ECOLOGY LETTERS

Ecology Letters, (2018) **21**: 253–263 doi: 10.1111/ele.12890

LETTER

The impact of host genetic diversity on virus evolution and emergence

Cristina Rodríguez-Nevado,[1] Tommy T-Y Lam,[2] Edward C. Holmes[3] and Israel Pagán[1]*

Abstract

Accumulating evidence indicates that biodiversity has an important impact on parasite evolution and emergence. The vast majority of studies in this area have only considered the diversity of species within an environment as an overall measure of biodiversity, overlooking the role of genetic diversity within a particular host species. Although theoretical models propose that host genetic diversity in part shapes that of the infecting parasite population, and hence modulates the risk of parasite emergence, this effect has seldom been tested empirically. Using *Rabies virus* (RABV) as a model parasite, we provide evidence that greater host genetic diversity increases both parasite genetic diversity and the likelihood of a host being a donor in RABV cross-species transmission events. We conclude that host genetic diversity may be an important determinant of parasite evolution and emergence.

Keywords

Biodiversity, host genetic diversity, parasite emergence, rabies virus, virus evolution.

Ecology Letters (2018) **21**: 253–263

INTRODUCTION

Increasing evidence indicates that changes in ecosystem biodiversity may be a key determinant of parasite evolution and emergence (Ostfeld & Keesing 2012; Pagán *et al*. 2016). The vast majority of these studies have used species diversity as the sole measure of host biodiversity. However, biodiversity not only includes the diversity of species, but also the diversity of genotypes within species (Wilson 1992). This intraspecies component of biodiversity may have different effects on parasite evolution and emergence than the interspecies component (Keesing *et al*. 2010; Pepin *et al*. 2010; Ostfeld & Keesing 2012). Hence, to fully understand the determinants of these processes, it is of fundamental importance to determine how they are affected by different levels of host genetic diversity.

Theoretical models of the relationship between host population genetic diversity and parasite evolution and emergence assume that host-parasite interactions are characterised by genotype-specific resistance and infectivity, usually with no single host or parasite genotype having the highest fitness among its conspecifics (Haldane 1949; Hamilton 1980; Agrawal & Lively 2010). Under this assumption, a genetically heterogeneous host population would possess more polymorphisms for resistance than a genetically homogeneous one. This, in turn, would result in more polymorphisms in parasite traits that counter such resistance, thereby selecting for greater genetic diversity in the parasite population. Thus, theory predicts a positive association between the genetic diversity of hosts and their parasites populations. Genotype-specific resistance and infectivity have been widely documented in plant- and animal-parasite interactions (Altizer & Pedersen 2008; Pagán *et al*. 2016). To date, however, whether this translates into a positive association between host and parasite population genetic diversity has been only tested indirectly, and with contradictory results: parasite population genetic diversity tends to be higher in genetically diverse wild hosts compared to the same parasites sampled in domestic hosts with lower genetic diversity (Dugan *et al*. 2008; Lima *et al*. 2013), although the absence of such effect has been reported (Rodelo-Urrego *et al*. 2015). Also, analyses of host-parasite co-evolution have revealed a positive association between host and parasite genetic diversity (Paterson *et al*. 2010; Schulte *et al*. 2010), but not always (Lively *et al*. 2004; Rodelo-Urrego *et al*. 2013).

A possible explanation for the contradictory evidence is that the association between hosts and parasite population genetic diversity is time-dependent. If a parasite has only recently emerged in a host population it may not have had sufficient time to generate all the polymorphisms needed for host adaptation. Hence, parasite genetic diversity may be a function of time rather than of host genetic diversity (Streicker *et al*. 2012a), and it seems reasonable to assume that adaptation to more diverse host populations would require a larger number of positively selected substitutions. Finally, parasite genetic diversity might not be directly driven by host population genetic diversity, but by other host population traits that eventually cross-correlate with host genetic diversity. For instance, the population genetic diversity of an organism has

[1]*Centro de Biotecnología y Genómica de Plantas (Universidad Politécnica de Madrid – Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria), Madrid 28223 Spain*
[2]*State Key Laboratory of Emerging Infectious Diseases, Centre of Influenza Research, School of Public Health, The University of Hong Kong, Hong Kong, China*
[3]*Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, University of Sydney, Sydney, NSW 2006 Australia*
**Correspondence: E-mail: jesusisrael.pagan@upm.es*

been shown to be associated with population size/density, adult body size, and home range (Ellegren & Galtier 2016); all traits that may have significant impact in parasite genetic diversity (Barrett *et al.* 2008; Huang *et al.* 2015). However, the relative importance of these host traits in determining parasite evolution has not yet been analysed.

The relationship between host and parasite population genetic diversities may also affect parasite emergence. Emergence may be achieved through different processes, including spatial spread, increasing prevalence and cross-species transmission. Emergence through cross-species transmission (host jumping) underpins numerous cases of parasite emergence and is associated with many infectious disease pandemics. This complex process involves multiple steps including the initial infection in a new host, the production of propagules within the new host population, and the establishment of sustainable transmission networks (Hudson *et al.* 2008; Holmes 2009). Ecological factors, such as the frequency of contacts between reservoir and novel hosts, are of fundamental importance in determining the frequency with which host jumping occurs (Anderson *et al.* 2004; Keesing *et al.* 2010). However, many parasites often replicate poorly in new hosts and hence are inefficiently transmitted, such that genetic factors also need to be invoked to explain successful emergence (Pepin *et al.* 2010; Elena *et al.* 2014). For rapidly evolving parasites with large population sizes such as RNA viruses, genetic variants that increase fitness in the recipient host species and are present in the donor host species prior to transmission increase the probability of a successful jump (Holmes 2009), and might be detected, using genomic analyses of selection pressures (Streicker *et al.* 2012a). The frequency of these advantageous variants will be directly linked to the extent of genetic diversity in the virus population from the donor host (Cleaveland *et al.* 2001; Holmes 2009). Thus, if host and parasite population genetic diversities were positively associated then, all other things being equal, hosts with greater genetic diversity would more commonly act as donors in successful cross-species transmission events. Despite its relevance to understand parasite emergence, this hypothesis is yet to be tested.

To address these central questions in parasite emergence, we utilised *Rabies virus* (RABV) as model. RABV maintains successful transmission cycles within some host species of the mammalian orders Carnivora and Chiroptera (Jackson 2013). However, RABV infects a wider range of mammals, and examples of transient 'spill-over' infections from reservoir to novel host species are commonplace (Mollentze *et al.* 2014). For instance, RABV has been reported to infect humans, causing up to 60 000 deaths every year (Fooks *et al.* 2014), but with no onward human-to-human transmission. The determinants of RABV evolution and emergence are only partially understood, although the preponderance of spill-over infections compared to the far smaller number of successful host jumps indicates that ecological factors alone are insufficient to explain cross-species transmission (Mollentze *et al.* 2014).

Herein, we analyse the association between host and RABV population genetic diversity and the frequency of host jumps, utilising sequences collected from 34 species including both reservoir and spill-over hosts. Specifically, we analysed whether RABV genetic diversity, the number of synonymous and non-synonymous substitutions and selection pressures in the viral N and G genes, and the frequency of host jumps reconstructed from the RABV phylogeny, were associated with the population genetic diversity of the hosts they infect. Using multiple regression models, we also analysed the relative importance of co-evolutionary time-scale, host population size and density, body size, and host geographical and home range in shaping the relationship between host genetic diversity and RABV evolution and emergence.

METHODS

RABV sequences and host traits

Rabies virus nucleotide sequences of the viral nucleoprotein (N gene, 1353 nt) and glycoprotein (G gene, 1575 nt) were compiled from GenBank (www.ncbi.nlm.nih.gov/genbank/). The host and sampling location of each RABV sequence were obtained either from GenBank, from the associated publications, or were kindly provided by the relevant authors. Recombinant sequences as detected by at least four methods implemented in RDP4 (Martin *et al.* 2015), sequences from extensively passaged isolates, and from hosts associated with less than 10 (N) or 5 (G) RABV sequences, were excluded. This resulted in a data set of 1060 and 623 RABV sequences for the N and G genes respectively (Data S1). Sequence alignments were obtained using MUSCLE 3.7 (Edgar 2004). RABV sample size (number RABV sequences/host) and sample width (number of RABV populations sampled/host) were also calculated.

Genetic diversity values of the 34 (N) and 26 (G) hosts, for which RABV sequences were compiled, were measured as expected heterozygosity (H_e) and taken from the literature (Data S1). Measurements were directly extracted from relevant publications such that: (1) only microsatellite-based H_e values were considered; (2) at least three measures per host were obtained; and (3) measures for each host were from populations of the same biogeographical province [i.e. areas sharing similar biome and climatic conditions (Ladle & Whittaker 2011)], and time interval as the corresponding RABV sequences. Values of genetic diversity for each host were calculated as: (1) averaged H_e by averaging all the values obtained for a given host, and (2) weighted H_e by accounting for the different number of RABV sequences from each location, as: $\Sigma[(H_{ek} \times S_k)/S]$; where H_{ek} is the expected heterozygosity in the host population of the kth location, S is the total number of RABV sequences for a given host, and S_k is the number of RABV sequences in that host from the kth location. Data on host population size (n), density (individuals/km^2), geographic and home ranges (km^2), and adult body size (g) were extracted from the PanTHERIA database (Jones *et al.* 2009) or the IUCN Red List (http://www.iucnredlist.org/). Values of host traits for viral monophyletic groups including multiple host species were calculated as described above for weighted H_e, substituting H_e by each host trait and k being host species.

Genetic diversity and selection pressures in RABV populations

Virus genetic diversity (π) was estimated as average pairwise nucleotide difference between sequences, using the

Tamura-Nei, the Kimura-2-parameter and the composite likelihood nucleotide substitution models implemented in MEGA 6 (Tamura *et al.* 2011). Since all models led to the same conclusions, only values obtained using the Tamura-Nei model are presented. Standard errors of each measure were based on 1000 bootstrap replicates. Selection pressures, depicted as the ratio between the mean number of non-synonymous (d_N) and synonymous (d_S) nucleotide substitutions per site (d_N/d_S), and the number of positively selected sites were estimated using the fast unbiased Bayesian approximation (FUBAR), the fixed effect likelihood (FEL), and the random effects likelihood (REL) methods implemented in HyPhy (Kosakovsky-Pond & Frost 2005). Because all methods led to the same conclusions, only the FUBAR results are shown. We also obtained individual d_N and d_S values.

Phylogenetic analyses

Phylogenetic trees for RABV sequences were estimated using both the N and G genes. Bayesian trees were inferred using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in MrBayes v3.1.2 (Huelsenbeck & Ronquist 2001), utilising the general time-reversible substitution model with invariant sites and a gamma distribution of among-site rate variation (GTR + I + Γ_4). All analyses were run until relevant parameters converged, with 10% of the MCMC chains discarded as burn-in. Maximum clade credibility (MCC) trees, with Bayesian posterior probability values providing a measure of the robustness of each node, were summarised from the Mr Bayes tree samples. For each RABV monophyletic group, times to the most recent common ancestor (TMRCA) were estimated using the MCMC method available in BEAST v.1.8.2 (Drummond & Rambaut 2007). RABV sequence data sets were run as above and incorporating a relaxed (uncorrelated lognormal) molecular clock. Statistical confidence in the TMRCA estimates was represented by values for the 95% highest posterior density (HPD) intervals. The lower 95%HPD value was considered as the conservative TMRCA. TMRCAs were estimated using an empirical prior distribution on the nucleotide substitution rate based on substitution rates estimated by Streicker *et al.* 2012b for clades of bat viruses, and Troupin *et al.* 2016 for clades of terrestrial mammal hosts.

Analyses of genetic structure in RABV populations

We calculated the percentage of the total RABV genetic diversity explained by the between-host genetic diversity using the N_{ST} and the F_{ST} coefficients. N_{ST} was calculated as implemented in MEGA 6 following the expression: $N_{ST} = \frac{\delta_{ST}}{\pi_T}$, where ($\delta_{ST}$) is the between-host genetic diversity and π_T is the total genetic diversity Nei & Kumar 2000). The F_{ST} fixation index was calculated by analysis of molecular variance (AMOVA), as implemented in Arlequin v.3.11 (Excoffier *et al.* 2005). Statistical significance of these differences was obtained by performing 10 000 permutations.

Reconstruction of host jumps

We modelled RABV host jumps as a stochastic diffusion process among a set of discrete states (here, monophyletic groups) in a Bayesian framework. Following Faria *et al.* (2013), the phylogenetic diffusion model included different transition processes on external and internal branches in order to discriminate between recent jumps, likely to result in dead-end infections, and host shifts deeper in the evolutionary history that reflect successful jumps. Discrete phylogenetic diffusion analyses were performed under an asymmetric diffusion model (Edwards *et al.* 2011) as implemented in BEAST, using the settings described above. Bayesian stochastic search variable selection (BSSVS) procedures were used to identify significant pathways of host diffusion, which deliver a Bayes Factor (BF) as measure of statistical support for host jumps. We considered as statistically supported host jumps those with BF > 3 (Lemey *et al.* 2009). A robust counting procedure was used to estimate the posterior expectations of the number of host jumps along tree branches (O'Brien *et al.* 2009).

Statistical analyses

Differences in RABV genetic diversity, the number of synonymous and non-synonymous substitutions per site, of positively selected sites and selection pressures between host species and between reservoir and spill-over hosts were calculated using parametric (general linear models, GLM) and non-parametric (Permutation tests) methods. Since both approaches led to similar conclusions, only the GLM analyses are shown. The cross-correlation of host genetic diversity (as H_e or H_{ew}), population size and density, geographic and home ranges, adult body size, host-virus average or conservative co-evolutionary time-scales, and virus sample size and width was analysed using variance inflation factor (VIF). VIF values were smaller than 2 for all variables except host genetic diversities, co-evolutionary time-scales and sample size/width, but this was due to the high degree of correlation between H_e/H_{ew}, average/conservative time-scales and sample size/width. Dropping one variable from each of these three pairs resulted in VIF values smaller than 2 in all variables, indicating minimal cross-correlation. Hence, we performed model selection analyses using H_e, averaged time-scale and sample size, or H_{ew}, conservative time-scale and sample width in combination with the remaining host traits. Because both sets of variables yielded similar conclusions, we only present results for the former group of variables.

Mixed effect multiple regression tests were used to analyse the association between host traits and RABV evolutionary parameters (Burnham & Anderson 2002). Host traits were scaled and tested for normality using a Kolmogorov–Smirnov test. A set of models that included a global model containing all host traits as fixed predictors (except sample size/width that were considered as covariates), and nested models that contained all possible combinations of these predictors/covariates, was fitted for each RABV evolutionary parameter using general linear mixed models (R-library: asreml4). Models were constructed using an unstructured variance-covariance matrix for fixed and residual terms to avoid assumptions on the variance-covariance structure, and with a Gaussian linked function. Despite minimal cross-correlation, we built reduced rank models by incorporating Principal Components and Factor Analysis to the previously described model analyses. These

analyses yielded similar results (available upon request). Global and nested models were ranked according to Akaike's information criteria (AIC), and the model with the lowest AIC score was selected as the best-ranked. The relative importance of the predictors included in each model was calculated by residual maximum likelihood (R-library: asreml4). Bivariate tests were used to analyse the association between evolutionary parameters of the RABV populations and host genetic diversity. GLM and bivariate analyses were performed using SPSS 21 (SPSS Inc., Chicago, IL, USA).

RESULTS

Genetic differentiation of RABV according to host

We analysed whether RABV was genetically structured according to host species by estimating N gene-based phylogenetic trees (Fig. 1). RABV sequences from most host species formed strongly supported monophyletic groups (Fig. 1). Exceptions were RABV sequences from Vespertilionidae bats, coyotes and red foxes that formed two phylogenetically distant monophyletic groups each. In these three hosts, each of the two RABV monophyletic groups exhibited significant differences in virus genetic diversity ($F \leq 465.57$; $P > 1 \times 10^{-4}$). Given these differences, we considered each of these monophyletic groups separately (Fig. 1 and Table 1). RABV sequences from Molossidae bats, cows, dogs and humans formed several, but closely related, monophyletic groups. As the analysis of individual monophyletic groups within these four species/families did not result in significant differences in RABV genetic diversity compared to when they were grouped together ($F \leq 0.28$; $P \geq 0.596$), they were considered as a single monophyletic group to avoid artificial over-representation of these host–virus interactions. Overall, we identified a total of 22 host-specific monophyletic groups (Fig. 1). Accordingly, F_{ST} ($F_{ST} = 0.65$; $P < 1 \times 10^{-5}$) and N_{ST} ($N_{ST} = 0.60$; $P < 10^{-5}$) values showed significant genetic structure according to host. Equivalent results were obtained, using the G gene (Fig. S1).

Host traits affecting RABV genetic diversity

Since RABV populations were genetically structured according to host, we performed more detailed analyses of the association between host population traits and RABV genetic diversity. To this end, we created two data sets including virus sequences from: (1) 'broadly' monophyletic groups that considered all the virus genetic diversity sampled in a given host as long as they were monophyletic ($n = 1011$), and (2) 'tight' monophyletic groups that excluded the more divergent members of a host-specific clade that could conceivably represent independent cross-species transmission events ($n = 820$). Because both data sets yielded similar results (Tables 1 and

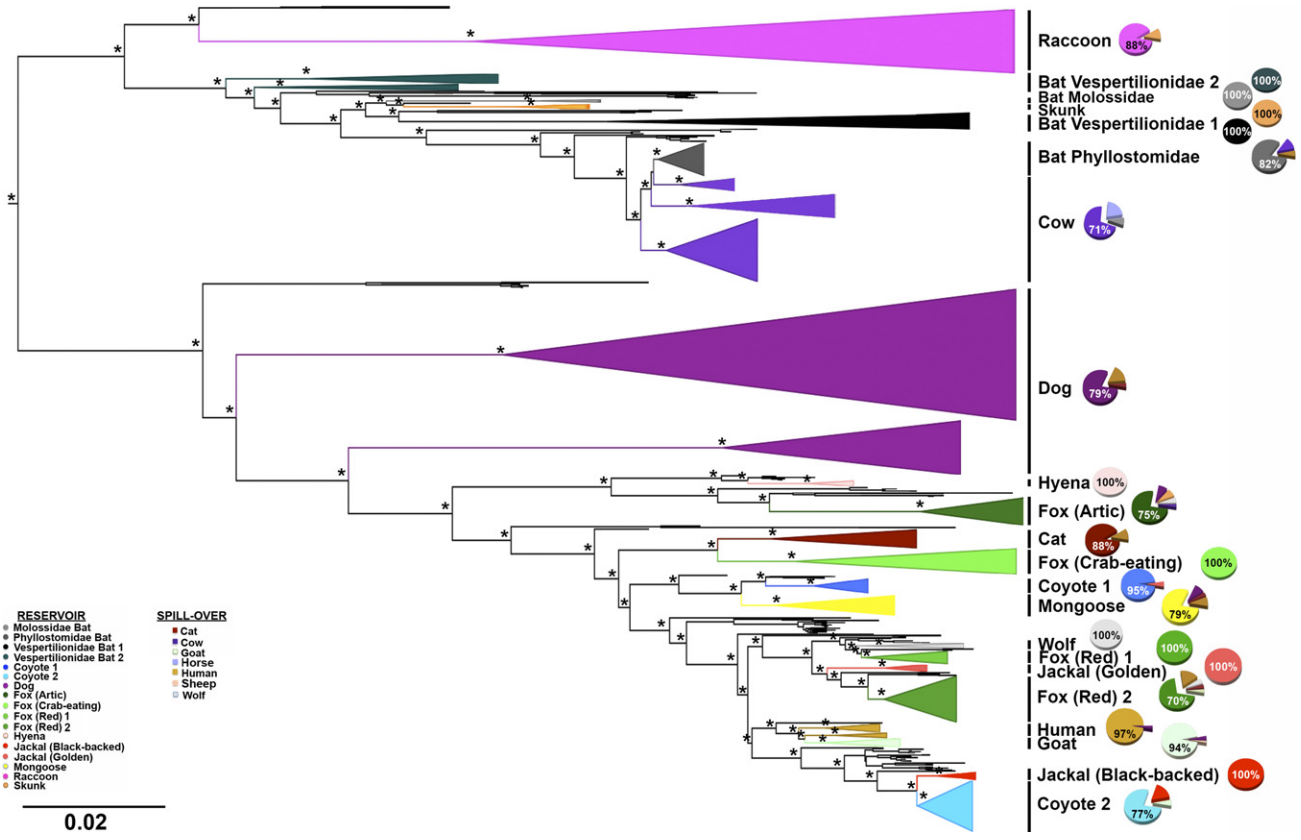

Figure 1 Bayesian *Rabies virus* (RABV) phylogeny based on the virus N gene. Nodes are collapsed and colored based on the host of origin. For a node to be collapsed, at least 80% of the RABV sequences in that node had to come from the same host as indicated by pie charts. Asterisks indicate nodes with posterior probabilities of ≥ 0.80. The tree is mid-point rooted for clarity only. Horizontal branches reflect the number of nucleotide substitutions per site. Pie chart colours indicate the hosts of origin of the RABV sequences in each monophyletic cluster with the corresponding percentages.

Table 1 Host genetic diversity (H_e), and host-specific *Rabies virus* (RABV) populations genetic diversity (π), d_N, d_S, and d_N/d_S, number of positively selected sites and time to the most recent common ancestor (TMRCA) based on the N gene (n = 1060)

Host	H_e*	H_{ew}*	$n^{\dagger}$	$\pi^{\ddagger}$	$d_N{}^{\S}$	$d_S{}^{\S}$	$d_N/d_S{}^{\S}$	(+) sites	D/R[¶]	TMRCA**
Bat (Molossidae)	0.901 ± 0.034	0.924 ± 0.034	15	0.114 ± 0.007	0.012 ± 0.001	0.508 ± 0.021	0.023 ± 0.019	2	3.1/2.8	314 (261-373)
Bat (Phyllostomidae)	0.657 ± 0.011	0.673 ± 0.010	46	0.033 ± 0.003	0.004 ± 0.000	0.161 ± 0.004	0.026 ± 0.005	1	1.0/1.2	219 (181-259)
Bat (Vespertilionidae) 1	0.822 ± 0.031	0.837 ± 0.025	17	0.062 ± 0.005	0.008 ± 0.000	0.259 ± 0.016	0.029 ± 0.004	1	2.9/1.0	270 (217-323)
Bat (Vespertilionidae) 2	0.850 ± 0.025	0.855 ± 0.036	32	0.100 ± 0.006	0.014 ± 0.000	0.406 ± 0.007	0.034 ± 0.001	5	2.1/0.9	218 (190-247)
Coyote 1	0.425 ± 0.100	0.425 ± 0.100	18	0.011 ± 0.002	0.001 ± 0.000	0.047 ± 0.010	0.023 ± 0.043	0	0.0/0.9	149 (113-186)
Coyote 2	0.425 ± 0.100	0.425 ± 0.100	51	0.001 ± 0.000	0.001 ± 0.000	0.004 ± 0.000	0.153 ± 0.017	0	0.9/1.1	84 (64-105)
Dog	0.730 ± 0.045	0.730 ± 0.045	230	0.134 ± 0.008	0.009 ± 0.000	0.668 ± 0.002	0.013 ± 0.001	6	3.2/2.0	550 (490-733)
Fox (Artic)	0.672 ± 0.026	0.683 ± 0.032	27	0.020 ± 0.002	0.002 ± 0.000	0.081 ± 0.002	0.023 ± 0.002	2	0.0/1.1	304 (221-392)
Fox (Crab-eating)	0.616 ± 0.115	0.616 ± 0.115	34	0.049 ± 0.004	0.011 ± 0.000	0.172 ± 0.003	0.064 ± 0.008	2	0.0/1.0	96 (67-128)
Fox (Red) 1	0.561 ± 0.013	0.561 ± 0.013	42	0.011 ± 0.002	0.002 ± 0.000	0.041 ± 0.001	0.057 ± 0.009	2	1.6/1.0	123 (100-146)
Fox (Red) 2	0.692 ± 0.056	0.721 ± 0.040	20	0.050 ± 0.004	0.006 ± 0.000	0.199 ± 0.010	0.031 ± 0.007	1	2.0/1.1	84 (68-101)
Hyena	0.592 ± 0.087	0.592 ± 0.087	12	0.063 ± 0.014	0.006 ± 0.001	0.254 ± 0.022	0.022 ± 0.008	0	0.0/1.0	263 (195-339)
Jackal (Black-backed)	0.699 ± 0.043	0.699 ± 0.043	26	0.042 ± 0.006	0.007 ± 0.000	0.169 ± 0.012	0.042 ± 0.007	1	1.0/0.9	98 (79-117)
Jackal (Golden)	0.670 ± 0.042	0.629 ± 0.059	13	0.054 ± 0.007	0.006 ± 0.001	0.240 ± 0.031	0.026 ± 0.006	1	1.1/1.6	111 (87-133)
Mongoose[††]	0.517 ± 0.043	0.514 ± 0.014	24	0.054 ± 0.007	0.005 ± 0.000	0.266 ± 0.015	0.019 ± 0.008	0	0.0/1.2	98 (73-124)
Raccoon	0.735 ± 0.046	0.790 ± 0.044	74	0.078 ± 0.006	0.007 ± 0.000	0.471 ± 0.013	0.016 ± 0.002	2	0.0/1.5	390 (226-563)
Skunk[††]	0.828 ± 0.033	0.828 ± 0.033	28	0.159 ± 0.010	0.017 ± 0.000	0.779 ± 0.026	0.022 ± 0.002	4	2.5/1.2	217 (122-321)
Reservoir hosts	*0.670 ± 0.034*	*0.677 ± 0.035*	*709*	*0.061 ± 0.011*	*0.007 ± 0.001*	*0.278 ± 0.053*	*0.037 ± 0.007*	*1.765 ± 0.373*	*21.4/21.4*	*NA*
Cat	0.633 ± 0.068	0.640 ± 0.060	35	0.123 ± 0.007	0.011 ± 0.000	0.613 ± 0.018	0.017 ± 0.005	0	–	133 (95-171)
Cow	0.706 ± 0.017	0.744 ± 0.031	100	0.109 ± 0.008	0.009 ± 0.000	0.576 ± 0.008	0.016 ± 0.001	3	–	252 (154-361)
Goat	0.673 ± 0.026	0.632 ± 0.035	38	0.125 ± 0.013	0.016 ± 0.000	0.673 ± 0.019	0.024 ± 0.004	2	–	90 (67-114)
Horse[1]	*0.735 ± 0.005*	*0.735 ± 0.005*	*32*	*0.033 ± 0.003*	*0.004 ± 0.000*	*0.145 ± 0.012*	*0.029 ± 0.016*	*3*	–	*NA*
Human	0.703 ± 0.038	0.665 ± 0.038	114	0.173 ± 0.010	0.014 ± 0.000	0.971 ± 0.006	0.014 ± 0.001	1	–	156 (124-194)
Sheep[1]	*0.714 ± 0.016*	*0.711 ± 0.020*	*17*	*0.156 ± 0.013*	*0.023 ± 0.001*	*0.810 ± 0.045*	*0.028 ± 0.002*	*1*	–	*NA*
Wolf	0.719 ± 0.022	0.730 ± 0.032	15	0.081 ± 0.007	0.018 ± 0.001	0.331 ± 0.027	0.057 ± 0.005	0	–	90 (71-107)
Spill-over hosts	*0.698 ± 0.013*	*0.694 ± 0.018*	*351*	*0.113 ± 0.018*	*0.013 ± 0.002*	*0.583 ± 0.107*	*0.027 ± 0.003*	*1.429 ± 0.271*	–	*NA*
ALL	*0.678 ± 0.024*	*0.682 ± 0.025*	*1,060*	*0.076 ± 0.010*	*0.009 ± 0.001*	*0.367 ± 0.056*	*0.034 ± 0.006*	*1.667 ± 0.342*	*21.4/21.4*	*631 (515-782)*

Notes *Average (H_e) and weighted (H_{ew}) expected heterozygosity. Values are mean±standard errors based in at least three different measures.
[†]Number of RABV sequences per host.
[‡]Values are mean ± standard errors based in 1000 bootstrap replicates.
[§]Values are mean ± standard error based on pairwise determination of d_N, d_S and d_N/d_S.
[¶]Number of events in which a host was detected as donor(D)/recipient(R) in RABV host jumping at the root of host-specific monophyletic groups.
[**]Time to the Most Recent Common Ancestor (years before present). 95%HPD intervals are shown in parenthesis.
[††]Mongoose: *Herpetes javanicus*; Skunk: *Mephitis mephitis*.
NA: Not applicable. [1]Hosts in italics were not considered in statistical analyses, due to absence of monophyletic clusters in the phylogenetic tree.

S1) we only describe the results for the larger data set in detail. For the N gene, RABV population genetic diversity (π), d_N, d_S and d_N/d_S and the number of positively selected sites greatly varied depending on the host (Table 1). Host population genetic diversity values also varied widely between species for both averaged (H_e) and weighted (H_{ew}) expected heterozygosity (Table 1).

We considered the following host traits as predictors of RABV evolution: genetic diversity (H_e), average host-virus co-evolutionary time-scale (TMRCA) for each monophyletic group, population size and density, adult body size, and geographic and home ranges. We also included RABV sample size as a covariate to control for its influence in our estimates of RABV genetic diversity (Tables 1 and S2). To analyse the association between host traits and RABV genetic diversity, we used multiple regression model selection analyses (Data S2). The model containing H_e, home range and TMRCA gave the best prediction of RABV genetic diversity ($r = 0.83$; $P < 1.0 \times 10^{-4}$) and d_N ($r = 0.79$; $P < 1.0 \times 10^{-4}$); and in both H_e had the highest relative importance (84.8 and 90.6% respectively) (Table 2). The same predictors were the components of the best-ranked model explaining RABV d_S ($r = 0.74$; $P = 2.0 \times 10^{-3}$), having more similar relative importance (47.6, 24.0 and 28.4% for H_e, home range and TMRCA respectively) (Table 2). Although none of the models accurately predicted RABV d_N/d_S ($r \leq -0.44$; $P \geq 0.101$), H_e was again the best predictor (relative importance = 86.8%) of the number of positively selected sites in the corresponding best-ranked model ($r = 0.73$; $P = 1.0 \times 10^{-3}$), which also included TMRCA (Table 2).

Since H_e was the chief predictor of RABV evolutionary parameters, we performed bivariate analyses to explore such associations in greater detail. H_e was positively associated with RABV population π, d_N, d_S and number of positively selected sites ($r \geq 0.53$; $P \leq 0.011$) (Fig. 2a–d), but not with d_N/d_S ($r = -0.44$; $P = 0.101$) (Fig. 2e). Together, these results indicate that higher RABV genetic diversity and number of positively selected sites, but not d_N/d_S, are associated with higher host species genetic diversity. Similar results were obtained using the G gene (Fig. S2; Tables S3 and S4).

Importantly, our bat monophyletic groups included RABV sequences from more than one species, which could result in overestimates of host and virus genetic diversities. To control for this potential bias, we repeated model selection analyses considering only bat species-specific monophyletic groups. These analyses, for which sequences from one bat species per monophyletic group were retained (Table S5), yielded similar conclusions as those with multi-species monophyletic groups (Fig. S3; Table S6; Data S3).

Table 2 Model selection analyses for *Rabies virus* (RABV) population genetic diversity (π), d_N, d_S, d_N/d_S, number of positively selected sites and number of host jumps. Model structures included averaged host genetic diversity, co-evolutionary time-scale, host population size and density, and host geographic and home ranges as predictors and RABV sample size as covariate. Best-ranked models are shown

Model structure*	$r^{\dagger}$	logLik	AIC‡	Δ_i[§]	ω_i[¶]
π					
A: H_e (84.8) + Home range (13.3) + TMRCA (1.9)	0.83*	48.38	−92.77	1	0.68
R: H_e (93.8) + Home range (6.2)	0.81*	38.16	−72.31	1	0.73
S: H_e (61.2) + TMRCA (38.8)	0.29	6.32	−6.64	0	1.00
d_N					
A: H_e (90.6) + Home range (8.1) + TMRCA (1.3)	0.79*	93.65	−181.30	2	0.48
R: H_e (100.0)	0.78*	71.20	−138.40	0	1.00
S: Home range (34.2) + H_e (31.9) + TMRCA (26.4) + Population density (2.0)	0.98	13.61	−21.21	3	0.36
d_S					
A: H_e (47.6) + TMRCA (28.4) + Home range (24.0)	0.74*	13.96	−23.92	2	0.55
R: H_e (94.0) + Home range (3.2) + TMRCA (2.8)	0.79*	13.66	−23.32	0	1.00
S: H_e (81.8) + TMRCA (18.2)	0.32	−0.46	6.92	6	0.19
Positively selected sites					
A: H_e (86.8) + TMRCA (13.2)	0.73*	−19.63	43.25	2	0.57
R: H_e (80.4) + Home range (17.2) + TMRCA (2.4)	0.80*	−15.00	34.00	0	1.00
S: Population density (100.0)	0.82	−7.31	18.63	4	0.23
d_N/d_S					
A: H_e (100)	−0.44	53.68	−103.37	3	0.39
R: H_e (53.4) + TMRCA (46.6)	0.54	37.21	−72.07	3	0.46
S: H_e (83.9) + Home range (16.1)	0.92	7.67	−11.35	1	1.00
Host jumps					
R: H_e (53.1) + Home range (28.8) + Range size (9.2) + TMRCA (8.9)	0.87*	−89.46	202.93	0	1.00

Notes *The relative importance (%) of each predictor variable is shown in parenthesis. A: all hosts; R: reservoir hosts; S: spill-over hosts.
[†]Correlation coefficient. Asterisks indicate significant correlations ($P < 0.05$).
[‡]Akaike's Information Criterion.
[§]Number of models closely competing with the best-ranked ($\Delta_i < 2$ out of 127 models tested). Δ_i, is the difference between the AIC of a given model and that of the best-ranked model, and quantifies how models compete (best-ranked model: $\Delta_i = 0$; substantial empirical support: $\Delta_i = 1$–2; considerable less support: $\Delta_i = 2$–7; and no support; $\Delta_i > 10$) (Burnham & Anderson 2002).
[¶]AIC model weight as $\omega_i = \exp(-0.5\Delta_i)/\Sigma\exp(-0.5\Delta_i)$. The larger the ω, the greater the likelihood of the model relatively to the competing models. Maximum $\omega_i = 1$.

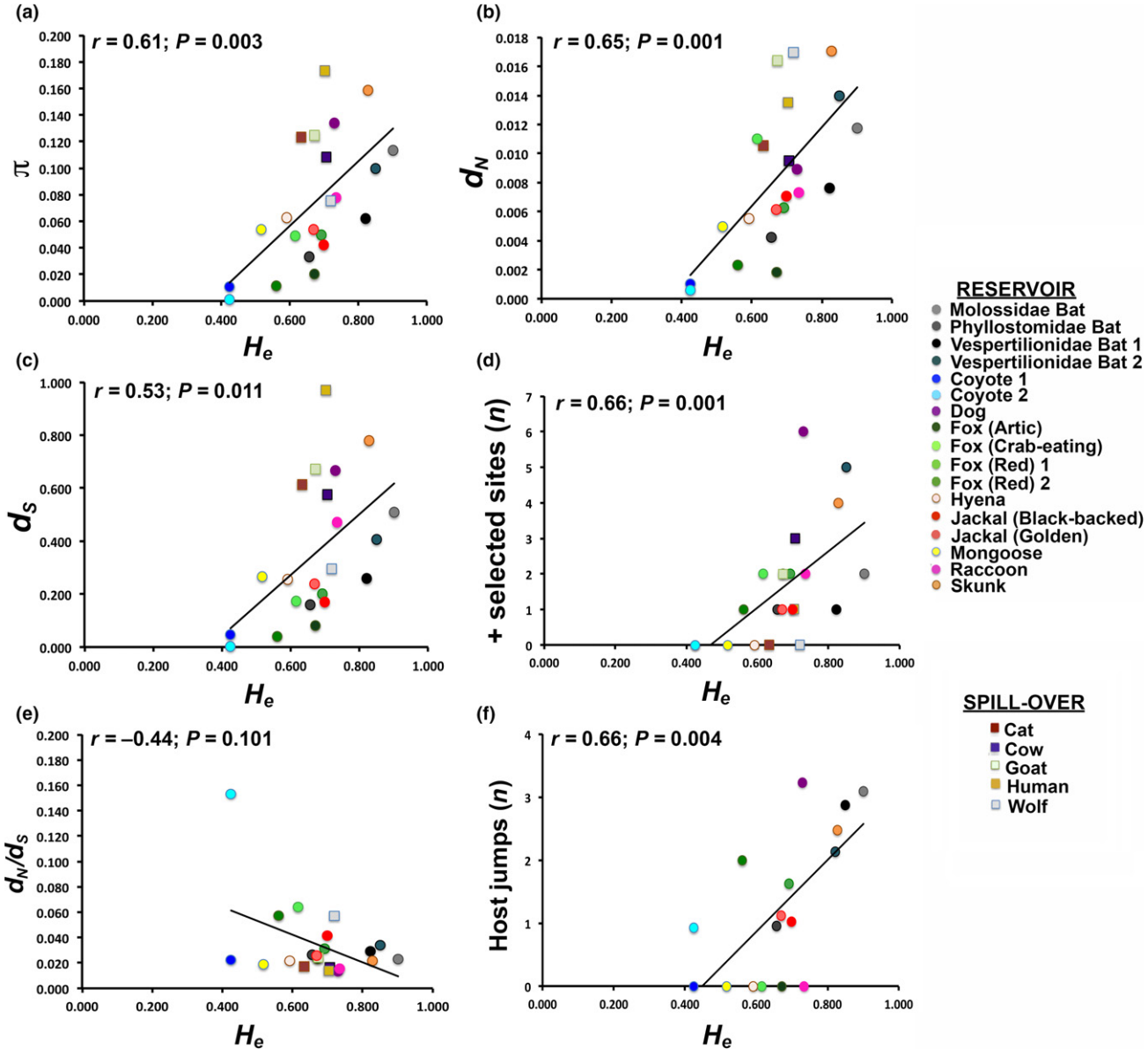

Figure 2 Bivariate relationships between host genetic diversity and N gene-based *Rabies virus* (RABV) evolution and the frequency of host jumps. Regressions of host genetic diversity (H_e, expected heterozygosity) on RABV genetic diversity, π (a); number of non-synonymous mutations per site, d_N (b); number of synonymous mutations per site, d_S (c), number of positively selected sites (d), overall selection pressures, d_N/d_S (e) and frequency of each reservoir host as donor in RABV host jumps (f) are represented. Dots indicate values for reservoir hosts, and squares indicate values for spill-over hosts. Note the different scales on the *y*-axis depend on the specific parameter analysed.

Effect of sustained virus transmission on the association between host and RABV genetic diversity

For a host species to be considered as reservoir, RABV must establish permanent transmission cycles between conspecifics. Not every host considered here act as reservoir; rather, a number are spill-over hosts in which RABV results in dead-end infections (Jackson 2013). Indeed, some of the monophyletic clusters defined in Fig. 1 mostly represented virus sequences from spill-over hosts, perhaps reflecting recurrent jumps from the same reservoir. This provided an important control to test the central hypothesis of this study: as RABV populations don't evolve in spill-over hosts, their genetic diversity should not be associated with that of RABV. We therefore divided the RABV populations into: (1) reservoir hosts (17/22 monophyletic groups) and (2) spill-over hosts (5/22 monophyletic groups) (Table 1). We then analysed host and RABV genetic diversities in both categories, and the association between them. N gene-based RABV π, d_N and d_S were higher in spill-over than in reservoir hosts ($F_{1,21} \geq 7.76$; $P \leq 0.011$), whereas d_N/d_S and the number of positively selected sites did not significantly vary among host types ($F_{1,21} \leq 0.51$; $P \geq 0.485$) (Table 1). Average H_e was similar in both spill-over and reservoir hosts ($F_{1,21} = 0.26$; $P = 0.614$) (Table 1).

Model selection analyses using data from reservoir hosts indicated that H_e had the highest relative importance in the best-ranked models predicting RABV N gene-based π, d_N, d_S and the number of positively selected sites (Table 2).

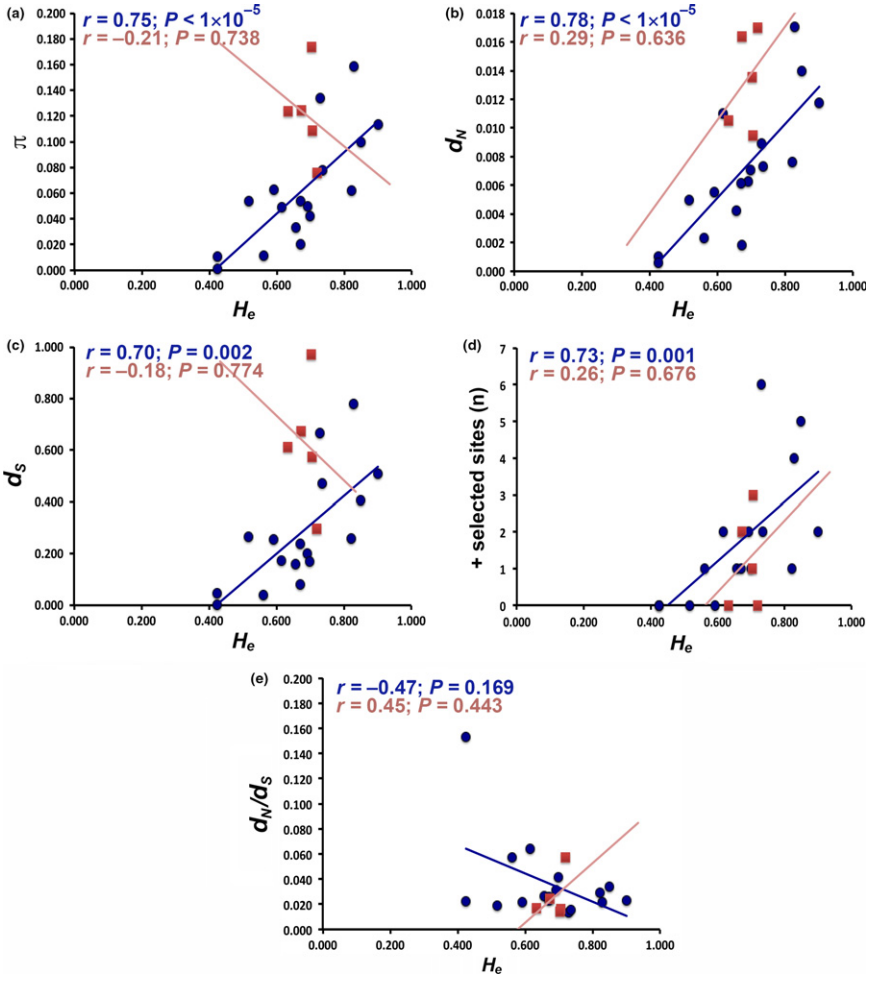

Figure 3 Bivariate relationships between host genetic diversity in reservoir and spill-over hosts and N gene-based *Rabies virus* (RABV) evolutionary parameters. Regressions of host genetic diversity (H_e, expected heterozygosity) on RABV genetic diversity, π (a); number of non-synonymous substitutions per site, d_N (b); number of synonymous substitutions per site, d_S (c), number of positively selected sites (d) and overall selection pressures, d_N/d_S (e), for reservoir (dark blue) and spill-over (light red) hosts are represented. Dark blue dots indicate values for reservoir hosts, and light red squares indicate values for spill-over hosts. Note the different scales on the *y*-axis depend on the specific parameter analysed.

Accordingly, in reservoir hosts there was a positive association of H_e with π, d_N, d_S and the number of positively selected sites ($r \geq 0.70$; $P \leq 2 \times 10^{-4}$) (Fig. 3a–d), but not with d_N/d_S ($r = -0.47$; $P = 0.169$) (Fig. 3e). In contrast, in spill-over hosts none of the tested models accurately predicted RABV evolution (Data S2), and bivariate analyses indicated that none of the RABV evolutionary parameters was associated with H_e ($r \leq 0.45$; $P \geq 0.443$) (Fig. 3). Parallel analyses using the G gene (Fig. S4; Tables S3 and S4) or N gene-based species-specific monophyletic clusters (Fig. S5; Table S6 and Data S3) led to the same conclusions. Thus, our results reveal that RABV genetic diversity is correlated with that of reservoir hosts and suggests that no such relationship occurs in spill-over hosts.

Association between host and RABV genetic diversity and the frequency of host jumps

On average, the probability of host-adaptive variants arising should be positively correlated with the overall level of genetic diversity in RABV populations (Mollentze *et al.* 2014). Our results indicated that reservoir hosts with higher population genetic diversity harbour greater virus genetic diversity. Thus, it could be hypothesised that more genetically diverse hosts should be better donors in RABV host jumps. To test this possibility, we reconstructed host transitions in internal branches of reservoir host-based RABV phylogeny. We used only the N gene as previous analyses using the N and G genes yielded equivalent results, and the N gene data set included sequences for a larger number of hosts.

Our reconstruction identified 33 statistically supported host jumps, 21 of them at the root of the defined monophyletic groups (Table 1 and Data S1). We analysed the association between the instances each host was a donor in these 21 RABV jumps and host traits by constructing multiple regression models containing as predictors the same host traits as described above (Data S2). The best-ranked model included host H_e, home range, geographic range and TMRCA ($r = 0.87$; $P = 1 \times 10^{-3}$), H_e being again the chief predictor (Table 2). Accordingly, the number of jumps in which each host acted as a donor was significantly associated with RABV genetic diversity ($r = 0.63$; $P = 6 \times 10^{-3}$) (Fig. S6). If the best donor hosts are also the most frequent recipients, the observed associations might be the consequence of the larger number of RABV genotypes migrating into these hosts, rather than their capacity to act as donors. However, no significant association was observed between the number of events as recipient host and the corresponding values of H_e or RABV genetic diversity ($r \leq 0.31$; $P \geq 0.176$). Similar results were obtained with species-specific monophyletic groups (Figs S3, S6 and Data S3). Hence, these results indicate that reservoir

hosts with higher population genetic diversity tend to be more frequent donors in RABV host jumps.

DISCUSSION

The potential impact of host population genetic diversity on parasite evolution and emergence remains largely unexplored (Keesing *et al*. 2010; Ostfeld & Keesing 2012). Using RABV as a model parasite, we provide evidence that greater host population genetic diversity is associated with both increased parasite population genetic diversity and likelihood of cross-species transmission, and may therefore be an important determinant of parasite evolution and emergence. Importantly, the methodology used here avoids artificial correlations and allowed us to disentangle the effects of host genetic diversity in RABV evolution without the confounding effects of potentially cross-correlating host traits.

Multiple regression analyses indicated that host population genetic diversity was a good predictor of RABV evolution, whereas other host traits had much poorer predictive power. Indeed, the positive association between host and virus population genetic diversity observed here accords with both theoretical models (Haldane 1949; Hamilton 1980; Agrawal & Lively 2010), and previous studies considering narrower host ranges (Schulte *et al*. 2010, 2013; Lima *et al*. 2013; Rocha *et al*. 2013). Theory and these previous experimental analyses proposed that the association between host and parasite population genetic diversity was the result of host adaptation by the parasite. Two lines of evidence suggest that this may also be the case of RABV. First, increasing host genetic diversity was associated with higher d_N and number of positively selected sites in the virus N and G genes, both of which have been associated with host adaptation (Streicker *et al*. 2012a). Second, the association between host and RABV genetic diversity was only observed in reservoir hosts, where the virus is obviously likely to be better host-adapted (Jackson 2013). Moreover, RABV genetic diversity in reservoir hosts was associated with increasing d_N and number of positively selected sites, and RABV sequences strongly clustered according to host species, both of which are compatible with host adaptation. Indeed, patterns suggestive of RABV adaptation have been observed in some reservoir hosts (Srithayakumar *et al*. 2011; Goldsmith *et al*. 2016), although not always (Talbot *et al*. 2013; Kyle *et al*. 2014). Conversely, in spill-over hosts no such associations were observed and RABV sequence clustering was weaker. This may in part be due to the small number of spill-over hosts considered ($n = 5$), which limited statistical power. Although enlarging this number by including RABV sequences that did not clustered according to host (i.e. horse and sheep) did not change our conclusions, more data are needed to reliably assess these relationships. It should be noted that host population genetic diversity was not associated with RABV d_N/d_S, which could be taken to mean that d_N and the number of positively selected sites simply accumulates as a function of time and not due to host adaptation. If so, we would expect co-evolutionary time-scale, rather than host H_e, to be the chief predictor of RABV π, d_N and the number of positively selected sites. However, in bivariate and multiple regression models H_e always had greater predictive power than co-evolutionary time-scale (see Data S2 and S3). In these models part of the variation in RABV evolutionary parameters remained unexplained. Therefore, other host factors not considered here, such as variation in host population structure (Goldsmith *et al*. 2016), could also play a role in RABV evolution.

We also present tentative evidence that reservoir hosts with greater population genetic diversity are more efficient donors in RABV host jumps. Such hosts are also those with greater RABV genetic diversity, supporting the hypothesis that higher virus population genetic diversity facilitates cross-species virus transmission, likely by increasing the chance that a donor host possesses genetic variants that are advantageous in a recipient host (Holmes 2009). Although testing whether host-adaptive mutations are more common in hosts with higher population genetic diversity is beyond the scope of this work, it is compatible with the observed trend towards a higher number of non-synonymous mutations and positively selected sites in reservoir hosts with higher population genetic diversity. Irrespective of the underlying mechanism, our data suggest that host population genetic diversity may play an important role in RABV emergence. Obviously, this does not exclude that other host factors also impact RABV emergence, particularly as there is clearly a large behavioural component (i.e. biting) to viral transmission among terrestrial mammals. Indeed, our bivariate analyses indicated that host population genetic diversity explained about a third of the variation in the number of RABV jumps, and multiple regression models indicated that host ecological factors such as home and geographic range also have an impact. Moreover, some of our results suggest that additional ecological factors, such as the extent of overlap between the geographic distribution of the donor and recipient hosts, might affect emergence. For instance, cattle were the best recipients of jumps from Phyllostomidae bats, and RABV sequences in both hosts had the same geographical origin. Also, host jumps with dogs as donors primarily had humans as recipients (Data S1). Indeed, previous analyses demonstrated that geographic range overlap and genetic relatedness between hosts are important predictors of rabies cross-species transmission (Streicker *et al*. 2010; Faria *et al*. 2013). Thus, our results are compatible with the notion that RABV emergence is determined by a combination of genetic and ecological factors. Interestingly, virus genetic diversity was higher in spill-over than in reservoir hosts. This is likely the consequence of spill-over hosts acting as recipients in host jumps from multiple donors as observed in our phylogenetic reconstructions. Together, these findings may be relevant to understanding the role of biodiversity in parasite emergence through host range expansion, particularly that intraspecies genetic diversity may be an important determinant of cross-species transmission.

It is important to note that host and RABV population genetic diversity were not estimated from exactly the same locations but from the same biogeographic region, such that our results may be biased if host-virus co-evolution occurs only at the local population scale (Thompson 2005): we would be comparing host and virus populations that do not necessarily interact with each other. However, such bias would have prevented detecting an association between host and virus genetic diversity, which was not the case. Also, weighting host genetic diversity according to geographical

origin did not affect our results (both H_e and H_{ew} were highly correlated), even though RABV sequences from some hosts were collected from geographically distant host populations. This suggests that either: (1) host-virus co-evolution operates at landscape scales other than the local population (Meentemeyer *et al*. 2012) such that the use of biogeographical region has sufficient discriminatory power; or (2) that our data are sufficiently representative of the interacting host and virus populations, even if this interaction occurs only at the local population scale. In either case our analyses seem robust to the potential effects of the co-evolutionary landscape scale.

In sum, the results presented here provide evidence that host genetic diversity plays an important role in parasite evolution and emergence, and highlights the necessity of considering all the components of biodiversity to understand the factors driving these evolutionary processes.

ACKNOWLEDGEMENTS

We thank the three anonymous reviewers for extremely useful comments on the manuscript. Alberto Cobos-Piñuela provided excellent technical assistance. This work was supported by a Career Integration Grant (PCIG11-GA-2012-322100) from the Marie Curie Actions and grant BIO2016-79165-R, Plan Nacional de I+D+i, from Ministerio de Economía y Competitividad, Spain to IP. CRN was supported by a UPM PhD fellowship. IP was supported by a Ramón y Cajal contract from the Ministerio de Economía y Competitividad de España (RYC-2011-08574). ECH was supported by an Australia Fellowship from the National Health and Medical Research Council, Australia (GNT1037231).

AUTHORSHIP

IP conceived and designed the study. CRN, IP and TL performed analyses with input from ECH. IP wrote the first draft of the manuscript and all authors contributed comments.

DATA ACCESSIBILITY STATEMENT

All data are available from the Dryad Digital Repository: https://doi.org/10.561/dryad.s59v1.

REFERENCES

Agrawal, A. & Lively, C.M. (2010). Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evol. Ecol. Res*., 4, 79–90.

Altizer, S. & Pedersen, A. (2008). Host-pathogen evolution, biodiversity and disease risk for natural populations. In: *Conservation Biology: Evolution in Action* (eds Carroll, S.P. & Fox, C.W.). Oxford University Press, Oxford, UK, pp. 259–277.

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.

Barrett, L.G., Thrall, P.H., Burdon, J.J. & Linde, C.C. (2008). Life history determines genetic structure and evolutionary potential of host–parasite interactions. *Trends Ecol. Evol*., 23, 678–685.

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multi-Model Inference: A Practical Information Theoretic Approach*, 2nd edn. Springer-Verlag, New York, NY.

Cleaveland, S., Laurenson, M.K. & Taylor, L.H. (2001). Diseases of humans and their domestic mammals: pathogen characteristics, host range and risk of emergence. *Phil. Trans. R. Soc. Lond. B*, 356, 991–999.

Drummond, A.J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol*., 7, 214.

Dugan, V.G., Chen, R., Spiro, D.J., Segamalay, N., Zaborsky, J., Ghedin, E. *et al*. (2008). The evolutionary genetics and emergence of Avian Influenza viruses in wild birds. *PLoS Pathog*., 4, e1000076.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*., 32, 1792–1797.

Edwards, C.J., Suchard, M.A., Lemey, P., Welch, J.J., Barnes, I., Fulton, T.L. *et al*. (2011). Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr. Biol*., 21, 1251–1258.

Elena, S.F., Fraile, A. & García-Arenal, F. (2014). Evolution and emergence of plant viruses. *Adv. Virus Res*., 88, 184–192.

Ellegren, H. & Galtier, N. (2016). Determinants of genetic diversity. *Nat. Rev. Genet*., 17, 422–433.

Excoffier, L., Laval, G. & Schneider, S. (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online*., 1, 47–50.

Faria, N.R., Suchard, M.A., Rambaut, A., Streicker, D.G. & Lemey, P. (2013). Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Phil. Trans. R. Soc. Lon. B*, 368, 20120196.

Fooks, A.R., Banyard, A.C., Horton, D.L., Johnson, N., McElhinney, L.M. & Jackson, A.C. (2014). Current status of rabies and prospects for elimination. *Lancet*, 384, 1389–1399.

Goldsmith, E.W., Renshaw, B., Clement, C.J., Himschoot, E.A., Hundertmark, K.J. & Hueffer, K. (2016). Population structure of two rabies hosts relative to the known distribution of rabies virus variants in Alaska. *Mol. Ecol*., 25, 675–688.

Haldane, J.B.S. (1949). Disease and evolution. *Ric. Sci*., 19, S68–S76.

Hamilton, W.D. (1980). Sex versus non-sex versus parasite. *Oikos*, 35, 282–290.

Holmes, E.C. (2009). *The Evolution and Emergence of RNA Viruses*. Oxford University Press, Oxford, UK.

Huang, S., Bininda-Edmonds, O.R.P., Stephens, P.R. & Gittleman, J.L.A.S. (2015). Phylogenetically related and ecologically similar carnivores harbour similar parasite assemblages. *J. Anim. Ecol*., 83, 671–680.

Hudson, P., Perkins, S. & Cattadori, I. (2008). The emergence of wildlife disease and the application of ecology. In: *Infectious Disease Ecology: Effects of Ecosystems on Disease and of Disease on Ecosystems* (eds Ostfeld, R.S., Keesing, F. & Eviner, V.T.). Princeton University Press, Princeton, NJ, pp. 347–367.

Huelsenbeck, J.P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754–755.

Jackson, A.C. (2013). *Rabies: Scientific Basis Of The Disease And Its Management*, 3rd edn. Elsevier Inc., San Diego, CA.

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.

Keesing, F., Belden, L.K., Daszak, P., Dobson, A., Harvell, C.D., Holt, R.D. *et al*. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468, 647–652.

Kosakovsky-Pond, S.L. & Frost, S.D. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, 21, 2531–2533.

Kyle, C.J., Rico, Y., Castillo, S., Srithayakumar, V., Cullingham, C.I., White, B.N. *et al*. (2014). Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (*Procyon lotor*) populations exposed to raccoon rabies. *Mol. Ecol*., 23, 2287–2298.

Ladle, R. & Whittaker, R.J. (2011). *Conservation Biogeography*. John Wiley & Sons, Chichester, UK.
Lemey, P., Rambaut, A., Drummond, A.J. & Suchard, M.A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput. Biol.*, 5, e1000520.
Lima, A.T.M., Sobrinho, R.R., González-Aguilera, J., Rocha, C.S., Silva, S.J.C., Xavier, C.A.D. *et al.* (2013). Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. *J. Gen. Virol.*, 94, 418–431.
Lively, C.M., Dybdahl, M.F., Jokela, J., Osnas, E.E. & Delph, L.F. (2004). Host sex and local adaptation by parasites in a snail-trematode interaction. *Am. Nat.*, 164, S6–S18.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.*, 1, vev003.
Meentemeyer, R.K., Haas, S.E. & Vaclavik, T. (2012). Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. *Annu. Rev. Phytopathol.*, 50, 379–402.
Mollentze, N., Biek, R. & Streicker, D.G. (2014). The role of viral evolution in rabies host shifts and emergence. *Curr. Op. Virol.*, 8, 68–72.
Nei, M. & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York, NY.
O'Brien, J.D., Minin, V.N. & Suchard, M.A. (2009). Learning to count: robust estimates for labeled distances between molecular sequences. *Mol. Biol. Evol.*, 26, 801–814.
Ostfeld, R.S. & Keesing, F. (2012). Effects of host diversity on infectious diseases. *Annu. Rev. Ecol. Evol. Syst.*, 43, 157–182.
Pagán, I., Fraile, A. & García-Arenal, F. (2016). Evolution of the interactions of viruses with their plant hosts. In: *Virus Evolution: Current Research and Future Directions* (eds Weaver, S.C., Denison, M., Roossinck, M. & Vignuzzi, M.). Caister Academic Press, Poole, UK, pp. 127–153.
Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A.J., Thomson, N.R. *et al.* (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, 464, 275–278.
Pepin, K.M., Lass, S., Pulliam, J.R.C., Read, A.F. & Lloyd-Smith, J.O. (2010). Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat. Rev. Microbiol.*, 8, 802–813.
Rocha, C.S., Castillo-Urquiza, G.P., Lima, A.T., Silva, F.N., Xavier, C.A., Hora-Júnior, B.T. *et al.* (2013). Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. *J. Virol.*, 87, 5784–5799.
Rodelo-Urrego, M., Pagán, I., González-Jara, P., Betancourt, M., Moreno-Letelier, A., Ayllón, M.A. *et al.* (2013). Landscape heterogeneity shapes host-parasite interactions and results in apparent plant–virus codivergence. *Mol. Ecol.*, 22, 2325–2340.
Rodelo-Urrego, M., García-Arenal, F. & Pagán, I. (2015). The effect of ecosystem biodiversity on virus genetic diversity depends on virus species: a study of chiltepin-infecting begomoviruses in Mexico. *Virus Evol.*, 1, vev004.
Schulte, R.D., Makus, C., Hasert, B., Nico, K.M. & Schulenburg, H. (2010). Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. *Proc. Natl Acad. Sci. USA*, 107, 7359–7364.
Schulte, R.D., Makus, C. & Schulenburg, H. (2013). Host-parasite coevolution favours parasite genetic diversity and horizontal gene transfer. *J. Evol. Biol.*, 26, 1836–1840.
Srithayakumar, V., Castillo, S., Rosatte, R.C. & Kyle, C.J. (2011). MHC class II DRB diversity in raccoons (*Procyon lotor*) reveals associations with raccoon rabies virus (*Lyssavirus*). *Immunogenetics*, 63, 103–113.
Streicker, D.G., Turmelle, A.S., Vonhof, 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.
Streicker, D.G., Altizer, S.M., Velasco-Villa, A. & Rupprecht, C.E. (2012a). Variable evolutionary routes to host establishment across repeated rabies virus host shifts among bats. *Proc. Natl Acad. Sci. USA*, 109, 19715–19720.
Streicker, D.G., Lemey, P., Velasco-Villa, A. & Rupprecht, C.E. (2012b). Rates of viral evolution are linked to host geography in bat rabies. *PLoS Pathog.*, 8, e1002720.
Talbot, B., Garant, D., Paquette, S.R., Mainguy, J. & Pelletier, F. (2013). Genetic structure and diversity among rabid and nonrabid raccoons. *Ecoscience*, 20, 345–351.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, 28, 2731–2739.
Thompson, J.N. (2005). *The Geographic Mosaic of Coevolution*. University of Chicago Press, Chicago, IL.
Troupin, C., Dacheux, L., Tanguy, M., Sabeta, C., Blanc, H., Bouchier, C. *et al.* (2016). Large-scale phylogenetic analysis reveals the complex evolutionary history of rabies virus in multiple carnivore hosts. *PLoS Pathog.*, 12, e1006041.
Wilson, E.O. (1992). *The Diversity of Life*. Harvard University Press, Harvard, MA.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Editor, Vanessa Ezenwa
Manuscript received 21 July 2017
First decision made 23 August 2017
Second decision made 19 October 2017
Manuscript accepted 2 November 2017