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

Kimberly L. VanderWaal[1,2,3]*, Edward R. Atwill[4,5], Lynne. A. Isbell[1,6] and Brenda McCowan[1,2,4]

[1]Animal Behavior Graduate Group, University of California, Shields Avenue, Davis, CA 95616, USA; [2]International Institute for Human-Animal Networks, University of California, Davis, CA, USA; [3]Wangari Maathai Institute for Peace and Environmental Studies, University of Nairobi, PO Box 29053-00635, Nairobi, Kenya; [4]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, CA, USA; [5]Western Institute for Food Safety and Security, University of California, Davis, CA, USA; and [6]Department of Anthropology, University of California, Davis, CA, USA

Summary

1. 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) within social networks.

2. Using a novel approach, we 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 (*Giraffe camelopardalis*). Individuals were considered to be a part of the same transmission chain and were interlinked in the transmission network if they shared genetic subtypes of *E. coli*.

3. By using microbial genetics to quantify who transmits to whom independently from the behavioural data on who is in contact with whom, we were 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 with two separate contact networks defined from the behavioural data: a social network based on association patterns, and a spatial network based on patterns of home-range overlap among individuals.

4. We found that links in the transmission network were more likely to occur between individuals that were strongly linked in the social 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 the spatial 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.

5. 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 analyse transmission networks in a wildlife population and highlights the potential utility of an approach integrating microbial genetics with network analysis.

Key-words: bacterial genotyping, disease ecology, epidemiology, infection dynamics, social network analysis, social structure, space-use patterns, wildlife disease

*Correspondence author. E-mail: kvanderwaal@ucdavis.edu

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, population spatial and social structure create heterogeneity in transmission patterns (Keeling & Eames 2005; Bansal, Grenfell & Meyers 2007; Craft & Caillaud 2011). Network theory provides a set of tools for analysing such heterogeneity by not only taking into account direct connections, but also indirect links between individuals, providing a sophisticated method to analyse relations among individuals (Wasserman & Faust 1994; Croft, James & Krause 2008). Recently, there has been considerable effort to incorporate network theory into epidemiological models (Keeling 2005; May 2006; Bansal, Grenfell & Meyers 2007). Compared with traditional mass-action models, network models tend to result in reductions in the early growth rate, number of secondary infections produced by each infected individual, and final size of an epidemic (Keeling & 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 & Eames 2005; 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, Smith & Hutchings 2009; Craft & 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 behavioural 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; VanderWaal *et al.* 2013a). While modelling has proven to be a useful approach to study the importance of these wildlife contact networks for disease spread (Cross *et al.* 2004; Perkins, Ferrari & Hudson 2008; Craft *et al.* 2010; Griffin & 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, Marion & Hutchings 2009). In many cases, 'transmission networks' have been constructed based on social interactions between individuals without assessing any pathogens (Böhm, Hutchings & White 2009; Grear, Perkins & Hudson 2009; Hamede *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 influences the likelihood of being infected. Generally, it has been shown that infected animals tend to have more connections in the contact network (Corner,

Pfeiffer & Morris 2003; Godfrey *et al.* 2009, 2010), or that individuals are at greater risk of becoming infected if they are connected to infected animals or engage in certain types of interactions (Otterstatter & Thomson 2007; Drewe 2009; Porphyre, McKenzie & Stevenson 2011; MacIntosh *et al.* 2012; VanderWaal *et al.* 2013a). Because the 'transmission networks' in these studies are not independently defined from behaviour-based contact networks, it has not been possible to truly assess how social patterns influence the structure of transmission networks.

In contrast, we use the genetics of a diverse microbe, *Escherichia coli*, to reveal where transmission has already occurred and construct a transmission network based on quantifiable transmission events in a wild giraffe population (*Giraffa camelopardalis*). If two individuals share the same genetic subtype of *E. coli*, then we infer that either direct transmission has occurred through social interactions or indirect transmission has occurred due to exposure to a common environmental source (Archie, Luikart & Ezenwa 2008). Recent work by Bull, Godfrey & Gordon (2012) analysed how lizard social networks influenced the likelihood of two individuals sharing the same genetic subtype of *Salmonella*, but the strain-sharing data were not conceptualized as a transmission network that was distinct from the social network. In our transmission networks, individual giraffe are interlinked based upon patterns of *E. coli* subtype sharing. Therefore, the transmission network is defined independently from behaviour-based contact networks. This allows us to directly compare the structure of transmission and contact networks.

Using this novel approach, we examine how well the transmission network overlays onto social and spatial networks. This allows us to ascertain the relative importance of social vs. environmental transmission mechanisms for enteric microbes such as *E. coli*. If social contact is crucial for transmission to occur, then the occurrence of links in the transmission network should be highly correlated with the social network. However, if environmental transmission is more common, then the transmission network should map more closely onto a spatial network in which individuals are linked according to the extent to which their home-ranges overlap.

In addition, individuals that are highly connected in the transmission network may be potential super-spreaders (Lloyd-Smith *et al.* 2005). Do individuals become super-spreaders because they are highly social or because they have a large number of spatial contacts? By comparing individual-level connectivity across networks, we investigate which mechanisms give rise to individual-level variation in transmission network connectivity. If highly social animals are potential super-spreaders, then individual-level connectivity in the social and transmission networks should be correlated. However, if widely ranging animals tend to be super-spreaders, then transmission network connectivity should be correlated with large home-range sizes and high connectivity in the spatial network.

Materials and methods

STUDY SITE AND ORGANISMS

Escherichia coli is an excellent model organism for examining microbe transmission pathways in relation to contact networks. It can readily be cultured from faecal samples and exhibits immense genetic diversity. Furthermore, the tools for genotyping *E. coli* are well-established (Dombek *et al.* 2000; Goldberg, Gillespie & Singer 2006). *Escherichia coli* is frequently used as an indicator for environmental faecal contamination (Tallon *et al.* 2005), and subtyping is a common method for tracing *E. coli* to its source (Simpson, Santo Domingo & Reasoner 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, Clabots & Kuskowski 2008). Recently, serotypes of shiga-toxin producing *E. coli* have become a major public health concern and are considered an emerging infectious disease (Beutin 2006). Transmission of *E. coli* generally occurs via a faecal–oral route, although flies can function as mechanical vectors (Förster *et al.* 2007; Fetene & Worku 2009).

This study was conducted at Ol Pejeta Conservancy (OPC), a 364 km^2 semi-arid savanna woodland ecosystem located in Laikipia, Kenya (0°N, 36°56′E). All giraffe within OPC were recognized using individually unique spot patterns on their necks. At the conclusion of this study, OPC had a population of 212 reticulated giraffe. Immigration and emigration were relatively negligible for this population because, except for a few narrow gaps, OPC is enclosed by a perimeter fence. Indeed, in the last 5 months of the study, only two new adults were discovered. Giraffes were aged according to height estimates and age-associated behaviours (VanderWaal *et al.* 2013b). 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. OPC's giraffe population consisted of 160 adults, 20 subadults and 32 juveniles. The population exhibited a 50 : 50 sex ratio.

FIELD OBSERVATIONS

From 21 January 2011 to 2 August 2011, giraffe group composition and membership were recorded for all giraffe groups sighted while driving daily survey routes. Routes were pre-determined so that different regions of the study area were surveyed in rotation, allowing for most of the study area to be surveyed once every 3 days. Routes were approximately 100 km in length, covered approximately 115 km^2 each, and traversed all habitat types. Giraffe groups observed from 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 solitary individual or set of individuals engaged in the same behaviour, or moving in the same direction or towards 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, we 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 observations 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 km^2 for adult males, 64·2 ± 3·4 km^2 for adult females and 51·0 ± 7·7 km^2 for juveniles. Giraffe were excluded from subsequent analyses if they were seen fewer than five times ($n = 2$) or if we failed to collect a faecal sample from them ($n = 14$).

FAECAL SAMPLE COLLECTION AND GENETIC ANALYSIS

We collected faecal samples from 194 giraffe. Because there can be significant turnover of *E. coli* subtypes in the gut (Anderson, Whitlock & Harwood 2006), faecal samples were collected during a brief time period to ensure comparability (10 August 2011 – 11 September 2011) Four faecal samples were collected after this period, but within 3 weeks of the end of the primary sampling period. Faecal 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 chromogenic agar that exhibits high specificity for *E. coli*. After overnight incubation at 37 °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), we 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). Genetic subtypes were determined using BOX-PCR and gel electrophoresis, which is a well-established method for discriminating between genetically similar *E. coli* subtypes (Cesaris *et al.* 2007; Mohapatra & Mazumder 2008). See Appendix S1 in the online supporting information for detailed laboratory procedures.

Densitometric profiles for each isolate were generated from the banding patterns revealed by BOX-PCR and gel electrophoresis (Johnson & O'Bryan 2000; Goldberg, Gillespie & Singer 2006). Similarity of each isolate to all others was determined through pairwise comparisons of densitometric curves (Fig. 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 our laboratory, this threshold value minimizes Type I errors in matching to 1% while limiting the Type II error rate to <5%. Among the 776 *E. coli* isolates that were analysed, 134 genetically distinct subtypes were found in the OPC giraffe population. On average, an individual giraffe hosted approximately 1·7 subtypes.

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 (i.e. two individuals either shared subtypes [1], or they did not [0]).

A social network was constructed from observed association patterns. We 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 ties 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-km^2 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 ties weighted according to the extent to which their home ranges overlapped.

Because water points may serve as an environmental source of bacteria, we 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. We 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 ties being weighted according to the number of water sources shared.

NETWORK-LEVEL ANALYSIS – TO WHAT EXTENT DO THE CONTACT NETWORKS PREDICT THE TRANSMISSION NETWORK?

If subtype sharing is dependent on social contact, then AS in the social network should be correlated with the occurrence of links in the transmission network. However, if environmental transmission dominates transmission dynamics, then correlations should be highest between the transmission and spatial networks. 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 with the other contact networks.

To determine the extent to which the contact networks (social, spatial, water-sharing) influenced the transmission network, we used the multiple regression quadratic assignment procedure (MR-QAP), a method of matrix regression developed for network data (Krackhardt 1988; Dekker, Krackhardt & Snijders 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 a link 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. Therefore, 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, Krackhardt & Snijders 2007). Using MR-QAP with double Dekker semipartialling and 1000 permutations (Dekker, Krackhardt & Snijders 2007), we investigated the effect of tie strength (AS, home-range overlap, shared water sources) in the contact networks on the log odds of a tie occurring in the transmission network. The transmission network was regressed against each contact network separately in univariate models and against the contact networks in combination with multivariate analyses (Table 1). Analysis was performed using the 'sna' package of R (Butts 2010).

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 individual-level connectivity: degree, betweenness, closeness, eigenvector centrality and information centrality (see Appendix S2 for metric definitions). Because many of these metrics were highly correlated with each other, we chose to focus our 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 animal 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 & 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 flow of pathogens through a network. Individuals with high betweenness can be considered to 'mediate' or 'regulate' flow of pathogens between sections of the network that would otherwise by poorly connected. These individuals have the potential to play a large role in regulating pathogen spread in the network (Borgatti 1995).

We also calculated these measures for each individual in each contact network. Because links in the social and spatial networks were weighted according to tie strength, we used

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

Model	Covariate	Coefficient	*P*-value
A	Association strength	1·22	0·02
B	Spatial overlap	0·03	0·91
C	Shared water sources	0·02	0·92
D	Association strength	2·26	0·07
	Spatial overlap	−0·50	0·25
E	Association strength	2·13	0·03
	Shared water sources	−0·41	0·19

weighted versions of degree and betweenness (Newman 2001; Opsahl 2009). Weighted degree (hereafter: social/spatial degree) simultaneously accounts for the number of individuals 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 links connected to the focal animal, regardless of the total number of neighbours. To illustrate the distinction between degree and overall tie strength, envision two individuals: individual A has only one neighbour, but they are linked with a tie strength of 10; individual B has 10 neighbours, but the tie strength with each neighbour is only 1. Overall tie strength does not distinguish between A and B's connectivity (summed weight of links = 10), but B is scored more highly for weighted degree because it has a greater diversity of neighbours. Thus, overall tie strength and weighted degree capture different elements of an individual's connectivity. Weighted measures were calculated in the 'tnet' package in R (Opsahl 2009).

To address whether an individual's position in a contact network influenced its position in the transmission network, we used general linear models (GLMs) to examine correlations between transmission degree and betweenness, home-range size, and connectivity in the contact networks. Measures of social and spatial network connectivity included weighted betweenness, weighted degree and overall tie strength. We 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 (Hanneman & Riddle 2005). The regression coefficient was recalculated after each permutation of the data (3000 total permutations), 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 coefficients more extreme than the observed value. Because degree exhibited a bimodal distribution, degree was ranked for GLM analyses. Higher ranks indicate higher degree.

Individual-level analyses of the water-sharing network were not included because this network was nearly identical to the spatial 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

We tested the extent to which the transmission network was predicted by three types of contact networks. Only AS 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 *E. coli* transmission chain (Table 1, Fig. 1). Contact among individuals via shared space use, whether measured by home-range overlap or shared water sources, was not significantly
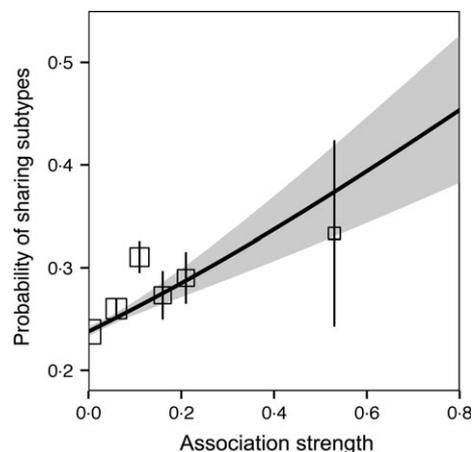

Fig. 1. Relationship between association strength (AS) in the social network and the probability of sharing genetic subtypes of *E. coli* in the transmission network. Shading indicates a 95% confidence interval around the regression line. Data points indicate the proportion of dyads (±SE) that shared subtypes. Dyads were binned by association strength (AS ≤ 0.05, n = 15 151 dyads; 0.05 < AS ≤ 0.1, n = 1991 dyads; 0.1 < AS ≤ 0.15, n = 873 dyads; 0.15 < AS ≤ 0.2, n = 355 dyads; 0.2 < AS ≤ 0.5, n = 324 dyads; AS > 0.5, n = 27 dyads).

correlated with sharing *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 Fig. 2, which depicts a subset of the transmission network (N = 30 randomly selected giraffe) alongside the spatial and social networks.

At the individual level, we investigated the extent to which an animal's position in one network was correlated with its position in another network (Fig. 3). Only models with significant coefficients are presented in Fig. 3 (See 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 MR-QAP analysis.

An individual's transmission betweenness was positively correlated with social betweenness, although this trend was not significant (P = 0.075). Transmission betweenness was unaffected by connectivity in the spatial network or measures of social degree. It is also interesting to note that social betweenness is positively correlated with spatial degree, 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 (social degree), but not higher social tie strength.
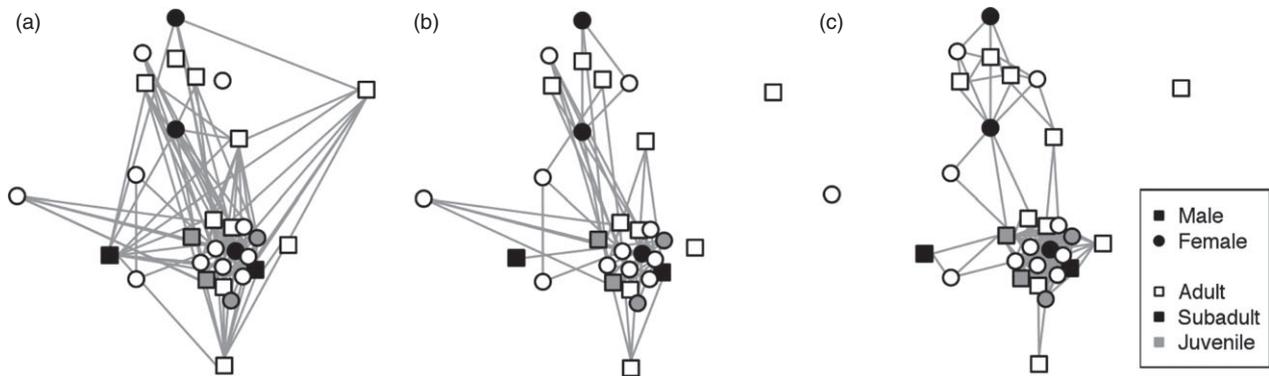
Fig. 2. Comparison of the (a) spatial, (b) social and (c) transmission networks. For visualization purposes, only links that exceed the mean tie strength + 1 SD are pictured in the social and spatial network (Home-range overlap > 36·4%, AS > 8·9%), and only a randomly selected subset of 30 individuals is shown.
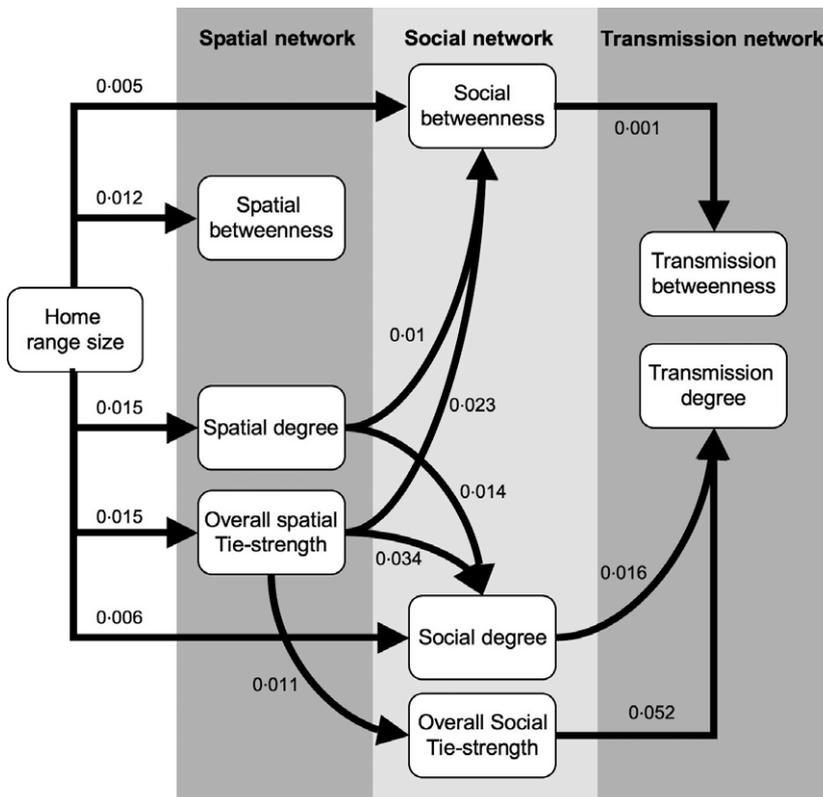

Fig. 3. Inter-relationships among individual-level measures of connectivity. All relationships depicted are positive and significant in univariate general linear models ($P < 0·05$), except for the relationships between transmission and social betweenness ($P = 0·075$) and between social betweenness and spatial degree ($P = 079$). 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 found 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 paralleled 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. Furthermore, giraffe with high social betweenness may bridge socially distinct clusters of individuals. These same individuals also tended to serve as bridges in the transmission network by occupying

positions of high flow (high transmission betweenness) and were more likely to be positioned along potential 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 & Jones 2010).

While there seems to be greater relative importance of social patterns over space use in determining the transmission network, our results do not negate a role for shared space use. Rather, the importance of spatial patterns on transmission seemed to be mediated through the role that the spatial network played in determining the social network. Specifically, individuals with high degree in the spatial network tended to have higher social degree and betweenness, which were 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 and social betweenness (Fig. 3). In addition, pairs of individuals that had greater home-range overlap had significantly higher AS (VanderWaal *et al.* 2013b), and AS was positively correlated with the likelihood of sharing *E. coli* subtypes. These results indicate that space-use patterns play a key role in determining the social network, which in turn affects pathogen transmission patterns.

Indeed, when the contact networks are visually compared (Fig. 2a,b), the social network appears to be a pared-down, or filtered, version of the spatial network. While overlapping in space was a prerequisite for social interaction, some giraffe that shared space did not associate. The social network can be considered to simply apply a spatiotemporal filter to the spatial overlap data: dyads must not only share space, but also be at the same place at the same time to be linked in the social network. The spatial ties that were eliminated by this spatiotemporal (social) filter were ones that provided little information about transmission opportunities. It is also apparent that not all social contacts resulted in sharing *E. coli* subtypes (Fig. 2c). 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 *E. coli* is primarily environmentally transmitted through the ingestion of faecal contaminated water and forage, although 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 *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 & Isbell 1991), and forage contamination by gastrointestinal parasites is minimized at higher feeding heights (Apio, Plath & Wronski 2006). Perhaps more likely, however, is

that social associations led to synchronous space use. Even though *E. coli* can persist in the environment for months, densities rapidly drop off in the first 2 weeks (Medema, Bahar & Schets 1997; Avery *et al.* 2008) and inactivation by sunlight further reduces populations (Sinton, Hall & Braithwaite 2007). 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 waterborne *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, previous studies have typically used an individual's social network connectivity 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, McKenzie & Stevenson 2011), or animals that engage in certain types of behaviour are at higher risk (Drewe 2009). In a recent study that incorporated microbial genetics, the approach was similar to previous risk factor analyses in that social network connectivity was analysed as a risk factor for being infected with specific genetic strains of *Salmonella* (Bull, Godfrey & Gordon 2012). Bull, Godfrey & Gordon (2012) did perform a dyad-level analysis on the likelihood of two individuals sharing *Salmonella* strains, and similar to our study, they found that strain sharing was influenced by the strength of the pair's social relationship rather than spatial proximity. Our approach builds upon their work by conceptualizing the strain-sharing data as a distinct network and then quantifying individual-level connectivity in this transmission network. Thus, our approach allows us to directly compare contact networks to transmission networks. 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, characterizing super-spreaders in transmission networks, or the effect of environmental change on the structure and connectivity of transmission networks.

In the novel approach used here, we integrated host behavioural 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 faecal–oral pathogens that are epidemiologically similar. Although this study focused on a commensal organism, this is the first study to our knowledge 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 interspecific routes of transmission.

Acknowledgements

References

Ames, G.M., George, D.B., Hampson, C.P., Kanerek, A.R., McBee, C.D., Lockwood, D.R. *et al.* (2011) Using network properties to predict disease dynamics on human contact networks. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **278**, 3544–3550.

Anderson, M.A., Whitlock, J.E. & Harwood, V.J. (2006) Diversity and distribution of *Escherichia coli* genotypes and antibiotic resistance phenotypes in feces of humans, cattle, and horses. *Applied and Environmental Microbiology*, **72**, 6914–6922.

Apio, A., Plath, M. & Wronski, T. (2006) Foraging height levels and the risk of gastro-intestinal tract parasitic infections of wild ungulates in an African savannah eco-system. *Helminthologia*, **43**, 134–138.

Archie, E.A., Luikart, G. & Ezenwa, V.O. (2008) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution*, **24**, 21–30.

Avery, L.M., Williams, A.P., Killham, K. & Jones, D.L. (2008) Survival of *Escherichia coli* O157:H7 in waters from lakes, rivers, puddles and animal-drinking troughs. *Science of the Total Environment*, **389**, 378–385.

Bansal, S., Grenfell, B.T. & Meyers, L.A. (2007) When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*, **4**, 879–891.

Bashaw, M.J., Bloomsmith, M.A., Maple, T.L. & Bercovitch, F.B. (2007) The structure of social relationships among captive female giraffe (*Giraffa camelopardalis*). *Journal of Comparative Psychology*, **121**, 46–53.

Besser, T.E., Richards, B.L., Rice, D.H. & Hancock, D.D. (2001) *Escherichia coli* O157:H7 infection of calves: infectious dose and direct contact transmission. *Epidemiology and Infection*, **127**, 555–560.

Beutin, L. (2006) Emerging enterohaemorrhagic *Escherichia coli*, causes and effects of the rise of a human pathogen. *Journal of Veterinary Medicine Series B – Infectious Diseases and Veterinary Public Health*, **53**, 299–305.

Böhm, M., Hutchings, M.R. & White, P.C.L. (2009) Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine TB among badgers and cattle. *PLoS One*, **4**, e5016.

Borgatti, S.P. (1995) Centrality and AIDS. *Connections*, **18**, 112–115.

Bull, C.M., Godfrey, S.S. & Gordon, D.M. (2012) Social networks and the spread of *Salmonella* in a sleepy lizard population. *Molecular Ecology*, **21**, 4386–4392.

Butts, C.T. (2010) *sna: tools for social network analysis*. R package version 2.2-0.

Caley, P., Marion, G. & Hutchings, M.R. (2009) Assessment of transmission rates and routes, and the implications for management. *Management of Disease in Wild Mammals* (eds R.J. Delahay, G.C. Smith & M.R. Hutchings), pp. 31–52. Springer, New York.

Cesaris, L., Gillespie, B.E., Srinivasan, V., Almeida, R.A., Zecconi, A. & Oliver, S.P. (2007) Discriminating between strains of *Escherichia coli* using pulse-field gel electrophoresis and BOX-PCR. *Foodborne Pathogens and Disease*, **4**, 473–480.

Corner, L.A.L., Pfeiffer, D.U. & Morris, R.S. (2003) Social-network analysis of *Mycobacterium bovis* transmission among captive brushtail possums (*Trichosurus vulpecula*). *Preventive Veterinary Medicine*, **59**, 147–167.

Craft, M.E. & Caillaud, D. (2011) Network models: an underutilized tool in wildlife epidemiology. *Interdisciplinary Perspectives on Infectious Diseases*, **2011**, 1–12.

Craft, M.E., Volz, E., Packer, C. & Meyers, L.A. (2010) Disease transmission in territorial populations: the small-world network of Serengeti lions. *Journal of the Royal Society Interface*, **8**, 776–786.

Croft, D.P., James, R. & Krause, J. (2008) *Exploring Animal Social Networks*. Princeton University Press, Princeton, NJ.

Cross, P.C., Lloyd-Smith, J.O., Bowers, J.A., Hay, C.T., Hofmeyr, M. & Getz, W.M. (2004) 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.

Dekker, D., Krackhardt, D. & Snijders, T.A.B. (2007) Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, **72**, 563–581.

Delahay, R.J., Smith, G.C. & Hutchings, M.R. (2009) *Management of Disease in Wild Mammals*. Springer, New York.

Dombek, P.E., Johnson, L.K., Zimmerley, S.T. & Sadowsky, M.J. (2000) Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Applied and Environmental Microbiology*, **66**, 2572–2577.

Drewe, J.A. (2009) Who infects whom? Social networks and tuberculosis transmission in wild meerkats. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **277**, 633–642.

Fennessy, J. (2004) *Ecology of Desert-Dwelling Giraffe Giraffa camelopardalis angolensis* in Northwestern Namibia. pp. 223. University of Sydney, Sydney.

Fetene, T. & Worku, N. (2009) Public health importance of non-biting cyclorrhaphan flies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **103**, 187–191.

Förster, M., Klimpel, S., Mehlhorn, H., Sievert, K., Messler, S. & Pfeffer, K. (2007) Pilot study on synanthropic flies (e.g. *Musca, Sarcophaga, Calliphora, Fannia, Lucilia, Stomoxys*) as vectors of pathogenic microorganisms. *Parasitology Research*, **101**, 243–246.

Godfrey, S.S., Bull, C.M., James, R. & Murray, K. (2009) Network structure and parasite transmission in a group living lizard, gidgee skink, *Egernia stokesii*. *Behavioral Ecology and Sociobiology*, **63**, 1045–1056.

Godfrey, S.S., Moore, J.A., Nelson, N.J. & Bull, C.M. (2010) Social network structure and parasite infection patterns in a territorial reptile, the tuatara (*Sphenodon punctatus*). *International Journal for Parasitology*, **40**, 1575–1585.

Goldberg, T.L., Gillespie, T.R. & Singer, R.S. (2006) Optimization of analytical parameters for inferring relationships among *Escherichia coli* isolates from repetitive-element PCR by maximizing correspondence with multilocus sequence typing data. *Applied and Environmental Microbiology*, **72**, 6049–6052.

Goldberg, T.L., Gillespie, T.R., Rwego, I.B., Estoff, E.L. & Chapman, C.A. (2008) Forest fragmentation as cause of bacterial transmission among nonhuman primates, humans, and livestock, Uganda. *Emerging Infectious Diseases*, **14**, 1375–1382.

Grear, D.A., Perkins, S.E. & Hudson, P.J. (2009) Does elevated testosterone result in increased exposure and transmission of parasites? *Ecology Letters*, **12**, 528–537.

Griffin, R.H. & Nunn, C.L. (2012) Community structure and the spread of infectious disease in primate social networks. *Evolutionary Ecology*, **26**, 779–800.

Hamede, R.K., Bashford, J., McCallum, H. & Jones, M. (2009) Contact networks in a wild Tasmanian devil (*Sarcophilus harrisii*) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecology Letters*, **12**, 1147–1157.

Hanneman, R.A. & Riddle, M. (2005) *Introduction to Social Network Methods*. University of California, Riverside, CA.

Harris, S., Cresswell, W.J., Forde, P.G., Trewhella, W.J., Woollard, T. & Wray, S. (1999) Home-range analysis using radio-tracking data – a review of problems and techniques particularly as applied to the study of mammals. *Mammal Review*, **29**, 97–123.

Henderson, H. (2008) Direct and indirect zoonotic transmission of Shiga toxin-producing *Escherichia coli*. *Journal of the American Veterinary Medicine Association*, **232**, 848–859.

Johnson, J.R., Clabots, C. & Kuskowski, M.A. (2008) Multiple-host sharing, long-term persistence, and virulence of *Escherichia coli* clones from human and animal household members. *Journal of Clinical Microbiology*, **46**, 4078–4082.

Johnson, J.R. & O'Bryan, T.T. (2000) Improved repetitive-element PCR fingerprinting for resolving pathogenic and nonpathogenic phylogenetic groups within *Escherichia coli*. *Clinical and Diagnostic Laboratory Immunology*, **7**, 265–273.

Keeling, M.J. (1999) The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **266**, 859–867.

Keeling, M.J. (2005) The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, **67**, 1–8.

Keeling, M.J. & Eames, K.T.D. (2005) Networks and epidemic models. *Journal of the Royal Society Interface*, **2**, 295–307.

Krackhardt, D. (1988) Predicting with networks: nonparametric multiple regression analysis of dyadic data. *Social Networks*, **10**, 359–381.

Leuthold, B.M. (1979) Social organization and behaviour of giraffe in Tsavo East National Park. *African Journal of Ecology*, **17**, 19–34.

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

MacIntosh, A.J.J., Jacobs, A., Garcia, C., Shimizu, K., Mouri, K., Huffman, M.A. *et al.* (2012) Monkeys in the middle: parasite transmission through the social network of a wild primate. *PLoS One*, **7**, e51144.

May, R.M. (2006) Network structure and the biology of populations. *Trends in Ecology & Evolution*, **21**, 394–399.

Medema, G.J., Bahar, M. & Schets, F.M. (1997) Survival of *Cryptosporidium parvum, Escherichia coli,* feacal enterococci and *Clostridium perfringens* in river water: influence of temperature and autochthonous microoogransism. *Water Science Technology*, **35**, 249–252.

Mohapatra, B.R. & Mazumder, A. (2008) Comparative efficacy of five different rep-PCR methods to discriminate *Escherichia coli* populations in aquatic environments. *Water Science and Technology*, **58**, 537–547.

Newman, M.E.J. (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, **64**, 06132.

Opsahl, T. (2009) *Structure and Evolution of Weighted Networks*. pp. 104–122. University of London (Queen Mary College), London, UK.

Otterstatter, M.C. & Thomson, J.D. (2007) Contact networks and transmission of an intestinal pathogen in bumble bee (*Bombus impatiens*) colonies. *Oecologia*, **154**, 411–421.

Perkins, S.E., Ferrari, M.F. & Hudson, P.J. (2008) The effects of social structure and sex-biased transmission on macroparasite infection. *Parasitology*, **135**, 1561–1569.

Perkins, S.E., Cagnacci, F., Stradiotto, A., Arnoldi, D. & Hudson, P.J. (2009) Comparison of social networks derived from ecological data: implications for inferring infectious disease dynamics. *Journal of Animal Ecology*, **78**, 1015–1022.

Porphyre, T., McKenzie, J. & Stevenson, M.A. (2011) Contact patterns as a risk factor for bovine tuberculosis infection in a free-living adult brushtail possum *Trichosurus vulpecula* population. *Preventive Veterinary Medicine*, **100**, 221–230.

Salathé, M. & Jones, J.H. (2010) Dynamics and control of diseases in networks with community structure. *PLoS Compuational Biology*, **65**, e1000736.

Simpson, J.M., Santo Domingo, J.W. & Reasoner, D.J. (2002) Microbial source tracking: state of the science. *Environmental Science and Technology*, **36**, 5279–5288.

Singer, R.S., Johnson, W.O., Jeffrey, J.S., Chin, R.P., Carpenter, T.E., Atwill, E.R. *et al.* (2000) A statistical model for assessing sample size for bacterial colony selections: a case study of *Escherichia coli* and avian cellulitis. *Journal of Veterinary Diagnostic Investigation*, **12**, 118–125.

Sinton, L., Hall, C. & Braithwaite, R. (2007) Sunlight inactivation of *Campylobacter jejuni* and *Salmonella enterica*, compared with *Escherichia coli*, in seawater and river water. *Journal of Water and Health*, **5**, 357–365.

Tallon, P., Magajna, B., Lofranco, C. & Leung, K.T. (2005) Microbial indicators of faecal contamination in water: a current perspective. *Water, Air, and Soil Pollution*, **166**, 139–166.

Tompkins, D.M., Dunn, A.M., Smith, M.J. & Telfer, S. (2011) Wildlife diseases: from individuals to ecosystems. *Journal of Animal Ecology*, **80**, 19–38.

Turner, J., Bowers, J.A., Clancy, D., Behnke, M.C. & Christley, R.M. (2008) A network model of *E coli* O157 transmission within a typical UK dairy herd: the effect of heterogeneity and clustering on the prevalence of infection. *Journal of Theoretical Biology*, **254**, 45–54.

VanderWaal, K.L., Atwill, E.R., Hooper, S., Buckle, K. & McCowan, B. (2013a) Network structure and prevalence of Cryptosporidium in Belding's ground squirrels. *Behavioral Ecology and Sociobiology*, doi: 10.1007/s00265-013-1602-x

VanderWaal, K.L., Wang, H., McCowan, B., Fushing, H. & Isbell, L.A. (2013b) Multilevel social organization and space use in reticulated Giraffe (*Giraffa camelopardalis*). *Behavioral Ecology*, Advance Access, doi: 10.1093/beheco/art061

Wasserman, S. & Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

Young, T.P. & Isbell, L.A. (1991) Sex differences in giraffe feeding ecology: energetic and social constraints. *Ethology*, **87**, 79–89.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Laboratory methods and genetic analysis.

Appendix S2. Definitions of individual-level network metrics.

Fig. S1. Diagram depicting BOX-PCR banding patterns of 10 *E. coli* isolates corresponding with ten different giraffe.

Table S1. Results of univariate Poisson GLM regressions examining the correlations between individual connectivity values across networks ($N = 194$ individuals).