Linking Social and Pathogen Transmission Networks Using Microbial Genetics in East African Ungulates

By

Kimberly L. VanderWaal
B.S. (University of Minnesota) 2007

DISSERTATION
Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Animal Behavior

in the
OFFICE OF GRADUATE STUDIES
of the

UNIVERSITY OF CALIFORNIA
DAVIS

Approved:

____________________________
Lynne A. Isbell, Chair

____________________________
Edward R. Atwill

____________________________
Brenda McCowan

Committee in Charge
2013
# Table of Contents

Table of Contents........................................................................................................iii

Abstract.......................................................................................................................... vi

Acknowledgments........................................................................................................ viii

Chapter 1: Heterogeneity in pathogen transmission: mechanisms and methodology ...... 1

  Introduction.................................................................................................................. 1
  Mechanisms promoting heterogeneity ..................................................................... 5
  Scales of heterogeneity ............................................................................................ 14
  Quantifying heterogeneity ....................................................................................... 19
  Literature cited ......................................................................................................... 24

Chapter 2: Multi-level social organization and space use in reticulated giraffe (*Giraffa camelopardalis*) ........................................................................................................ 35

  Abstract..................................................................................................................... 35
  Introduction................................................................................................................ 36
  Methods....................................................................................................................... 40
  Results......................................................................................................................... 48
  Discussion................................................................................................................... 54
  Literature cited ......................................................................................................... 62

Chapter 3: Linking social and pathogen transmission networks using microbial genetics in giraffe (*Giraffa camelopardalis*) ........................................................................... 72
Network analysis has been lauded as the next key tool in wildlife disease ecology, yet because of limitations in detecting the occurrence of transmission, past studies have been based on the possibility that transmission might occur between interacting individuals. I used the genetics of a diverse microbe, *Escherichia coli*, to reveal that transmission has already occurred. If two animals shared the same genetic subtype of *E. coli*, then it can be inferred that these animals were part of the same transmission chain. Data on subtype sharing were used to link individuals in a transmission network. Here, I combine a network approach with microbial genetics to assess pathways of (I) intraspecific transmission in giraffe (*Giraffa camelopardalis*), and (II) interspecific transmission between wild and domestic ungulates in Laikipia, Kenya.

I first quantified giraffe social networks and determined that the flexible fission-fusion association patterns exhibited by giraffe are embedded in higher levels of social organization. I then found that giraffe that were more strongly linked in the social network were more likely to share *E. coli* subtypes, while neither spatial overlap nor number of shared water sources was related to the likelihood of being linked in the
transmission network. These results emphasize the importance of association patterns in understanding transmission dynamics, even in multi-host, environmentally transmitted pathogens such as *E. coli*. I also used *E. coli* subtyping to construct a multi-species transmission network for ten wild and domestic species. By quantifying each individual’s position, I demonstrated that there is variation across species in their connectivity in the transmission network, with some contributing disproportionately to pathogen transmission. High levels of connectivity exhibited by individuals of key species indicate that there are potential “super-spreader” species in this ecosystem. Grant’s gazelles (*Gazella granti*), for example, are connected to a greater number of others, while plains zebra (*Equus burchelli*) may serve as bottlenecks for pathogen flow. The novel methods used here combine a network approach with microbial genetics to provide a sophisticated and effective tool for tracing transmission pathways in wildlife. These methods are broadly applicable and improve our ability to predict the spread of disease.
Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Lynne Isbell, and my dissertation committee, Drs. Brenda McCowan and Rob Atwill. Their confidence in me never wavered and I would never have been able to come this far without their support. As my advisor and friend, Lynne provided huge amounts of expertise and insight on animal behavior, the ecology of African ecosystems, and how to survive field research in Kenya. Both Brenda and Rob adopted me into their respective laboratories, providing me with both space to work and colleagues with similar interests. They each made valuable and diverse contributions to this dissertation, freely sharing their time and expertise during many in-depth conversations about social network theory, genetic analyses, and statistics.

I thank Dr. Kiama Gitahi of the University of Nairobi. He was responsible for my appointment as a research associate in the Wangari Maathai Institute for Peace and Environmental Studies, which greatly facilitated my research permit application process. I also thank the Kenya Wildlife Service, especially Vincent Obanda and Dr. Matthew Mutinda. I wish to also acknowledge the Earthwatch program at Ol Pejeta Conservancy, especially Joseph Makau, who shared their experience data with me. In addition, I thank all the staff at Ol Pejeta Conservancy, especially Nathan Gichohi and Dr. George Omondi Paul, who provided logistical aid throughout the project. I am also grateful to Catherine Mundia and Joseph Koskei, who made staying at the Research Centre a pleasure.
I gratefully acknowledge Nicole Sharpe, Phillip Buckham-Bonnet, Jordan Lillienstein, and Sophie Preckler-Quisquater for their invaluable assistance in the fieldwork. They helped with giraffe identification, fecal sample collection, lab work, and most importantly, consuming massive quantities of goat at our nyama choma parties. It was these people, together with research colleagues Blair Roberts, Jenny Schieltz, and George Omondi Paul, who made the 14 months I spent living in Ol Pejeta Conservancy some of the best months of my life.

I also want to acknowledge the Atwill laboratory for all their help in teaching me the laboratory techniques necessary to complete this work. I especially want to thank my undergraduate volunteers, Elaine Wang, Young Ha Suh, and Navreen Pandher. Without them, I truly believe it would have taken me a year longer to finish this dissertation. I also thank Missy Partyka for her patience in teaching me ArcGIS and Gail Patricelli’s laboratory for allowing me to conduct the GIS analysis using their computers and software.

I would also like to thank my co-authors, Fushing Hseih and Hui Wang. They brought new and innovative statistical tools to Chapter 2, and this substantially increased the quality of the analysis. I also thank Drs. Vanessa Ezenwa, Andy Sih, and Truman Young for valuable comments during the development of the project, and the faculty of the Animal Behavior Graduate Group for their overwhelming commitment to educating young scientists such as myself.

I want to thank all the people who have supported me through my graduate work. In particular, I want to thank my family who been there every step of the way throughout my education and life; for that, I will be forever grateful. My parents, Joel and Debby
VanderWaal, are a large reason that I entered the field of wildlife ecology in the first place. We visited national parks and other wild places throughout my childhood, which fostered my love of wildlife from an early age. In addition, they have been incredibly supportive throughout my graduate work and even came to visit me at my field site in Kenya.

My sister and best friend, Kristy VanderWaal Mills, and her husband Travis Mills, are invaluable and irreplaceable companions in my life. I owe so much to them for their advice, support, and companionship.

I also want to thank my cohort, Daniel Gottlieb, Alli Heagerty, Sarah Laredo, Erin Mintline, and Jenn Phillips (and Jamie Bunting), as well as all my other friends in the Animal Behavior Graduate Group, especially Vivek Thuppil, Nicole Sharpe, and Beth Schultz. I have learned so much from these people. We have been through a lot together, sharing in all the triumphs and frustrations of graduate school.

I would like to give special thanks to Dr. George Omondi Paul. He has been a steadfast source of support, especially in the final months of my dissertation. In addition, he provided considerable aid in logistical aspects of the project. He also has been a treasure-trove of knowledge about infectious diseases and epidemiology in wildlife. More importantly, he was a constant source of support, encouragement and motivation.

This research was approved by Kenya’s National Council for Science and Technology (Permit NCST/RRI/12/1/MAS/147) and the UC Davis Institutional Animal Care and Use Committee (protocol no. 15887). This project would not have been possible without many organizations that supported this research. I was supported by the ARCS
Foundation and a National Science Foundation Graduate Research Fellowship. This research was supported by the National Science Foundation (Doctoral Dissertation Improvement Grant IOS-1209338), UC Davis Wildlife Health Center, Phoenix Zoo, Explorer’s Club, Oregon Zoo Future for Wildlife Program, American Society of Mammalogists, Cleveland Metroparks Zoo, Cleveland Zoological Society, Sigma Xi, Animal Behavior Society, Northeastern Wisconsin Zoo, and the UC Davis Animal Behavior Graduate Group.
Chapter 1: Heterogeneity in pathogen transmission: mechanisms and methodology

Introduction

In the past 15 years, there has been increasing evidence that understanding heterogeneity in pathogen transmission is key to predicting the dynamics of infectious diseases. Heterogeneity is commonly observed in epidemics, where most individuals infect few others while a small subset of the population is responsible for the majority of new cases (Woolhouse et al. 1997; Lloyd-Smith et al. 2005). The classic example of this is “Typhoid Mary”, who was responsible for 28 outbreaks of typhoid fever in the early 20th century (Hudson et al. 2008), but there is evidence for “super-spreaders” in many other human disease outbreaks (Lloyd-Smith et al. 2005). Numerous theoretical modeling papers have considered the implications of individual-level heterogeneity for the dynamics of infectious disease (Keeling 1999; Cross et al. 2004; Keeling 2005; Keeling and Eames 2005; Lloyd-Smith et al. 2005; Tildesley et al. 2010; Ames et al. 2011). However, there has yet to be a cohesive discussion on the mechanisms that give rise to heterogeneity. Most discussion so far has been on a case-by-case basis with little attempt to ground the discussion in a common framework (Paull et al. 2012). Specifically, the questions that need to be addressed are what mechanisms give rise to individual heterogeneity in pathogen transmission, at what scales does heterogeneity occur, and how can we quantify heterogeneity.
The ability of a pathogen to spread in a population is commonly summarized by the basic reproduction number, $R_0$, which is defined as the mean number of secondary cases produced by a single initial case in a naïve population (Anderson and May 1992). For an infection to spread in a population, $R_0$ must be greater than one. Thus, effective pathogen control strategies aim to reduce $R_0$ to below one. The proportion of the population that must be covered by control measures (i.e., treatment, vaccination, etc) in order to achieve this reduction is higher when the population exhibits heterogeneity than when homogeneity is assumed. In traditional models, $R_0$ is defined as a population average and does not account for the fact that the number of secondary cases produced by an individual may vary. More realistically, Lloyd-Smith et al. (2005) defines an individual’s reproduction number, $V$, as the number of new infections produced by a particular individual. In a population, $V$ is distributed around the population mean, $R_0$, and super-spreading individuals can be considered individuals that fall into the extreme right tail of this distribution (Lloyd-Smith et al. 2005). The implications of heterogeneity in $V$ is that pathogens are less likely to invade a population, but outbreaks have the potential to be explosive if key individuals become infected (Lloyd-Smith et al. 2005). Control measures targeted at this highly infectious subset can lead to more effective and efficient control strategies than population-wide measures (Woolhouse et al. 1997; Lloyd-Smith et al. 2005).

Although control strategies targeted at super-spreaders are potentially a very effective tool for managing disease (Woolhouse et al. 1997; Craft and Caillaud 2011), such strategies are not currently feasible for wildlife because logistical and diagnostic limitations make it difficult to identify super-spreaders. These limitations necessitate the
development of new tools (Cross et al. 2009), and the best method for identifying super-
spreaders will depend on the mechanisms that give rise to heterogeneity. In this review, I
focus on factors that influence the number of new infections produced by an individual,
though I acknowledge that heterogeneities in the likelihood of acquiring an infection
(exposure and susceptibility) also have the potential to impact disease dynamics.

To understand the role that heterogeneity plays in transmission dynamics, I
introduce the SIR compartmental modeling framework in which every individual in the
population is classified as Susceptible, Infected, or Recovered (Anderson and May 1992).
The transmission rate ($\beta$) is defined as the rate at which animals move from the
susceptible to the infected class. The recovery rate ($\gamma$) is the rate at which infected
animals enter the recovery class, and the disease-related mortality rate ($\alpha$) is the rate at
which infected animals die (Figure 1). Individual heterogeneity can play a role in each of
these processes. The infectious period, for example, is determined by the combination of
the recovery and mortality rate, and these factors may vary across individuals. An
individual that has poor nutrition may recover from an infection more slowly than an
individual in good health, which increases the length of time that the animal is capable of
infecting others. Similarly, if the animal is in sufficiently poor condition, it may die much
sooner than a healthier animal, thus decreasing the period of time that it is capable of
infecting others.
Figure 1. Diagram of compartmental SIR model where every individual in the population is classified as Susceptible, Infected, or Recovered, and movement between classes is determined by the transmission rate ($\beta$), recovery rate ($\gamma$), and mortality rate ($\alpha$).

The transmission rate can be decomposed into two primary components that reflect the rate at which an infected animal transmits to susceptible animals. To transmit an infection, an individual must first contact a susceptible individual (behavioral component of transmission, $\beta_c$). Once an encounter between an infected and susceptible individual has occurred, the probability that the susceptible individual becomes infected depends on the efficiency of transmission, or transmissibility (Bansal et al. 2007; Craft et al. 2010). This physiological component of transmission ($\beta_p$) relies on both the infectiousness of the infected host and the susceptibility of the susceptible host (Hawley et al. 2011). Sometimes there may be covariation between $\beta_p$ and $\beta_c$, but that is not considered here because it is discussed at length elsewhere (Hawley et al. 2011).

Within the SIR framework, there are two primary pathways that give rise to variation in $V$. First, an animal may be a super-spreader because it has greater overall number of encounters during which transmission can occur. The total number of encounters will be determined by the rate at which contacts occur ($\beta_c$) and the total length
of time in which the animal is capable of infecting others (infections period, determined by $\alpha$ and $\gamma$). Thus, both behavioral and physiological components determine total encounters. Second, an animal may be a super-spreader because it is highly infectious. Given an equal number of encounters, a more infectious animal is more likely transmit a pathogen. For example, an animal that sheds above average numbers of infective agents (viruses, eggs, spores, etc) into its surroundings are more likely to infect susceptible animals. These so-called “super-shedders” can be considered a special case of super-spreaders. In summary, the number of new infections produced by an individual, $V$, can be defined as

$$V = \text{contact rate} \times \text{infectious period} \times \text{infectiousness}$$

Here, I first consider mechanisms that produce heterogeneity in transmission, then highlight scales (individual, types of individuals, species, etc) at which heterogeneity can be assessed, and conclude by discussing methods by which heterogeneity can be quantified.

**Mechanisms promoting heterogeneity**

Individual-level heterogeneity can arise from multiple sources. Here, I primarily sub-divide these processes into behavioral and physiological mechanisms. This dichotomy is somewhat arbitrary and by no means mutually exclusive. Behavioral differences among individuals, for example, can result from differences in physiology. Nonetheless, it is a useful dichotomy because each mechanism tends to affect distinct
parameters in the SIR model. Behavioral differences will most likely affect $V$ through variation in contact rates ($\beta_c$), while physiological differences will influence infectiousness ($\beta_p$) and the duration of the infection ($\alpha$ and $\gamma$). At the end of this section, I discuss interactions between behavioral and physiological mechanisms, and how certain factors, such as reproductive state, simultaneously affect both behavioral and physiological outcomes.

**Behavioral mechanisms**

The primary means by which behavior affects $V$ is through contact rate with susceptible individuals. Individual variation in contact patterns is influenced by many factors including age, reproductive state, social position, and personality. Personality, also known as behavioral syndromes or temperament, can be described as behavioral consistency across contexts and time (Sih et al. 2010). One common axis along which personality varies is shy to bold (Sih et al. 2010). Bold individuals are less averse to risk and may engage in more exploratory behavior compared to shyer individuals. These types of behaviors alter an individual’s contact rate with other individuals and alter where the animal sheds infectious agents into the environment. Social/asocial, another axis of personality, modifies host contact rates whereby highly social individuals engage in more interactions, creating more opportunities for transmission to occur. Behavioral patterns are important even for environmentally transmitted pathogens if patterns of space use and habitat choice affect the number of individuals that they contact in space. For example, boldness was demonstrated to influence tick levels in chipmunks (*Tamias sibiricus*) by increasing space use and exposure to ticks (Boyer et al. 2010). Sunfish (*Lipomis*)
*gibbosus*) with bolder personalities carry a different set of parasites than shyer sunfish, which reflects the fact that they tend to use different habitats (Coleman and Wilson 1998).

Inter-individual and sex differences in mating behavior can also affect heterogeneity in pathogen transmission. Sexually transmitted diseases are probably the most apparent example of how mating patterns influences the ability of pathogens to spread. Males in particular can show wide variation in number of sexual contacts (Trivers 1972). In humans, for example, a single promiscuous man, who is considered patient zero for the emergence of HIV in North America, transmitted HIV to at least 248 other men (Hudson et al. 2008). Even for pathogens with other transmission modes, aggregations of individuals for breeding purposes vastly changes in contact rates (Altizer et al. 2006). Males also vary in their ability to control both space and females, and that can affect pathogen transmission. Male Grant’s gazelle (*Gazella granti*), for example, vary in their resource-holding potential, and territorial males tend to exhibit higher parasite burdens than less sedentary bachelors (Ezenwa 2004a).

Individuals may also be disproportionately important in the spread of pathogens if they occupy key positions in the network that, if disrupted, hinders dissemination of pathogens across different subgroups. The important point here is that these individuals are not necessarily important because they have high $V$, but because they can potentially serve as bottlenecks for pathogen flow. Targeting such individuals may fracture the connectivity of the population. I draw from the network literature here and term these individuals as potential “cutpoints” because their removal may fragment the population in such a way as to prevent the flow of pathogens through the population. Targeted
interventions at cutpoints has the potential to increase the overall level of clustering in the population, which has been shown in numerous theoretical papers to hinder the spread of disease (Keeling 1999; Newman 2003; Turner et al. 2008; Wu and Liu 2008; Badham and Stocker 2010; Ames et al. 2011).

Individuals may serve as cutpoints for a variety of reasons related to individual space use and social interactions. There is some empirical evidence demonstrating that dispersing individuals, or those individuals which are not sedentary within a home range or within a social group, experience greater infection risk (Drewe 2009; Fenner et al. 2011). Compared to uninfected individuals, pygmy bluetongue lizards (Tiliqua adelaidensis) that were infected with nematodes had more contacts with dispersing lizards than with resident lizards, than uninfected individuals, implying that dispersing individuals might be key to pathogen spread (Fenner et al. 2011). In addition, ground squirrel colonies (Spermophilus beldingi) that had fewer widely ranging juvenile males were characterized by more clustered spatial networks, a feature that was correlated with lower prevalence of Cryptosporidium within the colony (VanderWaal et al. Submitted). Nomadic lions (Panthera leo) also increase population-level connectivity, but in this case, models show that their epidemiological impact is marginal (Craft et al. 2010).

Perhaps the best example highlighting the potential importance of cutpoints in disease control strategies comes from Ethiopian wolves (Canis simensis), a highly endangered species threatened by rabies. Vaccination strategies that targeted wolf packs occupying key spatial bottlenecks within their geographic range had the potential to prevent the transmission of rabies between metapopulations (Haydon et al. 2006). Although the importance of cutpoints cannot be quantified through examining variation in \( V \), the
concept of cutpoints is intimately related to heterogeneity in transmission. Because they have the potential to be disproportionately important in the spread of pathogens, cutpoints can be considered an important type of super-spreader.

*Physiological mechanisms*

Although there are numerous heterogeneities that arise through variation in susceptibility, I continue to focus on variation in the ability of a host to transmit infections. Susceptibility and infectiousness often covary (Hudson et al. 2008; Paull et al. 2012), and physiological risk factors for acquiring an infection are generally easier to measure than $V$. Physiological mechanisms can lead to variation in $V$ either through enhancing variation in individual infectiousness (i.e., some individuals excrete more pathogens), or through enhancing variation in the infectious period (i.e., some individuals shed infective stages of a pathogen for longer). The former leads to increased transmissibility of an infection while the latter increases the number of contacts that an infected individual may have with susceptible individuals. Both of these have important implications for disease spread.

Highly infectiousness individuals are sometimes called super-shedders. Super-shedding has sometimes been considered as distinct from super-spreading in that super-spreading refers specifically to behavioral mechanisms (Chase-Topping et al. 2008). However, I emphasize that super-spreading should be defined as any situation where a minority of individuals are responsible for a disproportionate amount of transmission, and that super-shedding can be considered a special case of super-spreading because super-
shedders are highly infectious and are thus disproportionately involved in disease transmission.

Perhaps the simplest way in which physiology contributes to heterogeneity is through body condition (Beldomenico and Begon 2010). Body condition can be affected by a number of factors that also may affect the immune system directly, but in the simplest case, poor body condition caused by inadequate nutrition may increase infectiousness or the infectious period because there is less opposition to the pathogen’s survival and proliferation within the host. Poor nutrition and stress can both contribute to poor body condition. Chronic stress in particular has direct effects on the immune system by impairing both cell-mediated and humoral immune responses and can impact an infection’s duration (Glaser and Kiecolt-Glaser 2005). Indirectly, stress can also increase an animal’s infectiousness because cortisol and epinephrine, two hormones associated with stress, may increase mucous secretion and vasodilation or increase symptoms such as sneezing (Cohen et al. 1997). Therefore, variation across individuals in their exposure and susceptibility to stressors will promote heterogeneity in transmission dynamics.

Coinfection by multiple pathogens concurrently can also play a significant role in generating heterogeneity. Several studies show that infection by one pathogen modifies the likelihood of secondary infections (Kamal and El Sayed Khalifa 2006; Pederson and Fenton 2007; Telfer et al. 2010), though relatively fewer examine how coinfection influences the ability of the host to pass on the infection (Ezenwa and Jolles 2011). Helminth parasites, for example, have strong effects on immune function, and consequently, host responses to secondary infection by microparasites (viruses and some bacteria). Helminth coinfections not only suppress immune function in general, but also
can make the host more susceptible to microparasites by indirectly inhibiting the humoral immune response, which is crucial for combating microparasites. Coinfection with helminths may create heterogeneity among African buffalo (*Syncerus caffer*) in their ability to transmit bovine tuberculosis by influencing either the transmission rate, the infectious period, or both (Ezenwa et al. 2010; Ezenwa and Jolles 2011). Similarly, rabbits (*Sylvilagus brasiliensis*) infected with myxomavirus are more susceptible to helminths and consequently release more infective stages of helminth parasites into their surroundings (Hudson et al. 2008).

In addition to individual heterogeneity in infectiousness, $V$ is also determined by the host’s efficiency at spreading the pathogen. For example, experimental work has shown that there is substantial variation among humans in the amount of exhaled bio-aerosols generated during normal breathing, which has implications for aerosol-transmitted pathogens (Edwards et al. 2004). In addition, coinfection with other pathogens may lead to increased coughing, thus increasing the likelihood of transmission for respiratory bacteria. When humans who are nasal carriers for *Staphylococcus aureus* are coinfectected with mild respiratory viruses, such as adenovirus or rhinovirus, they convert from being *S. aureus* carriers to “cloud” patients, and are highly infectious because they are surrounded by clouds of aerosolized bacteria (Bassetti et al. 2005b). This mechanism has been suggested to be responsible for some super-spreading events in SARS outbreaks (Bassetti et al. 2005a). For vector-borne diseases, the attractiveness of the host to the vector has potential to introduce heterogeneity. Highly attractive individuals not only are more likely to acquire the infection through being bitten by an infected vector, they are also more likely to pass it on to subsequent vectors with whom
they are in contact. In humans, variation in mosquito attractiveness is related to skin surface area and age (Burkot 1988). Hosts already infected by malaria also become more attractive to vectors, which has important implications for transmission (Lacroix et al. 2005; Cornet et al. 2012).

**Interactions between behavioral and physiological mechanisms**

There are significant interactions between behavioral and physiological mechanisms; physiological mechanisms can mediate behavior, and behavior can affect an animal’s physiological state. Personality, for example, not only affects exposure and contact rates, but also can lead to physiological differences. Decades of study of human personality have consistently demonstrated that sociable humans experience less sickness (Cohen et al. 1997; Friedman 2008). Sociability was also negatively correlated with Simian Immunodeficiency Virus (SIV) viral load in experimentally-inoculated rhesus macaques (*Macaca mulatta*) (Capitanio et al. 1999). Furthermore, personality can interact with social context in that asocial individuals in unstable social contexts demonstrated increased basal cortisol and viral load compared to social individuals or asocial individuals in stable contexts (Capitanio et al. 2008). It is important to note that personality may not cause for different stress responses to the same situation. Rather, the same underlying physiological mechanism may lead both to behavioral differences and immunological differences. For example, mice with proactive versus reactive personalities exhibited different neuroendocrine profiles (hypothalamic-pituitary-adrenal axis reactivity to stress) that not only influenced behavior but also affected disease susceptibility and stress pathology (Koolhaas et al. 1999).
Within social groups, individuals vary in their contact patterns not only based on
group membership, but also based on their dominance within the group. Dominance may
be correlated with personality factors, stress levels, testosterone, and contact patterns, all
of which may affect their ability to transmit infections. In chimpanzees (*Pan troglodytes*),
gastrointestinal helminth richness was negatively correlated with dominance rank
(Muehlenbein and Watts 2010). In contrast, in dairy goats (*Capra aegagrus*), higher-
ranking females had fewer gastrointestinal parasite eggs in their feces than lower-ranking
individuals (Ungerfeld and Correa 2008), and in male alpine ibex (*Capra ibex*),
dominance was unrelated with fecal parasite egg counts (Decristophoris et al. 2007). This
suggests that the relationship between social status and infection risk is complex, likely
because physiological responses to social interactions (e.g., changes in cortisol or
testosterone) are situation-dependent and not always expressed in the same direction.

Reproductive state is perhaps the best example of when one factor has large
impacts on both behavior and physiology. Gonadal hormones associated with
reproduction have been demonstrated to have immuno-modulatory effects. Elevated
levels of progesterone, especially in pregnant females, inhibit humoral immunity, while
estrogens stimulate humoral immunity (Tait et al. 2008). In general, though, females are
considered to have more robust immune responses. Males, in contrast, exhibit higher
rates of parasitism, which may be the result from a combination of sex-specific stressors,
behaviors, and physiology (Zuk and McKean 1996). Elevated testosterone is generally
associated with immunosuppression (Zuk and McKean 1996). In chimpanzees, parasite
richness was positively correlated with both testosterone and cortisol (Muehlenbein
2006). In addition to physiological and immunological effects, variation in testosterone
levels may also alter contact rates, activity patterns, space use and home range size, which may enhance or reduce opportunities for transmission (Chandler et al. 1994; Denardo and Sinervo 1994; Grear et al. 2009). In Grant’s gazelles, a male’s ability to defend a territory and thus gain access to females is correlated with testosterone, and territorial males tend to exhibit higher parasite burdens than bachelor males (Ezenwa 2004a; Ezenwa et al. 2012). In addition, mice (*Peromyscus leucopus*) from sites where male testosterone levels were elevated maintained more contacts than mice from control sites (Grear et al. 2009).

**Scales of heterogeneity**

Heterogeneity in transmission potential can occur at several scales. The most basic scale is variation in $V$ among individuals. However, it may be that certain classes or types of individuals generally have similar $V$ and that transmission potential can be approximated simply by knowing a few key attributes about the individual. For pathogens that infect multiple hosts within an ecosystem, it is also possible that certain species have far higher transmission potential than others.

**Individuals**

Individual-level heterogeneity is produced by mechanisms where specific information about the transmission potential of each individual must be known. There is no clear pattern predicting $V$ that can be accounted for by basic attributes (e.g., age and sex). Behavioral mechanisms often produce individual-level heterogeneity. Typhoid
Mary, for example, was thought to be a super-spread because of her unhygienic habits and tendency to undercook food (Hudson et al. 2008). A more general example of this is variation in $V$ that arises from individual personality, which cannot be predicted by age, sex, reproductive status, or any other easily assessable attribute. Thus, any targeted control strategy aimed at “high-risk” personalities, such as sociable, must assess each individual in the population to determine who are potential super-spreaders. Some physiological mechanisms, such as stress and coinfection, can also produce individual-level heterogeneity because these are properties that must be measured in each animal individually.

**Classes of individuals**

Although interventions targeted at super-spread individuals have the potential to limit disease spread much more effectively than non-targeted control measures (Lloyd-Smith et al. 2005), identifying *specific individuals* that are super-spreaders is time-consuming and population-specific because each individual’s location within its social network must be quantified. Even if certain behaviors were identified as high-risk (Drewe 2009), the application of strategies targeting animals that frequently display high-risk behavior would require behavioral monitoring of the population by a trained observer. It is more efficient to identify *classes of individuals* that are super-spreaders. Control strategies that target classes of individuals, or perhaps species within communities, could then be generalized across populations because the typical physiology or behaviors exhibited by that class is likely similar across populations. Adult males, for example, have often been shown to play a crucial role in the maintenance of infectious disease at
the population level (Ferrari et al. 2003; Grear et al. 2009). Transmission rates can also exhibit strong age dependence (Grenfell and Anderson 1985), which can arise both from age-related differences in behavior and contact rates and from age-related differences in physiology and immunological responses. Aggregation of children in schools enhances transmission of disease, and control measures targeted at school children, such as vaccination or school closure, may be very effective at reducing population-wide epidemics (Heymann et al. 2004; Weycker et al. 2005).

Some high-risk behaviors may be more prevalent in certain age-classes. In meerkats (Suricata suricatta), overall level of social activity was not the best predictor of Mycobacterium bovis infection risk. Rather, certain types of interactions were correlated with risk, and individuals varied in how often they engaged in these interactions based on their social status. Subordinate individuals, for example, more frequently groomed others, which is a behavior associated with increased risk of acquiring a pathogen, while dominant animals more frequently were recipients of grooming (Drewe 2009). Targeted vaccination of younger male meerkats, who are often subordinate and frequently engage in behaviors with high risk for transmission, could disrupt chains of transmission (Drewe et al. 2011). In Belding’s ground squirrels, juvenile males tend to be more exploratory than other age-classes of individuals, and removing these individuals from spatial networks leads to higher levels of clustering within the colony, a feature that is associated with lower parasite prevalence (VanderWaal et al. Submitted).

Beyond age and sex, body size is an easy attribute by which to categorize individuals. In some cases, body size is correlated with transmission potential. The transmission of tick-borne encephalitis is contingent on ticks co-feeding on the same
host. In yellow-necked mice (*Apodemus flavicollis*), host body size increases the frequency at which co-feeding occurs. Over 90% of transmission potential was attributed to the top 20% of mice with the highest number of co-feeding ticks. These mice tended to be large-bodied, sexually active males, and control strategies aimed at this class of individuals have the potential to effectively eliminate tick-borne encephalitis at fraction of the cost of widespread treatment (Perkins et al. 2003).

*Species within communities*

Because of interspecific variation in physiology and behavior, certain species may function as super-spreaders for multi-host pathogens if individuals of that species tend to have higher $V$ on average. Species whose contribution to community-wide transmission is disproportionate to their relative abundance can be considered “super-spreader species” (Paull et al. 2012). American robins (*Turdus migratorius*), for example, comprised over 40% of mosquito blood meals as compared to other songbird species despite the fact that they accounted for <4% of the songbird population. This makes American robins disproportionately important in the spread of West Nile Virus, a mosquito-borne pathogen (Kilpatrick et al. 2006). In this case, vector preference for certain host species is likely responsible for interspecific differences in transmission potential, but such differences can also arise for behavioral and physiological reasons as well.

Social behavior is a key factor to the spread of disease. Highly social species experience higher contact rates and more transmission opportunities, creating situations where a pathogen can rapidly proliferate within the species (Coté and Poulin 1995; Altizer et al. 2003). However, territoriality or infrequent contact between social groups
may create conditions in which population-wide pathogen dissemination is hindered, even if within-group transmission rates are high (Loehle 1995; Altizer et al. 2003; Cross et al. 2005). Ultimately, the degree to which social behavior facilitates versus hinders population-wide transmission depends on the insularity of social groups, and the frequency of contact between groups relative to the infectious period (Cross et al. 2005; Craft et al. 2010). When social behavior does facilitate population-wide transmission and thus elevates overall prevalence of the pathogen in the ecosystem, this may increase disease risk for other susceptible species (Paull et al. 2012).

While sociality is the most apparent behavior affecting contact rates, other behaviors may be important especially in the case of environmentally transmitted pathogens. For example, species that tend to defecate in middens are potentially creating highly localized hotspots of fecal-oral pathogens in the environment, which other species may avoid (Altizer et al. 2003; Ezenwa 2004b). This may reduce the transmission potential of these species. More generally, space use patterns may affect the total number of inter- and intra-specific contacts an animal may make, and species with large home ranges or that utilize a greater diversity of habitats may have the potential to disperse pathogens more widely (Chapter 4).

Clearly, interspecific variability in transmission is enormously influenced by physiological differences in host competence, which can be defined as the efficiency with which a host acquires and transmits a pathogen (Paull et al. 2012). Variation in host competence has important implications for community-level infection patterns. For example, white-footed mice (*Peromyscus leucopus*) are the most competent host for Lyme disease. The presence of less competent host species within a community reduces
the prevalence of Lyme disease because these less competent hosts feed, but rarely infect, ticks. Fragmented habitats tend to have reduced host diversity, yet they have high abundances of mice. Without other hosts to feed on, a greater proportion of ticks will feed on the most competent host (mice) and become infected, which increases the prevalence of infected ticks and also the risk of infection to humans. This is an example of how variation in host competence leads to the “dilution effect,” whereby the presence of less competent hosts “dilutes” the transmission potential of the most competent host (LoGiudice et al. 2003).

Quantifying heterogeneity

Variation in $V$ is notoriously hard to quantify given that it is nearly impossible to determine who transmits to whom. Many of the studies discussed in this review have focused on factors that influence the risk of acquiring an infection rather than the number of individuals that were infected by the host. While infection risk may or may not be correlated with $V$, it still gives some insight into factors producing heterogeneity in transmission dynamics. In addition, if the risk of acquiring an infection is grounded in immunological mechanisms, the same mechanism may also translate into higher infectiousness or slower recovery (e.g., longer infectious period), although these factors are less often measured. Some studies do quantify variation in pathogen shedding, such as viral load or fecal parasite concentrations, and this may serve as a useful proxy for $V$ if shedding rate is directly correlated with $V$. In many cases, this assumption may not be valid because susceptible contacts may actively avoid highly infected animals (Hawley et al. 2011). Quantifying contact rates may also be useful but is similarly plagued by the
possibility that contact rates change once an animal becomes infected. Because contact rate, infectiousness and infectious period all contribute to $V$, quantifying only one of these dimensions may not yield a full picture. Nonetheless, they do serve as useful starting points in quantifying heterogeneity in transmission.

*Experimental approaches*

Much of the work on how physiological mechanisms influence infectiousness or the infectious period has been conducted in experimental set-ups. It is possible to quantify variation in infectiousness in the field by measuring some aspect of shedding (Atwill et al. 2001), but regular and repeated sampling of the same individuals is necessary to quantify the duration of the infection. In addition, experimental manipulation is generally required to link observed variation with physiological mechanisms. For example, the effect of diet on shedding of pathogenic *Escherichia coli* was investigated through experimentally inoculating sheep assigned to different diet treatments (Kudva et al. 1997). Although correlative studies are possible in the natural settings, such as relating hormonal levels to parasite richness in chimpanzees (Muehlenbein 2006), causation is more easily determined with controlled experiments.

Experimental approaches are most commonly used to quantify physiological mechanisms, as natural behavior (e.g., contact rates) is difficult to reliably reproduce in a laboratory setting. However, there are some notable exceptions. Capitanio et al. (1999; 2008), for example, studied the implications of personality and social context on viral load in a controlled experiment with rhesus macaques. However, personality by definition refers to behavioral consistencies across contexts, which may make personality
a more tractable topic in a laboratory as compared to attempting to manipulate contact rates.

**Social networks**

Social network theory has been widely used to quantify variation in contact patterns. In a social network, individuals are interlinked according to who has been in contact with whom, and contact is usually quantified through direct observations of behavior or through measures of shared space use (Croft et al. 2008). Within a network framework, each individual’s level of connectivity can be evaluated using established network metrics. Average connectivity can be calculated for individuals in the same age-class or species, providing a straightforward way to quantify class- or species-level heterogeneity. One commonly used summary metric is “degree,” which is the total number of individuals the focal animal is linked to in the network (Wasserman and Faust 1994; Wey et al. 2008). The “degree distribution” essentially quantifies heterogeneity in connectivity among individuals in the network. Quantifying the degree distribution has received enormous attention in the literature because it essentially serves as a proxy for variation in $V$ if it is assumed that the number of animals that an individual is able to infect is directly proportional to the number of animals with which it in contact. However, caution should be exercised if using degree as a proxy for $V$, and conversely in using degree as a proxy for an individual’s infection risk. There may be positive or negative correlations between physiological and behavioral components of transmission, which means that contact rates by themselves only illuminate one dimension of $V$ (Hawley et al. 2011).
The degree distribution of a social network can be generalized to physiological mechanisms if we consider a “transmission network,” where individuals are interlinked based on who transmitted to whom rather than social interactions. An individual’s degree in the transmission network is by definition the total number of infections produced by that individual. When the network is defined in this way, the degree distribution becomes the distribution of $V$ with mean $R_0$. In these transmission networks, a super-shedder will have high centrality and a large number of connections. Network theory, therefore, is a useful approach for conceptualizing physiological heterogeneity in transmission as long as who transmitted to whom can be reliably determined.

*Combining genetics with network theory: A new tool for quantifying heterogeneity*

Although network analysis has been lauded as the next key tool for examining how heterogeneous transmission patterns impact disease spread (Wey et al. 2008; Sih et al. 2009), few studies have related empirically derived wildlife networks to data on pathogen presence (Corner et al. 2003; 2007; Drewe 2009; Godfrey et al. 2009). In many cases, “transmission networks” were constructed based on social interactions between individuals without assessing any pathogens. Because of limitations in detecting the occurrence of transmission, conclusions about disease spread have been based on the possibility that transmission could occur between individuals (Perkins et al. 2008; Böhm et al. 2009; Craft et al. 2009; Grear et al. 2009; Hamede et al. 2009; Perkins et al. 2009).

As discussed in this review, quantifying contact rates only accounts for one aspect of $V$. Data on who transmitted to whom is required to construct a true transmission network. One possible way to obtain such data is using the genetics of the pathogen itself. If two
individuals share genetically identical subtypes of a pathogen, then transmission can be inferred (Archie et al. 2008). These data can then be used to construct a transmission network based on quantifiable transmission events.

The advantage of this method is that it allows the transmission network to be defined independently of the social network, allowing for the question on how heterogeneity in contact patterns affects heterogeneity in transmission patterns to be explicitly addressed. In this dissertation, I explore integrating microbial genetics with network theory to investigate a number of topics discussed in this review. In Chapter 2, I quantify social and spatial contact networks for a population of wild giraffe (*Giraffa camelopardalis*). In Chapter 3, I relate these contact networks to a transmission network based on patterns of *E. coli* genetic subtype sharing, explicitly addressing behavioral mechanisms that produce heterogeneity in transmission. In Chapter 4, I broaden this novel method to address larger species-level scales of heterogeneity. After constructing a multi-species transmission network for ten wild and domestic ungulate species, I examine how species vary in their connectivity in the transmission network and whether any species may function as super-spreaders. My intent in this dissertation is to demonstrate the potential utility of an approach that combines microbial genetics with network theory for quantifying heterogeneity in transmission, thus enhancing our understanding of how infectious diseases spread through populations.
Literature cited


Bassetti S, Bischoff WE, Walter M, Bassetti-Wyss BA, Mason L, Reoussin BA, D'Agostino RBJ, Gwaltney JMJ, Pfaller MA, Sherertz RJ, 2005b. Dispersal of Staphylococcus aureus into the air associated with a rhinovirus infection. Infection Control and Hospital Epidemiology 25:196-203.


Hawley DM, Etienne RS, Ezenwa VO, Jolles AE, 2011. Does animal behavior underlie covariation between hosts' exposure to infectious agents and susceptibility to
infection? Implications for disease dynamics. Integrative and Comparative Biology 5:528-539.


behaviour: a promising tool for the study of sociality. Animal Behaviour 75:333-
344.

Population-wide benefits of routine vaccination of children against influenza.
Vaccine 23:1284-1293.

Heterogeneities in the transmission of infectious agents: implications for the
design of control programs. Proceedings of the National Academy of Sciences
94:338-342.


International Journal for Parasitology 26:1009-1023.
Chapter 2: Multi-level social organization and space use in reticulated giraffe

(*Giraffa camelopardalis*)

Abstract

It is increasingly recognized that association patterns of most gregarious animals are non-random. However, non-random patterns can emerge in any population that exhibits spatial structure, even if individuals associate randomly. In species that lack clearly differentiated social relationships characteristic of socially complex mammals, space use patterns must be considered alongside association patterns in order to establish whether non-random association patterns are determined by underlying social structure or are merely an artifact of spatial structure. Here, I simultaneously consider space use and association patterns for a wild population of reticulated giraffe. I examined whether the giraffe’s flexible fission-fusion association patterns were embedded in higher levels of social organization. I identified multi-level social organization in which individuals were members of social cliques. Cliques were embedded in larger sub-communities, which in turn were embedded in communities. The frequency with which two individuals were observed together was positively correlated with the extent to which their home range overlapped, implying an underlying role of shared space use in determining association patterns. However, membership in cliques and sub-communities was relatively unrelated to space use patterns for males. For females, space use played a much larger role in
determining multi-tiered social organization, which is consistent with a matrilineal-based society characterized by female philopatry. While giraffe social interactions are highly fluid in nature, it is apparent that association patterns in giraffe are not the result of random fission-fusion events, but are embedded within a structured social network characterized by multiple levels of organization.

Introduction

The ecological basis of social organization has had a long history of investigation in behavioral ecology, and it is now understood that ecological factors, such as predation and the distribution of resource, play a crucial role in shaping social structure (Alexander 1974; Rubenstein and Wrangham 1986; Isbell and Young 2002). Animal social structures that are characterized by fission-fusion dynamics exhibit frequent coalescing and dividing of group members into smaller sub-groups (Langman 1977; Leuthold 1979; Couzin 2006; Aureli et al. 2008). Fission-fusion dynamics are thought to allow animals to respond to changing environmental conditions and flexibly balance conflicting demands. They allow animals to form larger groups when there are reproductive, foraging, or anti-predator benefits, but to minimize costs of intra-group competition if the benefits of grouping change. Flexible grouping dynamics are exhibited in a broad range of taxa, including shoaling fish (Hoare et al. 2004; Kelley et al. 2011), bats (Popa-Lisseanu et al. 2008; Kerth et al. 2011), primates (Kummer 1968; Symington 1990), carnivores (Schaller 1972; Wolf et al. 2007; Smith et al. 2008), ungulates (Aycrigg and Porter 1997; Cross et al.
2005; White et al. 2010), elephants (Wittemyer et al. 2005; Archie et al. 2006b), and marine mammals (Lusseau 2003; Pearson 2009).

Understanding how fission-fusion dynamics influence population structure has important implications for disease transmission (Keeling 1999; Craft et al. 2010; Griffin and Nunn 2012), information flow (McComb et al. 2001; Vital and Martins 2009), mating opportunities (Hashimoto et al. 2001), and gene flow (Altmann et al. 1996). Although fission-fusion societies are sometimes thought to occur only in species with higher cognitive abilities due to the need to maintain social bonds (Aureli et al. 2008), it is increasingly recognized that fission-fusion dynamics can emerge through self-sorting of individuals with similar needs and motivations (Couzin 2006; Ramos-Fernández et al. 2006). Fission-fusion dynamics can result from simple foraging models with no description of how actors should behave socially (Ramos-Fernández et al. 2006). Thus, in species with less obvious social relationships than species such as chimpanzees (*Pan troglodytes*: Goodall 1986) or elephants (*Loxodonta africana*: Moss 1988; Archie et al. 2006a), it can be difficult to determine whether observed fission-fusion dynamics are socially mediated or merely an artifact of foraging patterns or space use.

plains zebra, *Equus burchelli*: Rubenstein and Hack 2004). For example, zebra social
groups temporarily and non-randomly fuse with other groups to form larger herds
(Rubenstein and Hack 2004), and white-tailed deer fission-fusion dynamics are rooted
within a genetically-related community of females (Miller et al. 2010). As in deer, social
structure in some ungulate species may be based on female locational philopatry, which
may lead to bonds based on matrilineal relationships (Greenwood 1980; Murray 1982). In
other ungulate species, there is little evidence of preferential bonds or higher levels of
structure (Lott and Minta 1983; Schulte and Klingel 1991; Le Pendu et al. 1995), though
it is possible that the higher levels of social organization are undetected in less studied
species.

As yet, there is a little agreement in the literature as to the existence and extent of
social structure in giraffe (*Giraffa camelopardalis*). Some studies concluded that giraffes
lack social organization and that association patterns are random (Foster and Dagg 1972;
Dagg and Foster 1976; Leuthold 1979; Le Pendu et al. 2000). This conclusion stemmed
primarily from perpetually shifting grouping patterns observed among giraffe. Groups
daily or even hourly coalesce into larger groups or break apart into smaller groups. An
individual’s associates may shift numerous times in the course of a day (Leuthold 1979;
Pratt and Anderson 1982; Pellew 1984; Pratt and Anderson 1985; Bercovitch and Berry
2009a). Dagg and Foster (1976) and le Pendu and Coifiolo (2000) stated that individual
interactions were ephemeral and bonds non-existent.

Recent work suggests that giraffe populations have more complex structure than
previously thought (Pratt and Anderson 1982; Fennessy 2004; Bashaw et al. 2007;
Shorrock and Croft 2009; Bercovitch and Berry 2012; Carter et al. 2012). Association
patterns appear to be non-random and female giraffe exercise social preferences, which are partly determined by shared space use and genetic relatedness (Carter et al. 2009; Bercovitch and Berry 2012). Although it has been suggested that giraffe sociality may be characterized by fission-fusion grouping dynamics embedded within a larger social community (Pratt and Anderson 1982; Bashaw et al. 2007; Bercovitch and Berry 2009a, 2012; Carter et al. 2012), it is still unclear whether giraffe exhibit higher levels of social structure, as seen in some other fission-fusion species (Symington 1990; Wittemyer et al. 2005; Wolf et al. 2007; Fortuna et al. 2009; Mourier et al. 2012). While much of the previous work has focused on quantifying variation in pairwise association indices (Leuthold 1979; Pratt and Anderson 1985; Le Pendu et al. 2000; Fennessy 2004; Carter et al. 2012), I apply social network analysis to uncover multi-level social organization.

Social network analysis provides a more sophisticated technique for analyzing association patterns because it not only takes into account direct (dyadic) interactions, but also indirect connections between individuals (Wey et al. 2008; Sih et al. 2009; Makagon et al. 2012). Even though dyadic association indices among giraffe are low (Leuthold 1979; Le Pendu et al. 2000), network analysis may reveal social organization if it allows for the detection of clusters of individuals within the network that interact more frequently with one another but more rarely with others.

In this study, I analyze a giraffe social network consisting of over 1,000 observations of giraffe groups and over 200 known individuals to gain an unprecedented level of detail into the social structure of the giraffe, with a specific focus on identifying multiple levels of social organization. I then compare the social structure to the spatial distribution of the population to determine whether the observed social structure arises as
an artifact of individual space use or whether it is a social phenomenon in that aspects of social structure cannot be explained from space use alone.

Methods

Study site and population

This study was conducted in Ol Pejeta Conservancy (OPC), a 364 km² semi-arid wildlife reserve located on the equator (0° N, 36°56’ E) approximately 220 km north of Nairobi, Kenya. It is part of the larger Laikipia plateau (altitude 1800 m), which extends from Mount Kenya to the Aberdares Mountains. The reserve is a grassland-woodland mosaic, with the dominant woody species being *Acacia drepanolobium* and *Euclea divinorum*. OPC receives on average 900 mm of rainfall per year (Birkett 2002), with peak rainfall occurring in March – April and October – November. During the study period (2011), however, monthly rainfall during the dry season was above average and exhibited less seasonal variation than in normal years (Ol Pejeta Conservancy, unpublished data). Large mammals found on Ol Pejeta include lions (*Panthera leo*), spotted hyena (*Crocuta crocuta*), leopards (*Panthera pardus*), elephants (*Loxodonta africana*), cheetah (*Acinonyx jubatus*), buffalo (*Syncerus caffer*), black and white rhinoceros (*Diceros bicornis, Ceratotherium simum*), Grevy’s and plains zebra (*Equus grevyi and E. burchelli*), Thomson’s and Grant’s gazelle (*Gazella thomsonii* and *G. granti*), impala (*Aepyceros melampus*), and oryx (*Oryx gazella*).

Across their geographic range, there are several features that emerge as characteristic of giraffe sociality. Giraffes are usually found in groups of on average three to five individuals, with approximately 20-25% of groups being larger than six. Groups
larger than 30 are extremely rare but do occur. Adult males are commonly found alone, while adult females are almost always in groups (Dagg and Foster 1976; Leuthold 1979; Le Pendu et al. 2000; van der Jeugd and Prins 2000; Fennessy 2004; Bercovitch and Berry 2009a; Shorrocks and Croft 2009). Adult males adopt a roaming strategy searching for females in estrus (Dagg and Foster 1976; Pratt and Anderson 1985). Breeding tends to be relatively aseasonal, though birthing peaks sometimes occur (Fennessy 2004; Bercovitch and Berry 2009b). Females with calves often pool their young into crèches that may persist for several months (Langman 1977; Pratt and Anderson 1979).

All giraffe at OPC were recognized using individually unique spot patterns along their necks. At the time of this study, OPC had a population of 212 reticulated giraffe. I believe that this population estimate represents a complete census of the population because all giraffe were observed approximately once per week and only two new adults were discovered in the entire last five months of the study. Giraffes were aged according to height estimates and age-associated behaviors. This aging scale was adapted from the literature (Langman 1977; Pratt and Anderson 1979; Fennessy 2004). Neonates (<3 months old) still had attached umbilical cords, and the length of the neck was short relative to the height of the shoulder. Juveniles (3 months to 1.5 years) were larger than neonates, but still accompanied their mother. Activity budgets of juveniles begin to resemble an adult’s. Subadults (1.5 – 4 years) no longer consistently accompany their mothers, but were smaller than adults. Adults (>4 years) coincide approximately with the onset of sexual maturity and adult size. At the conclusion of the study period, OPC’s giraffe population consisted of 160 adults, 20 subadults, 21 juveniles, and 11 neonates. The number of giraffe classified as subadults was likely underestimated because of older
individuals being erroneously classified as adults. The population exhibited a 50:50 sex ratio.

Disappearances can usually be attributed to death rather than emigration because OPC is enclosed by a perimeter fence, eliminating immigration and emigration from the giraffe population except through a few narrow gaps in the fence. Of the six adults that disappeared between July 2010 and August 2011, three were observed in very poor condition prior to disappearing. The remaining three disappeared during the last few months of 2010 when the population was not being monitored.

Because of the large size of the reserve (364 km$^2$), I do not expect ranging patterns to be significantly influenced by the fact that this population was fenced. In Kenya, cows and bulls live in overlapping home ranges that vary from 13 to 162 km$^2$, and 16.5 to 164 km$^2$, respectively (Foster and Dagg 1972; Leuthold and Leuthold 1978). Home range sizes from the upper end of this range were from Tsavo National Park, which is significantly drier than OPC. In the other two locations for which average home range sizes have been estimated, mean home range size was under 85 km$^2$ (Foster and Dagg 1972; Leuthold and Leuthold 1978).

Field observations

From January 21 to August 2, 2011, giraffe group composition and membership were recorded for all giraffe groups sighted while driving pre-determined survey routes. Routes were determined so that a different part of the study area was surveyed each day, allowing for most of the study area to be surveyed once every three days. Routes were approximately 100 km in length, covered approximately 115 km$^2$ each, and traversed all
habitat types. Observed giraffe groups were followed off-road until a complete census of the individuals present was accomplished. A group was defined as a set of individuals engaged in the same behavior, or moving in the same direction or toward a common destination, as long as each giraffe was no more than 500 m from at least one other group member. All individuals observed within a group were considered to be in association with every other member of the group. As group membership is constantly shifting (Le Pendu et al. 2000; Fennessy 2004), independence of observations was ensured by using only the first observation of an individual’s group on a given day for social network analysis. Group sizes do not vary with time of day (Bercovitch and Berry 2009a). During the study period, K. L. V. collected a total of 1089 observations of giraffe groups. On average, 30.7 giraffe were observed per day, distributed between four to six groups. Each individual giraffe was observed on average 31.1 times (approximately once per week). Given that nearly all giraffe were sighted approximately once per week, any giraffe that had not been seen between July 1 and August 2 was assumed to have died or left the study area and was excluded from analysis. Giraffe that were seen fewer than five times were also excluded (n=1).

Network construction

A social network was constructed from observed association patterns. Instead of the absolute number of times the animals were seen together, I controlled for varying re-sighting frequencies among individuals by calculating association strength (AS) as follows:
\[ AS = \frac{Y_{ij}}{Y_{ij} + Y_i + Y_j}, \]

where \( Y_{ij} \) is the number of times that individuals \( i \) and \( j \) were observed in the same group. \( Y_i \) and \( Y_j \) represent the number of times each individual was observed in a group where the other was absent. An association matrix was first constructed with each element \( AS_{ij} \) representing the AS between the \( i \)th and \( j \)th individuals, and then used to construct a network of associations between individuals. Pairs with non-zero AS were linked in the network, with links being weighted according to the AS value.

We also constructed a “home range network” using the extent of home range overlap to connect individuals. Individual home ranges were mapped using the GPS locations recorded for each giraffe sighting. Home range boundaries were determined using a fixed-kernel utilization-distribution of sightings (Worton 1989; Harris et al. 1999). Due to potential sensitivity to sample size, I used a 75% probability contour (kernel density isopleth) to exclude outlying observations and produce a core home range. Home range overlap between two individuals was defined as the number of 1 km\(^2\) grid squares that fell within both individuals’ home ranges, divided by the total size of both individuals’ home ranges. These dyadic home range overlap values were used to connect individuals in the home range network.

**Network and statistical analysis**

We constructed three datasets to examine social organization in adults and subadults: female only (N=86), male only (N=84), and combined sex (N=170). All statistical analyses were performed separately on each of these three datasets. Neonates
and juveniles were excluded from all datasets used to examine social structure because animals of these age groups are nearly always found near or accompanying their mothers.

Data cloud geometry (DCG) methods (Fushing and McAssey 2010) were applied to the association matrix to identify social community structure at multiple scales. The DCG method utilizes the tendency of random walks to remain within clusters of highly connected nodes to quantify community structure in the network. Detailed methods are described elsewhere (Fushing et al. 2013). Essentially, this method employs regulated random walks with recurrence-time dynamics to detect information about the number of clusters and the corresponding cluster membership of each individual at multiple scales based on local information provided by the similarity measure AS; hierarchical levels \{T_1, T_2, \ldots, T_K\} correspond to phase transitions, which are then utilized to build the geometric hierarchy of the data cloud into a tree with \{T_1, T_2, \ldots, T_K\} hierarchical levels. The resulting hierarchical tree is termed a DCG tree (Fushing et al. 2013). I term the highest hierarchical level as a community (Level A). Each community consists of multiple sub-communities (Level B), and each sub-community consists of multiple cliques (Level C). Although cliques were identified using DCG algorithms, it is worth noting that the term “clique” also has a specific definition in network theory: a completely connected set of nodes (Wasserman and Faust 1994). In my analysis, only 3.5% of within-clique dyads were not connected (AS=0) in females. For males, only 2.4% of within-clique dyads were not connected. Thus, social cliques defined by DCG very nearly meet the classic definition.

There are several advantages of the DCG technique as compared to other commonly used community-finding algorithms. Unlike some algorithms, it does not
require ties to be binary and instead utilizes the weighted nature of the association index. As compared to traditional hierarchical clustering methods, DCG trees are more robust, less sensitive to measurement errors, and provide information on the intrinsic scales embedded within the data cloud (Fushing and McAssey 2010; Fushing et al. 2013). Fushing et al. (2013) showed that DCG algorithms more accurately accounted for spatial clustering in the population as compared to classic hierarchical clustering. Specifically, hierarchical clustering grouped a number of individuals dwelling on the west side of the river into the same spatial cluster as animals on the eastern side. In contrast, the DCG method accurately grouped these animals with other animals living on the same side of the river. In this study, DCG algorithms were applied separately to the male and female social networks, as well as to the combined sex dataset. The algorithm was also performed on the home range network to identify neighborhoods of individuals that clustered spatially. Clustering trees generated by DCG can be used to examine multi-level social organization.

Significance of network clustering configurations (DCG trees) was determined using Monte Carlo tests. Multiple levels of clustering are present in a DCG tree. Thus, to test the significance of the DCG tree involves testing the clustering structure at each hierarchical level $T_k$. I achieved this by sampling 1000 random clustering configurations at each level $T_k$ ($k = 1, \ldots, K$). In these random permutations, the number and size of clusters matched the observed DCG clustering configuration, but individual membership was randomly allocated. At each hierarchical level $T_k$, within-cluster tie strength ($WCTS_k$) was calculated for each cluster by taking the average AS among animals within the same cluster. The observed mean $WCTS_k$ of all clusters at level $T_k$ was compared to the
permuted distribution of mean $WCTS_K$. This distribution was generated by calculating mean $WCTS_K$ for the randomized clustering configurations. P-values were calculated as the percentage of permuted $WCTS_K$ that were more extreme than the observed $WCTS_K$. $WCTS_K$ was considered significant if it fell in the 95% percentile of the permuted distribution of $WCTS_K$ ($p < 0.05$). When $p < 0.05$, it indicates that at level $T_K$, the DCG approach produced clusters that were significantly denser than randomly generated clusters.

In addition, I examined the significance of each distinct cluster in each level by comparing $WCTS_{KL}$ with the 95% percentile of 1000 permuted $WCTS_{KL}$ values, where $WCTS_{KL}$ is the within-cluster tie strength of the $l$th cluster identified within level $T_K$. For example, are members of a particular cluster more densely connected than those in the random clusters generated in the procedure above? Although the population could exhibit significant clustering at a given level $T_K$, not all individuals must be involved in “significant clusters” that are significantly denser than random clusters.

Finally, I analyzed whether the association strength between two individuals was correlated with the extent to which their home ranges overlapped. Because of the non-independent nature of network data, I used MR-QAP (multiple regression quadratic assignment procedure, (Krackhardt 1988; Dekker et al. 2007) to determine the effect of shared space use on AS (arcsine-transformed). MR-QAP is a variation of the Mantel test that allows a dependent matrix (in this case, the association matrix) to be regressed against independent matrices (home range overlap matrix). After performing a standard regression analysis across the corresponding cells of each matrix, the procedure randomly permutes the rows and columns of the dependent matrix and re-computes regression
coefficients 1000 times. This generates a distribution of coefficients against which the observed coefficients are compared in order to generate p-values (Krackhardt 1988). Because early analysis indicated that age may be a critical factor influencing grouping patterns in males, 0/1 dummy variables were included in the male MR-QAP representing the age combination present in the dyad: younger-younger, younger-older, and older-older. The younger class included both subadult males and younger adult males. Ages were subjectively determined by height, coloration, and ossicone size (Pratt and Anderson 1982).

Results

Giraffe group sizes ranged from 1 to 44 individuals (mean: $5.42 \pm 0.19$ individuals, mode: 1). Giraffe observed alone were nearly always adult males (84% of lone giraffe sightings), although females were sometimes observed alone in the days prior to being observed with very young calves. Approximately 27.4% groups were greater than six animals, and 12.1% of groups were greater than 12. Average home range size was $95.7 \pm 3.3$ km$^2$ for adult males, $64.2 \pm 3.4$ km$^2$ for adult females, $110.0 \pm 8.8$ km$^2$ for subadult males, $70.5 \pm 15.2$ km$^2$ for subadult females, $51.0 \pm 7.7$ km$^2$ for juveniles, and $17.9 \pm 3.4$ km$^2$ for neonates. Population density was approximately $0.6$ giraffe/km$^2$.

In each subset of the data (adult and subadult females, adult and subadult males, males and females combined), community structure algorithms identified three hierarchical levels of clustering in which several social cliques were embedded in higher-order sub-communities, which were in turn embedded in larger communities (Figures 1-2). In the female network, for example, five cliques were embedded in three sub-
communities, and two of these sub-communities (B.2 and B.3) were embedded into a single community cluster (A.2). In the combined sex analysis, seven out of eight social cliques consisted primarily of animals of the same sex (>70% of membership was one sex). Therefore, I focused on single-sex networks in subsequent analyses.

**Figure 1.** Social network and map of social clique home ranges for (A) females, and (B) males. For visualization purposes, network edges have been filtered at the population mean AS + 1 SD. Edges with lower AS are not included in this figure. Some males were involved in no dyads that exceeded this threshold, and thus became isolates in the filtered network. Isolated individuals (31 males) are not pictured. White nodes indicate males that were not members of significant clusters (≥4 members and p < 0.05 in randomization tests).
Figure 2. Multi-tiered DCG trees depict social clustering for female giraffes (a-b) and male giraffes (c-d). Heatmap matrices show AS (a and c) and home range overlap (b and d) for each pair of individuals. The social DCG tree is used to determine the order of individuals for both the association strength and home range overlap matrices. Darker shading indicates higher values of AS or overlap. Trees on the horizontal and vertical axes are identical. Organizational levels are denoted with letters (Communities=A, Sub-communities=B, Cliques=C), and asterisks indicate clusters that were significantly better connected than random expectations.
To visualize the significance of the DCG trees, I depicted the AS matrix as a heatmap in which darker shading indicates higher AS values for the $i$th and $j$th giraffe (Figure 2). Darkly shaded blocks in the heatmap highlight clusters with denser connections among individuals. In Figure 2a, for example, females are grouped into two clusters at the community level, A.1 and A.2. Each of these clusters corresponds to a darker portion of the heatmap of the AS matrix, which means the within-cluster tie strength is high for both A.1 and A.2. In contrast, the area of the heatmap corresponding to associations between individuals of different communities is relatively lighter in color, indicating low tie-strength between A.1 and A.2. At the sub-community level, A2 is split into two clusters (B.2 and B.3). Associations within B.2 and B.3, respectively, correspond to even darker blocks in the heatmap, which illustrates that there are two densely-connected finer-scale clusters within community A.2. These patterns are well captured by the multi-leveled DCG tree. Further hierarchical levels can be observed and interpreted similarly. Additional heatmaps were constructed to illustrate whether individuals in the same social clusters also tended to have high levels of home range overlap (Figure 2b, d).

For females, clustering at each hierarchical level was significant. In addition, each distinct community, sub-community, and clique was significantly denser than randomly generated clusters (Figure 2a-b). For males, only clustering at the social clique level was significant, and not all cliques were significantly denser than randomly generated clusters (Figure 2c-d).

The percentage of observations in which two individuals were seen together was highly correlated with shared space use (Figure 3). When age was included in the male
MR-QAP, younger adults and subadults had significantly higher AS with males of the same age group than with older males ($\beta = 0.06, p < 0.01$), while older males had lower AS with other older males than with younger males ($\beta = -0.02, p < 0.01$). Female social organization appeared to closely correspond with shared space use; when pairwise home range overlap values were organized according to the social clustering trees, the block-like patterns depicted for home range overlap were highly similar to the social patterns (Figure 2a-b). This was not the case for male communities; there was little correspondence between social clustering and home range overlap (Figure 2c-d).

Moreover, when community-finding algorithms were performed on the home range network to assess spatial clustering, only 32.6% of male pairs assigned to the same spatial neighborhood were also in the same social clique. Thus, social and spatial communities were not analogous for males. In contrast, 89.3% of female dyads that were assigned to the same spatial neighborhood were also in the same social clique.

This difference between males and females is also apparent through a comparison of the degree to which social clique home ranges overlap (Figure 1). Figure 4 shows the average home range overlap between animals in the same versus different clusters. If social clustering were determined by space use alone, I would expect home range overlap to be higher for individuals in the same community than across communities. Consequently, grid-cells along the table’s diagonal should display higher values, as seen for female communities and sub-communities. At finer scales of social organization (cliques), however, females of different communities exhibited home range overlap >35% (cliques C.1 and C.2). Some female cliques were almost completely encompassed within the home range of another clique (C.5 by C.1 and C.4, Figure 1a). The home
ranges of male cliques were also highly overlapping (Figures 1, 4). Thus, although home range overlap and association strength are correlated, space use did not seem to determine social structure in males or to completely determine clique-level social organization in females.

**Figure 3.** Effect of home range overlap on association strength for (A) adult and subadult females, (B) adult and subadult males, and (C) combined. MR-QAP regression of the effect of home range overlap on association strength (arcsine-transformed) for the adult and subadult male ($\beta = 0.58$, $p < 0.01$), female ($\beta = 0.70$, $p < 0.01$), and combined datasets ($\beta = 0.61$, $p < 0.01$).
Figure 4. Average percent home range overlap of individuals within the same social community versus between communities. Shading is darker for higher levels of overlap. Community and sub-community matrices are not shown for males because social clustering was not significant at these levels. For male cliques, only significant clusters are shown. Therefore, I generated an out-group metric which is the average overlap between individuals of a clique to other giraffe in the same sub-community but not the same clique. No out-group was provided for females because all females were part of significant clusters. Darker shading of cells on the diagonal of each table reflects where space use patterns correspond with social structure. For males, in contrast, average home range overlap for pairs of males in the same social clique is no higher than overlap that for pairs across clique.

Discussion

The social organization structure identified in this study is the first quantitative evidence of multiple levels of social organization in the giraffe and confirms earlier suggestions that giraffe fission-fusion dynamics are embedded in higher levels of
organization (Pratt and Anderson 1982; Bashaw et al. 2007; Bercovitch and Berry 2009a; Carter et al. 2012). Individual giraffe exhibit the strongest social ties with a core group of others (their clique), but still maintain relatively high AS with members of their sub-community. Individuals may sustain moderate amounts of association with members of their community, but AS is typically low with giraffe outside their community (Figure 2). My analyses do not mean that members of cliques are always observed together or do not associate with other giraffe, but rather that these animals tend to be together more often than with animals that are not members of their clique. Association strength within clusters may not necessarily be substantially higher than across clusters, but there is high transitivity within clusters, indicating that individuals linked to a common neighbor are themselves likely to be linked (i.e. the friend of my friend is also my friend, Wasserman and Faust 1994).

While several authors suggest that giraffe social organization is characterized by fission-fusion dynamics embedded in higher-level communities (Bashaw et al. 2007; Bercovitch and Berry 2009a; Carter et al. 2012), these conclusions have stemmed from observed variation in group sizes, shifting group membership, and pairwise association indices without putting such interactions into the context of a larger social network. Thus, they suggested that community structure might exist without defining the multi-leveled nature of the organization, as I was able to demonstrate here. Pratt and Anderson (1982) described an apparent social division in their study populations, but the division was not quantified rigorously. Such divisions have not been reported elsewhere, but that may be due to small sample sizes. Community structure in a loosely social species like the giraffe may be difficult to readily observe without large sample sizes and the application of new
analytical tools. Shorrocks and Croft (2009) conducted a preliminary analysis of giraffe social networks. However, the conclusions they were able to make about the nature of giraffe social organization were limited. The maximum number of observations that any individual was re-sighted was four instances, and the maximum number of times they observed repeated associations among individuals was twice (Shorrocks and Croft 2009). In their network, an average individual in their network was connected to a total of five other giraffe. This number is far below the total number of connections exhibited by an average individual in my population (~98 connections), but similar to the average group size observed both here and in other populations (Dagg and Foster 1976; Leuthold 1979; Le Pendu et al. 2000; van der Jeugd and Prins 2000; Fennessy 2004; Bercovitch and Berry 2009a; Shorrocks and Croft 2009). Thus, the network observed in their four-week study was more of a snapshot of association patterns in time rather than an overall description of giraffe social organization. In comparison, I recorded nearly 1100 groups and observed each individual approximately 30 times with some pairs recorded together up to 34 times, allowing giraffe social organization to be rigorously analyzed.

Males and females can be considered to occupy separate social networks, given that seven of eight social cliques were primarily single-sex in the combined analyses. Although most observed giraffe groups contain both males and females, the repeated observations and transitivity that are necessary to classify individuals into clusters tend to occur only within sexes. Giraffe populations also exhibit sexual segregation by habitat, with cow-calf groups preferring open habitats and bulls being more commonly observed in denser habitats (Young and Isbell 1991; Ginnett and Demment 1999; Bercovitch and Berry 2009a).
Male social cliques are akin to “bachelor herds” described in other ungulate species. Bachelor cliques observed in the study population do not appear to be random collections of young males, but rather consist of males that are familiar with each other and are repeatedly observed together. Younger males tended to be observed in larger groups of other males, which may or may not also be accompanied by females. Older males tended to be found alone, in pairs, or with mostly female groups. Indeed, the MR-QAP analysis suggests that younger males actively prefer to associate with animals of similar age, while older males, which tend to dominate mating opportunities (Pratt and Anderson 1982), avoided associating with rivals. This follows the observations of Pratt and Anderson (1982). Carter et al. (2012) demonstrated a non-significant trend for males to have preferred social partners only when subadult males were included in the analysis, suggesting that older males do not exhibit social preferences. The most readily apparent clusters within our male social network (Figure 2c) tended to consist of younger males. Four of the five significant male cliques were comprised of at least >70% younger bulls, and 100% of cliques C.3 and C.7 were younger.

Any population that exhibits spatial structure and loose aggregations of individuals will likely exhibit social network structure, even if individuals merely associate randomly with individuals in spatial proximity. Indeed, association strength was highly correlated with home range overlap, which is unsurprising given that animals cannot possibly socially interact if they do not share space. Highly overlapping home ranges should lead to at least some interaction, while discontiguous home ranges rarely lead to interaction. However, there is a broad range of intermediate overlap where variation in association strength cannot be explained purely by home range overlap.
While home range overlap does explain a substantial amount of variance in association strength (adjusted $R^2 = 0.61$ for combined sex analysis), it is evident that there are numerous dyads with home range overlaps greater than 60% that associate no more frequently than dyads with only 20% overlap (Figure 3). These results mirror those of Carter et al. (2012).

Random associations among individuals with overlapping home ranges should lead to high levels of correlation between the structure of social and home range networks. Differences in the social network structure relative to the home range network emerge from behavioral choices by individuals. The decision to associate with a preferred set of individuals within one’s spatial neighborhood will cause the home range network to be overlaid with an additional social layer representing spatiotemporal overlap (sharing the same space at the same time). Community-finding algorithms were performed on the home range network to assess if social clusters simply consisted of individuals that were clustered in the same spatial neighborhood. Only a third of male pairs assigned to the same spatial neighborhood were also in the same social clique. Thus, social and spatial communities were not analogous for males. In contrast, nearly 90% of female dyads that were assigned to the same spatial neighborhood were also in the same social clique. This suggests that female social organization has an underlying spatial basis. Female communities A.1 and A.2 were geographically separated by a river, which females only rarely crossed. However, spatial separation between female sub-communities was evident despite no geographic barriers to movement (sub-communities B.2 and B.3, Figure 1), and the home ranges of female social cliques overlapped substantially in some cases (cliques C.1 and C.2; cliques C.1, C.4, and C.5, Figure 1). These results suggest although
space use is a primary contributor to association strength among females, finer-scale social structure (i.e. membership in sub-communities and cliques) does not emerge solely from patterns of space use.

Female social organization was much more strongly influenced by shared space use than male community structure (Figure 2). This relationship between female community structure and space use is consistent with matrilineal-based structure if two conditions are met: female philopatry and female-female bonds. Preferential associations among kin are not required to produce spatially-based kin structure. Even in solitary species, female locational philopatry may lead to clustering of related animals on the landscape (Waser and Jones 1983). If females associate randomly with other individuals within their home range, philopatry alone may produce kin-based social structure even if bonds are not maintained between female kin (Waser and Jones 1983; Isbell 2004; Wolf and Trillmich 2008).

The fact that social cliques, in some cases, occupied highly overlapping home ranges suggests that females do not randomly associate with other females within their own home ranges. Recent work by Bercovitch and Berry (2012) indicates that female-female bonds do exist. Sister-sister pairs were more likely to associate, as were mothers with their adult daughters (Bercovitch and Berry 2012). Mothers were observed alone with their adult offspring of up to 10 years, and giraffe groups were observed with up to three generations of maternal kin (Bercovitch and Berry 2009a). Further evidence for kin-based bonds among female giraffe arises from the fact that female calves remain with their mothers longer than male calves (Pratt and Anderson 1979), though not all studies confirm this (Bercovitch and Berry 2009a). In a captive study, adult females were most
affiliative with their subadult daughters (Bashaw et al. 2007). Fennessy (2004) described a small population of 17 giraffe that appeared to be divided into two core groups, each consisting of a cow and her calves. Finally, Carter et al. (2012) showed that female giraffe exhibiting social preference for one another were more related than expected by chance.

While these studies suggest the existence of kinship-based social structure, genetic and long-term behavioral studies will be required to confirm whether multi-level social structure is matrilineally based and how this may influence the evolution and expression of other social behaviors. For example, mothers frequently pool their calves in nursery crèches in which one female often remains with the calves while the other mothers forage elsewhere (Langman 1977; Pratt and Anderson 1979). Even in the absence of an inclusive fitness explanation, female social cliques may set the stage for evolution of crèches by reciprocal altruism.

Giraffe social structure is highly fluid, which has historically made it difficult to identify social structure. The frequent fissioning and fusioning of giraffe groups has led to interpretations that giraffe social interactions are random and ephemeral in nature. Through the application of network analysis and DCG community-finding methods, however, I was able to discern structural social organization. The results presented here, taken together with other recent work on giraffe social behavior (Bercovitch and Berry 2012; Carter et al. 2012), lead me to reject the notion that giraffe lack social organization and that associations are simple random aggregations. While social interactions are highly fluid in nature, it is becoming increasingly clear that association patterns in giraffe are not the result of random fission-fusion events, but are embedded within a structured
social network characterized by multiple levels of organization. This has implications for
giraffe conservation as well as for the evolution of social behavior. More generally, these
results give us new insights into how social structure in fission-fusion species is
influenced by underlying patterns of space use. Observations such as these are key for
further elucidating how socioecological factors give rise to social structure in fission-
fusion species.
Literature cited


Chapter 3: Linking social and pathogen transmission networks using microbial genetics in giraffe (*Giraffa camelopardalis*)

Abstract

Although network analysis has drawn considerable attention as a promising tool for disease ecology, empirical research has been hindered by limitations in detecting the occurrence of pathogen transmission (who transmitted to whom) in social networks. Using a novel approach, I utilize the genetics of a diverse microbe, *Escherichia coli*, to infer where direct or indirect transmission has occurred and use these data to construct transmission networks for a wild giraffe population (*Giraffa camelopardalis*). Individuals were considered to be part of the same transmission chain and were interlinked in the transmission network if they shared genetic subtypes of *E. coli*. By using microbial genetics to quantify who transmits to whom independently of behavioral data on who is in contact with whom, I was able to directly investigate how the structure of contact networks influences the structure of the transmission network. To distinguish between the effects of social and environmental contact on transmission dynamics, the transmission network was compared to two separate contact networks defined from behavioral data: a social network based on association patterns, and a spatial network based on patterns of home range overlap among individuals. I found that links in the transmission network were more likely to occur between individuals that were strongly linked in the social
network. The occurrence of links was unrelated to the space-use network. Furthermore, individuals that had more numerous connections or that occupied “bottleneck” positions in the social network tended to occupy similar positions in the transmission network. No similar correlations were observed between individual position in the space-use and transmission networks. This indicates that an individual’s social network position is predictive of transmission network position, which has implications for identifying individuals that function as super-spreaders or transmission bottlenecks in the population. These results emphasize the importance of association patterns in understanding transmission dynamics, even for environmentally-transmitted microbes like *E. coli*. This study is the first to use microbial genetics to construct and measure variation in individual connectivity in a transmission network for a transmissible agent in wildlife and highlights the potential utility of an approach integrating microbial genetics with network analysis.

**Introduction**

Epidemiological models have traditionally assumed that the probability of contact is equal for every pair of individuals in the population. Thus, every individual is equally likely to acquire a pathogen from an infected animal. In reality, however, an animal’s risk of infection is dependent on local patterns of interaction. Therefore, a population’s spatial and social structure create heterogeneity in transmission patterns (Keeling and Eames 2005; Bansal et al. 2007; Otterstatter and Thomson 2007; Ostfeld et al. 2008; Perkins et al. 2008; Craft and Caillaud 2011). Network theory provides a set of tools for analyzing such heterogeneity by not only taking into account direct connections, but also indirect links between individuals, providing a sophisticated method to analyze relations among
individuals (Wasserman and Faust 1994; Croft et al. 2008). Recently, there has been considerable effort to incorporate network theory into epidemiological models (Keeling 2005; May 2006; Bansal et al. 2007; Craft and Caillaud 2011). These models tend to result in reductions in the early growth rate, number of secondary infections for each infected individual, and final size of an epidemic as compared to traditional, mass-action models (Keeling and Eames 2005; Turner et al. 2008). Thus, accounting for heterogeneity in transmission dynamics improves our ability to understand and predict the dynamics of infectious disease (Keeling 1999; Keeling and Eames 2005; Turner et al. 2008; Ames et al. 2011).

Although network analysis has been lauded as the next key tool for examining how heterogeneous transmission patterns affect disease spread (Delahay et al. 2009; Craft and Caillaud 2011; Tompkins et al. 2011), the structure of transmission networks in wildlife is relatively unknown and difficult to quantify. In empirical studies, contact networks are generally constructed using behavioral data on space use or social interactions, both of which have the potential to create transmission opportunities (Godfrey et al. 2009; Perkins et al. 2009). While modelling has proven useful as an approach to study the importance of wildlife contact networks in disease spread (Cross et al. 2004; Perkins et al. 2008; Craft et al. 2010; Griffin and Nunn 2012), it is difficult to empirically study transmission routes in wild populations because data on who transmitted a pathogen to whom is almost impossible to obtain using current methods, such as commonly used serological techniques (Caley et al. 2009). In many cases, “transmission networks” are constructed based on social interactions between individuals without assessing any pathogens. Because of limitations in detecting the occurrence of
transmission, conclusions about disease spread are often based on the possibility that transmission could occur between interacting individuals (Perkins et al. 2008; Böhm et al. 2009; Craft et al. 2009; Grear et al. 2009; Hamede et al. 2009; Perkins et al. 2009). Studies that do integrate networks with empirical data on pathogens typically treat an individual’s infection status as an individual attribute and then assess how one's connectivity in the contact network affects the likelihood of being infected. Generally, it has been shown that infected animals tend to have more connections in the contact network (Corner et al. 2003; Godfrey et al. 2009; Godfrey et al. 2010; Fenner et al. 2011), or that individuals are more at risk of becoming infected if they are connected to infected individuals or engage in certain types of interactions (Otterstatter and Thomson 2007; Drewe 2009; Porphyre et al. 2011). Because it is generally not possible to determine who transmitted to whom outside of detailed longitudinal studies, the transmission networks in these studies are not independently defined from a behavior-based contact network. Therefore, it is not possible to truly assess how social patterns influence the structure of the transmission network.

In contrast, I use the genetics of a diverse microbe, Escherichia coli, to reveal where transmission has occurred and construct a transmission network based on quantifiable transmission events. If two individuals share the same genetic subtype of E. coli, then I infer that either direct transmission has occurred through social interactions or indirect transmission has occurred due to exposure to a common environmental source (Archie et al. 2008). In this transmission network, individuals are interlinked based upon patterns of E. coli subtype sharing. Thus, the transmission network is defined independently from contact networks, which are defined by behavioral data on who has
social or spatial contact with whom. This allows us to examine how well the transmission network overlays onto social and spatial networks and ascertain the relative importance of direct versus indirect transmission mechanisms for enteric microbes such as *E. coli*.

*Escherichia coli* is an excellent model organism for examining microbe transmission pathways in relation to contact networks. It can readily be cultured from fecal samples and exhibits immense genetic diversity. Furthermore, the tools for genotyping *E. coli* are well established (Dombek et al. 2000; Simpson et al. 2002; Goldberg et al. 2006; Mohapatra and Mazumder 2008). *Escherichia coli* is commonly used as an indicator for environmental fecal contamination (Tallon et al. 2005) and subtyping is a common method for tracing *E. coli* to its source (Simpson et al. 2002). *Escherichia coli* subtype sharing has been used to demonstrate that transmission regularly occurs between humans and their pets, and between humans, livestock, and wild primates (Goldberg et al. 2008; Johnson et al. 2008; Damborg et al. 2009). Recently, serotypes of shiga-toxin producing *E. coli* have become a major public health concern and are considered an emerging infectious disease (Beutin 2006). Here, I combine network analysis with microbial genetics to construct transmission networks in reticulated giraffe (*Giraffa camelopardalis reticulata*). I further investigate how strongly social and spatial structure within the population influence transmission patterns.

**Methods**

**Study site and organisms**

This study was conducted at Ol Pejeta Conservancy (OPC). OPC is a 364 km$^2$ semi-arid savanna woodland ecosystem located in Laikipia, Kenya (0º N, 36º56’ E). At
the conclusion of the study period, OPC’s giraffe population consisted of 160 adults, 20 subadults, and 32 juveniles. The population exhibited a 50:50 sex ratio. All giraffe were recognized using individually unique spot patterns on their necks. Giraffes were aged according to height estimates and age-associated behaviors (see Chapter 2 for full detail). Animals were considered juveniles if they were <1.5 years and adult at approximately 4 years. In this study, subadults (1.5 – 4 years) were grouped with adults because they no longer constantly accompanied their mothers and they exhibited adult-like social and space-use patterns.

Field observations

From Jan. 21 to Aug. 2, 2011 giraffe group composition and membership were recorded for all giraffe groups sighted while driving pre-determined survey routes. Routes were determined so that a different part of the study area was surveyed each day, allowing for most of the study area to be surveyed once every three days. Routes were approximately 100 km in length, covered approximately 115 km² each, and traversed all habitat types. Giraffe groups observed from the survey routes were followed off-road until a complete census of the individuals present was accomplished.

All individuals observed within a group were recorded as “in association” with every other member of the group. A group was defined as a set of individuals engaged in the same behavior, or moving in the same direction or toward a common destination, as long as each giraffe was no more than 500 m from at least one other group member. This definition was adapted from the literature (Leuthold 1979; Fennessy 2004). During the study period, I collected a total of 1089 sightings of giraffe groups. Each individual
giraffe was observed on average 31.1 ± 7.6 SD times (approximately once per week). Group sizes at OPC ranged from 1 to 44 giraffe (mean: 5.42, mode: 1 giraffe).

Each individual’s home range was mapped using the GPS locations recorded for each sighting. Home range boundaries were determined using a fixed-kernel utilization-distribution of sightings. A 75% contour (kernel density isopleth) was used to produce a core home range for each animal (Harris et al. 1999). We found that there was no correlation between number of sightings and home range size when the total number of sightings for an individual was greater than five. Average home range size was 95.7 ± 3.3 km² for adult males, 64.2 ± 3.4 km² for adult females, 51.0 ± 7.7 km² for juveniles.

Giraffe were excluded from subsequent analyses if they were seen fewer than five times (n=2) or if I failed to collect a fecal sample from them (n=14).

Fecal sample collection and genetic analysis

We collected fecal samples from 194 giraffe. Because there can be significant turnover of *E. coli* subtypes in the gut (Anderson et al. 2006), fecal samples were collected during the brief time period between Aug. 10 – Sept. 11, 2011 to ensure comparability. Four fecal samples were collected after this period, but within three weeks of the end of the primary sampling period. Fecal samples were collected immediately after defecation was observed and transported on ice to the field laboratory. Samples were streaked for bacterial isolation onto CHROMagar EC agar (CHROMagar, Paris France), a selective chromagenic agar that exhibits high specificity for *E. coli*. After overnight incubation at 37º C, four randomly selected *E. coli* colony isolates were cultured and then frozen. Using data from captive giraffe, which hosted approximately
two subtypes of *E. coli* per individual (unpublished data), I calculated that there was a <10% probability of failing to capture subtype diversity in the gut if four isolates were taken (Singer et al. 2000). *E. coli* genetic subtypes were determined using BOX-PCR and gel electrophoresis, which is a well-established method for discriminating between genetically similar *E. coli* subtypes (Goldberg et al. 2006; Cesaris et al. 2007; Mohapatra and Mazumder 2008). Detailed laboratory procedures are listed in the Supplementary Information.

Densitometric profiles for each isolate were generated from the banding patterns revealed by BOX-PCR and gel electrophoresis (Johnson and O’Bryan 2000; Goldberg et al. 2006). Similarity of each isolate to all others was determined through pairwise comparisons of densitometric curves (Supplementary Figure S1). Subtypes were considered to be matching if their densitometric curves were >90% similar (Pearson’s correlation coefficient). Based on a reproducibility analysis conducted in my lab, this cutoff value minimizes Type I errors in matching to 1% while limiting the Type II error rate to <5%.

**Network construction**

We constructed one transmission network, and three contact networks representing three types of contact: i) a social network, ii) spatial network, and iii) water-sharing network. All networks contained the exact same set of individuals (N=194). For the transmission network, two individuals were considered to be part of the same transmission chain if they shared at least one *E. coli* subtype. In this network, individuals
were linked according to subtype sharing. Connections between individuals, called “ties,” were unweighted.

A social network was constructed from observed association patterns. I calculated the association strength (AS) between every pair of individuals as the total number of observations they were seen together divided by the total number they were seen together or apart. Pairs with non-zero AS were linked in the social network, with links being weighted according to the value of their AS.

We also constructed a spatial network using the extent of home range overlap to connect individuals. Home range overlap between two giraffe was defined as the number of 1 km² grid squares that fell within both individuals’ home ranges, divided by the total size of both individuals’ home ranges. These dyadic spatial overlap values were used to connect giraffe in the spatial network, with links being weighted according to the extent to which their home ranges overlapped.

Because water points may serve as an environmental source of bacteria, I also constructed a network based on shared water points. Water points could be rivers, streams, dammed streams, or concrete water troughs constructed for cattle. No observations of individual usage of water points were made. Rather, it was assumed that animals have a high probability of using water points within their home range. I counted the number of permanent water points that fell within each animal’s home range, and then calculated the proportion of water points that were shared between each pair. These dyadic values were used to connect giraffe in a water-sharing network, with links being weighted according to the number of water sources shared.
**Statistical analysis**

1) *Network-level analysis - To what extent do the contact networks predict the transmission network?*

To determine the extent to which the contact networks (social, spatial, water-sharing) influenced the transmission network, I used the multiple regression quadratic assignment procedure (MR-QAP), a method of matrix regression developed for network data (Krackhardt 1988; Dekker et al. 2007). Each network was represented as an adjacency matrix, where each cell denoted the relationship between the $i^{th}$ and $j^{th}$ giraffe. Essentially, MR-QAP coerces matrices into vectors, and then performs a standard logistic regression on the log-odds of an edge occurring in the transmission network given the dyad’s social and spatial relationships. Relational data used to construct networks have the potential to be autocorrelated, given that every element in the $j^{th}$ row of the matrix is associated with a single individual. Thus, traditional standard error and p-value estimates are potentially biased because the assumption of independence is violated; a single individual appears in multiple dyadic relationships. Thus, MR-QAP uses a Monte Carlo method, in which rows and columns are randomly permuted within matrices, to determine the significance of regression coefficients (Dekker et al. 2007). Using MR-QAP with double Dekker semi-partialling and 1,000 permutations (Dekker et al. 2007), I investigated the effect of tie-strength (association strength, home range overlap, shared water sources) in the contact networks on the log-odds of a tie in the transmission network. Analysis was performed using the ‘sna’ package of R (Butts 2010).

If social contact is crucial for transmission to occur, then the MR-QAP analysis should reveal high correlations between AS in the social network and the occurrence of
subtype sharing in the transmission network. However, if environmental transmission dominates transmission dynamics, then correlations should be highest between the transmission network and the spatial or water-sharing network rather than the social network. If drinking from the same water source is the most critical component leading to environmental transmission, then individuals that have a higher probability of drinking from the same water sources would be more likely to share *E. coli* subtypes, and the water-sharing network should yield the highest correlation with the transmission network compared to the other contact networks.

2) Individual-level analysis – Does an individual’s position in the contact networks predict its position in the transmission network?

We calculated the centrality and connectivity of each individual in the transmission network using five established measures of nodal connectivity: degree, betweenness, closeness, eigenvector centrality, and information centrality (see the Supplementary Text for metric definitions). Because many of these metrics were highly correlated with each other, I chose to focus my analysis on degree and betweenness, which are two of the most common and easily interpretable metrics used in the literature. Degree is defined as the number of individuals to which the focal node is linked. Betweenness is a measure of centrality that is based on how many paths pass through the focal individual if the shortest paths between every other pair of individuals are traced (Wasserman and Faust 1994). Thus, in the context of pathogen transmission, it quantifies the extent to which an individual serves as a conduit or bottleneck for the spread of pathogens through a network. An individual that lies on the path connecting two other
individuals can be considered to “mediate” or “regulate” flow of pathogens between those two individuals. Individuals with high betweenness lie on many such paths and thus can be considered to have high flow. These individuals have the potential to play a large role in regulating pathogen spread in the network (Borgatti 1995).

I also calculated these measures for each individual in the contact networks. Because edges in the social and spatial networks were weighted according to tie strength, I used weighted versions of degree and betweenness (Newman 2001; Opsahl 2009). Weighted degree (hereafter: social/spatial degree) simultaneously accounts for the number of nodes the focal animal is connected to and the strength of those connections (tie-strength). Overall tie-strength, which is closely related to degree, is the sum of the weights of all edges connected to the focal node, regardless of the total number of neighbors. Weighted measures were calculated in the ‘tnet’ package in R (Opsahl 2009).

Individuals that are connected to a large number of others in the transmission network of others may be potential super-spreaders (Lloyd-Smith et al. 2005). I am interested in the mechanisms that give rise to individual-level variation in transmission network connectivity; do individuals become super-spreaders because they are highly social or because they have a large number of spatial contacts? If highly social animals are potential super-spreaders, then social and transmission degree should be correlated. However, if widely ranging animals tend to be super-spreaders, than transmission degree should be correlated with large home range sizes and high spatial degree.

To address this, I examined how transmission degree and betweenness were correlated with home range size and connectivity in the spatial and social networks using general linear models. Measures of social and spatial network connectivity included
weighted betweenness, weighted degree, and overall tie-strength. I also used GLMs to examine the effect of home range size on an individual’s connectivity in the contact networks, and the effect of the spatial network on social network connectivity. All GLMs were univariate because high levels of multicollinearity among the network metrics precluded multivariate analysis. Because of possible non-independence concerns in network data, regression coefficients were determined using GLMs while p-values were calculated via permutation methods in which the order of y was randomized relative to x. After each permutation of the data (3000 total permutations), the regression coefficient was recalculated, generating a distribution of coefficients in which the relationship of y to x was calculated on randomized data. P-values were defined as the proportion of permutations that produced slopes more extreme than the observed value. Because degree exhibited a bimodal distribution, degree was ranked for GLM analyses. High ranks indicate higher degree.

Individual-level analyses of the water-sharing network were not included because the network was nearly identical to the home range overlap network. Centrality metrics of individuals in the spatial overlap and water-sharing networks were highly correlated (r > 0.9), and the relationships of these values to transmission network centrality were very similar.

**Results**

I tested the extent to which the transmission network was predicted by three types of contact networks. Only association strength in the social network was significantly correlated with the presence of a transmission link (i.e., sharing *E. coli* subtypes);
individuals that associated more frequently were more likely to be part of the same \textit{E. coli} transmission chain (Table 1, Figure 1). Contact among individuals via shared space use, whether measured by home range overlap or shared water sources, was not significantly correlated with sharing \textit{E. coli} subtypes. Thus, the presence and strength of links in the social network best predicted the transmission network. The higher correspondence of the transmission network to the social network can be visually assessed in Figure 2, which depicts a subset (\( N = 30 \) randomly selected giraffes) of the transmission network alongside the spatial and social networks.

**Table 1.** Effects of association strength and spatial overlap on the probability (log-odds) that two animals are linked in the transmission network. Regression coefficients were estimated through standard logistic regression. P-values were based on MR-QAP permutation tests.

<table>
<thead>
<tr>
<th>Model</th>
<th>Covariate</th>
<th>Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Association strength</td>
<td>1.22</td>
<td>0.02</td>
</tr>
<tr>
<td>B</td>
<td>Spatial overlap</td>
<td>0.03</td>
<td>0.91</td>
</tr>
<tr>
<td>C</td>
<td>Shared water sources</td>
<td>0.02</td>
<td>0.92</td>
</tr>
<tr>
<td>D</td>
<td>Association strength</td>
<td>2.26</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Spatial overlap</td>
<td>-0.50</td>
<td>0.25</td>
</tr>
<tr>
<td>E</td>
<td>Association strength</td>
<td>2.13</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Shared water sources</td>
<td>-0.41</td>
<td>0.19</td>
</tr>
</tbody>
</table>

**Figure 1.** Relationship between association strength in the social network and the probability of sharing genetic subtypes of \textit{E. coli} in the transmission network. Shading indicates a 95\% confidence interval around the regression line. Data points indicate the proportion of dyads that share subtypes.
Figure 2. Comparison of the (a) spatial, (b) social, and (c) transmission networks. For visualization purposes, only links that exceed the mean tie-strength + 1 SD are pictured in the social and spatial network (Home range overlap > 36.4%, AS > 8.9%) and only a randomly selected subset of 30 individuals is shown.
At the individual level, I investigated the extent to which an animal’s position in one network was correlated with its position in another network (Figure 3). Only models with significant coefficients are presented in Figure 3 (See Supplementary Table S1 for full model output). A giraffe’s degree in the transmission network was positively correlated with its degree in the social network, indicating that individual giraffe with higher social degree and overall social tie-strength tended to have a greater number of connections in the transmission network. Home range size and measures of spatial degree and tie-strength were not correlated with transmission degree, which mirrors the results of the spatial MR-QAP analysis.

An individual’s transmission betweenness was positively correlated with social betweenness, though this trend was not quite significant (p = 0.065). Transmission betweenness, was unaffected by connectivity in the spatial network or measures of social degree. It is also interesting to note that spatial degree is positively correlated with social betweenness, likely because animals that are well connected in the spatial network are positioned ideally to connect social communities. Individuals whose home ranges overlapped with a large number of others tended to have a greater diversity of associates, but not higher social tie-strength.
**Figure 3.** Correlations across individual-level measures of connectivity. All relationships depicted are positive and significant in univariate GLMs ($p < 0.05$), except for the relationship between transmission and social betweenness ($p = 0.065$). Regression coefficients are noted next to each arrow. Individual connectivity measures in the transmission network (betweenness / degree) were correlated with the same measure in the social network, but not correlated with spatial connectivity measures or home range size. Social connectivity was influenced by home range size and spatial connectivity.

Discussion

Our results clearly indicate that patterns of social interaction play a stronger role in determining patterns of *E. coli* transmission than space use. Both network- and individual-level analyses find greater correspondence of the transmission network to the social network as compared to the spatial network. Association strength between dyads in the social network was strongly correlated with the occurrence of a transmission link between that pair of individuals. In contrast, shared space use had no effect on the
occurrence of links in the transmission network, regardless of whether shared space use was measured as home range overlap or shared water sources (Table 1). Measures of individual connectivity parallel these findings in that an individual’s position (degree / betweenness) in the transmission network was highly correlated with its position in the social network, but not the spatial network. Individuals with greater numbers of social connections (social degree) tended to have more connections in the transmission network. In other words, individuals that functioned as social hubs were also transmission hubs. Giraffe with high social betweenness tend to bridge socially-distinct clusters of individuals. These same individuals also tend to serve as bridges in the transmission network by occupying positions of high flow. Individuals with high betweenness are potentially positioned along bottlenecks for pathogen spread in the network. In an epidemic, bottlenecks may function as firebreaks for pathogen spread, but when these individuals do become infected, the pathogen may spread to new regions of the network (Salathé and Jones 2010).

While there seems to be greater relative importance of social patterns over shared space use in determining the transmission network, my results do not negate a role of shared space use. Rather, the importance of spatial patterns on transmission seems to be mediated through the role that the spatial network plays in determining the social network. Specifically, individuals with high degree in the spatial network tended to have higher social degree and betweenness, which are in turn strongly correlated with higher transmission degree and betweenness. Giraffe whose home ranges overlapped with a greater diversity of individuals also tended to have higher social degree (Figure 3). In addition, pairs of individuals that have greater home range overlap have significantly
higher association strength (Chapter 2), and association strength is positively correlated with the likelihood of sharing \textit{E. coli} subtypes. These results indicate that the spatial network plays a key role in determining the social network, which in turn affects pathogen transmission patterns.

Indeed, when the spatial network (Figure 2a) is visually compared to the social network (Figure 2b), the social network appears to be a pared-down version of the spatial network. This indicates that while overlapping in space is a prerequisite for social interaction, some giraffe that share space do not frequently associate. This paring down appears to continue when comparing the social network to the transmission network (Figure 2c) in that not all social contacts resulted in sharing \textit{E. coli} subtypes. The social network can be viewed as an additional level of filtering of spatial network data: dyads must not only overlap spatially, but also spatiotemporally to be linked in the social network. High correlations between the social and transmission networks are produced because the social network filters out many of the extraneous ties in the spatial network that are relatively meaningless for the transmission of \textit{E. coli}. These findings about the inter-relationship between social, spatial, and transmission patterns within a population have broad applicability to other host-pathogen systems.

Other studies indicate that \textit{E. coli} is primarily environmentally transmitted through the ingestion of fecal contaminated water and forage, though there is evidence that direct transmission is important in certain situations (Besser et al. 2001; Henderson 2008; Turner et al. 2008). While giraffe do engage in some tactile contact during greetings and fights, direct contact is not a substantial feature in social interactions (Bashaw et al. 2007). Thus, the importance of social interactions in transmitting \textit{E. coli}
seems counter-intuitive at first glance. It is possible that the role of environmental transmission may be reduced in giraffe because they rarely forage on the ground (Young and Isbell 1991), and forage contamination by gastrointestinal parasites is minimized at higher feeding heights (Apió et al. 2006). Perhaps more likely, however, is that social associations lead to synchronous space use. For example, if two giraffe frequently associate, then they may drink from the same water source at the same point in time. Each would then be exposed to the same water-borne *E. coli*. Thus, frequent association may enhance spatiotemporal synchrony in exposure to environmental sources of *E. coli*.

Because of the difficulty of quantifying who transmitted to whom in wildlife populations, it has previously not been possible to link social networks with transmission networks. Instead, an individual’s social network connectivity has been used more as a risk factor for infection. For example, individuals with higher degree in the social network are more likely to be infected (Godfrey et al. 2009; Porphyre et al. 2011), or animals that engage in certain types of behavior are at higher risk (Drewe 2009). In a recent study that did utilize microbial genetics, the data were treated similarly to the risk factor analyses. Social network connectivity was analyzed as a risk factor for being infected with specific *Salmonella* strains rather than infected with *Salmonella* in general (Bull et al. 2012). They did conduct a dyad-level analysis on the likelihood of two individuals sharing *Salmonella* strains. Similar to my study, they found that strain sharing was influenced by the strength of the pair’s social relationship rather than spatial proximity. However, they did not conceptualize the strain sharing data as a distinct network and thus did not calculate individual-level connectivity metrics on transmission data (Bull et al. 2012). The approach used in my study allows me to compare contact
networks to transmission networks directly. This opens the door to new lines of research in network epidemiology, such as the extent to which animals occupy similar positions in contact and transmission networks, identification of characteristics of highly-connected super-spreaders, or the effect of environmental change on the structure and connectivity of transmission networks.

In the novel approach used here, I integrated host behavioral data with microbial genetic data to provide a detailed picture of how contact and transmission patterns are related. While not usually pathogenic, *E. coli* is a useful proxy for pathogen transmission because it allows us to study transmission pathways without waiting for a clinical epidemic or making post-hoc conclusions about transmission patterns after an epidemic has occurred. Transmission routes demonstrated by *E. coli* are most applicable to other fecal-oral pathogens that are epidemiologically similar. Though this study focused on transmission patterns of a commensal organism, this is the first study to utilize this integrative approach to construct a network for any transmissible agent in an animal population. These methods can be employed to demonstrate possible routes of transmission through an ecosystem and are broadly applicable across studies of both intra- and inter-specific routes of transmission.
Literature cited


Butts CT, 2010. sna: tools for social network analysis. R package version 2.2-0.


disease spillover in a wildlife population. Proceedings of the Royal Society Series
B 276:1777-1785.

populations: the small-world network of Serengeti lions. Journal of the Royal
Society Interface 8:776-786.

Princeton University Press.

Integrating association data and disease dynamics in a social ungulate: bovine
tuberculosis in African buffalo in the Kruger National Park. Annales Zoologici
Fennici 41:879-892.

humans and dogs: insights into within-household transmission of phylotypes
associated with urinary tract infections. Epidemiology and Infection 137:1457-
1464.

Dekker D, Krackhardt D, Snijders TAB, 2007. Sensitivity of MRQAP tests to collinearity

Mammals. New York: Springer.

Dombek PE, Johnson LK, Zimmerley ST, Sadowsky MJ, 2000. Use of repetitive DNA
sequences and the PCR to differentiate *Escherichia coli* isolates from human and


and clustering on the prevalence of infection. Journal of Theoretical Biology 254:45-54.


**Supplementary Information**

**Laboratory methods and genetic analysis**

Samples were shipped on dry ice to the UC Davis School of Veterinary Medicine. DNA was extracted from cultured cells using DNeasy Blood and Tissue kits (QIAGEN, Valencia, CA) and genetic subtypes were determined using BOX-PCR and gel electrophoresis, which is a well-established method for discriminating between genetically similar *E. coli* subtypes (Goldberg et al. 2006, Cesaris et al. 2007, Mohapatra and Mazumder 2008). BOX-PCR amplifies repetitive DNA sequences dispersed throughout the bacterial chromosome (Goldberg et al. 2006). PCR reactions were performed in 25 µL volumes containing 12.5 µL Promega MasterMix (Promega, Madison, WI), 10.5 µL water, 2 µL DNA template, and 2 µL primer BOX-AIR (5’-CTACGGCAAGGCGACGCTGACG-3’). Reaction mixtures were cycled in a model 2720 thermocycler (Applied Biosystems, Carlsbad, CA) at 94º for 5 minutes, and then for 35 cycles at 94º for 1 min, 53º for 1 min, and 72º for 1 min, followed by a 72º final extension step for 10 min and an indefinite 4ºC soak.

PCR products underwent gel electrophoresis on a 20x30cm 1.5% agarose gel made of 200 mL TAE (tris-acetate-EDTA) buffer. 10 µL of sample was mixed with 2 µL of loading dye, and 10 µL of this mixture was loaded into gel wells. Gels ran for 20 hours at 35V. Gels were pre-stained with GelRed nucleic acid stain (Phenix, Candler, NC) and photographed under UV transillumination.

Gel images were analyzed using GelCompare II (Applied Maths, Austin, TX). Densitometric curves from each lane were normalized with respect to a DNA ladder (O’geneRuler Plus 100 bp, Fermantas, Carlsbad, CA), which was included every 10 lanes.
on every gel. Regions of the gel >3000 or <400 bp were excluded from the analysis because bands in these ranges tended to be poorly resolved.

**Definitions of individual-level network metrics**

1. Degree: An unweighted measure of the number of individuals that the focal individual is connected to in a given network (Wasserman and Faust 1994). This method was used to calculate *Transmission degree*.

2. Weighted degree: For the spatial and social network, ties were weighted according to the strength of the relationship. Weighted degree takes into account not only how many individuals the focal node was connected to, but also how strength of those connections (Opsahl et al. 2010). This method was used to calculate *Spatial degree* and *Social degree*.

3. Overall tie-strength: A modification of degree for weighted networks defined for an individual as the sum of the weight of its ties rather than the total the number of ties it possesses. It does not take into account he total number of ties, as weighted degree does (Barrat et al. 2004, Opsahl et al. 2010). This method was used to calculate *Overall social tie-strength* and *Overall spatial tie-strength*.

4. Betweenness: An unweighted measure of the number of paths that pass through the focal individual if the shortest paths between every other pair of individuals are traced. Put simply, individuals with high betweenness have “high-flow” in the network because
they frequently lie on paths connecting others (Wasserman and Faust 1994). This method was used to calculate Transmission betweenness.

5. Weighted betweenness: This measure generalizes unweighted betweenness to weighted networks by modifying the definition of “shortest path.” In unweighted cases, the shortest path is simply the path between two nodes that passes through the least number of ties. In the weighted case, path length is no longer defined as the number of ties passed through, but rather the summed weight of those ties (Newman 2001, Opsahl et al. 2010). This allows for the fact that disease or information will flow through a network along paths with the least resistance, even if they are longer, because transmission is probabilistically more likely to occur across ties with higher weights (e.g. more frequent association in the social network). This method was used to calculate Social betweenness and Spatial betweenness.

6. Closeness: An unweighted measure of the inverse sum of the shortest paths between the focal node and all others (Wasserman and Faust 1994).

7. Information centrality (IC): IC is an unweighted measure calculated by taking the harmonic mean of all possible paths in the network that start from the focal node. Nodes with high IC can be considered to have greater amounts of control over information flow (Wasserman and Faust 1994).
8. Eigenvector centrality (EC): An unweighted measure corresponding to the first eigenvector of the adjacency matrix. A node’s EC is proportional to the sum of the ECs of its neighbors. Thus, EC factors in not just how many neighbors the focal node possesses, but also how connected those neighbors are (Bonacich 2007).
Supplementary literature cited


Supplementary Figure S1: BOX-PCR banding patterns of ten *E. coli* isolates corresponding with ten different giraffe. The similarity matrix to the right depicts the correlation coefficient of *i*th and *j*th individuals’ densitometric curves. Curves that were >90% similar were considered “matching” subtypes and are shaded in black in the similarity matrix.
Supplementary Table S1: Results of univariate Poisson GLM regressions examining the correlations between individual connectivity values across networks (N=194 individuals). Regression coefficients were calculated using GLMs while p-values were based on 3000 permutations of x relative to y. For any GLMs that included either Spatial or Social Betweenness as a variable, a single outlier with high influence was excluded (Social/Spatial Betweenness value >1500). (**) indicates p<0.01. (*) indicates p <0.05. (*#) indicates a non-significant trend (<0.065). P-values are two-tailed.

<table>
<thead>
<tr>
<th>Model</th>
<th>Response Variable</th>
<th>Independent Variable(s)</th>
<th>Coefficient</th>
<th>Psuedo-R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Trans. degree</td>
<td>Home range size</td>
<td>&lt;0.0001</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>2</td>
<td>Trans. degree</td>
<td>Spatial betweenness</td>
<td>0.0001</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>3</td>
<td>Trans. degree</td>
<td>Spatial degree</td>
<td>0.001</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>Trans. degree</td>
<td>Overall spatial tie-strength</td>
<td>-0.001</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>Trans. degree</td>
<td>Social betweenness</td>
<td>0.0003</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>Trans. degree</td>
<td>Social degree</td>
<td>0.016</td>
<td>* 0.59</td>
</tr>
<tr>
<td>7</td>
<td>Trans. degree</td>
<td>Overall social tie-strength</td>
<td>0.052</td>
<td>* 0.66</td>
</tr>
<tr>
<td>8</td>
<td>Trans. betweenness</td>
<td>Home range size</td>
<td>0.21</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Home range size^2</td>
<td>-0.0002</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Trans. betweenness</td>
<td>Spatial betweenness</td>
<td>0.001</td>
<td>0.95</td>
</tr>
<tr>
<td>10</td>
<td>Trans. betweenness</td>
<td>Spatial degree</td>
<td>0.139</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spatial degree^2</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Trans. betweenness</td>
<td>Spatial tie-strength</td>
<td>0.123</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spatial tie-strength^2</td>
<td>-0.002</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Trans. betweenness</td>
<td>Social betweenness</td>
<td>0.001</td>
<td>*# 0.51</td>
</tr>
<tr>
<td>13</td>
<td>Trans. betweenness</td>
<td>Social degree</td>
<td>0.005</td>
<td>0.57</td>
</tr>
<tr>
<td>14</td>
<td>Trans. betweenness</td>
<td>Social tie-strength</td>
<td>0.03</td>
<td>0.25</td>
</tr>
<tr>
<td>15</td>
<td>Social betweenness</td>
<td>Home range size</td>
<td>0.005</td>
<td>* 0.99</td>
</tr>
<tr>
<td>16</td>
<td>Social betweenness</td>
<td>Spatial betweenness</td>
<td>0.001</td>
<td>0.59</td>
</tr>
<tr>
<td>17</td>
<td>Social betweenness</td>
<td>Spatial degree</td>
<td>0.010</td>
<td>* 0.97</td>
</tr>
<tr>
<td>18</td>
<td>Social betweenness</td>
<td>Spatial tie-strength</td>
<td>0.023</td>
<td>* 0.99</td>
</tr>
<tr>
<td>19</td>
<td>Social degree</td>
<td>Home range size</td>
<td>0.002</td>
<td>** 0.16</td>
</tr>
<tr>
<td>20</td>
<td>Social degree</td>
<td>Spatial betweenness</td>
<td>0.0001</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>21</td>
<td>Social degree</td>
<td>Spatial degree</td>
<td>0.005</td>
<td>** 0.14</td>
</tr>
<tr>
<td>22</td>
<td>Social degree</td>
<td>Spatial tie-strength</td>
<td>0.014</td>
<td>** 0.31</td>
</tr>
<tr>
<td>23</td>
<td>Social tie-strength</td>
<td>Home range size</td>
<td>-0.001</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>24</td>
<td>Social tie-strength</td>
<td>Spatial betweenness</td>
<td>-0.0003</td>
<td>0.01</td>
</tr>
<tr>
<td>25</td>
<td>Social tie-strength</td>
<td>Spatial degree</td>
<td>-0.002</td>
<td>0.01</td>
</tr>
<tr>
<td>26</td>
<td>Social tie-strength</td>
<td>Spatial tie-strength</td>
<td>0.006</td>
<td>* 0.02</td>
</tr>
<tr>
<td>27</td>
<td>Spatial betweenness</td>
<td>Home range size</td>
<td>0.012</td>
<td>** 0.99</td>
</tr>
<tr>
<td>28</td>
<td>Spatial degree</td>
<td>Home range size</td>
<td>0.015</td>
<td>** 0.99</td>
</tr>
<tr>
<td>29</td>
<td>Spatial tie-strength</td>
<td>Home range size</td>
<td>0.015</td>
<td>** 0.99</td>
</tr>
</tbody>
</table>
Chapter 4: Heterogeneity in transmission network connectivity across species: 
Implications for pathogen management and control

Abstract

Multi-host wildlife pathogens are an increasing concern for conservation and livestock husbandry. Although social network analysis has been lauded as a key new tool in wildlife epidemiology, few studies have combined a network approach with real data on pathogen transmission because of limitations in detecting the occurrence of transmission. Here, I combine a network approach with microbial genetics to assess routes of interspecific transmission between wild and domestic ungulates in Kenya. I use the genetics of a diverse microbe, *Escherichia coli*, to reveal where transmission has occurred. If two individuals share the same genetic subtype of *E. coli*, then I can infer that these individuals are part of the same transmission chain and these individuals are interlinked in a transmission network. Given interspecific variation in physiology and behavior, some species may function as “super-spreaders” if individuals of that species are consistently central in the transmission network. Management strategies targeted at key super-spreader species are theoretically more effective at limiting disease spread than conventional strategies, and my methods provide a means to identify candidate super-spreaders in wild populations. I found that Grant’s gazelle typically occupy central positions in the network and are connected to a large number of other individuals in the
network. Zebra, in contrast, seem to function as bridges between regions of the network that would otherwise be relatively separated. Their removal increases the level of fragmentation in the network. Although not usually pathogenic, \textit{E. coli} transmission routes give insight into transmission dynamics by demonstrating where contact between species is sufficient for transmission to occur and which species are potential super-spreaders.

**Introduction**

Understanding the dynamics of pathogen transmission is important for predicting the potential impact of wildlife diseases and developing disease control strategies. Approximately 77\% of livestock pathogens are multi-host (Cleaveland et al. 2001), and pathogens shared among livestock and wild ungulates may have adverse effects on both populations. Indeed, most endangered species at risk from disease acquire their pathogens from domestic populations (Altizer et al. 2003). Pathogen transmission is of particular concern in sub-Saharan Africa because of the close proximity of wildlife to livestock and the high prevalence and diversity of pathogens (Cleaveland et al. 2005; Wambwa 2005). Current data are inconclusive about risks associated with wildlife-livestock disease transmission, and better data on the dynamics of interspecific transmission are urgently needed.

Critical questions concerning pathogen transmission have largely gone unanswered because the field is limited by available methodology. It is difficult to study transmission routes in wild populations using current methods, such as commonly-used serological techniques, because data on who transmitted an infection to whom is almost
impossible to obtain (Caley et al. 2009). However, such data can be obtained by assessing the genetics of the pathogen itself: if two individuals share similar genetic subtypes of a pathogen, then transmission can be inferred (Goldberg et al. 2007; Archie et al. 2008). Here, I use genetic data to infer transmission patterns in a community of African herbivores, and use these data to construct interspecific transmission networks to further our understanding of how pathogens spread through multi-species communities.

Theoretical epidemiological models often assume that the probability of contact is equal for every pair of individuals in the population, even though spatial and social structure create heterogeneity in transmission patterns (Keeling and Eames 2005; Bansal et al. 2007; Otterstatter and Thomson 2007; Perkins et al. 2008; Craft and Caillaud 2011). Incorporating contact networks into epidemiological models provides one mechanism to account for such heterogeneity (Keeling 2005; May 2006; Bansal et al. 2007; Craft and Caillaud 2011). As compared to traditional mass-action models, these models tend to result in reductions in the early growth rate, number of secondary infections for each infected individual, and final size of an epidemic (Keeling and Eames 2005; Turner et al. 2008). Thus, failing to account for heterogeneity reduces our ability to understand and predict the spread of infectious diseases (Keeling 1999; Keeling and Eames 2005; Ames et al. 2011). However, the structure of transmission networks in wildlife is relatively unknown and difficult to quantify.

Empirical transmission networks in wildlife are often constructed based on interactions between individuals, which may be quantified through direct observation, proximity-logging collars, or records of shared space use (Corner et al. 2003; Otterstatter and Thomson 2007; Drewe 2009; Godfrey et al. 2009; Hamede et al. 2009; Porphyre et
Limitations in detecting the occurrence of transmission have meant that conclusions about disease spread are based on the possibility that transmission could occur between interacting individuals (Corner et al. 2003; Otterstatter and Thomson 2007; Perkins et al. 2008; Böhm et al. 2009; Craft et al. 2009; Drewe 2009; Godfrey et al. 2009; Grear et al. 2009; Hamede et al. 2009; Perkins et al. 2009). In contrast, I reveal where transmission has already occurred using the genetics of a diverse microbe, *Escherichia coli*, allowing me to construct a transmission network based on quantifiable transmission events. Individuals that share genetically similar subtypes of *E. coli* are inferred to be a part of the same chain of transmission, either though direct transmission through interactions or indirect transmission due to exposure to a common environmental source. Individuals are interlinked in the transmission network based upon patterns of *E. coli* subtype sharing.

While not usually pathogenic, *E. coli* is a useful proxy for pathogen transmission because it allows us to study transmission pathways without waiting for a true epidemic or making post-hoc conclusions after it has occurred. Because of its immense genetic diversity, subtyping is a common method for tracing *E. coli* to its source (Simpson et al. 2002); *E. coli* subtype sharing has been used to demonstrate that transmission regularly occurs between humans and their pets, and between humans, livestock, and wild primates (Goldberg et al. 2008; Johnson et al. 2008; Damborg et al. 2009). Recently, the emergence of highly virulent forms of *E. coli* have become a significant public health concern (Beutin 2006). For these reasons, *E. coli* is an excellent model organism for examining enteric pathogen transmission pathways in wildlife.
The epidemiological term “super-spreader” refers to individuals who are disproportionately involved in transmission due to either high pathogen shedding rates or high levels of sociality (Lloyd-Smith et al. 2005). Given interspecific variation in physiology and behavior, some species may function as super-spreaders if individuals of that species are consistently central in the transmission network. Although control strategies targeted at super-spreaders are potentially a very effective tool for managing disease (Woolhouse et al. 1997; Craft and Caillaud 2011), such strategies are not currently feasible because logistical and diagnostic limitations make it difficult to identify super-spreaders. These limitations necessitate the development of new tools (Cross et al. 2009). Here, I combine microbial genetics with network analyses to examine heterogeneity in transmission patterns in a community of East African ungulates.

Methods

Study site and species

This study was conducted in Ol Pejeta Conservancy (OPC), a 364 km² wildlife reserve in central Kenya that integrates commercial cattle ranching with wildlife conservation. OPC is a fenced, semi-arid savanna ecosystem located on the equator (0° N, 36°56’ E). The reserve is a Acacia-woodland/grassland mosaic and receives on average 900 mm of rainfall per year (Birkett 2002). Species included in this study are the African buffalo (*Syncerus caffer*), eland (*Taurotragus oryx*), Grant’s gazelle (*Gazella granti*), Thomson’s gazelle (*G. thomsonii*, status: Near threatened), reticulated giraffe (*Giraffa camelopardalis reticulata*, sub-species status: Lower risk – conservation dependent), Jackson’s hartebeest (*Alcelaphus buselaphus jacksonii*, sub-species status: Endangered),
impala (*Aepyceros melapus*), black rhinoceros (*Diceros bicornis*, status: Critically endangered), plains zebra (*Equus burchelli*), and domestic cattle (*Bos indicus*). These species account for 99% of ungulate biomass and 97% of the ungulate population in OPC (OPC Ecological Monitoring department, unpublished data).

Because dietary niche is likely to affect both pathogen exposure and the micro-environment within the gut (Dehority and Odenyo 2003; Apio et al. 2006), study species included three browsers, four grazers, and three mixed feeders whose diets consist of a combination of browse and grass (Table 1). Species whose diets consist of >70% grass were defined as grazers, 30-70% grass as mixed feeders, and <30% grass as browsers. Values for diet composition were taken from the literature (Spinage et al. 1980; Hoffmann 1989; Gagnon and Chew 2000; Watson and Owen-Smith 2000; Cerling et al. 2003; Sponheimer et al. 2003; Codron et al. 2007; Copeland et al. 2009). Digestive physiology may also have an effect on the types of *E. coli* fostered in the gut, and studies of *E. coli* population genetics demonstrate that host order is among the most important factors differentiating *E. coli* (Souza et al. 1999). This set of species includes eight ruminants (Order Artiodactyla) and two non-ruminants (Order Perissodactyla, Table 1). The inclusion of buffalo is especially relevant to livestock owners because they are closely related to cattle and are considered a major reservoir for both Foot and Mouth Disease and bovine tuberculosis (Kock 2005). The black rhinoceros is endangered and vulnerable to disease-related population declines (Emslie and Brooks 1999).
Table 1. Sample size, population size, home range size (HR), and other attributes of study species.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Pop. Size</th>
<th>Order</th>
<th>Family</th>
<th>Sub-family</th>
<th>Digest</th>
<th>Forage Niche</th>
<th>HR - km²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Buffalo</strong></td>
<td>29</td>
<td>1200</td>
<td>Artiodactyla</td>
<td>Bovidae</td>
<td>Bovinae</td>
<td>Rumen</td>
<td>Grazer</td>
<td>60a</td>
</tr>
<tr>
<td><strong>Cattle</strong></td>
<td>30</td>
<td>6500</td>
<td>Artiodactyla</td>
<td>Bovidae</td>
<td>Bovinae</td>
<td>Rumen</td>
<td>Grazer</td>
<td>6b</td>
</tr>
<tr>
<td><strong>Eland</strong></td>
<td>31</td>
<td>400</td>
<td>Artiodactyla</td>
<td>Bovidae</td>
<td>Bovinae</td>
<td>Rumen</td>
<td>Browser</td>
<td>38a</td>
</tr>
<tr>
<td><strong>Grant's gazelle</strong></td>
<td>17</td>
<td>900</td>
<td>Artiodactyla</td>
<td>Bovidae</td>
<td>Antilopinae</td>
<td>Rumen</td>
<td>Mixed</td>
<td>6c</td>
</tr>
<tr>
<td><strong>Thomson's gazelle</strong></td>
<td>32</td>
<td>1400</td>
<td>Artiodactyla</td>
<td>Bovidae</td>
<td>Antilopinae</td>
<td>Rumen</td>
<td>Mixed</td>
<td>3d</td>
</tr>
<tr>
<td><strong>Hartebeest</strong></td>
<td>31</td>
<td>140</td>
<td>Artiodactyla</td>
<td>Bovidae</td>
<td>Alcelaphinae</td>
<td>Rumen</td>
<td>Mixed</td>
<td>5e</td>
</tr>
<tr>
<td><strong>Impala</strong></td>
<td>32</td>
<td>3600</td>
<td>Artiodactyla</td>
<td>Bovidae</td>
<td>Aepycerotinae</td>
<td>Rumen</td>
<td>Mixed</td>
<td>2e</td>
</tr>
<tr>
<td><strong>Giraffe</strong></td>
<td>32</td>
<td>200</td>
<td>Artiodactyla</td>
<td>Giraffidae</td>
<td>n/a</td>
<td>Rumen</td>
<td>Browser</td>
<td>73f</td>
</tr>
<tr>
<td><strong>Black rhino</strong></td>
<td>14</td>
<td>90</td>
<td>Perissodactyla</td>
<td>Rhinocerotidae</td>
<td>n/a</td>
<td>Hindgut</td>
<td>Browser</td>
<td>12e</td>
</tr>
<tr>
<td><strong>Plains Zebra</strong></td>
<td>31</td>
<td>4500</td>
<td>Perissodactyla</td>
<td>Equidae</td>
<td>n/a</td>
<td>Hindgut</td>
<td>Grazer</td>
<td>120a</td>
</tr>
</tbody>
</table>

*(Jones et al. 2009); *(OPC Cattle Department, personal communication); *(Vanessa Ezenwa, unpublished data); *(Walther 1973); *(OPC Ecological Monitoring Department, unpublished data); *(this dissertation, Chapter 2).  

Synchronized exposure to environmental sources likely increases the probability of two individuals to share subtypes (Chapter 2). *E. coli* can persist in natural water sources for months, but densities rapidly drop off in the first two weeks (Medema et al. 1997; Avery et al. 2008). Inactivation by sunlight further reduces the populations (Sinton et al. 2007). In soil, *E. coli* can persist eight to 25 weeks (Habteselassie et al. 2008) and can even increase over two-week time periods in all but the driest soils (Berry and Miller 2005), but survival is reduced by low moisture content and warm temperatures (Habteselassie et al. 2008). It is also possible for flies to function as mechanical vectors and transmit *E. coli* when they move from one animal to another (Ahmad et al. 2007; Förster et al. 2007; Fetene and Worku 2009).
Interspecific associations

To quantify the extent to which each species aggregated with other species, I recorded the proximity of each species to other species while driving pre-determined road transects through OPC. Transects were designed so the majority of the Conservancy was surveyed once every three days (N = 2143 observations). Observations of interspecific association patterns were recorded between March 17 and August 2, 2011. Association was defined as the percentage of observations that species x was observed within 50 m of species y relative to the total number of times those species were observed together or apart.

Sample collection, genetic analysis and network construction

We collected a total of 279 fecal samples from ten species (Table 1). Sample collection was stratified across ten spatial blocks in the study area. Because there may be significant monthly turnover of E. coli subtypes in the gut (Anderson et al. 2006), fecal samples were collected during the brief period between August 28, 2011 and October 7, 2011 to ensure comparability. A few giraffe samples were collected as early as August 14. Black rhino samples were collected from the rectum during routine immobilizations conducted by the Kenya Wildlife Service. Fecal samples for other species were collected from the ground immediately after defecation was observed and transported on ice to the field laboratory. Samples were then diluted in sterilized water, streaked onto CHROMagar EC agar plates (CHROMagar, Paris France), and incubated overnight at 37° C. CHROMagar is a selective chromogenic agar that exhibits high specificity for E. coli. After incubation, four randomly selected E. coli colony isolates were cultured and then
frozen. Using data from captive giraffe, which hosted approximately two subtypes of *E. coli* per individual (unpublished data), I calculated that there was a <10% probability of failing to capture subtype diversity existing in the gut if four isolates were taken (Singer et al. 2000).

Samples were shipped on dry ice to the UC Davis School of Veterinary Medicine. DNA was extracted from cultured cells using QIAGEN DNeasy Blood and Tissue kits (QIAGEN, Valencia, CA) and genetic subtypes were determined using BOX-PCR and gel electrophoresis, which is a well-established method for discriminating between genetically similar *E. coli* subtypes (Goldberg et al. 2006; Cesaris et al. 2007; Mohapatra and Mazumder 2008). BOX-PCR amplifies repetitive DNA sequences dispersed throughout the bacterial chromosome (Goldberg et al. 2006). PCR reactions were performed in 25 µL volumes containing 12.5 µL Promega MasterMix (Promega, Madison, WI), 10.5 µL water, 2 µL DNA template, and 2 µL primer BOX-AIR (5’-CTACGGCAAGGCGACGCTGACG-3’). Reaction mixtures were cycled in a model 2720 thermocycler (Applied Biosystems, Carlsbad, CA) at 94º for 5 minutes, and then for 35 cycles at 94º for 1 min, 53º for 1 min, and 72º for 1 min, followed by a 72º final extension step for 10 min and an indefinite 4ºC soak. PCR products underwent gel electrophoresis on a 20x30cm 1.5% agarose gel made of 200 mL TAE (tris-acetate-EDTA) buffer. 10 µL of sample was mixed with 2 µL of loading die, and 10 µL of this mixture was loaded into gel wells. Gels ran for 19 hours at 35V. Gels were pre-stained with GelRed (Phenix, Candler, NC) and photographed under UV transillumination.

Images were analyzed using GelCompare II (Applied Maths, Austin, TX). Densitometric curves from each lane were normalized with respect to a DNA ladder.
(O’geneRuler Plus 100 bp, Fermantas, Carlsbad, CA). Regions of the gel >3000 bp or <400 bp were excluded from the analysis because bands in these ranges tended to be poorly resolved and irreproducible. Similarity of each isolate to all others was determined through pairwise comparisons of densitometric curves (see Supplementary Figure S1 in Chapter 3). Isolates were considered to be matching if their densitometric curves were >90% similar (Pearson’s correlation coefficient). Based on a reproducibility analysis conducted in my lab, this cutoff value minimizes Type I errors in matching while limiting the Type II error rate to <5%. A transmission network was constructed from patterns of *E. coli* subtype sharing. Sampled individuals were represented as nodes and nodes were linked if they shared at least one *E. coli* subtype.

**Dyad-level analysis: What factors influence the likelihood of two individuals sharing *E. coli* subtypes?**

To determine which factors made the likelihood of a transmission link between dyads more or less likely, I performed generalized linear mixed models (GLMMs) with binomial distributions. The outcome variable was if a pair of individuals were linked (1) or unlinked (0) in the transmission network. Random effects were included for each individual in the dyad to account for the fact that individuals occurred in many dyads. Dyad-level covariates included whether the individuals in the dyad were from the same species, sub-family, or family, whether they exhibited the same foraging niche (browser, mixed feeder, or grazer) or digestive system (rumination or hindgut fermentation), and whether they were sampled from the same spatial block. Taxonomic order was not considered because it covaried perfectly with digestive system: all Perissodactylys are
hindgut fermenters and all Artiodactyls in this study are ruminants. Models were compared using Akaike’s Information Criterion (AIC).

We did not use MR-QAP here because it does not allow for interactions. In the giraffe analysis, it was important to use the MR-QAP because autocorrelations in y (transmission links) were expected to relate to autocorrelations in x (e.g. more social individuals that had higher AS were also expected to have more links). Thus, treating individual as a random effect would not allow me to examine relationships between autocorrelations in x and y because it removes the autocorrelation of y by accounting for that variation through random effect. Unlike the giraffe analysis, I do not expect autocorrelations in x to impact autocorrelations in y. For example, individuals do not vary in the extent to which they are part of same-species dyads. Thus, autocorrelations in x and y are not expected to covary. For this analysis, using a random effect to account for autocorrelations in y should not hinder us from looking at the relationship of same-species and the occurrence of transmission links.

Species-level comparisons

We calculated four measures of connectivity for each individual (Wasserman and Faust 1994). “Degree” is the number of individuals, or neighbors, to which the focal node was connected. “Information centrality” essentially measures the distance of the focal node to all others, providing a measure of the extent to which an animal is located at the center of the network. It is calculated by taking the harmonic mean of all possible paths in the network that originate from the focal node. Nodes with high information centrality
can be considered to have greater amounts of control over flow in the network (Wasserman and Faust 1994).

We calculated an individuals’ “betweenness” using Newman’s (Newman 2005) random-walk definition. Betweenness is defined as the number of paths that pass through the focal individual if all paths between every other pair of individuals are traced (Newman 2005). Individuals can have high betweenness (i.e. large number of paths pass through them) either because they are connected to a large number of other individuals or because they lie on paths that are bottlenecks for flow. The former leads to high correlations between betweenness and degree (r = 0.57). The latter is more relevant for disease management because these individuals are potential “cutpoints” in the network; removal of individuals that serve as bottlenecks may divide the network into multiple disconnected sub-networks. I developed a measure to disentangle betweenness from its correlations with degree. A regression line was fitted that related betweenness to degree. From this regression equation, I calculated each individual’s expected betweenness given its degree. A residual was then calculated by subtracting observed betweenness from expected betweenness. I refer to this residual as an individual’s “cutpoint potential.” Positive cutpoint potentials indicate individuals that are candidate bottlenecks for pathogen flow in the network. Degree and information centrality were calculated using R’s “sna” package (Butts 2010), and random-walk betweenness was calculated using NetMiner (NetMiner 2.6, Cyram Corporation, Seoul, Korea). All statistical analyses were performed using R.

Species connectivity measures (degree, information centrality, betweenness, cutpoint potential) were compared using Kruskal-Wallis tests followed by pairwise
comparisons with a Bonferroni correction. I hypothesized that cutpoint potential would be related to species-typical home range size because animals with large home ranges may connect different regions in the study area. Therefore, I regressed cutpoint potential (ranked) on species-typical home range size (km$^2$), and used permutation methods to calculate p-values. This method first calculates the slope for the regression using ordinary least squares, then it recalculates the slope for 5000 iterations in which y is randomly permuted relative to x. P-values are defined as the proportion of random permutations that have slopes more extreme than the observed value (Good 2000). Ranked values were used because permutation methods are susceptible to outliers (Croft et al. 2008).

To assess whether differences in species connectivity were an artifact of the fact that I sampled the same number of individuals (~30) from species of vastly different population sizes (e.g. 85 rhinos versus 6500 cattle), I constructed five large 4700 node random Bernoulli networks with the same density as the observed transmission network and species represented proportionally to their abundance in the ungulate community. From each of these networks, I randomly drew nodes with each species’ sample size proportional to my real-life sampling strategy. Using these randomly sampled individuals, I constructed a sub-network. For each random graph, I performed 200 sampling iterations for a total of 1000 sub-networks. These sub-networks were considered the null expectations for connectivity patterns produced from a random network. Connectivity patterns were compared for the random sub-networks and observed network. All species, regardless of how well sampled they were relative to their population size, were not shown to differ in their connectivity in any of the four measures in the random networks.
We also analyzed the effect of the removal of certain species on the overall connectivity of the transmission network. I summarized overall network connectivity using two metrics that are theoretically highly important in determining pathogen spread (Keeling 1999; Newman 2003; Turner et al. 2008; Wu and Liu 2008; Badham and Stocker 2010; Ames et al. 2011). Density is proportion of ties that occur in the network relative to the total possible number of ties. Transitivity is defined as the number of triangles in the network (A is linked to B, B is linked to C, and C is also linked to A) relative to the number of triplets (e.g. A is linked to B, B is linked to C, but C is not linked to A). Networks characterized by lower density or higher transitivity are theoretically predicted to exhibit reduced transmission (Keeling 2005; Ames et al. 2011).

I calculated the change in density ($\Delta$-density) and transitivity ($\Delta$-transitivity) when individuals of a given species were removed. The observed $\Delta$-values were compared to a permuted distribution of $\Delta$-values. This distribution was generated by removing an equal number of random individuals and calculating the resulting $\Delta$-values. P-values were calculated as the percentage of permuted $\Delta$-values that were more extreme than the observed $\Delta$-value. A p-value of $< 0.05$ indicates that the removal of a given species produces significantly greater change in network connectivity than removing an equal number of random individuals.

**Sampling effort and network robustness**

We examined how robust the species-level comparisons were relative to sampling effort. To assess the robustness of the results if only 90% of the network had been sampled, I randomly deleted 10% of individuals in each species. I then re-calculated two
of the node-level measures (betweenness and degree) on the resulting networks and re-ranked each species according to their mean connectivity. I calculated the extent to which the rankings in the sub-sample correlated with the rankings in the full network (Spearman’s rank correlation). I repeated this sub-sampling 100 times to generate a distribution of Spearman’s rank correlations produced by a 90% sampling effort. This process was repeated at 80%, 70%, 60%, 50%, and 40% sampling effort.

Results

What factors influence the likelihood of two individuals sharing E. coli subtypes?

The density of the full transmission network was 0.16, indicating that 16% of possible links existed in the network. The likelihood of a transmission link forming between any pair of individuals was highly dependent on host relatedness, with conspecific links being approximately 1.5 times more likely to occur than heterospecific links (Table 2). Despite having significant positive coefficients in univariate models, family and sub-family were not included in multivariate models due to the nested nature of the taxonomic covariates; dyads that are the same species are always same sub-family, which are always in the same family. Although digestive system is equivalent to host order, I kept digestive system in the multivariate models because it espouses an important physiological as well as taxonomic difference between hosts. Pairs of individuals with the same digestive physiology were approximately 1.4 times more likely to share E. coli subtypes than pairs with different physiologies. The effect of sharing a foraging niche also had a small positive effect, but multivariate models that included foraging niche did not perform better than models that did not (Table 2). Dyads were more likely to be
linked if they were from the same sampling region. This effect of region was amplified if they were also the same species. When the dataset was truncated to include only interspecific dyads, individuals of species that associated more frequently were more likely to share *E. coli* subtypes ($\beta = 0.013$, $p = 0.01$).

**Table 2.** Best-fit models for the probability (log-odds) that two animals are linked in the transmission network. Coefficients are exponentiated and reported as odds ratios. (*) indicate relationships that were significant at $p < 0.05$.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Multivariate models</th>
<th>Univariate models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
</tr>
<tr>
<td>Same Species</td>
<td>1.42*</td>
<td>1.48*</td>
</tr>
<tr>
<td>Same Region</td>
<td>1.14*</td>
<td>1.14*</td>
</tr>
<tr>
<td>Same Foraging Niche</td>
<td>1.06*</td>
<td>1.04</td>
</tr>
<tr>
<td>Same Digestion</td>
<td>1.35*</td>
<td>1.34*</td>
</tr>
<tr>
<td>Same Family</td>
<td>1.06*</td>
<td>1.04</td>
</tr>
<tr>
<td>Same Subfamily</td>
<td>1.26*</td>
<td>1.26*</td>
</tr>
<tr>
<td>Species*Region</td>
<td>1.27*</td>
<td>1.27*</td>
</tr>
<tr>
<td>Delta AIC</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**How does network connectivity vary across species?**

There were significant differences among species for degree, information centrality, and betweenness (Kruskal-Wallis tests, $p < 0.05$), and near significant differences for cutpoint potential ($p = 0.088$). Pairwise comparisons revealed that Grant’s gazelles were consistently the most well-connected in all measures except cutpoint potential (Figure 1). Grant’s gazelle information centrality and degree were significantly higher than all species except hartebeest. Buffalo and cattle consistently among the least connected in all measures except cutpoint potential. Cutpoint potential was positively correlated with species-typical home range size ($\beta = 0.36$, $p < 0.01$), while other network metrics were unrelated to home range. These differences between species were not an
artifact of the fact that I sampled the same number of individuals (~30) from species of
different population sizes. When observed connectivity patterns were compared to the
randomized sub-networks, species were not predicted to differ in their connectivity in the
random networks, regardless of how well sampled they were relative to their population
size.

**Figure 1.** Bar graphs summarizing species connectivity values for (a) degree, (b)
information centrality, (c) betweenness, and (d) cutpoint potential. Ranked values are
represented and used in tests because of the highly non-normal distributions produced by
some connectivity values. Axis labels indicate species: B-Buffalo, C-Cattle, TG-
Thomson’s gazelle, R-Black rhino, G-Giraffe, I-Impala, Z-Zebra, E-Eland, H-Hartebeest,
and GG-Grant’s gazelle. Species in the same letter group (lower-case) were not
significantly different in pairwise comparisons.
What are the implications of “targeted interventions” on network structure?

Targeting highly connected species during disease interventions may be very effective because individuals of those species are on average better connected in the network. Thus, their removal has greater consequences for reducing the connectivity of the network that choosing an individual at random. The only species whose removal altered network density was the Grant’s gazelle ($\Delta$-density = -0.01, $p = 0.04$) and hartebeest ($\Delta$-density = -0.02, $p < 0.01$). These species are also the two highest ranked species for degree (Figure 1). Their removal reduced network density significantly more than the removal of an equal number of random animals. To illustrate the implications of removing certain species, I constructed a transmission network that removed the two species ranking the most highly in degree (Figure 2b).

In comparison, I also constructed a network that removed the two species with the highest cutpoint potential (zebra and buffalo, Figure 2c). Because betweenness is highly dependent on tracing paths through the network, betweenness values may be drastically altered after the removal of one species. Therefore, I first removed the species with the highest cutpoint potential (zebra) and then re-calculated betweenness and cutpoint potentials on the resulting network. In this network, buffalo had the highest cutpoint potential. The removal of these two species together produced networks that were significantly more clustered (higher transitivity) than the removal of an equal number of random individuals ($\Delta$-transitivity = 0.03, $p = 0.05$). The removal of the two highest-ranked species in cutpoint potential without the re-calculation step (eland and zebra) did not produce more transitive networks, highlighting the importance of re-calculating path-based measures after the removal of the first species.
Figure 2. Transmission networks (a) inclusive of all species. For visualization purposes, only 100 randomly selected nodes are shown. Transmission network with (b) the two species with the highest degree removed (Grant’s gazelle and hartebeest), and (c) the two species with the highest cutpoint potential removed (zebra and buffalo).
Sampling effort and network robustness

Species rank-order was relatively conserved when sampling effort was reduced (Supplementary Figure S1). Degree was more robust than betweenness. Sampling effort could be as low as 60% and sub-networks still maintained high correlations in species rank-order when compared to the full network (Spearman’s rank correlation, $r > 0.9$). Spearman’s rank correlations for betweenness, in contrast, fell much more rapidly with reduced sampling effort. Sampling effort of >70% allowed correlations to remain above 0.8, but even 90% sampling effort failed to produce correlations of greater than 0.9. However, if our aim is to identify the most connected species, rank-order switches may not be that relevant as long as the top few ranks are unaffected. More than 90% of sub-samples had the same two species ranked the highest for betweenness when sampling effort was set at 90%.

Discussion

Species-level variation in transmission network connectivity

Our transmission network quantifies patterns of interspecific transmission of a fecal-orally transmitted pathogen in wildlife. *E. coli* transmission routes give insight into transmission dynamics by demonstrating where contact between species is sufficient for transmission to occur and which species are potential super-spreaders. There were significant differences among species in their connectivity in the transmission network. While interspecific differences in connectivity are unsurprising, we previously lacked
tools to quantify this heterogeneity despite the theoretical importance of such heterogeneity in pathogen spread.

The Grant’s gazelle is one candidate super-spreader species. Individuals of this species not only share transmission links with a large number of others (high degree), but also tend to occupy central positions in the network with high flow (high information centrality and betweenness). They also are among the species most frequently found in multi-species aggregations (Supplementary Table S1). Ezenwa (2003) found that Grant’s gazelles tended have higher prevalence and loads of strongyle nematodes, another fecal-oral generalist pathogen, as compared to sympatric ungulates. Furthermore, they shed significantly more strongyle eggs per gram of feces, and their shedding intensities were well more than double that of nearly all other species (Ezenwa 2003). I was unable to determine directionality in transmission using my methods and thus cannot distinguish whether an individual is a frequent transmitter or frequent recipient of *E. coli*. Nonetheless, my results indicate that highly connected species, such as the Grant’s gazelle, may play an integral role in fecal-oral transmission chains.

The plains zebra ranked the highest in cutpoint potential and second highest in betweenness, though not significantly. Betweenness, put simply, quantifies the extent to which an individual serves as a conduit for pathogen flow through a network (Wey et al. 2008). Because paths through the network are more likely to pass through nodes with high degree, betweenness does not allow us to differentiate between individuals that have high betweenness because they have high degree, such as the case for Grant’s gazelle, or because they lie along paths that are bottlenecks for flow. Cutpoint potential, in contrast, quantifies the extent to which individuals are potential bottlenecks. Individuals that lie on
a larger number of paths than expected given their degree are scored more highly for this measure, and the removal of such individuals has higher potential for fragmenting the transmission network. Here, I found that cutpoint potential was significantly correlated with species-typical home range size. Zebra may have high cutpoint potential because they are the most widely ranging species, connecting geographically separated clusters of animals. The literature-reported home range size for zebras is 120 km$^2$ (Jones et al. 2009). This is 60% larger than the second widest ranging species giraffe, and 100% larger than the third most widely ranging species (Table 1). Interestingly, hartebeest and impala had the smallest cutpoint potential and also have among the smallest home ranges (5 and 2 km$^2$ for hartebeest and impala, respectively). Thus, zebras serve as "bridges" between regions of the network that would otherwise be relatively separated. Animals such as impala and hartebeest, whose home ranges are much more localized, tend not to function as bridges.

Interestingly, livestock were among the least well-connected species (Figure 1). Although my results indicate that cattle are not particularly important disseminators of fecal-oral pathogens within this ecosystem, cattle remain an important factor in the emergence of wildlife diseases due to their role in introducing novel diseases into ecosystems and transporting diseases between geographically-distinct wildlife populations.

Factors influencing the likelihood of two individuals sharing E. coli subtypes

Individuals from the same sampling region were more likely to be connected in the transmission network, which indicates that transmission patterns are spatially
structured. Animals that are utilizing the same area are more likely to be in proximity to each other, creating opportunities for transmission. Perhaps more importantly, they are likely exposed to the same environmental sources of *E. coli*. Individuals of the same species sampled from the same region are even more likely to be linked, a pattern that could arise through social interactions or because of similarity in physiology.

*E. coli* subtypes were far more likely to be shared in same-species dyads than in interspecific dyads (Table 2). When only interspecific dyads were considered, however, individuals from species that associated more frequently were more likely to share *E. coli*. Indeed, species that associated 15% of the time were 1.2 more likely to share *E. coli* that species that very rarely associated.

The fact that individuals sharing the same digestive physiology (or host taxonomic order) were more likely to share *E. coli* subtypes suggests that physiological similarity plays an important role in determining patterns of subtype sharing. In this ecosystem, species with the same digestive physiology did not have any more opportunities for transmission to occur. Indeed, host order did not covary with foraging niche (Table 1), nor did species within the same order associate more frequently (Table A.1). Other studies have similarly shown that host order and digestive physiology are among the important factors for genetically differentiating *E. coli* (Souza et al. 1999). Digestive physiology, taxonomic grouping, and host diet are the primary factors influencing the internal environment and growth substrate (sugars) present within the gut. Subtypes that are well adapted for one host species may not be competitive in hosts offering different internal environments. Part of the reason for this is that different *E. coli* strains vary in the numbers and kinds of sugars they utilize (Souza et al. 1999). Thus,
even if species of different orders ingest the same *E. coli* subtypes from the environment, not all subtypes are equally likely to establish themselves in the herbivore’s gut.

In addition to digestive physiology, I found that dyads sharing the same foraging niche were slightly more likely to share subtypes of *E. coli*. This could arise through either similarity in exposure or selective establishment of *E. coli* subtypes based on similar sugar profiles within the gut. Animals with mixed diets, such as the Grant’s gazelle, are predicted to support both the highest and most diverse microbial populations. As the proportion of browse in the diet increases, there is more readily available energy for microbial growth. Past a certain point, however, the decline in gut pH and shorter gut retention times that accompanies high-browse diets leads to reduced growth and diversity (Dehority and Odenyo 2003). While the high connectivity of the Grant’s gazelle is consistent with the idea that mixed feeders house greater diversity, there was no overall trend for mixed feeders to be more connected in the transmission network. Nelson (2003) similarly found no clear divergence in microbial flora by feeding ecology in a phylogenetic analysis of microbial communities.

*Implications for management*

While I recognize that *E. coli* is generally not a disease that needs management (at least for now), my main intent in this section is to show the potential utility of an approach combining network theory with microbial genetics and to outline possible courses of analysis that could be beneficial to disease management. This approach could be readily applied to other pathogens to quantify heterogeneity in transmission patterns. Models suggest that control strategies targeted at super-spreaders have the potential to
limit disease spread much more effectively than conventional methods, such as mass vaccination (Lloyd-Smith et al. 2005), but the implementation of targeted interventions in wildlife is limited by our inability to identify super-spreaders (Dalahay et al. 2009). The development of methods to identify super-spreaders, such as the one demonstrated here, is the first step in making targeted intervention strategies feasible. The advantage of focusing on species-level rather than individual-level differences in network connectivity is that it is not necessary to quantify each unique individual’s network position in every new population, which is often nearly impossible due to high turnover of individuals, large population sizes, and the intensity of monitoring necessary to acquire such data. Rather, knowledge of the characteristic role that individuals of a species play in the transmission network allows results to be more generalizable across populations and over time.

We examined the effect of “removing” key species from the transmission network to demonstrate the importance of targeted interventions. I do not mean that these animals would be culled, but rather that they are removed from transmission chains through some measure of intervention (vaccination, treatment, etc). Targeted removal of the species with the highest degree (Grant’s gazelle and hartebeest) reduces the density of the transmission network. However, it does little to fragment the network, primarily because their removal merely eliminates some individuals within dense clusters without affecting many between-cluster paths (Figure 2b). In contrast, the removal of the species with the highest cutpoint potential increases the overall level of transitivity in the network (Figure 2c). Transitivity can be interpreted as the extent to which a network is clustered. High levels of clustering are correlated with reduced ability for a pathogen to spread in a
population (Keeling 1999; Newman 2003; Turner et al. 2008; Wu and Liu 2008; Badham and Stocker 2010; Ames et al. 2011; VanderWaal et al. Submitted). Lower density is also associated with slower spread of disease (Otterstatter and Thomson 2007; Ames et al. 2011), although the differences in density and transitivity observed here are not large. Combining transmission networks with quantitative epidemiological models would help determine the relative merits of reducing density versus increasing transitivity for disease control. This would be a fruitful next step in translating these sorts of data into management strategies.

Transmission networks provide different insights than other genetic approaches. Population genetic studies tend to focus on the gene flow between pathogen metapopulations found in different host species/populations (Brown et al. 2008; Rwego et al. 2008). Phylogenetic approaches study the evolutionary relationships among genetic lineages, inferring historical host-switching events or evolutionary relationships on large geographic scales (Wallace et al. 2007; Liu et al. 2010). Haplotype networks, for example, focus on the evolutionary relationships between subtypes as measured by mutations (e.g., Beja-Pereira et al. 2009; Archie and Ezenwa 2011), whereas my approach focuses on connectivity between individuals as measured by shared subtypes. I do not make any assumptions on the evolutionary relatedness of various E. coli subtypes. It may be possible to use relatedness data to weight network edges; if subtypes were more closely related, links would receive higher weight. To implement this, however, additional assumptions must be made about the rate at which mutations occur. Evolutionary changes may not be occurring on a time scale that is epidemiological relevant for inter-individual transmission chains.
Our methods can be employed to demonstrate possible routes of transmission through an ecosystem and are broadly applicable across studies of both intra- and interspecific routes of transmission. Even though transmission routes demonstrated by *E. coli* are likely only applicable to fecal-oral pathogens that are epidemiologically similar (e.g., pathogenic *E. coli*, *Cryptosporidium spp.*, and *Clostridia spp.*, some helminthes, etc), this approach allows us to quantify heterogeneity in transmission patterns. Transmission networks of other pathogens could be examined in a similar way, although consideration must be taken to select pathogens with suitable amounts of genetic variation. Too much variation may lead to a disconnected network, while too little will lead to all individuals being connected to all others. Also, it would be difficult to construct networks for pathogens with low prevalence due to the fact that many sampled individuals would not be infected and thus would not yield any new information about transmission network structure.

In conclusion, my results suggest that some species may play a disproportionately important role in the spread of fecal-oral pathogens. Species whose individuals typically exhibit a high number of connections or occupy positions with high flow are potential super-spreaders in this system. This study demonstrates the utility of an approach that combines genetics and network theory both for quantifying interspecific heterogeneity in transmission patterns and as a first step in making targeted control strategies feasible for the management of infectious diseases.
Literature cited


Butts CT, 2010. sna: tools for social network analysis. R package version 2.2-0.


Hamede RK, Bashford J, McCallum H, Jones M, 2009. Contact networks in a wild Tasmanian devel (*Sarcophilus harrisii*) population: using social network analysis
to reveal seasonal variability in social behaviour and its implications for

Hoffmann R, 1989. Evolutionary steps of ecophysical adaptation and diversification of

Johnson JR, Clabots C, Kuskowski MA, 2008. Multiple-host sharing, long-term
persistence, and virulence of *Escherichia coli* clones from human and animal

Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, Safi K, Sechrest W,
Boakes EH, Carbone C, Connolly C, Cutts MJ, Foster JK, Grenyer R, Habib M,
Plaster CA, Price SA, Rigby EA, Rist J, Teacher A, Bininda-Emonds ORP,
of life history, ecology, and geography of extant and recently extinct mammals.
Ecology 90:2648.


Kock RA, 2005. What is this infamous "wildlife/livestock disease interface?" A review
of current knowledge for the African continent. In: Conservation and
Development Interventions at the Wildlife/Livestock Interface: Implications for


Supplementary Information

**Supplementary Table S1:** Percent of observations each pair of species was observed <50m of each other. Species below the dotted line were not actively surveyed and were only recorded when observed with the study species.

<table>
<thead>
<tr>
<th></th>
<th>Black Rhino (n=56)</th>
<th>Buffalo (n=109)</th>
<th>Cattle (n=113)</th>
<th>Eland (n=218)</th>
<th>Giraffe (n=262)</th>
<th>Gr. Gazelle (n=650)</th>
<th>Hartbeest (n=184)</th>
<th>Impala (n=688)</th>
<th>Th. Gazelle (n=532)</th>
<th>Zebra (n=739)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black rhino</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buffalo</td>
<td>2.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cattle</td>
<td>0.6% 0.5%</td>
<td>1.8% 0.9%</td>
<td>1.1% 5.5%</td>
<td></td>
<td>0.9% 4.1%</td>
<td>6.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eland</td>
<td>1.9% 2.8%</td>
<td>2.5% 1.1%</td>
<td>5.2% 4.1%</td>
<td>1.8% 6.5%</td>
<td>3.0% 6.9%</td>
<td>5.5%</td>
<td>2.6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giraffe</td>
<td>2.6% 2.5%</td>
<td>1.1% 5.5%</td>
<td>4.1% 5.5%</td>
<td>1.8% 6.5%</td>
<td>3.0% 6.5%</td>
<td>6.9%</td>
<td>5.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gr. gazelle</td>
<td>0.6% 1.7%</td>
<td>1.7% 5.2%</td>
<td>1.7% 5.2%</td>
<td>1.8% 2.5%</td>
<td>3.0% 6.5%</td>
<td>6.9%</td>
<td>5.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hartbeest</td>
<td>0.4% 1.0%</td>
<td>0.3% 1.7%</td>
<td>3.0% 6.9%</td>
<td>1.8% 5.5%</td>
<td>3.0% 6.9%</td>
<td>6.9%</td>
<td>5.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impala</td>
<td>1.4% 2.2%</td>
<td>0.6% 1.7%</td>
<td>4.3% 7.9%</td>
<td>1.8% 5.5%</td>
<td>3.0% 6.9%</td>
<td>6.9%</td>
<td>5.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Th. gaz.</td>
<td>1.6% 2.9%</td>
<td>1.4% 4.3%</td>
<td>6.4% 4.3%</td>
<td>1.8% 5.5%</td>
<td>3.0% 6.9%</td>
<td>6.9%</td>
<td>5.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebra</td>
<td>1.0% 3.2%</td>
<td>2.4% 14.2%</td>
<td>2.4% 6.6%</td>
<td>6.6% 13.4%</td>
<td>6.9% 13.4%</td>
<td>7.5%</td>
<td>13.4%</td>
<td>7.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elephant</td>
<td>0.0% 0.0%</td>
<td>0.7% 1.2%</td>
<td>0.0% 0.0%</td>
<td>0.9% 0.0%</td>
<td>0.0% 0.6%</td>
<td>0.6%</td>
<td>0.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gerenuk</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grevy's zebra</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oryx</td>
<td>0.0% 2.6%</td>
<td>0.6% 1.6%</td>
<td>4.0% 3.9%</td>
<td>1.6% 5.5%</td>
<td>2.8% 4.7%</td>
<td>5.5%</td>
<td>2.6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheep</td>
<td>0.0% 0.0%</td>
<td>4.4% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0% 0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warthog</td>
<td>1.5% 4.7%</td>
<td>1.0% 2.6%</td>
<td>8.7% 2.6%</td>
<td>2.6% 4.7%</td>
<td>7.4% 5.1%</td>
<td>5.7%</td>
<td>5.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterbuck</td>
<td>1.1% 0.4%</td>
<td>0.4% 5.6%</td>
<td>1.3% 1.6%</td>
<td>1.3% 5.2%</td>
<td>2.7% 2.7%</td>
<td>1.1%</td>
<td>1.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White rhino</td>
<td>1.2% 3.0%</td>
<td>0.0% 0.0%</td>
<td>0.3% 0.3%</td>
<td>0.3% 1.0%</td>
<td>0.5% 0.5%</td>
<td>0.7%</td>
<td>1.6%</td>
<td></td>
<td></td>
<td>0.8%</td>
</tr>
</tbody>
</table>

Key: 0-2% 2-5% 5-8% 8-12% >12%
Supplementary Figure S1: Robustness of species connectivity measures given reduced sampling effort. Mean correlation in the rank-order of species given reduced sampling effort for (a) degree and (b) betweenness. Proportion of samples with the same two top ranked species for (c) degree and (d) betweenness.
Chapter 5: Conclusion

The importance of individual-level heterogeneity in pathogen transmission dynamics has been established in theoretical models over the past two decades (Cross et al. 2004; Keeling 2005; Keeling and Eames 2005; May 2006; Bansal et al. 2007; Perkins et al. 2008; Turner et al. 2008; Craft et al. 2010; Craft and Caillaud 2011; Griffin and Nunn 2012). However, empirical research has lagged behind theoretical models, especially in wildlife, because although we acknowledge that heterogeneity exists, we have limited ability to quantify it (Caley et al. 2009; Cross et al. 2009). While network analysis has drawn considerable attention as a new tool for quantifying heterogeneity in pathogen transmission dynamics, few studies have combined a network approach with real data on pathogen transmission (Otterstatter and Thomson 2007; Drewe 2009; Godfrey et al. 2009; Godfrey et al. 2010; Drewe et al. 2011; Fenner et al. 2011; Bull et al. 2012). Because of limitations in detecting the occurrence of transmission, social connections between individuals have often used as a proxy for inter-individual transmission potential. Thus, conclusions of past studies were based on the possibility that transmission might occur between interacting individuals (Perkins et al. 2008; Böhm et al. 2009; Craft et al. 2009; Grear et al. 2009; Hamede et al. 2009; Perkins et al. 2009). An individual’s social network position has generally been used as a risk factor for acquiring an infection rather than actually being able to determine who transmitted to whom (Drewe 2009; Godfrey et al. 2009; Porphyre et al. 2011).
The overarching goal of this dissertation was to develop and assess new methodology for investigating heterogeneity in pathogen transmission dynamics. I used the genetics of a diverse microbe, *Escherichia coli*, to reveal that transmission has already occurred and construct transmission networks based on quantifiable transmission events. If two animals shared the same genetic subtype of *E. coli*, then it can be inferred that these animals were part of the same transmission chain. These genetic data were used to construct pathogen transmission networks in which individuals were interlinked based upon patterns of subtype sharing. In this dissertation, I showed how microbial genetics can be used to create a transmission network that is defined independently from the social network, allowing for social and transmission networks to be directly compared.

One key goal in this dissertation was to determine how closely association patterns impact pathogen transmission patterns. To address this, I first quantified association patterns in giraffe (Chapter 2). The frequency in which two individuals were observed together was positively correlated with the extent to which their home range overlapped, implying an underlying role of shared space use in determining association patterns. New data cloud geometry algorithms were employed to identify multi-level social organization. Individual giraffe were members of social cliques, which were embedded in larger sub-communities, which were embedded in even larger communities. Shared space use played a much larger role in determining this multi-tiered organization in females than in males, which is consistent with a matrilineally based society characterized by female locational philopatry. From a pathogen transmission perspective, correlations between association and shared space use suggest that environmental and social exposure to *E. coli* will be correlated.
The third chapter of my dissertation thus investigated how giraffe association and space use patterns influence transmission dynamics. To distinguish between social and environmental contact, three networks were compared: a social network based on observed association patterns, a spatial network based on the degree of home range overlap between individuals, and a transmission network based on patterns of *E. coli* subtype sharing. *E. coli* was cultured from fecal samples and genotyped using BOX-PCR, an established method for discriminating among closely related *E. coli* subtypes (Goldberg et al. 2006; Cesaris et al. 2007; Mohapatra and Mazumder 2008). I found that giraffe that were more strongly linked in the social network were more likely to share *E. coli* subtypes, while neither spatial overlap nor number of shared water sources was related to the likelihood of being linked in the transmission network. In addition, an individual’s connectivity in the transmission network (degree and betweenness) was highly correlated with the same measure in the social but not spatial network. Although the spatial and the social network were correlated with one another, I suggest that synchronous exposure to a common environmental source of *E. coli* (i.e., being in the same place at the same time as a result of social associations) may be critical in determining patterns of subtype sharing.

These results emphasize the importance of association patterns in understanding transmission dynamics, even in multi-host, environmentally transmitted pathogens such as *E. coli*. In the one other study that combined microbial genetics with networks, genetic data were used to identify which individual lizards (*Tiliqua rugosa*) hosted which dominant strains of *Salmonella* (Bull et al. 2012). At the dyad level, they also found that association strength, and not spatial proximity, was correlated with sharing strains.
However, they did not use the strain sharing data to explicitly define a transmission network, and thus did not identify correlations in an individual’s network position across networks. Instead, they utilized analytical approaches similar to past studies in which measures of individual-level connectivity in the social network were used as risk factors for hosting specific strains (Bull et al. 2012). Thus, they did not address questions such as whether highly connected individuals in the social network were also potential super-spreaders.

Multi-host wildlife pathogens are an increasing concern for conservation and livestock husbandry (Cleaveland et al. 2001; Leendertz et al. 2006; Tompkins et al. 2011), but lack of conclusive information regarding the dynamics of interspecific pathogen transmission hampers our ability to quantify risk. Therefore, my fourth chapter extended the framework developed in Chapter 3 to assess interspecific transmission networks comprised of wild and domestic ungulates. Fecal samples were collected from 279 individuals from ten different species, and E. coli subtyping was used to construct a transmission network for wildlife and livestock. By quantifying each individual’s position in the network using network metrics, I demonstrated that there is variation across species in their connectivity in the transmission network, with some contributing disproportionately to pathogen transmission. High levels of connectivity exhibited by individuals of key species indicate that there are potential “super-spreader” species in this ecosystem. Grant’s gazelles, for example, not only are connected to a greater number of other individuals, but also occupy positions that are potential bottlenecks for pathogen flow in the network. In contrast, plains zebras may be super-spreaders because they range widely and overlap with geographically separated regions of the network. This seems to
allow them to function as bridges in the network and makes them potential bottlenecks for pathogen flow. Given interspecific variation in physiology and behavior, it is unsurprising that species contribute disproportionately to transmission, but it has previously been difficult to actually quantify these differences. The novel approach developed in this dissertation provides a new framework to investigate questions concerning heterogeneity in pathogen transmission.

**Future directions**

The major advantage of using genetic data to define the transmission network is that it is constructed independently of other factors, such as the social network. The goal of many empirical studies in network epidemiology is to analyze the extent to which social networks influence pathogen transmission. However, to truly address this question, pathogen transmission patterns must be independently quantified from the social network. Otherwise, the conclusions that can be drawn are limited to simply how the social network affects the risk of acquiring an infection without assessing pathways of transmission. The method developed in this dissertation opens the door to new lines of research in network epidemiology, such as the extent to which animals occupy similar positions in contact and transmission networks, identification of characteristics of highly connected “super-spreaders,” or the effect of environmental change on the structure and connectivity of transmission networks.

Addressing how seasonality and environmental change affect transmission networks would be a fruitful line for future research. Transmission dynamics vary seasonally because of changes in host behavior, host contact rates, and the ability of
pathogens to persist in the environment (Altizer et al. 2006). It is critically important to examine seasonal changes in transmission dynamics because they will impact not only how we manage disease during annual cycles, but also how we manage disease during longer-term climate change (Altizer et al. 2006). Most research in this area has focused on observed and projected shifts in the global geographic ranges of pathogens and vectors, climate effects on host susceptibility and pathogen life cycles, and the effect of over-wintering conditions on disease dynamics (Harvell et al. 2002; Atwill et al. 2004; Mills et al. 2010). On a landscape scale, however, fluctuations in precipitation modify the distribution of water and forage, resulting in changes in the way animals use the landscape and associate with other species (Altizer et al. 2006). This shift in landscape use and inter-specific associations will alter transmission dynamics within ecosystems where the pathogen already occurs. African ungulates, for example, generally decrease diet and habitat overlap during the dry season due to increased competition (Sinclair 1985; Voeten and Prins 1999), yet limited water availability may increase contact rates among species (Dalhay et al. 2009). Little is known about how seasonality and climate change will affect the ability of pathogens to spread and persist within a single ecosystem, yet seasonal changes in behavior have the potential to substantially affect the structure of the transmission network. Network analysis may prove to be an ideal tool for tracking seasonal changes in transmission dynamics and has yet to be applied in any wildlife system.

In this dissertation, BOX-PCR was used to identify subtypes of E. coli. One drawback of this method is that genetic data produced was of insufficient resolution to determine inter-individual transmission. Rather, a single subtype was often identified in
numerous individuals. These animals could be considered part of the same chain of transmission, but there was no way to determine who transmitted to whom within that chain. Ideally, I would like to pursue using a genetic tool that is more discriminatory. DNA sequencing, for example, would allow for the subtypes identified in BOX-PCR to be further sub-divided into genotypes based on basepair sequences. This higher resolution in the genetic data would lead to transmission networks that more accurately illustrate transmission pathways.

Although not usually pathogenic, *E. coli* transmission routes give insight into transmission dynamics by demonstrating where contact is sufficient for transmission to occur. The advantage of using *E. coli* as a proxy is that it allows me to study transmission dynamics without waiting for a true epidemic or making post-hoc conclusions about transmission after an epidemic has occurred. Control strategies targeted at super-spreaders are theoretically more effective at limiting disease spread than conventional strategies (Lloyd-Smith et al. 2005), but there are no existing methods to identify super-spreaders in wild populations (Cross et al. 2009). However, to apply knowledge of *E. coli* transmission pathways to control strategies for other pathogens, we first must investigate how closely *E. coli* transmission pathways resemble patterns exhibited by other types of pathogens. A comparison of *E. coli* transmission networks to the transmission networks of other diseases would be a critical step in validating the usefulness of *E. coli* as a proxy for pathogen transmission. More generally, though, genetic data could be utilized to construct transmission networks for a diversity of pathogens, and doing so would facilitate new research on the causes and consequences of heterogeneity in pathogen transmission. Thus, my novel approach combines a network approach with microbial
genetics to provide a sophisticated and effective tool for tracing transmission pathways and identifying super-spreaders in wildlife. The methods developed here are broadly applicable and promises to improve our ability to predict the impacts of disease. I have shown how these methods can be used to understand social processes in the transmission of fecal-oral pathogens in giraffe, assess species-level heterogeneity in pathogen transmission, and identify potential super-spreader species in a community of east African ungulates.


to reveal seasonal variability in social behaviour and its implications for
Harvell CD, Mitchell CE, Ward JR, Altizer S, Dobson AP, Ostfeld RS, Samuel MD,
296:2158-2162.
Leendertz FH, Pauli G, Maetz-Rensing K, Boardman W, Nunn C, Ellerbrok H, Jenson
SA, Junglen S, Boesch C, 2006. Pathogens as drivers of population declines: the
importance of systematic monitoring in great apes and other threatened mammals.
Biological Conservation 131:325-337.
Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM, 2005. Superspreading and the effect
May RM, 2006. Network structure and the biology of populations. Trends in Ecology and
Mills JN, Gage KL, Khan AS, 2010. Potential influence of climate change on vector-
borne and zoonotic diseases: a review and proposed research plan. Environmental
Health Perspectives 118:1507-1514.
Mohapatra BR, Mazumder A, 2008. Comparative efficacy of five different rep-PCR
methods to discriminate Escherichia coli populations in aquatic environments.
Water Science and Technology 58.3:537-547.


design of control programs. Proceedings of the National Academy of Sciences 94:338-342.
Appendix A: Demography of the Giraffe in Ol Pejeta Conservancy

A report submitted to the OPC Ecological Monitoring Department on 10 December, 2011

Methods

From January 21 to August 2, 2011, I collected 1089 observations of giraffe groups. A group was defined as a set of individuals engaged in the same behavior, or moving in the same direction or toward a common destination, as long as each giraffe was no more than 500 m from at least one other group member. As group membership is constantly shifting, independence of observations was ensured by only using the first observation of an individual’s group on a given day. Giraffe were recognized by unique spot patterns on their necks. Each individual was assigned an ID number as well as a nickname. A photo ID card was created for each of 230 individuals observed throughout the field season, although ID cards were rarely used once I became familiar with the animals. On average, I observed 30.7 giraffe day, distributed between four to six groups. I observed each giraffe approximately 31.1 times and saw each giraffe about once a week.

Giraffe group composition and membership were recorded for all groups sighted while driving pre-determined survey routes. Routes were determined so that a different part of the study area was traversed each day, allowing for most of the study area to be surveyed once every three days. Each route was approximately 100 km and traversed both major and minor roads in OPC.
East Route: Sweetwaters, Oryx Plain, Serat Plain, Zebra Plain, Morani Plain, Grant’s Plain, Buffalo Plain, and surrounding areas.

Southwest Route: Research airstrip, Scott’s Plain, Loidien area, Ngobit, Lerai Ndogo, Black Tank (Windmill Plain), Kicheche, and surrounding areas.

Northwest Route: Lodru, Muturu, Kamok, Kamok airstrip, Lairugurugu, Porini, and surrounding areas.

**Group size trends**

Group sizes ranged from 1 to 44 giraffes (Figure 1). The average was $5.42 \pm 0.198$ individuals, while the mode was 1. Giraffe observed alone were nearly always adult males, although females were sometimes observed alone in the days prior to being observed with very young calves.

**Figure A1:** Frequency of observed group sizes.
**Population size**

The most recent population estimate was 204 giraffe. This estimate was based on a mark-recapture analysis that I conducted during my pilot field season in July/August 2010.

As of August 1, 2011, the population was 212 individuals. This estimate was based on a census of known individuals. Giraffe monitoring ended on August 2. Given that nearly all giraffe were sighted approximately once a week, any giraffe that had not been seen between July 1 and August 2 was assumed to have died or left the study area. I consider it extremely unlikely that unknown individuals are living in OPC without my knowledge; only two new adults were discovered in the last five months of my study. One was likely a transient male because he was observed only a few times near one of the wildlife corridors. Thus, I believe that my population estimate of 212 individuals represents a complete census.

**Sex ratio and age structure**

Giraffes were aged according to validated height estimates and age-associated behaviors. This aging scale was adapted from scales used in the literature (Pratt and Anderson 1979; Langman 1977; Fennessy 2004).

*Infant:* <3 months old, umbilical cord still attached, the length of the neck is short relative to the height of the shoulder

*Juvenile:* 3 months to 1.5 years. Larger than a baby but still accompanies mother. Activity budget begins to resemble an adult’s.
**Subadult:** 1.5 – 4 years. No longer accompanies mother, but smaller than an adult. The number of giraffe in this age class is likely underestimated because of older individuals being erroneously classified as adults.

**Adult:** >4 years. Reaches sexual maturity and adult size.

### Table A1. Age structure and sex ratio of the OPC giraffe population.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult</strong></td>
<td>78</td>
<td>82</td>
<td>160</td>
</tr>
<tr>
<td><strong>Subadult</strong></td>
<td>12</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td><strong>Juvenile</strong></td>
<td>11</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td><strong>Baby</strong></td>
<td>11</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>101</td>
<td>100</td>
<td>212</td>
</tr>
</tbody>
</table>

**Births**

I recorded the births of 19 calves during the 2011 field season. Births peaked in the month following the cessation of the rains (Figure 2). The timing of this peak following the rainy season is similar to the birthing peak recorded Thornicroft’s giraffe (N = 35 births) in the Luangwa Valley, Zambia (Bercovitch and Berry 2009). As in Zambia and other giraffe populations, the calving peak occurred during a time of favorable environmental conditions.
Mortality

Mortality during the first months of life is high. Of 16 calves born at least one month before the end of the study period, only 50% of them survived to one month. First-month survivorship is probably overestimated, as some calves likely died before being observed. Given survival to the age of one month, the probability of survival to two months was 87.5%. Calves that reached two months of age had a very high likelihood of survival, though estimates were based on increasingly small sample sizes given the short duration of the study. Because most calves were born towards the end of my study, I was only able to follow a few calves for longer than three months.

Table A2. Survivorship of giraffe calves.

<table>
<thead>
<tr>
<th></th>
<th>1 mo.</th>
<th>2 mo.</th>
<th>3 mo.</th>
<th>4 mo.</th>
<th>5 mo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>16</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P(survival)</td>
<td>0.5</td>
<td>0.875</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Using the data I collected during my pilot field season in July/August 2010, I was able to calculate annual mortality for males and females. Of 87 adult females observed in July/August 2010, all but one of them were still alive in August 2011. Of 89 adult males observed in 2010, only four of them died or disappeared. The female and two of the males disappeared for unknown causes sometime between August 2010 and January 2011. One male disappeared after being observed with a mange-like skin condition (Figure 3). He appeared increasing malnourished prior to his disappearance. Another male was observed with a severely swollen ankle that inhibited his ability to walk. Shortly after, his carcass was found being eaten by lions, although it is unclear if the lions acquired the carcass through hunting or scavenging (Figure 4). The last male disappeared after being observed twice in an extremely malnourished condition (Figure 5).

Adult female annual mortality (N = 87): 2.2%
Adult male annual mortality (N=89): 4.5%

Based on photo records of 113 individuals provided by Alan Birkett (1998), 21 giraffe in the current population are at least 13 years old. While this exceeds the average life expectancy for giraffe, it falls well below the maximum life span for either wild or captive individuals (Dagg and Foster 1976).
Figure A3. Mange-like skin condition in a male giraffe. This male routinely was seen in and around the Morani visitor center, but disappeared in March 2011.

Figure A4. Adult male that died in March 2011 after being observed with a severe limp. Note that the animal can still be individually identified from the skin.
Figure A5. Giraffe carcass found in July 2011. Other giraffe appeared extremely interested in the carcass and investigated it for 20 – 30 minutes. This sort of reaction to the death of a conspecific has never been recorded in giraffe before.

Range Use

The goal of this study was not to assess giraffe habitat use. Thus, all results about giraffe habitat usage should be interpreted with extreme caution. I calculated the percentage of observations occurring in each vegetation class (Table 3, Figure 6). Giraffe were observed most often in open grassland, followed by Acacia drepanolobium bushland. These two habitat types account for over 90% of observations. Because my intent was not to accurately record the habitat use of giraffe, there was extreme sampling bias towards observing giraffes in open areas. That being said, A. drepanolobium is the most important forage species of the giraffe. Thus, it is unsurprising that a large proportion of sightings occurred in this habitat type.
Figure 7 shows the intensity of range use in Ol Pejeta. One of the most highly used areas was Oryx Plain, which was consistently a location for mothers to form crèches with their calves. To a lesser extent, Muturu Plain was also used as a crèche location. Crèche locations are conserved over time, with females returning repeatedly to calve in the same location (Langman 1977).

Table A3. Percentage of giraffe sightings in each habitat type.

<table>
<thead>
<tr>
<th>Vegetation Class</th>
<th>Number of Observations</th>
<th>% of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense Bushland (Euclea dominated)</td>
<td>37</td>
<td>3.45%</td>
</tr>
<tr>
<td>Open Bushland (Acacia drepanolobium dominated)</td>
<td>449</td>
<td>41.81%</td>
</tr>
<tr>
<td>Open Bushland (Mixed Bushland)</td>
<td>15</td>
<td>1.40%</td>
</tr>
<tr>
<td>Open Grassland</td>
<td>564</td>
<td>52.51%</td>
</tr>
<tr>
<td>Riverine Forest</td>
<td>9</td>
<td>0.84%</td>
</tr>
<tr>
<td>Swamp</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

Figure A6. Locations of giraffe sightings according to habitat type.
**Literature cited**


### Appendix B: Pathogens of the giraffe

**Table B1.** Pathogens documented in giraffe, their transmission mode and the length of time they persist in the environment.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Agent</th>
<th>Transmission mode</th>
<th>Environmental persistence*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anthropoda</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bot flies</td>
<td>Rhinoestrus spp.</td>
<td>Tick-borne</td>
<td>n/a</td>
</tr>
<tr>
<td>Sarcoptic mange</td>
<td>Sarcoptes scabiei</td>
<td>Direct and indirect</td>
<td>Several weeks</td>
</tr>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaplasmosis</td>
<td>Anaplasma spp.¹</td>
<td>Tick-borne</td>
<td></td>
</tr>
<tr>
<td>Anthrax</td>
<td>Bacillus anthracis²</td>
<td>Inhalation, mechanical vector</td>
<td>Years</td>
</tr>
<tr>
<td>Bovine tuberculosis</td>
<td>Mycobacterium tuberculosis³</td>
<td>Close contact, ingestion, inhalation, fecal-oral</td>
<td>4-8 weeks</td>
</tr>
<tr>
<td>Brucellosis</td>
<td>Brucella spp.⁴</td>
<td>Close contact, contact with fetal-contaminated material</td>
<td>Several months</td>
</tr>
<tr>
<td>Pathogenic E. coli</td>
<td><strong>Escherichia coli O157⁷</strong></td>
<td>Fecal-oral, mechanical vector</td>
<td>Soil: 8 -25 weeks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Water: 6 days – 2 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- 2log reduction in numbers after 6-18 days</td>
</tr>
<tr>
<td>Listeriosis</td>
<td>Listeria spp.⁸</td>
<td>Fecal-oral, ingestion, inhalation, close contact</td>
<td>7 – 311 days</td>
</tr>
<tr>
<td>Heartwater</td>
<td>Cowdria ruminantium⁹</td>
<td>Tick-borne</td>
<td>Several hours</td>
</tr>
<tr>
<td>Salmonellosis</td>
<td>Salmonella spp.¹⁰</td>
<td>Fecal-oral</td>
<td>50 days</td>
</tr>
<tr>
<td><strong>Cestoda</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tape worm</td>
<td>Echinococcus spp.¹¹</td>
<td>Ingestion</td>
<td>Years</td>
</tr>
<tr>
<td><strong>Nematoda</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongyle parasites</td>
<td>Cooperia spp.¹³</td>
<td>Ingestion</td>
<td>Several months</td>
</tr>
<tr>
<td>Parasite Type</td>
<td>Species</td>
<td>Route of Infection</td>
<td>Duration</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Other nematodes</td>
<td>Haemonchus spp.</td>
<td>Ingestion</td>
<td>Several months</td>
</tr>
<tr>
<td></td>
<td>Oesphagostomum sp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ostertagia spp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gongylonema spp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monodontella giraffae</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parabronema skrjabini</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Skrjabinema spp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trichostrongylus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platyhelminthes</td>
<td>Fasciola gigantica</td>
<td>Ingestion</td>
<td>&gt;3 months</td>
</tr>
<tr>
<td>Liver flukes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protozoa</td>
<td>Babesiosis</td>
<td>Babeisa spp.</td>
<td>Tick-borne</td>
</tr>
<tr>
<td></td>
<td>Cryptosporidiosis</td>
<td>Cryptosporidium spp.</td>
<td>Fecal-oral</td>
</tr>
<tr>
<td></td>
<td>Giardiasis</td>
<td>Giardia</td>
<td>Fecal-oral</td>
</tr>
<tr>
<td></td>
<td>Hepatozoonosis</td>
<td>Hepatozoon spp.</td>
<td>Ingestion of infected arthropods</td>
</tr>
<tr>
<td></td>
<td>Sarcocystosis</td>
<td>Sarcocystis spp.</td>
<td>Ingestion</td>
</tr>
<tr>
<td></td>
<td>Theileriosis</td>
<td>Theileria spp.</td>
<td>Tick-borne</td>
</tr>
<tr>
<td></td>
<td>Trypanosomiasis</td>
<td>Trypanosoma spp.</td>
<td>Vector-borne - tsetse flies</td>
</tr>
<tr>
<td></td>
<td>Toxoplasmosis</td>
<td>Toxoplasma gondii</td>
<td>Ingestion, mechanical vectors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>18 months</td>
</tr>
<tr>
<td>Viruses</td>
<td>Bluetongue</td>
<td>Orbivirus spp.</td>
<td>Vector-borne - midges</td>
</tr>
<tr>
<td></td>
<td>Bovine herpesvirus-2</td>
<td>Simplexvirus spp.</td>
<td>Mechanical vector</td>
</tr>
<tr>
<td></td>
<td>Bovine parainfluenza</td>
<td>Paramyxovirus spp.</td>
<td>Inhalation</td>
</tr>
<tr>
<td></td>
<td>Bovine respiratory syncytial virus</td>
<td>Pneumovirus spp.</td>
<td>Inhalation</td>
</tr>
<tr>
<td></td>
<td>Bovine viral diarrhea</td>
<td>Pestivirus spp.</td>
<td>Close contact, mechanical vector, fecal-oral</td>
</tr>
<tr>
<td></td>
<td>Coronavirus</td>
<td>Coronovirus spp.</td>
<td>Gastrointestinal and respiratory excretions</td>
</tr>
<tr>
<td></td>
<td>Crimean-Congo haemorragic</td>
<td>Nairovirus spp.</td>
<td>Tick-borne</td>
</tr>
<tr>
<td>Disease</td>
<td>Pathogen</td>
<td>Mode of transmission</td>
<td>Incubation Period (days)</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Febrile</td>
<td>Coronavirus Cardiovirus spp. 25</td>
<td>Close contact, fecal-oral</td>
<td>25</td>
</tr>
<tr>
<td>Foot and Mouth Disease</td>
<td>Aphthovirus spp. 26</td>
<td>Close contact, fecal-oral</td>
<td>14 – 200</td>
</tr>
<tr>
<td>Lumpy skin disease</td>
<td>Capripoxvirius spp. 27</td>
<td>Vector-borne – biting insects, close contact</td>
<td>79 – 172</td>
</tr>
<tr>
<td>Rinderpest</td>
<td>Morbillivirus spp. 30</td>
<td>Close contact</td>
<td>6 – 48</td>
</tr>
</tbody>
</table>

### Unknown agents

- Giraffe ear disease
- Giraffe skin disease

* Persistence depends on conditions (pH, moisture, temperature). For most pathogens, cool temperatures and moist conditions facilitate survival.

1. Davidson and Goff 2001, Kottler 1984
2. Turnbull et al. 1992
3. Lewerin et al. 2005, Cleaveland et al. 2005
4. Thorne 2001
6. Mörner 2001
10. Chakraborty 1004
11. Melbourne 1978
12. Kreczek et al. 1990
14. Kodádková et al. 2010
15. Bengis et al. 1998
16. Vanderplank 1942
17. Home and Mukaratirwa 2005
18. Kimber et al. 2002
19. Castro 2001
20. Van Campen et al. 2001
21. Hasoksuz et al. 2007
22. Burt et al. 1993
23. Thomson et al. 2001
25. Robinson and Kerr 2001
26. Kock 2005
27. Mlengeya and Lyaruu 2005
29. Goossens et al. 2005
30. Olson 2001
31. Craig 2001
32. Pybus 2001
33. Colwell 2001
34. Bornstein et al. 2001
Literature cited


Lewerin SS, Eld K, Bölske G, Olsson S-L, Röken B, Ghebremichael S, Koivula T, 
captive Asian elephants in a Swedish zoo. Veterinary Record 156:171-175.


Mlengeya T, Lyaruu V, 2005. Experiences with and the challenges of wildlife health 
management in the National Parks of Tanzania. In: Conservation and 
Development Interventions at the Wildlife/Livestock Interface: Implications for 
Wildlife, Livestock and Human Health (Osofsky SA, ed). Gland, Switzerland: 
IUCN; 51-54.


University Press; 399-416.

Oosthuizen MC, Allsopp BA, Troskie M, Collins NE, Penzhorn BL, 2009. Identification 
of novel *Babesia* and *Theileria* species in South African giraffe (*Giraffa 
camelopardalis*, Linnaeus, 1758) and roan antelope (*Hippotragus equinus, 
Desmarest 1804*). Veterinary Parasitology 163:49-46.
for *Cowdria ruminantium* in wild ruminants from Africa. Journal of Wildlife
Diseases 34:567-575.


Mammals, 3rd ed (Willems ES, Barker IK, eds). Ames: Iowa State University
Press.

State University Press; 119-130.


Turnbull PC, Doganay M, Lindeque PM, Aygen B, McLaughlin J, 1992. Serology and
anthrax in humans, livestock and Etosha National Park wildlife. Epidemiology
and Infection 108:299-313.

State University Press; 232-244.

Vanderplank FL, 1942. A note on the trypanosomiasis of game from tsetse areas at
Shinyanga and Ukerewe peninsula. Transactions of the Royal Society of Tropical
Medicine and Hygiene 35:319-322.
Appendix C: Fecal water content in East African ungulates

Table C1: To determine how fecal water content varied across species, fecal samples were dried in a toaster oven for approximately eight hours. Samples were weighed before and after the drying process. The difference in weights was used to calculate the percent water content in each sample.

<table>
<thead>
<tr>
<th>Species</th>
<th>N</th>
<th>Mean water Content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black rhino</td>
<td>9</td>
<td>67.1</td>
</tr>
<tr>
<td>Buffalo</td>
<td>10</td>
<td>76.3</td>
</tr>
<tr>
<td>Cattle</td>
<td>10</td>
<td>76.0</td>
</tr>
<tr>
<td>Eland</td>
<td>11</td>
<td>58.2</td>
</tr>
<tr>
<td>Giraffe</td>
<td>10</td>
<td>55.1</td>
</tr>
<tr>
<td>Grant’s gazelle</td>
<td>10</td>
<td>44.2</td>
</tr>
<tr>
<td>Hartebeest</td>
<td>10</td>
<td>54.0</td>
</tr>
<tr>
<td>Impala</td>
<td>9</td>
<td>48.9</td>
</tr>
<tr>
<td>Thomson’s gazelle</td>
<td>9</td>
<td>41.7</td>
</tr>
<tr>
<td>Zebra</td>
<td>12</td>
<td>65.5</td>
</tr>
</tbody>
</table>

Table C1: Variation in fecal water content across species.
Appendix D: Range use by East African ungulates

Figures D1-D12 show the intensity of range use by each large ungulate species. As described in Chapter 4, species locations were observed while driving road transects every three days throughout the study period. Observations were pooled across the entire study period with no division by season. Range use was determined using kernel density estimation techniques in ArcGIS (ESRI, version 9.3).

**Figure D1**: Buffalo range use intensity. Darker color represents higher intensity usage (N = 109 observations).
Figure D2: Black rhino range use intensity. Darker color represents higher intensity usage (N = 56 observations).

Figure D3: Cattle range use intensity. Darker color represents higher intensity usage (N = 113 observations).
**Figure D4:** Eland range use intensity. Darker color represents higher intensity usage (N = 218 observations).

**Figure D5:** Giraffe range use intensity. Darker color represents higher intensity usage (N = 1089 observations).
Figure D6: Grant’s gazelle range use intensity. Darker color represents higher intensity usage (N = 650 observations).

Figure D7: Hartebeest range use intensity. Darker color represents higher intensity usage (N = 184 observations).
Figure D8: Impala range use intensity. Darker color represents higher intensity usage (N = 688 observations).

Figure D9: Oryx range use intensity. Darker color represents higher intensity usage (N = 58 observations).
**Figure D10:** Thomson’s gazelle range use intensity. Darker color represents higher intensity usage (N = 532 observations).

![Thomson’s gazelle range use intensity](image1.png)

**Figure D11:** White rhino range use intensity. Darker color represents higher intensity usage (N = 30 observations).

![White rhino range use intensity](image2.png)
**Figure D12:** Plains zebra range use intensity. Darker color represents higher intensity usage (N=739 observations).