.2 *Reassortment and fitness differences*

Different evolutionary rates of antigenic change have been shown to result in different phylodynamic regimes and associated patterns of strain diversity, namely the successive strain replacement characteristic of influenza in contrast to the emergence of discordant strains occupying different ecological niches possibly applicable to rotavirus (Zinder et al. 2013). However, this previous study considered antigenic change resulting from mutation but did not address other evolutionary mechanisms by which novelty may arise. For rotavirus, reassortment combined with zoonotic introductions is one of the main sources of genomic novelty. The RNA segments can reassort, producing different combinations of the parental types. Although low reassortment rates could explain the coexistence of discordant serotypes (Zinder et al. 2017), work thus far has not considered the role of fitness constraints in the coexistence of these serotypes. Theory has largely focused on niche differences of the associated segments

that comprise alternative serotypes without accounting for their fitness differences resulting from functional differences and reproductive isolation. Fitness differences in serotypes can arise when their G-P associated proteins are able to persist and thus coevolve over time if the reassortment rate is low and there is enough isolation. The outcome of this fitness variation may lead to a limited success of certain reassortants, as well as to weakening the strength of the selective pressure imposed by the immune selection to exclude serotypes that overlap in their G or P types. If true, this hypothesis could explain, in part, the dominance of different serotypes in different regions.

References

Ayouni S., K. Sdiri-Loulizi, A. de Rougemont, M. Estienney, K. Ambert-Balay, S. Aho, S. Hamami, M. Aouni, M. Neji-Guediche, P. Pothier, G. Belliot. 2015. Rotavirus P [8] infections in persons with secretor and nonsecretor phenotypes, Tunisia. Emerging infectious diseases 21: 2055–2058.

Bányai, K., Gentsch, J.R., Martella, V., Bogdán, A., Havasi, V., Kisfali, P., Szabó, A., et al. 2009. Trends in the epidemiology of human G1P [8] rotaviruses: a Hungarian study. The Journal of infectious diseases, 200: S222–S227.

Bányai, K., László, B., Duque, J., Steele, A.D., Nelson, E.A.S., Gentsch, J.R. and Parashar, U.D. 2012. Systematic review of regional and temporal trends in global rotavirus strain diversity in the pre rotavirus vaccine era: insights for understanding the impact of rotavirus vaccination programs. Vaccine 30: A122–A130.

Blackhall, J., Fuentes a, Magnusson, G. 1996. Genetic stability of a porcine rotavirus RNA segment during repeated plaque isolation. Virology 225:181–190.

Bresee J., Z.Y. Fang, B. Wang, E.A.S. Nelson, J. Tam, Y *et al.* 2004. First report from the Asian Rotavirus Surveillance Network. Emerging Infectious Diseases 10: 988–995.

Buckee C.O., A. Wesolowski, N. Eagle, E. Hansen, R.W. Snow. 2013. Mobile phones and malaria: modeling human and parasite travel. Travel Med Infect Dis 11:15–22

Cash B.A., X. Rodó, J.L. Kinter, III. 2008. Links Between Tropical Pacific SST and the regional climate of Bangladesh: Role of the eastern and Central Tropical Pacific. J Clim 21: 4647–4663.

Cash B.A., X. Rodó, M. Emch, Md. Yunus, A.S.G Faruque, M. Pascual. 2014. Cholera and Shigellosis: Different epidemiology but similar responses to climate variability. . PLoS One 9:e107223.

Cazelles B, M. Chavez, A.J. McMichael, S. Hales. 2005. Nonstationary influence of El Niõ on the synchronous dengue epidemics in Thailand. PLoS Medicine 2:e106.

Chesson, P. 2000. Mechanisms of maintenance of species diversity. Annual review of Ecology and Systematics 31: 343–366.

Chiba S, S. Nakata, T. Urasawa, S. Urasawa, T. Yokoyama, Y. Morita, K. Taniguchi, T. Nakao. 1986. Protective effect of naturally acquired homotypic and heterotypic rotavirus antibodies. The Lancet 328: 417 – 421.

Colwell R. R.1996. Global Climate and Infectious Disease: The Cholera Paradigm. Science 274: 2025–2031.

Cook S.M., R.I. Glass, C.W. LeBaron, M.S. Ho. 1990. Global seasonality of rotavirus infections. Bulletin of the World Health Organization 68: 171–177.

Dalziel B.D., B. Pourbohloul, S.P. Ellner. 2013. Human mobility patterns predict divergent epidemic dynamics among cities. Proceedings of the Royal Society B: Biological Sciences 280: 1766.

Dewan, A., Corner, R. 2014. Dhaka Megacity. Geospatial Perspectives on Urbanization, Environment and Health. Springer Geography.

Dieckmann U., R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. Journal of mathematical biology 34: 579–612

Donker, N.C., Kirkwood, C.D.,. 2012. Selection and evolutionary analysis in the nonstructural protein NSP2 of rotavirus A. Infection, Genetics and Evolution 12: 1355–1361.

Dormitzer, P.R., Nason, E.B., Prasad, B.V. and Harrison, S.C. 2004. Structural rearrangements in the membrane penetration protein of a non-enveloped virus. Nature 430: 1053.

Dóró R., László B., Martella, V., Leshem, E., Gentsch, J., Parashar, U., Bányai K. 2014 Review of global rotavirus strain prevalence data from six years post vaccine licensure surveillance: is there evidence of strain selection from vaccine pressure? Infection, Genetics and Evolution 28: 446–461.

DSouza R.M., G. Hall, N.G. Becker. 2008. Climatic factors associated with hospitalizations for rotavirus diarrhoea in children under 5 years of age. Epidemiol. Infect 136: 56–64.

Egas, M., U. Dieckmann, M. W. Sabelis. 2004. Evolution restricts the coexistence of specialists and generalists: the role of trade-off structure. The American Naturalist 163: 518–531.

Faisal, I. M., Kabir, M. R., Nishat, A. 2003. The disastrous flood of 1998 and long term mitigation strategies for Dhaka City. In Flood Problem and Management in South Asia. Springer Netherlands. Natural Hazards 28: 85–99.

Ferdous F., S. Ahmed, F. D. Farzana, J. Das, M. A. Malek, S. K. Das, M. A. Salam and A. S. G. Faruque. 2015. Aetiologies of diarrhoea in adults from urban and rural treatment facilities in Bangladesh. Epidemiology and Infection 143: 1377–1387.

Gentsch J. R., A .R. Laird, B. Bielfelt, D. D. Griffin, K. Bányai, M. Ramachandran, V. Jain, N. A. Cunliffe, O. Nakagomi, C. D. Kirkwood, T. K. Fischer. 2005. Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. Journal of Infectious Diseases 192: S146–S159.

Geritz, S. A., E. Kisdi, G. Meszéna , J. A. Metz. 1997 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evolutionary ecology 12: 35–57.

Gog, J. R., B. T. Grenfell. 2002. Dynamics and selection of many-strain pathogens. Proceedings of the National Academy of Sciences 99: 17209–17214.

Gomes, M. G., G. F. Medley, D. J. Nokes. 2002. On the determinants of population structure in antigenically diverse pathogens Proceedings of the Royal Society of London B: Biological Sciences 269: 227–233.

Grenfell, B. T., O. G. Pybus, J. R. Gog, J. L. N. Wood,J. M. Daly,J. A. Mumford, et al. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303: 327–332.

Gupta, S., M. C. J, Maiden, I. M. Feavers, S. Nee, R. M. May, R. M. Anderson. 1996. The maintenance of strain structure in populations of recombining infectious agents. Nature medicine 2: 437–442.

Gupta, S., N. Ferguson, R. Anderson. 1998. Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. Science 280: 912–915.

Gupta, S., M. C. J, Maiden. 2001 Exploring the evolution of diversity in pathogen populations. Trends in microbiology 9: 181–185.

Hashizume M., B. Armstrong, Y. Wagatsuma, A.S.G. Faruque, T. Hayashi, D.A. Sack. 2008. Rotavirus infections and climate variability in Dhaka, Bangladesh: a time-series analysis. Epidemiology and Infection 136(09): 1281–1289.

Hieber J.P, S. Shelton, J.D. Nelson, J. Leon, E. Mohs. 1978. Comparison of human rotavirus disease in tropical and temperate settings. American Journal of Diseases of Children 132: 853–858.

Hoshen M. and A. Morse. 2004. A weather-driven model of malaria transmission. Malaria Journal 3: 32–46.

Hudson P. J. and I. M. Cattadori. 1999. The Moran effect: a cause of population synchrony. Trends in Ecology & Evolution 14: 1–2.

Hungerford, D., Vivancos, R., Read, J.M., Pitzer, V.E., Cunliffe, N., French, N. and Iturriza-Gmara, M. 2016. In-season and out-of-season variation of rotavirus genotype distribution and age of infection across 12 European countries before the introduction of routine vaccination, 2007/08 to 2012/13. Eurosurveillance, 21(2).

Koelle K., X. Rodó, M. Pascual, M. Yunus, G. Mostafa. 2005. Refractory periods and climate forcing in cholera dynamics. Nature 436: 696–700.

King A. A., D. Nguyen, E. L. Ionides. 2016. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. Journal of Statistical Software 69: 1 – 43.

Kirkwood, C. D., R. F. Bishop, B. S. Coulson. 1996. Human rotavirus VP4 contains strain-specific, serotype-specific and cross-reactive neutralization sites. Archives of virology 141: 587–600.

Konno T., H. Suzuki, N. Katsushima, A. Imai, F. Tazawa, T. Kutsuzawa, S. Kitaoka, M. Sakamoto, N. Yazaki, N. Ishida. 1983. Influence of temperature and relative humidity on human rotavirus infection in Japan. Journal of Infectious Diseases 147: 125–128.

Laneri K., A. Bahdra, E. Ionides, M.J. Bouma, R. Dhiman, Y. Rajpal, M. Pascual. 2010. Forcing versus feedback: epidemic malaria and monsoon rains in NW India. PLoS Computational Biology 6: e1000898.

Laneri, K., Paul, R. E., Tall, A., Faye, J., Diene-Sarr, F., Sokhna, C., Trape, J.F., Rodo, X. 2015. Dynamical malaria models reveal how immunity buffers effect of climate variability. Proceedings of the National Academy of Sciences 112: 8786–8791.

Larralde, G. I., B. G. Li, A. Z. Kapikian, M. A. Gorziglia. 1991. Serotype-specific epitope (s) present on the VP8 subunit of rotavirus VP4 protein. Journal of virology 65: 3213–3218.

Levy A., E. Hubbard, J.N.S. Eisenberg. 2009. Seasonality of rotavirus disease in the tropics: a systematic review and meta-analysis. International Journal of Epidemiology. 38: 1487–1496.

Lipp E.K., A. Huq, R.R. Colwell. 2002. Effects of Global Climate on Infectious Disease: the Cholera Model. Clin. Microbiol. Rev. 15: 757–770.

Lipsitch M., O'Hagan J. J. 2007. Patterns of antigenic diversity and the mechanisms that maintain them. Journal of the Royal Society Interface 4: 787–802.

Liu Y., T. A. Ramelot, P. Huang, Y. Liu, Z. Li, T. Feizi., W. Zhong, F. T. Wu, M. Tan, M. A. Kennedy, X. Jiang. 2016. Glycan Specificity of P [19] Rotavirus and Comparison with Those of Related P Genotypes. Journal of virology 90: 9983–9996.

Ma X., X. Sun, Y. Guo, J. Xiang, W. Wang, L. Zhang, Q. Gu, Z. Duan. 2015. Binding patterns of rotavirus genotypes P [4], P [6], and P [8] in China with histo-blood group antigens. PloS one 10: e0134584.

Macarthur, R., R. Levins. 1967. The limiting similarity, convergence, and divergence of coexisting species. The American Naturalist 101: 377–385.

Matthijnssens, J., Heylen, E., Zeller, M., Rahman, M., Lemey, P., Van Ranst, M. 2010. Phylodynamic analyses of rotavirus genotypes G9 and G12 underscore their potential for swift global spread. Mol. Biol. Evol. 27: 2431–2436.

Matthijnssens J., M. Ciarlet, S. M. McDonald, H. Attoui, K. Bányai, J. R. Brister, J. Buesa, M. D. Esona, M. K. Estes, J. R. Gentsch, M. Iturriza–Gómara M. 2011. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). Archives of virology 156:1397–1413.

McDonald, S.M., Matthijnssens, J., McAllen, J.K., Hine, E., Overton, L., Wang, S., et al., 2009. Evolutionary dynamics of human rotaviruses: balancing reassortment with preferred genome constellations. . PLoS Pathog. 5: e1000634.

Metz, J. A., S. A. Geritz, G. Meszéna, F. J. Jacobs, J. S. Van Heerwaarden. 1996 Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction Stochastic and spatial structures of dynamical systems 45:183–231.

Moran P. A. P. 1953. The statistical analysis of the Canadian lynx cycle. Australian Journal of Zoology 1: 291–298.

Nagao Y. and K. Koelle. 2008. Decreases in dengue transmission may act to increase the incidence of dengue hemorrhagic fever. Proceedings of the National Academy of Sciences 105: 2238–2243.

Nair, N., Feng, N., Blum, L.K., Sanyal, M., Ding, S., Jiang, B., Sen, A., Morton, J.M., He, X.S., Robinson, W.H. and Greenberg, H.B.. 2017. VP4-and VP7-specific antibodies mediate heterotypic immunity to rotavirus in humans. Science Translational Medicine 9:eaam5434.

Nakagomi T., O. Nakagomi, Y. Takahashi, M. Enoki, T. Suzuki, P.E. Kilgore. 2005. Incidence and burden of rotavirus gastroenteritis in Japan, as estimated from a prospective sentinel hospital study. Journal of Infectious Diseases 192: S106–S110.

Ndifon W., N. S. Wingreen, S. A. Levin. 2009. Differential neutralization efficiency of hemagglutinin epitopes, antibody interference, and the design of influenza vaccines. Proceedings of the National Academy of Sciences 106: 8701–8706.

Nordgren J., S. Sharma, F. Bucardo, W. Nasir, G. Günaydın, D. Ouermi, L. W. Nitiema, S. Becker-Dreps, J. Simpore, L. Hammarström, G. Larson. 2014. Both Lewis and secretor status mediate susceptibility to rotavirus infections in a rotavirus genotype–dependent manner. Clinical Infectious Diseases 59:1567–1573.

Parashar U.D., E.G. Hummelman, J.S. Bresee, M.A. Miller, R.I. Glass . 2003. Global illness and deaths caused by rotavirus disease in children. Emerging Infectious Diseases 9: 565–572.

Pascual M., X. Rodo, P. Ellner, R. Colwell, M.J. Bouma. 2000. Cholera dynamics and El Niño-Southern Oscillation. Science 289:1766–1769.

Pascual M. 2015. Climate and Population Immunity in Malaria Dynamics: Harnessing Information from Endemicity Gradients. Trends in parasitology 31:532–534.

Patel M.M., V.E., Pitze, W.J. Alonso, D. Vera, B. Lopman, J. Tate, C. Viboud, U.D. Parashar. 2013. Global seasonality of rotavirus disease. Pediatr Infect Dis J 32: e134–e147.

Pitzer V.E., C. Viboud, L. Simonsen, *et al.* 2009. Demographic variability, vaccination, and the spatiotemporal dynamics of rotavirus epidemics. Science 325: 290–294.

Pitzer, V.E., Viboud, C., Lopman, B.A., Patel, M.M., Parashar, U.D., Grenfell, B.T. 2011. Influence of birth rates and transmission rates on the global seasonality of rotavirus incidence. Journal of the Royal Society Interface, 64: 1584–1593.

Pitzer, V. E., J. Bilcke, E. Heylen, F. W. Crawford, M. Callens., F. De Smet, M. Van Ranst, M. Zeller, J. Matthijnssens, J. 2015. Did large-scale vaccination drive changes in the circulating rotavirus population in Belgium?. Scientific reports 5: 18585.

Purohit S. G., S. D. Kelkar, V. K. Simha. 1998. Time Series Analysis of Patients with Rotavirus Diarrhoea in Pune, India. J Diarrhoeal Dis Res 16:74–83.

Rahman M., R. Sultana, G. Ahmed, S. Nahar, Z.M. Hassan *et al.* 2007. Prevalence of G2P[4] and G12P[6] rotavirus, Bangladesh. Emerg Infect Dis 13: 18–24.

Ramani, S., Cortes-Penfield, N.W., Hu, L., Crawford, S.E., Czako, R., et al . 2013. The VP8* domain of neonatal rotavirus strain G10P [11] binds to type II precursor glycans. Journal of virology 87: 7255–7264.

Ramani S., L. Hu, B. V. V. Prasad, M. K. Estes. 2016. Diversity in Rotavirus–Host Glycan Interactions: A "Sweet" Spectrum. CMGH Cellular and Molecular Gastroenterology and Hepatology 2: 263–273.

Reiner R. C., A. A. Kinga, M. Emchc, M. Yunusd, A. S. G. Faruqued, M. Pascual . 2012. Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. Proceedings of the National Academy of Sciences 109: 2033–2036.

Richardson D.B., N.D. Volkow, M.P. Kwan, R.M. Kaplan, M.F. Goodchild, R.T. Croyle. 2013. Spatial turn in health research. Science 339: 1390–1391.

Ruggeri F. M., H. B. Greenberg. 1991 Antibodies to the trypsin cleavage peptide VP8 neutralize rotavirus by inhibiting binding of virions to target cells in culture. Journal of virology 65: 2211–2219.

Shirato, H., S. Ogawa, H. Ito, T. Sato, A. Kameyama, H. Narimatsu, Z. Xiaofan, T. Miyamura, T. Wakita, K. Ishii, N. Takeda. 2008. Noroviruses distinguish between type 1 and type 2 histo-blood group antigens for binding. Journal of virology 82: 10756–10767.

Stoll B.J., R.I. Glass, M.I. Huq, M.U. Khan, J.E. Holt, H. Banu. 1982. Surveillance of patients attending a diarrhoeal disease hospital in Bangladesh. British Medical Journal (Clinical Research Edition) 285: 1185–1188.

Sultan B., K. Labadi, J.F. Guégan, S. Janicot. 2005. Climate drives the meningitis epidemics onset in West Africa. PLoS Medicine 2: e6.

Sun X., N. Guo, D. Li, M. Jin, Y. Zhou, G. Xie, L. Pang, Q. Zhang, Y. Cao, Z. Duan. 2016a. Binding specificity of P [8] VP8* proteins of rotavirus vaccine strains with histo-blood group antigens. Virology 495: 129–135.

Sun X., D. Li, R. Peng, N. Guo, M. Jin, Y. Zhou, G. Xie, L. Pang, Q. Zhang, J. Qi, Z. Duan. 2016b. Functional and structural characterization of P [19] rotavirus VP8* interaction with histo-blood group antigens. Journal of Virology 90: 9758–9765.

Sun X., N. Guo, J. Li, X. Yan, Z. He, D. Li, M. Jin, G. Xie, L. Pang, Q. Zhang, N. Liu. 2016c. Rotavirus infection and histo-blood group antigens in the children hospitalized with diarrhoea in China. Clinical Microbiology and Infection 22: 740–e1.

Tate J.E., A.H. Burton, C. Boschi-Pinto, A.D. Steele, J. Duque, U.D. Parashar. 2012. 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: A systematic review and meta-analysis. Lancet Infect Dis 12: 136–141.

Taylor H. P., S. J. Armstrong, N. J. Dimmock. 1987. Quantitative relationships between an influenza virus and neutralizing antibody. Virology 159: 288–298.

Taylor H. P., N. J. Dimmock. 1994 Competitive binding of neutralizing monoclonal and polyclonal IgG to the HA of influenza A virions in solution: Only one IgG molecule is bound per HA trimer regardless of the specificity of the competitor. Virology 205: 360–363.

Tilman, D. 1982. Resource competition and community structure. Princeton university press.

Van Trang, N., H. T. Vu, N. T. Lee, P. Huang, X. Jiang, D. D. Anh. 2014. Association between norovirus and rotavirus infection and histo-blood group antigen types in Vietnamese children. Journal of clinical microbiology 52: 1366–1374.

Velazquez F.R., D.O. Matson, J.J. Calva, *et al.* 1996. Rotavirus infections in infants as protection against subsequent infections. N. Engl. J. Med. 335: 1022–1028.

Volz, E. M., K. Koelle, T. Bedford. 2013. Viral phylodynamics. PLoS Comput Biol 9: e1002947.

Wesolowski A., N. Eagle, A.J. Tatem, D.L. Smith, A.M. Noor, R.W. Snow, C.O. Buckee. 2012. Quantifying the impact of human mobility on malaria. Science 338: 267–270.

Wesolowski A., N. Eagle, A. M. Noor, R. W. Snow, C. O. Buckee. 2013. The impact of biases in mobile phone ownership on estimates of human mobility. J. R. Soc. Interface 10: 20120986.

Woods, R. J. 2015. Intrasegmental recombination does not contribute to the long-term evolution of group A rotavirus. Infection, Genetics and Evolution 32: 354–360.

Zaman, K., Sack, D.A., Neuzil, K.M., Yunus, M., Moulton, L.H., Sugimoto, J.D., Fleming, J.A., Hossain, I., El Arifeen, S., Azim, T. and Rahman, M. 2017. Effectiveness of a live oral human rotavirus vaccine after programmatic introduction in Bangladesh: A cluster-randomized trial. PLoS Medicine 14: e1002282.

Zhang X., Y. Long, M. Tan, T. Zhang, Q. Huang, X. Jiang, W. Tan, J. Li, G. Hu, S. Tang, Y. Dai. 2016. P [8] and P [4] Rotavirus Infection Associated with Secretor Phenotypes Among Children in South China. Scientific Reports 6: 34591.

Zhou G, Minakawa N, Githeko AK, Yan GY.. 2004. Association between climate variability and malaria epidemics in the East African highlands. Proceedings of the National Academy of Sciences 101: 2375–2380.

Zinder, D., Bedford, T., Gupta, S. and Pascual, M. 2013. The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. PLoS Pathogens 9: e1003104.

Zinder, D., Riolo, M.A., Woods, R.J. and Pascual, M. 2017. Role Of Competition In The Strain Structure Of Rotavirus Under Invasion And Reassortment. bioRxiv https://doi.org/10.1101/137661.