

LETTER

Phylogenetic host specificity and understanding parasite sharing in primates

Natalie Cooper,^{1,2,3*} Randi Griffin,¹ Mathias Franz,⁴ Moshood Omotayo⁵ and Charles L. Nunn¹

Abstract

Understanding how parasites are transmitted to new species is of great importance for human health, agriculture and conservation. However, it is still unclear why some parasites are shared by many species, while others have only one host. Using a new measure of 'phylogenetic host specificity', we find that most primate parasites with more than one host are phylogenetic generalists, infecting less closely related primates than expected. Evolutionary models suggest that phylogenetic host generalism is driven by a mixture of host–parasite cospeciation and lower rates of parasite extinction. We also show that phylogenetic relatedness is important in most analyses, but fails to fully explain patterns of parasite sharing among primates. Host ecology and geographical distribution emerged as key additional factors that influence contacts among hosts to facilitate sharing. Greater understanding of these factors is therefore crucial to improve our ability to predict future infectious disease risks.

Keywords

Generalist, helminth, intermediate host, phylogenetic comparative methods, protozoa, specialist, vector, virus.

Ecology Letters (2012) 15: 1370–1377

INTRODUCTION

Emerging infectious diseases (EIDs) present a major challenge to the health of humans and our domesticated animals and crops (Morens *et al.* 2004; Woolhouse *et al.* 2005). EIDs also affect wild animals and plants, and can be major drivers of wildlife population declines (e.g. amphibians and Tasmanian devils; McCallum & Jones 2006; Kilpatrick *et al.* 2010). The source of EIDs is often other species; for example, many recent human pandemics originated in wildlife, including human immunodeficiency virus (HIV) and severe acute respiratory syndrome (SARS), and parasite introduction via introduced species is a major driver of plant EID emergence (Anderson *et al.* 2004; Wolfe *et al.* 2005, 2007; Jones *et al.* 2008). If we understand the factors that increase the likelihood of parasites switching from one host species to another, we can identify problematic parasites before they are transmitted to new hosts. This would be advantageous for public health reasons, and for agriculture and the conservation of biodiversity.

The most reliable way to determine whether a parasite – here defined as any infectious organism – can switch from one species to another is to experimentally infect the new host in the laboratory. However, this is expensive, time consuming and raises ethical concerns for some host taxa. Another solution is to take a comparative approach that considers how parasites are shared among host species in nature. Patterns of present-day parasite sharing will reflect (among other things) historical transfers of parasites from one host to another. Thus, we can use the degree of parasite sharing among hosts to make inferences about the host and parasite characteristics

that promote host switching. To do this, we must first consider the steps involved in a host shift. Initially, the new host must make contact with the parasite, the likelihood of which will depend on factors that include whether the host and parasite have overlapping geographical ranges; the parasite's transmission mode (e.g. sexually transmitted diseases are less likely to be encountered by new hosts than environmentally transmitted diseases); the parasite's abundance in the original host; and similarities in the ecology and behaviour of the hosts. Once a new host is exposed, the probability that the parasite can replicate and spread to other individuals depends on the host's immune defences and the parasite's ability to circumvent these defences (Woolhouse *et al.* 2005; Parrish *et al.* 2008).

Previous authors have suggested that each stage of the host-shift process is more likely if the original host and the new host are closely related, because closely related species are likely to be similar in ecological, behavioural, physiological and distributional traits that increase the likelihood of exposure to a parasite and the ability of the parasite to circumvent the host's immune defences (Woolhouse *et al.* 2005; Engelstädter & Hurst 2006; Wolfe *et al.* 2007; Davies & Pedersen 2008). Indeed, in a range of host–parasite systems, parasites are more commonly shared among close relatives than among more distant relatives (e.g. Davies & Pedersen 2008; Krasnov *et al.* 2010; Poulin 2010; Streicker *et al.* 2010; Longdon *et al.* 2011). However, processes other than host switching also influence patterns of parasite sharing (summarised in Fig. S1; see Appendix S1 in Supporting Information). Close relatives also inherit parasites from their last common ancestor, and thus should share more parasites through common descent (this is referred to as cospeciation;

¹Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, 02138, USA

²School of Natural Sciences, Trinity College Dublin, College Green, Dublin, Ireland

³Trinity Centre for Biodiversity Research, Trinity College Dublin, Dublin, Ireland

⁴Courant Research Center Evolution of Social Behavior, Georg-August-Universität Göttingen, Kellnerweg 6, 37077, Göttingen, Germany

⁵Harvard School of Public Health, 677 Huntington Avenue, Boston, MA, 02115, USA

*Correspondence: E-mail: ncooper@tcd.ie

Engelstädter & Hurst 2006). In addition, parasites may be missing from some species simply because the host has been studied insufficiently to detect the parasite, or because the parasite has gone extinct.

Here, we use phylogenetic comparative methods to identify factors that influence parasite sharing among primate hosts. We use primates because they are well studied and also our closest relatives, allowing us to make inferences about parasite sharing among primates and humans. Using a measure of 'phylogenetic host specificity' (or 'phylospecificity'; Poulin *et al.* 2011), we first investigate whether parasites infect more closely related hosts than expected by chance. Our measure differs from standard measures of host specificity because it takes into account the relatedness of the hosts, rather than relying on taxonomic or other categories (e.g. Pedersen *et al.* 2005). With this measure, a parasite that infects distantly related hosts is a phylogenetic host generalist, whereas a parasite that infects closely related hosts is a phylogenetic host specialist. We predict that viruses will be phylogenetic host generalists because the fast rates of viral evolution should allow them to easily adapt to new hosts (Woolhouse & Gowtage-Sequeria 2005; Parrish *et al.* 2008; Elena & Froissart 2010). Conversely, we predict that helminths will be phylogenetic host specialists, due to their often complex lifecycles that may constrain their ability to adapt to new hosts (Woolhouse *et al.* 2005; Parrish *et al.* 2008). We also use evolutionary models to investigate whether patterns of sharing are driven by host shifts, host–parasite cospeciation, parasite extinction, or biases in sampling effort (summarised in Fig. S1).

Next, we investigate patterns of parasite sharing among primates in relation to primate phylogeny, ecology and geography, and parasite phylogenetic host specificity. We predict that parasite sharing will be highest among closely related, ecologically similar primates that live in close proximity, as these factors should favour parasite transmission. We also expect that the relationship between parasite sharing and host phylogeny will be stronger in phylogenetically host-specialised parasites. Finally, we use phylogenetic models to investigate levels of parasite sharing among primates and humans.

MATERIALS AND METHODS

Host–parasite data

We obtained 2867 host–parasite combinations, representing 128 primate species and 437 parasite species (60 viruses, 247 helminths, 91 protozoa, 35 bacteria and four fungi), from the *Global Mammal Parasite Database* (GMPD, accessed 25 July 2011; Nunn & Altizer 2005). We took our list of endemic human parasites from Taylor *et al.* (2001).

Primate data

For each species, we collated data on adult body mass (g) and social group size from various sources (de Magalhães & Costa 2009; Jones *et al.* 2009; Nowak 1999 plus references in Appendix S2). We also defined each species as terrestrial (> 90% of time on ground), semi-terrestrial (< 90%, but > 50% of time on ground), semi-arboreal (< 90%, but > 50% of time in trees) or arboreal (> 90% of time in trees).

We downloaded geographical range maps from the IUCN (2010) and used these to estimate the geographical range size of

each species (km²). For each pair of primates we then determined their geographical range overlap (km²) and the geodesic distance between their geographical range centroids (km). We obtained mean annual temperature (0.1 °C) and precipitation (mm) data from WorldClim (Hijmans *et al.* 2005) and extracted mean temperature and precipitation values for each primate's geographical range. We used the R packages *maptools* (Lewin-Koh & Bivand 2011), *PBS-mapping* (Schnute *et al.* 2010), *raster* (Hijmans & van Etten 2011), *rgdal* (Keitt *et al.* 2011) and *sp* (Pebesma & Bivand 2005) for all geographical data manipulation.

As a measure of sampling effort, we collected citation counts for each primate species, i.e. the number of references retrieved from the ISI Web of Knowledge (<http://wokinfo.com/>; accessed 5 July 2012), where the Latin binomial of the species appeared in either the title or topic fields, and either 'parasite', 'disease' or 'pathogen' also appeared in the topic field. Where the species binomial had changed between 1993 and 2005 (Wilson & Reeder 1993, 2005), we summed the number of citations for the species names from both taxonomies. We used the dated consensus phylogeny from *10kTrees* version 3 (Arnold *et al.* 2010) and the taxonomy of Wilson & Reeder (2005) for all analyses.

Parasite data

We defined each parasite's geographical range as the union of the geographical ranges of its hosts (GR1). However, it is unlikely that a parasite will be found across the whole geographical range of a host, especially where hosts have large ranges. Therefore, we derived a second parasite geographical range (GR2). GR2 was equal to the area encompassed by the minimum convex polygon created using the host–parasite latitude and longitude locality records for a given parasite, plus a five degree buffer. To estimate GR2, we obtained latitude and longitude coordinates for each host–parasite combination from the GMPD. We excluded localities that we were unable to unambiguously georeference, and localities with an extent greater than 10 000 km². We then extracted geographical ranges for our parasites using the R packages cited above plus *rgeos* (Bivand & Rundel 2011) and *wild1* (Sargeant 2011). Parasite geographical ranges are probably highly dynamic and vary seasonally; thus, we lack information on which of these ranges best represents the true geographical range of the parasite. Therefore, we performed analyses using both GR1 and GR2. We report only results using GR2; however, our GR1 results are qualitatively similar.

Sampling biases

Host–parasite data are sensitive to sampling effort: hosts which have been thoroughly sampled for parasites may appear to have more parasites than those which have been less well sampled (Gregory *et al.* 1996). To deal with this issue, we focused on well-sampled primates, defined as species with at least 10 different parasites, and evidence for saturation in their parasite accumulation curves (Fig. S2). This left 35 primate species, 2144 host–parasite combinations and 346 parasite species (52 viruses, 180 helminths, 78 protozoa, 33 bacteria and three fungi). We report results using the 35 well-sampled primate species; however, our results were qualitatively similar when we used all primate species. All data are available in Appendix S3.

Analyses

Phylogenetic host specificity

We first investigated whether primate parasites infect more closely related hosts than expected by chance by using the net relatedness index (NRI) to investigate the ‘phylogenetic host specificity’ (or ‘phylospecificity’; Poulin *et al.* 2011) of each parasite (Webb *et al.* 2002). NRI is based on the mean phylogenetic distance (MPD) between all possible pairs of hosts infected by a parasite (MPD_{obs}), where phylogenetic distance is defined as the sum of all intervening branch lengths between two hosts. To allow comparisons among multiple parasites, we standardised these MPD values by (i) subtracting the mean MPD expected for *n* hosts drawn at random from the host phylogeny across 999 iterations (MPD_n) and then (ii) dividing by the standard deviation of the MPD from these 999 randomly drawn *n* hosts [*s*(MPD_n)]. Finally, we multiplied these values by -1 so that positive values of NRI reflect phylogenetic host specialisation. Thus, NRI is calculated as follows:

$$\text{NRI} = -1 * ((\text{MPD}_{\text{obs}} - \text{MPD}_n) / s(\text{MPD}_n)) \quad (1)$$

To test for statistical significance in phylogenetic host specificity, we compared MPD_{obs} values with those from the 999 randomly generated MPD values. A parasite was considered significantly phylogenetically host specific if less than 5% of these random MPD values were larger than MPD_{obs} ($P < 0.05$).

We estimated NRI values for each parasite that infected at least two of our 35 primate hosts ($N = 141$) using the R package picante (Kembel *et al.* 2010). We used chi-squared tests to determine whether different groups of parasites (virus, protozoa or helminth), or parasites transmitted by vectors or intermediate hosts rather than by more direct means, had significantly different proportions of phylogenetic host-specialist parasites. We also ran simulations to investigate whether sampling effort affected our power to detect significant phylogenetic host specificity (Appendix S4).

Evolutionary models

Patterns of phylogenetic host specificity are the result of multiple processes including host shifts, parasite extinctions and host–parasite cospeciation (summarised in Fig. S1). To better understand these mechanisms, we used maximum-likelihood methods to compare three evolutionary models: (i) host shifts and parasite extinctions occur at different rates, (ii) host shifts and parasite extinctions occur at equal rates and (iii) only parasite extinctions occur. We computed the maximum likelihood for each model using the Multi-State option in BayesTraits (Pagel *et al.* 2004). This fits a continuous-time Markov model to discrete character data distributed on the tips of a phylogenetic tree by traversing the phylogeny and estimating rate parameters associated with transitions among character states (Pagel *et al.* 2004). We then used the Akaike Information Criterion (AIC) to select the best model for each parasite, defining models with greater than 2 units difference in AIC as different. We also compared estimated rates of host shifts (host gain) and parasite extinction (host loss) taken from model (i) to determine which process occurred at a higher rate.

Parasite sharing among primates

We estimated the similarity of parasite communities among pairs of primate hosts using either Jaccard’s index (Jaccard 1922):

$$\text{Parasite similarity} = a / (a + b + c) \quad (2)$$

or a modified version of Jaccard’s index, calculated as follows:

$$\text{Parasite similarity} = ((a + e) / \gamma) / ((a + b + c + e) / \gamma) \quad (3)$$

where *a* is the number of parasites shared by host species B and C, *b* is the number of parasites unique to species B, *c* is the number of parasites unique to species C and *e* is the number of parasites present in the shared parasite species pool (γ) but missing from species B and C. We defined the shared parasite species pool, γ , as all the parasites with geographical ranges that overlap with the geographical ranges of both species B and C. We divide by γ so that parasite similarity values range from 0 to 1, where 0 indicates that species B and C have no shared parasites nor any shared absences of parasites, to 1 where species B and C share all their parasites and/or absences of parasites. We include this modified version because Jaccard’s index does not include information on shared absences of species (Anderson *et al.* 2011), and in terms of parasite transmission likelihoods, whether two species both lack a parasite may be informative (e.g. it may indicate that the parasite lacks the ability to circumvent the immune system of both species). Because parasite similarity values are bounded between 0 and 1 and thus produce non-normal error distributions, we logit transformed our similarity indices, after first adding the minimum non-zero value of parasite similarity to each value (because zeros cannot be log transformed; Warton & Hui 2011). Note that where we report predicted values and intercepts, we subtract this value from our estimates.

After calculating parasite similarity for all primate pairs, we fit linear models of each similarity index against the time since the species’ most recent common ancestor, i.e. divergence time (Davies & Pedersen 2008). Previous studies have shown that divergence time does not fully explain variation in parasite sharing (e.g. Davies & Pedersen 2008), which could be the result of differences in the ecology or distribution of host pairs. We therefore fit linear models of parasite similarity against divergence time with all possible combinations of the following variables: the difference between each species pair in adult body mass, social group size, terrestriality (as a ranked ordinal variable), mean annual temperature and precipitation, the geographical range overlap of each pair of species and the distance between their geographical range centroids. All predictor variables except terrestriality were natural-log transformed prior to analysis. We then used model averaging with the R package MuMIn (Barton 2011) to summarise the 95% confidence set of models for explaining variation in parasite similarity with divergence time as a fixed variable in all models. We repeated all analyses using subsets of the data restricted to phylogenetic host specialist or generalist parasites, viruses, protozoa and helminths separately.

To determine whether sampling effort could bias our results and account for the relationship between parasite similarity and divergence time, we fit models of parasite similarity against the sum of citation counts from species B and C, and the difference in citation counts from species B and C. Both predictors were natural-log transformed prior to the analysis. We also note that because each species appears multiple times in our pairwise measures of parasite similarity, our models contain pseudoreplication. However, each pair of primates produces an independent value of parasite similarity and, as we are using these values in the models rather than species

values, we feel this is not a major issue. A more important issue is pseudoreplication in divergence times, i.e. each New World monkey vs. Old World monkey comparison has the same divergence time, leading to many points with the same value. To determine whether this influenced our results, we repeated the analyses 1000 times, randomly selecting only one species pair for each divergence time on each iteration, fitting the models and then averaging the results across all 1000 iterations. Our results are qualitatively similar (Table S1).

Human–primate parasite sharing predictions

We used the parasite sharing models described above to predict levels of parasite sharing among primates and humans. We used only human parasites which also occurred in our primate dataset so we could classify parasites as phylogenetic host specialists or generalists. We assumed that humans could potentially contract any parasite due to our cosmopolitan distribution.

RESULTS

Phylogenetic host specificity

Of 141 parasites, 58 were identified as phylogenetic host specialists, i.e. they infect more closely-related primate species than expected by chance (Table 1; Fig. 1; Appendix S3). As predicted, viruses had lower phylogenetic host specificity than protozoa or helminths ($\chi^2 = 9.745$, d.f. = 2, $P = 0.008$; Table 1), but there were no significant differences among transmission modes ($\chi^2 = 1.554$, d.f. = 1, $P = 0.213$; Table 1). When parasites with a single host were included as specialists, helminths had much higher phylogenetic host specificity than protozoa or viruses ($\chi^2 = 11.57$, d.f. = 2, $P = 0.003$; Table 1), but differences among transmission modes remained non-significant ($\chi^2 = 0.004$, d.f. = 1, $P = 0.9496$; Table 1). Simulations revealed that our method had very high power to detect significant phylogenetic host specificity even under heterogeneous host sampling (Appendix S4). Our results were qualitatively similar when we used only well-sampled parasites (i.e. those with > 4 hosts from our subset of 35 well-sampled primates).

Evolutionary models

Evolutionary model (i), where hosts were gained and lost at different rates, was the best fitting model for 89 of our 141 parasites,

Table 1 The distribution of phylogenetic host-specialist and -generalist parasites across different types of parasite

Parasite subset	<i>n</i>	Specialist	Generalist	Singletons
All	346	58	83	205
VI	187	38	44	105
noVI	164	19	41	104
Viruses	52	4	21	27
Protozoa	78	25	22	31
Helminths	180	28	31	121
Bacteria	33	1	8	24
Fungi	3	0	1	2

Singletons are parasites with only one host. VI = parasites transmitted by vectors or intermediate hosts; noVI = parasites transmitted sexually, vertically or by close or non-close contact.

suggesting that their transmission history included a mixture of host shifts, host–parasite cospeciation and parasite extinction (Appendix S3). The remaining 52 parasites showed support for more than one model, although model (i) was always one of the equally supported models. Rates of host gain in these models were much lower than rates of host loss, suggesting that the patterns are driven mainly by host–parasite cospeciation and parasite extinction, rather than extensive host switching (Appendix S3). These results did not differ when phylogenetic host generalist and specialist parasites were considered separately. We also found no significant differences in the host gain to host loss rate ratio between phylogenetic host generalist and specialist parasites (ANOVA: $F_{1,140} = 0.133$, $P = 0.716$), or among parasite types (ANOVA: $F_{4,136} = 1.868$, $P = 0.119$).

Parasite sharing among primates

As predicted, we found a significant negative relationship between divergence time and parasite similarity ($r^2 = 0.331$), and this relationship held across all transmission modes and parasite types (Fig. 2; Fig. S3; Table S2) using either of the Jaccard's indices (Table S5). This relationship is strongest in phylogenetic host-specialist parasites ($r^2 = 0.411$; phylogenetic host generalists: $r^2 = 0.161$), with the greatest scatter about the regression line in viruses ($r^2 = 0.116$), but with little difference between transmission modes (Table S2 and Table S5).

The estimated intercept of the relationship between divergence time and parasite similarity for all parasites was 0.793 ± 0.022 , whereas theoretically it should be one (i.e. when divergence time is zero parasite communities should be identical, although this will be true only when all parasites in a host are recorded without sampling error). The lower than expected intercept suggests that parasite sharing among primates is not simply the result of phylogenetic relationships. When phylogenetic host-specialist parasites alone were considered, the intercept was 0.903 ± 0.018 , which is much closer to one, vs. 0.449 ± 0.033 in phylogenetic host-generalist parasites, suggesting that phylogenetic relatedness is more important for parasite sharing in phylogenetic host specialists than in phylogenetic host generalists.

Statistical models including ecology and geography explained substantially more of the variation in parasite sharing than models including only divergence time ('best' model, all parasites, $r^2 = 0.613$; Table 2; Table S3 and Table S6). Other than divergence time, the most important predictors in our models were the distance between host geographical range centroids, the body mass difference between hosts and the difference in mean annual temperature across the species' ranges. As predicted, parasite sharing increased as differences decreased, indicating that greater geographical and ecological similarity increases parasite similarity (Table 2; Table S3). The relative importance of these factors varied with subsets of the data; only the distance between host geographical range centroids was consistently found in all of our 95% confidence sets of models. Thus, greater distances resulted in lower parasite similarity, consistent with opportunity for contact and/or habitat similarity as predictors of parasite sharing.

Although we found significant relationships among parasite similarity and host sampling effort, these associations were weak (mean $r^2 < 0.05$), suggesting that our results are not an artefact of differences in sampling effort (Table S4).

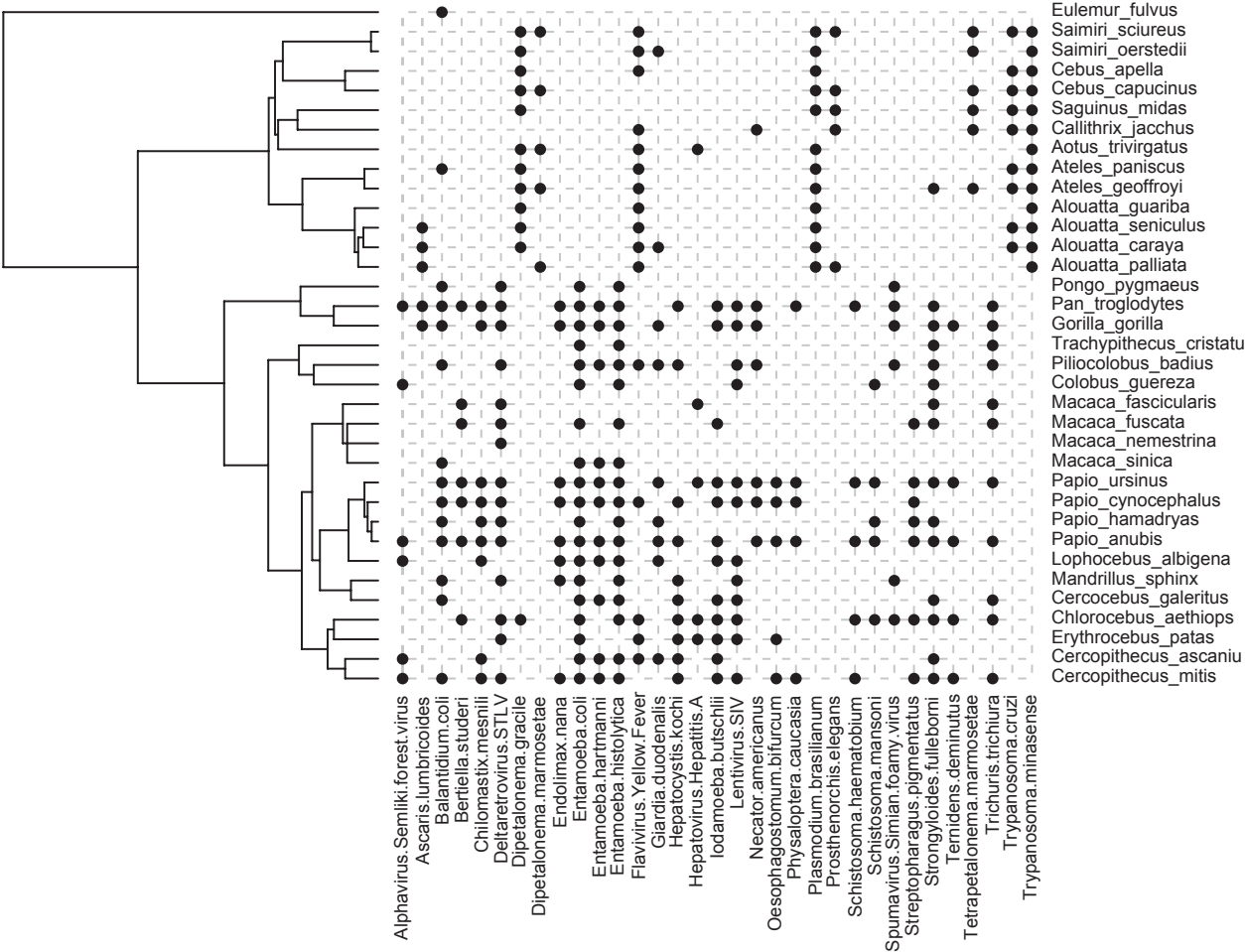


Figure 1 Phylogenetic distribution of the primate hosts for the most common parasites in our study. The primate phylogeny is on the left, primate species are on the right and parasites are under the plot. Black dots indicate parasite presence.

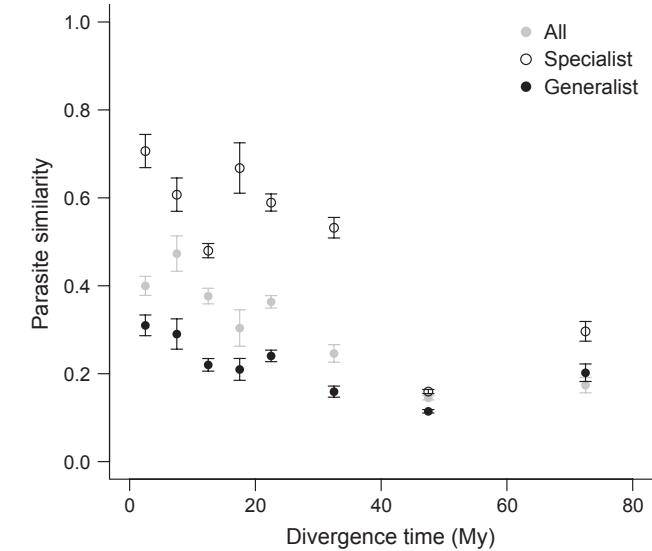


Figure 2 Scatter plots showing the relationship between parasite similarity and divergence time (millions of years) among pairs of primates for all parasites and for phylogenetic host specialist and generalist parasites separately. Results have been subdivided into 5-million year divergence time bins. Points are the mean values within these bins and error bars represent standard errors.

Table 2 Model averaging results from multiple regressions of parasite similarity against various predictors, for pairs of primates

Variable	Slope	Adjusted SE	Lower CI	Upper CI	Importance
Divergence time	−0.355	0.044	−0.441	−0.269	1.000
Body mass difference	−0.086	0.017	−0.120	−0.052	1.000
Group size difference	−0.036	0.028	−0.091	0.020	0.447
Terrestriality difference	0.038	0.033	−0.026	0.102	0.410
Temperature difference	−0.063	0.024	−0.110	−0.017	0.982
Precipitation difference	0.037	0.023	−0.009	0.082	0.558
Geographical range overlap	0.003	0.008	−0.013	0.019	0.274
Geographical range distance	−0.535	0.033	−0.600	−0.469	1.000

N = 595 primate species pairs. Coefficients are model averages from the 95% confidence set of models. SE = standard error; CI = confidence interval; importance = the relative importance of each variable.

Human–primate parasite sharing predictions

As predicted by the parasite-sharing models, we share most parasites with our closest relatives, chimpanzees and gorillas (Fig. 3). However, overall parasite sharing did not decrease predictably with

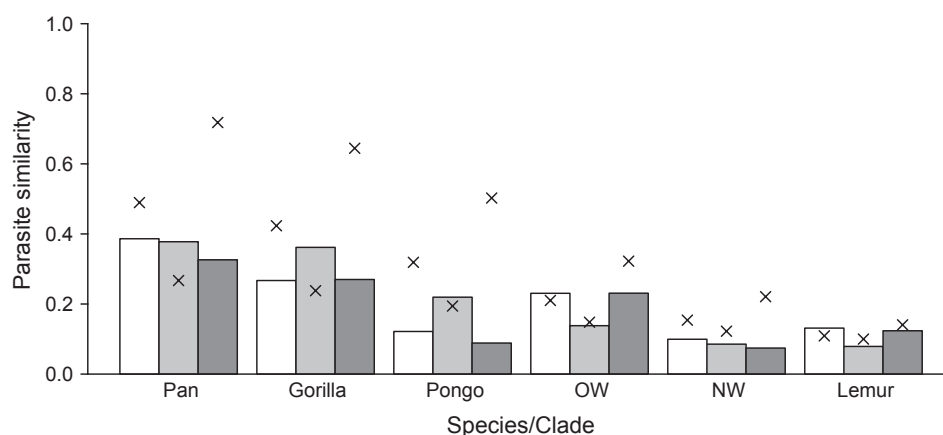


Figure 3 Average parasite similarity among humans and various species/clades of primates. White bars represent overall parasite similarity, light-grey bars represent parasite similarity for phylogenetic host-generalist parasites only, and dark-grey bars represent parasite similarity for phylogenetic host-specialist parasites only. Crosses represent predicted values from parasite-sharing models. Pan = *Pan troglodytes*; Gorilla = *Gorilla gorilla*; Pongo = *Pongo pygmaeus*; OW = Old World monkeys (Cercopithecidae); NW = New World monkeys (Platyrrhini); Lemur = Strepsirrhini.

phylogenetic distance: on average, humans share more parasites with Old World monkeys and lemurs than with orangutans. This appears to be because humans share far fewer phylogenetic host-specialist parasites with chimpanzees, gorillas and orangutans than predicted by our models (Fig. 3). When considering phylogenetic host generalists, however, parasite sharing among humans and primates decreased more consistently with phylogenetic distance (Fig. 3).

DISCUSSION

Over 60% of human EIDs are shared with animals (Jones *et al.* 2008). Thus, it is vitally important to understand the drivers of parasite sharing among species so we can predict which parasites may emerge in humans in the future. Previous authors found that close relatives should share more parasites than distant relatives (e.g. Davies & Pedersen 2008; Krasnov *et al.* 2010; Longdon *et al.* 2011). We also show that closely related primates share more parasites than more distantly related primates. Importantly, however, phylogenetic distance did not fully explain variation in parasite sharing; ecological and distributional differences among primates were also relevant predictors, particularly the distance between the species' geographical ranges. Davies & Pedersen (2008) also found that geographical range overlap was important for parasite sharing in primates, and Krasnov *et al.* (2010) found that environmental differences were important in ectoparasite sharing among rodents, particularly for ectoparasites that sometimes live independently of their hosts. Conversely, Longdon *et al.* (2011) found that phylogenetic distances among hosts explained almost all variation in the ability of viruses to infect new *Drosophila* species. However, they avoided any effects of host ecology and distribution by artificially infecting the hosts.

Taken together, these results suggest that host ecological and distributional differences are important through their effects on the rate and intensity of initial host-parasite contacts, while phylogeny is important for parasite sharing only after host-parasite contact has occurred. Other factors are also important for understanding parasite sharing. Using a measure of phylogenetic host specificity, for example, we found that more variance in parasite sharing was explained by phylogeny for host-specialist parasites.

Most of our parasites were phylogenetic host generalists, infecting more distantly related primates than expected. Evolutionary models suggest that these patterns are driven primarily by parasite extinction and host-parasite coevolution rather than host switching (although our method may not detect recent host switching among many close relatives, e.g. SIV; Charleston & Robertson 2002). This was true for phylogenetic host-generalist and -specialist parasites, although we expected to see higher rates of host switching in phylogenetic host generalists. These results suggest that phylogenetic host generalists infect many hosts by resisting extinction (rather than by regularly switching hosts). Parasite extinctions could occur through the evolution of host immune defences, changes in host ecology or distribution that make the host unsuitable (e.g. dispersal to an area without an appropriate vector), competitive exclusion by other parasites or simply by chance (e.g. 'missing the boat' at speciation; Krasnov *et al.* 2010). Phylogenetic host generalists may avoid extinction by quickly evolving adaptations to circumvent such host changes.

Most viruses in our dataset are phylogenetic host generalists. Interestingly, Longdon *et al.* (2011) found the opposite result in *Drosophila* sigma viruses, although this may reflect the fact that the hosts were infected artificially. Contrary to expectations, evolutionary models did not show that viruses had higher rates of host switching than other groups of parasites. This suggests that viruses are particularly good at resisting extinction, perhaps owing to higher rates of evolution (Woolhouse & Gowtage-Sequeria 2005; Woolhouse *et al.* 2005; Parrish *et al.* 2008; Elena & Froissart 2010). Bacteria were also predominantly phylogenetic host generalists, although we have little data on these parasites (Table 1). Protozoa and helminths, in contrast, have fairly equal numbers of phylogenetic host specialists and generalists. However, this changes if we consider parasites with just one host. Around 50% of viruses, 70% of bacteria, 40% of protozoa and 70% of helminths in our dataset infect only a single primate species, suggesting that we are underestimating phylogenetic host specificity. One explanation for this pattern involves constraints imposed by the parasite's lifecycle. For example, helminths often have complex lifecycles, which may reduce their ability to adapt to new hosts because they must retain adaptations to their

intermediate hosts (Woolhouse *et al.* 2005; Parrish *et al.* 2008). In addition, vector borne or intermediate host parasites are also constrained by availability of sufficient competent vectors or susceptible intermediate hosts, even in areas where susceptible hosts are abundant (Randolph & Rogers 2010; Kilpatrick 2011). However, we failed to find any differences in phylogenetic host specificity in relation to transmission mode. Parasites may also appear to infect only one host due to low sampling effort (see below).

Our parasite-sharing models predicted that humans should share most parasites with our closest primate relatives. However, parasite sharing did not decrease with phylogenetic distance as strongly as expected; humans share more host-specific parasites with Old World monkeys and lemurs than with our closer relative, the orangutan. This departure from expectations may reflect the greater research effort aimed at documenting parasites in Old World monkeys (see Fig. S2), especially as many macaque species are used in medical research, and because of greater contact between humans and these lineages of primates. For example, humans are likely to have more contact with Old World monkeys than orangutans because orangutans are rare and primarily arboreal, whereas Old World monkey species are more terrestrial and are often commensal with humans (Smuts *et al.* 1987). In addition, many human diseases appear to originate from human–wildlife contact during bushmeat hunting or butchering, including Old World monkeys (Wolfe *et al.* 2005; Betsem *et al.* 2011). Hence, these findings also suggest that the rate and intensity of parasite contacts is a better predictor of parasite sharing among species than phylogeny alone, which may explain why more parasites of humans are shared with rodents and domestic animals than with wild primates (Woolhouse & Goutage-Sequeria 2005; Engelstädter & Hurst 2006; Wolfe *et al.* 2007; Parrish *et al.* 2008). We omitted ecological differences in analyses involving humans because human ecological attributes are difficult to quantify. It would be useful to formalise ecological overlap between humans and other mammals more formally in future extensions of this research.

Sampling effort is a major issue in comparative analyses of parasites because poorly studied hosts will appear to have fewer parasites than well-known hosts (Gregory *et al.* 1996). We dealt with this issue by focusing on the 35 best-studied primates (of 376 species in the Order Primates; Wilson & Reeder 2005) and find that our analyses still had high power to detect significant phylogenetic host specificity even under heterogeneous parasite sampling. However, we also find that even in these well-studied primates, more effort is needed to sample the full complement of parasites found in most primates, as indicated by only slight saturation in the parasite accumulation curves (Fig. S2). In addition, 205 parasites are found only in one primate species. In some cases, the lack of saturation reflects a truly specialised parasite; in others it likely reflects under sampling of the parasite. This under sampling is also uneven across parasite groups: bacterial parasites are rarely reported, possibly due to lower severity of bacterial diseases compared with viral diseases (e.g. Ebola). Given the past and future implications of emerging parasites contracted from human–wildlife contacts, further research into wildlife parasites is crucial to fill these research gaps.

Several other methodological issues deserve mention. First, many parasite ‘species’ consist of multiple cryptic species that infect different hosts. Treating these cryptic species as one species could artificially inflate the phylogenetic host specificity of the parasite; however, if a cryptic species complex is phylogenetically host specific, it is likely that its constituent species will be too. In addition,

species definitions differ across types of parasite, which makes it difficult to compare phylogenetic host specificity across parasite types. Second, our measure of phylogenetic host specificity only captures one aspect of host specificity in parasites. Poulin *et al.* (2011) discuss other important factors, including structural (i.e. how the parasite prevalence and abundance vary among hosts) and geographical (i.e. how host use varies geographically) specificity. Third, the mechanisms leading to a pair of hosts sharing two closely related parasites or two distantly related parasites probably differ; however, our methods do not distinguish between these scenarios. If parasite relatedness was considered, we would expect parasite sharing of closely related parasites to be more strongly governed by host phylogeny. We did not consider parasite relatedness due to a lack of suitable parasite phylogenies; however, this should be considered in the future as more comprehensive parasite phylogenies are published (Poulin *et al.* 2011).

In conclusion, understanding the drivers of parasite sharing among species is vital to understanding which parasites will be more likely to emerge in humans, domesticated animals, crops or endangered wildlife. Our results confirm that more closely related hosts are more likely to share parasites, but also highlight the importance of host ecology and geographical distribution and of the phylogenetic host specificity of the parasites involved. We also highlight major gaps in our knowledge of wildlife parasites. Greater understanding of these gaps is essential to improve our ability to predict future infectious disease risks.

ACKNOWLEDGEMENTS

We thank the Nunn lab and three anonymous referees for helpful comments. Funding was provided by Harvard University and the NSF (BCS-0923791).

AUTHORSHIP

NC designed the study, performed the analyses and wrote and revised the manuscript. RG contributed to the study design and performed analyses. MF contributed to the study design. MO collected the georeferencing data. CLN designed the study and revised the manuscript.

REFERENCES

- Anderson, P.K., Cunningham, A.A., Patel, N.G., Morales, F.J., Epstein, P.R. & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.*, 19, 535–544.
- Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L. *et al.* (2011). Navigating the multiple meanings of diversity: a roadmap for the practicing ecologist. *Ecol. Lett.*, 14, 19–28.
- Arnold, C., Matthews, L.J. & Nunn, C.L. (2010). The 10kTrees website: a new online resource for primate phylogeny. *Evol. Anthropol.*, 19, 114–118.
- Barton, K. (2011). MuMIn: Multi-Model Inference. R package version 1.6.0.
- Betsem, E., Rua, R., Tortevoie, P., Froment, A. & Gessain, A. (2011). Frequent and recent human acquisition of Simian Foamy Viruses through apes’ bites in Central Africa. *PLoS Pathog.*, 7, e1002306.
- Bivand, R. & Rundel, C. (2011). rgeos: Interface to Geometry Engine – Open Source (GEOS). R package version 0.1-8.
- Charleston, M.A. & Robertson, D.L. (2002). Preferential host switching by primate Lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Syst. Biol.*, 51, 528–535.

- Davies, T.J. & Pedersen, A.B. (2008). Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proc. R. Soc. Lond. B Biol. Sci.*, 275, 1695–1701.
- Elena, S.F. & Froissart, R. (2010). New experimental and theoretical approaches towards the understanding of the emergence of viral infections. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 365, 1867–1869.
- Engelstädter, J. & Hurst, G.D.D. (2006). The dynamics of parasite incidence across host species. *Evol. Ecol.*, 20, 603–616.
- Gregory, R.D., Keymer, A.E. & Harvey, P.H. (1996). Helminth parasite richness among vertebrates. *Biodivers. Conserv.*, 5, 985–997.
- Hijmans, R.J. & van Etten, J. (2011). raster: Geographic Analysis and Modeling with Raster Data. R package version 1.8-39.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, 25, 1965–1978.
- IUCN (2010). IUCN Red List of Threatened Species. Version 2010.4. Available at: <http://www.iucnredlist.org>. Last accessed 25 July 2011.
- Jaccard, P. (1922). La chrologie sélective et sa signification pour la sociologie végétale. *Mem. Soc. Vand. Sci. Nat.*, 2, 81–107.
- Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. *et al.* (2008). Global trends in emerging infectious diseases. *Nature*, 451, 990–993.
- Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L. *et al.* (2009). PanTHERIA: a species-level database of life-history, ecology and geography of extant and recently extinct mammals. *Ecology*, 90, 2648.
- Keitt, T.H., Bivand, R., Pebesma, E.J. & Rowlingson, B. (2011). rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.7-1.
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D. *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463–1464.
- Kilpatrick, A.M. (2011). Globalization, land use, and the invasion of West Nile virus. *Science*, 334, 323–327.
- Kilpatrick, A.M., Briggs, C.J. & Daszak, P. (2010). The ecology and impact of chytridiomycosis: an emerging disease of amphibians. *Trends Ecol. Evol.*, 25, 109–118.
- Krasnov, B.R., Moullot, D., Shenbrot, G.I., Khokhlova, I.S., Vinarski, M.V., Korralo-Vinarskaya, N.P. *et al.* (2010). Similarity in ectoparasite faunas of Palaearctic rodents as a function of host phylogenetic, geographic or environmental distances: Which matters the most? *Int. J. Parasitol.*, 40, 807–817.
- Lewin-Koh, N.J. & Bivand, R. (2011). maptools: Tools for Reading and Handling Spatial Objects. R package version 0.8-9.
- Longdon, B., Hadfield, J.D., Webster, C.L., Obbard, D.J. & Jiggins, F.M. (2011). Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathog.*, 7, e1002260.
- de Magalhães, J.P. & Costa, J. (2009). A database of vertebrate longevity records and their relation to other life-history traits. *J. Evol. Biol.*, 22, 1770–1774.
- McCallum, H. & Jones, M. (2006). To lose both would look like carelessness: Tasmanian devil facial tumour disease. *PLoS Biol.*, 4, 1671–1674.
- Morens, D.M., Folkers, G.K. & Fauci, A.S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature*, 430, 242–249.
- Nowak, R.M. (1999). *Walker's mammals of the world*. The Johns Hopkins University Press, Baltimore.
- Nunn, C.L. & Altizer, S. (2005). The Global Mammal Parasite Database: an online resource for infectious disease records in wild primates. *Evol. Anthropol.*, 14, 1–2.
- Pagel, M., Meade, A. & Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.*, 53, 673–684.
- Parrish, C.R., Holmes, E.C., Morens, D.M., Park, E.-C., Burke, D.S., Calisher, C. H. *et al.* (2008). Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.*, 72, 457–470.
- Pebesma, E.J. & Bivand, R.S. (2005). *sp: Classes and Methods for Spatial Data in R*. R News, 5, 9–13.
- Pedersen, A.B., Altizer, S., Poss, M., Cunningham, A.A. & Nunn, C.L. (2005). Patterns of host specificity and transmission among parasites of wild primates. *Int. J. Parasitol.*, 35, 647–657.
- Poulin, R. (2010). Decay of similarity with host phylogenetic distance in parasite faunas. *Parasitology*, 137, 733–741.
- Poulin, R., Krasnov, B.R. & Moullot, D. (2011). Host specificity in phylogenetic and geographic space. *Trends Parasitol.*, 27, 355–361.
- Randolph, S.E. & Rogers, D.J. (2010). The arrival, establishment and spread of exotic diseases: patterns and predictions. *Nat. Rev. Microbiol.*, 8, 361–371.
- Sargeant, G.A. (2011). wild1: R Tools for Wildlife Research and Management. R package version 1.09.
- Schnute, J.T., Boers, N., Haigh, R. & Couture-Beil, A. (2010). PBSmapping: Mapping Fisheries Data and Spatial Analysis Tools. R package version 2.61.9.
- Smuts, B.B., Cheney, D.L., Seyfarth, R.M., Wrangham, R.W. & Struhsaker, T.T. (1987). *Primate Societies*. Chicago University Press, Chicago.
- Streicker, D.G., Turmelle, A.S., Vonnhof, M.J., Kuzmin, I.V., McCracken, G.F. & Rupprecht, C.E. (2010). Host phylogeny constrains cross-species emergence and establishment of Rabies virus in bats. *Science*, 329, 676–679.
- Taylor, L.H., Latham, S.M. & Woolhouse, M.E.J. (2001). Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 356, 983–989.
- Warton, D.I. & Hui, F.K.C. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, 92, 3–10.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Ann. Rev. Ecol. Syst.*, 33, 475–505.
- Wilson, D.E. & Reeder, D.A.M. (1993). *Mammal Species of the World: A Taxonomic and Geographic Reference*. 2nd ed. Smithsonian Institution Press, Washington DC.
- Wilson, D.E. & Reeder, D.A.M. (2005). *Mammal Species of the World: A Taxonomic and Geographic Reference*. 3rd ed. Smithsonian Institution Press, Washington DC.
- Wolfe, N.D., Peter, D., Marm, K.A. & Burke, D.S. (2005). Bushmeat hunting, deforestation, and prediction of zoonoses emergence. *Emerg. Infect. Dis.*, 11, 1822–1827.
- Wolfe, N.D., Dunavan, C.P. & Diamond, J. (2007). Origins of major human infectious diseases. *Nature*, 447, 279–283.
- Woolhouse, M.E.J. & Gowtage-Sequeria, S. (2005). Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.*, 11, 1842–1847.
- Woolhouse, M.E.J., Haydon, D.T. & Antia, R. (2005). Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.*, 20, 238–244.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor, John Fryxell

Manuscript received 23 May 2012

First decision made 4 July 2012

Manuscript accepted 30 July 2012