Virus Evolution, 2019, 5(1): vez009

doi: 10.1093/ve/vez009
Research article

Identifying the patterns and drivers of *Puumala hantavirus* enzootic dynamics using reservoir sampling

Lies Laenen,[1] Valentijn Vergote,[1] Bert Vanmechelen,[1] Katrien Tersago,[2,3] Guy Baele,[1] Philippe Lemey,[1,†] Herwig Leirs,[2] Simon Dellicour,[1,4,‡] Bram Vrancken,[1,§] and Piet Maes[1,*,**]

[1]KU Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Division of Clinical and Epidemiological Virology, Herestraat 49, 3000 Leuven, Belgium, [2]Evolutionary Ecology Group, Department of Biology, University of Antwerp, Antwerp, Belgium, [3]Epidemiology of Infectious Diseases, Belgian Institute of Health, Sciensano, Brussels, Belgium and [4]Spatial Epidemiology Lab (spELL), Université Libre de Bruxelles, Bruxelles, Belgium

*Corresponding author: E-mail: piet.maes@kuleuven.be

†http://orcid.org/0000-0003-2826-5353

‡http://orcid.org/0000-0001-9558-1052

§http://orcid.org/0000-0001-6547-5283

**http://orcid.org/0000-0001-9118-8608

Abstract

Hantaviruses are zoonotic hemorrhagic fever viruses for which prevention of human spillover remains the first priority in disease management. Tailored intervention measures require an understanding of the drivers of enzootic dynamics, commonly inferred from distorted human incidence data. Here, we use longitudinal sampling of approximately three decades of *Puumala orthohantavirus* (PUUV) evolution in isolated reservoir populations to estimate PUUV evolutionary rates, and apply these to study the impact of environmental factors on viral spread. We find that PUUV accumulates genetic changes at a rate of $\sim 10^{-4}$ substitutions per site per year and that land cover type defines the dispersal dynamics of PUUV, with forests facilitating and croplands impeding virus spread. By providing reliable short-term PUUV evolutionary rate estimates, this work facilitates the evaluation of spatial risk heterogeneity starting from timed phylogeographic reconstructions based on virus sampling in its animal reservoir, thereby side-stepping the need for difficult-to-collect human disease incidence data.

Key words: *Puumala orthohantavirus;* Bunyavirales; bank vole; evolutionary rate; landscape genetics; zoonosis.

1. Introduction

Hantaviruses are important zoonotic pathogens that can cause hemorrhagic fever with renal syndrome (HFRS) and hantavirus pulmonary syndrome, potentially life-threatening diseases in humans (Maes et al. 2004). Concomitant with the increasing public health relevance of hantaviruses (Kruger et al. 2015), there is a growing interest in uncovering the present-day patterns and drivers of the spatiotemporal dynamics of hantavirus circulation in its host reservoirs (Guivier et al. 2011; Barrios et al. 2013a,b; Haredasht et al. 2013a,b; Korva et al. 2013; Weber de Melo et al. 2015; Laenen et al. 2016; Tian et al. 2017) due to its potential in assisting risk determination and targeting surveillance efforts. Human infections result from inhaling infectious aerosols from excreta of infected animals. HFRS incidence is strongly correlated with the reservoir host population dynamics, resulting in seasonal and multi-annual fluctuations in incidence rates and sporadic local outbreaks (Kallio et al. 2009; Haredasht et al. 2014; Luis et al. 2015). Urbanization and landscape factors such as habitat composition, size and quality affect small mammal biodiversity and shape host distribution and metapopulation dynamics, consequently influencing hantavirus transmission within the host population and spillover to humans (Guivier et al. 2011; Ecke et al. 2017; Tian et al. 2018). Because of the difficulties associated with collecting reliable incidence data covering longer time periods and large geographic areas (e.g. due to subclinical infections (Clement et al. 1997) and differences in notification status of disease among regions (Heyman, Vaheri, and Members 2008)), the imprint of epidemic events that is contained in virus genetic data constitutes an attractive alternative source of information (see Biek and Real 2010; Bloomquist, Lemey, and Suchard 2010; Pybus, Tatem, and Lemey 2015; Holmes et al. 2016; for recent reviews on this topic). To use such imprints in genomic sequences for identifying the drivers of virus emergence and spread; however, it is crucial that the epidemiological history of the virus of interest that underlies all inferences, is accurately dated. This can be achieved by exploiting the temporal dimension in samples collected over the last few years or decades. Such temporally spaced virus genomic data can reliably inform evolutionary rate estimates when they foster a statistically significant number of genetic differences (Drummond, Pybus, and Rambaut 2003), and it is now well-recognized that this assumption should be thoroughly tested in order to avoid spurious results (Firth et al. 2010; Duchêne et al. 2015; Murray et al. 2016; Rambaut et al. 2016).

In Western and Northern Europe, *Puumala orthohantavirus* (PUUV) is the principal causative agent of hantavirus disease in humans (Maes et al. 2004; Kruger et al. 2015). The natural host of PUUV is the bank vole (*Myodes glareolus*), which has broadleaf forests with dense undergrowth as a preferred habitat in Western Europe (Mazurkiewicz 1994; Linard et al. 2007). Bank voles remain persistently infected with PUUV, but there can be remarkable dissimilarity in the persistence of PUUV variants over time within local vole populations, with both high rates of lineage turnover and preservation of identical variants for several years (Johansson et al. 2008; Weber de Melo et al. 2015). This extensive variation in the substitution rate among lineages likely results from a complex interplay between virus and host ecology. For example, Voutilainen et al. (2015) reported a strong variability in individual PUUV excretion dynamics of bank voles in natural settings, possibly reflecting differences in virus replication due to fluctuating immune responses as a result of changes in resource availability. Bank voles can get infected with PUUV either through direct contact or indirectly via contaminated food or environment (Sauvage et al. 2003). The relative importance of each transmission path depends on seasonal fluctuations in ecological factors and bank vole behavior (Tersago, Verhagen, and Leirs 2011) and on multiannual population density cycles (Voutilainen et al. 2016). Additionally, hantavirus *ex vivo* stability varies substantially depending on the environmental conditions (Kallio et al. 2006; Linard et al. 2007), thereby influencing the efficiency of the indirect transmission path. Thus, population-specific differences in transmission dynamics can affect the PUUV evolutionary rate as well as its constancy. This is important because substantial deviations of the linear accumulation of divergence through time can confound evolutionary rate estimations (Trovao et al. 2015; Raghwani et al. 2016). Furthermore, within-host life-history traits that counter the effects of short-sighted evolution may add to the high heterogeneity in evolutionary rates along different lineages, as is often observed in chronic viral infections (Vrancken et al. 2014; Lythgoe et al. 2017; Vrancken, Suchard, and Lemey 2017).

We here capitalize on the longest available sampling through time of well-defined PUUV populations infecting bank voles to estimate the rate of PUUV evolution at a timescale relevant for the present-day epidemic emergence and spread. We first exhaustively investigate the veracity of the temporal signal before estimating substitution rates and, in the time-structured clades, evaluate the impact of several landscape factors on the rate of PUUV spread.

2. Results

2.1 Dataset compilation

A total of 664 bank voles were trapped at eight locations in Belgium between 2006 and 2016, and ninety-eight (14.8%) were PUUV IgG positive. Complete S segment sequences were generated from eighty-two (83.7%) of these samples. In addition, 2 S segments from early passage PUUV isolates originating from bank voles captured between 1984 and 1986, were also sequenced. The new sequence data were complemented with the available PUUV S-segment data from NCBI GenBank.

2.2 Identifying isolated populations

We first focused on the fragment of the S segment that is most frequently sequenced (nucleotide (nt) positions 400–900 in PUUV strain Sotkamo, NC_005224) to frame the Belgian PUUV diversity in a global perspective. An unconstrained analysis (with no molecular clock enforced) using MrBayes showed that all Belgian PUUV lineages group in a well-supported clade that also includes strains from nearby locations in France (Charleville (Castel et al. 2015)), Germany (Cologne (Essbauer et al. 2007), Erft (Heiske et al. 1999), Koblenz (Schilling et al. 2007)), and The Netherlands (Middelbeers, Oirschot, and Kaatsheuvel (de Vries et al. 2016)) Supplementary Fig. S1. Next, we focused on this 'Belgian' cluster and extended the scope to the entire coding region of the S segment (1,299 nt) to more reliably identify PUUVs that have been evolving in isolated bank vole populations. This analysis showed that the Belgian lineages group in three well-defined clades (each having 100% posterior support), that can roughly be assigned to the north, middle, and south of Belgium (Fig. 1). These lineages will in this manuscript be further referred to as the Campine, Sonian Forest, and Ardennes clades, respectively.

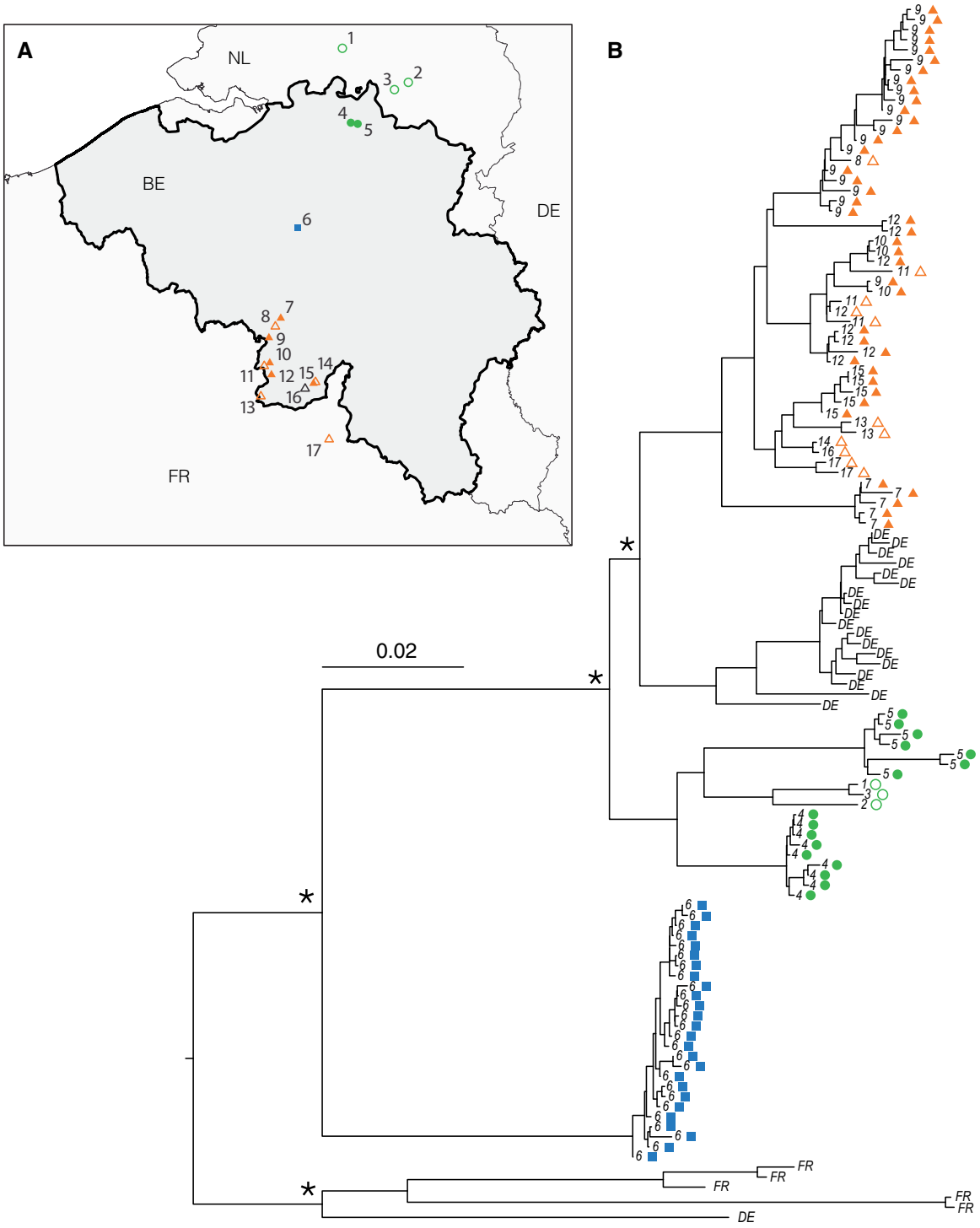

Figure 1. Spatial distribution and phylogenetic relationship of PUUV. (A) Map showing the trapping locations of the bank vole populations that were sampled. Locations with newly generated virus sequence data that were generated for this study ($n = 8$) are indicated with filled symbols. Locations with virus sequence data available in NCBI GenBank are indicated with open symbols. Trapping locations are denoted by numbers ranging from 1 to 17, with 1: Kaatsheuvel, 2: Oirschot, 3: Middelbeers, 4: Gierle, 5: Zevendonk (Turnhout), 6: Sonian Forest, 7: Fontaine l'Evêque, 8: Thuin, 9: Beaumont, 10: Rance, 11: Montbliart, 12: Chimay, 13: Mornignies, 14: Olloy-sur-Viroin, 15: Viroinval, 16: Couvin, and 17: Charleville. (B) MCC tree representation of the three PUUV populations estimated from the protein-coding part of the S segment. Nodes indicated with an asterisk have a posterior support of 1. The geographical clustering of lineages is indicated by colored symbols (orange triangles, Ardennes; green circles, Campine; blue squares, Sonian Forest). Sampling locations from other countries are marked by country codes (DE, Germany; FR, France). Branch lengths are expressed in substitutions per site as indicated by the scale bar.

2.3 Validation of temporal structure and estimating evolutionary rates

To explore the presence of temporal structure and the data quality for the Campine, Sonian Forest, and Ardennes clades (Supplementary Fig. S2) the genetic divergence was regressed against time. This revealed a complete lack of temporal structure in the Sonian Forest clade (Supplementary Fig. S2C), that only has 4 months between the earliest and the most recent available sequence. In contrast, there appeared to be a time-related accumulation of mutations in the Campine and Ardennes clades, and for this reason, we focused on these PUUV populations in subsequent explorations (Supplementary Fig. S2A and B). For neither clade, there was evidence of confounding between the genetic and temporal structure (Table 1), and the regression-based analyses indicate that the observed correlation between the root-to-tip distances and time only differs significantly from what is expected in the absence of a temporal structure for the Ardennes lineage (Murray et al. 2016). Additionally, the proportion of explained variation (R^2) is low for both the Ardennes and Campine datasets (Table 1), and large residuals indicate that the data may better fit a model where the substitution rate can vary along the branches (Supplementary Fig. S2A and B).

The suboptimal R^2 values of the regression analysis and the possible confounding impact of the non-independence of the data (Drummond, Pybus, and Rambaut 2003; Vrancken, Suchard, and Lemey 2017), motivated us to further evaluate the presence of temporal signal using a tip-date randomization test in conjunction with BEAST (Firth et al. 2010). In addition to appropriately accounting for the shared ancestry, this also allows to capture differences in the rate of evolution along lineages with a relaxed molecular clock (Drummond et al. 2006). Relaxed clock models provide an estimate of the variation in evolutionary rate among lineages, the coefficient of variation (CoV) statistic that can be used to evaluate the clocklikeness of evolution. Specifically, the CoV statistic depicts the scaled variance in evolutionary rate among lineages, meaning that a strict clock model may fit the data better than a relaxed clock model when

Table 1. Quantifying the temporal structure.

Clade	Sampling time range (years)	Mantel	Linear regression		
		P	Correlation	P	R^2
Ardennes	26.6	0.272	0.301	0.022	0.09
Campine	31.8	0.184	0.222	0.127	0.05

Each clade was tested for confounding between the genetic distances and exact sampling time differences using the Mantel test (Murray et al. 2016). *P*-values ≥ 0.05 indicate no evidence for confounding of the temporal with the genetic structure of the data was present. The heterochronicity of each clade was analyzed using a linear regression of the root-to-tip divergence against the sampling dates. The *P*-values relate to the correlation between root-to-tip genetic distance and sampling times.

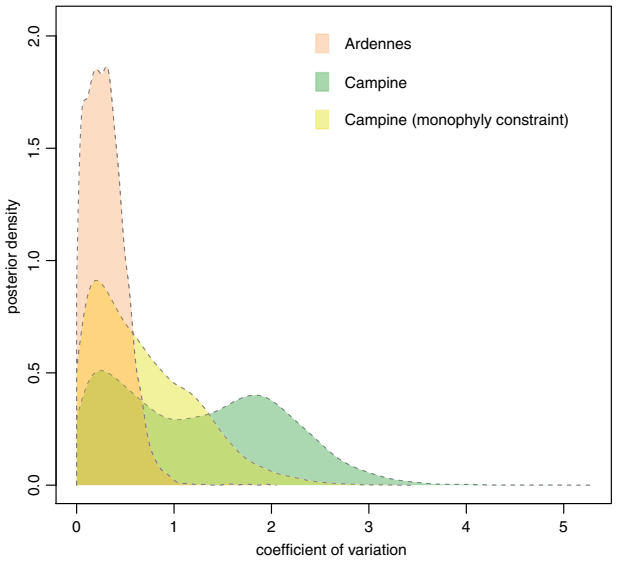

Figure 3. The levels of among-lineage rate heterogeneity differ between the Ardennes and Campine clades. The CoV represents the measured variance in evolutionary rate scaled by the mean. The CoV from the Ardennes and Campine clades was estimated using a relaxed clock model with branch rates drawn from an underlying discretized log-normal distribution (Drummond et al. 2006). For the Campine clade, the CoV was also assessed after imposing a monophyly constraint on the Gierle-Zevendonk-Turnhout subclade. Correspondence between the posterior densities and the datasets is indicated in this figure.

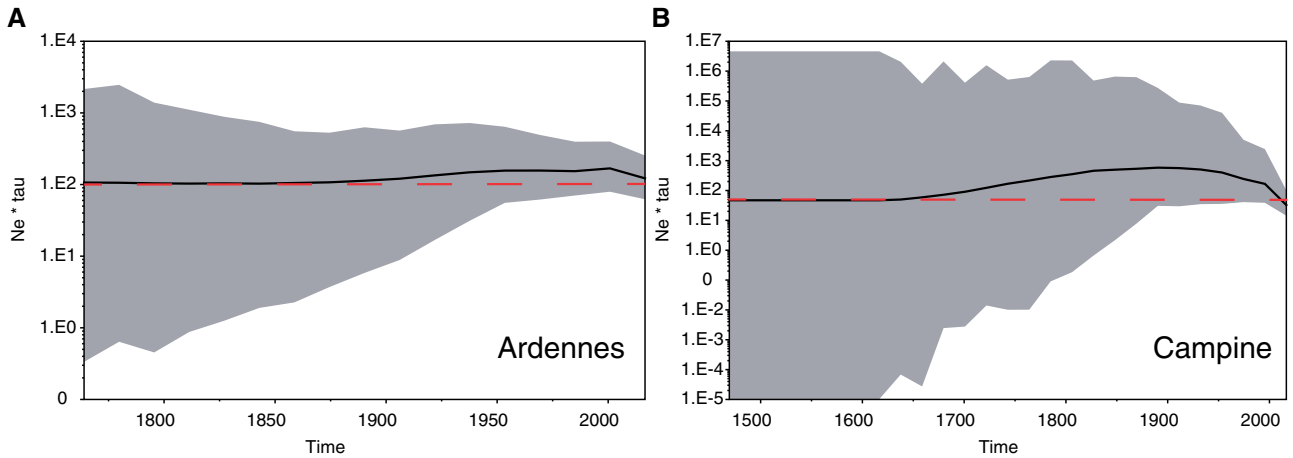

Figure 2. Population size histories for the Ardennes and Campine clades. Demographic histories were estimated under a flexible non-parametric coalescent model (Gill et al. 2013). The choice of the oldest change point (the 'cutoff') was based on the upper limit of the 95 per cent BCI of the time to the most recent common ancestor (t_{MRCA}) from a preliminary analysis with the skyline coalescent model (Drummond et al. 2005). The dashed red line marks a possible constant population size scenario and shows that for both the Ardennes clade (A) and the Campine clade (B) the demographic history mimics a constant population size scenario. Note that population structure can induce an artefactual signal of a population decline towards the most recent sampling time (Heller, Chikhi, and Siegismund 2013).

Table 2. Comparing the fit of strict and relaxed molecular clock models.

Clade	Strict clock[a]	Relaxed clock[a]	log(BF)relaxed[b]
Ardennes	−3,469.8	−3,468.6	0.9
Campine (w/o monophyly constraint)	−2,626.6	−2,624.6	2.0
Campine (w/ monophyly constraint)	−2,614.3	−2,610.4	3.9

[a]Numbers refer to the log likelihood of the data, averaged over the entire parameter space (log marginal likelihoods) estimated with the Generalized Stepping Stone estimator as implemented in BEAST (Baele et al. 2016). Higher values indicate a better model fit.
[b]log BF support in favor of the relaxed clock model wherein the evolutionary rates operating on the branches are drawn from a discretized lognormal distribution.

the 95 per cent highest posterior density (HPD) interval of the CoV statistic borders zero (Drummond et al. 2006; Gray et al. 2011). Reconstructions of the demographic histories of the Ardennes and Campine clade using a non-parametric coalescent model were in agreement with a constant population size scenario for both clades (Fig. 2). This was implemented in subsequent phylogenetic inferences.

The credible interval of the CoV for the Ardennes dataset (Fig. 3) buts to zero, but the rather liberal range of plausible CoV values indicates that a relaxed and strict clock model may have a comparable goodness of fit of the data. The Campine CoV posterior density (Fig. 3) also shows that the CoV cannot be used to unambiguously set apart different clock model hypotheses, albeit the support for rate heterogeneity is more outspoken. Because of these ambiguities, we formally compared the fit of strict and relaxed clock models to both datasets using a Generalized Stepping Stone marginal likelihood estimator in BEAST (Table 2) (Baele, Lemey, and Suchard 2016). In agreement with the patterns that emerge from Fig. 3, we find that both types of clock models fit the Ardennes data comparably well and the relaxed clock model fits the Campine data slightly better.

Additionally, the Campine CoV posterior density indicates that levels of rate variation comparable to those observed for the Ardennes clade are slightly better supported than higher levels (Fig. 3). Furthermore, a comparison of the maximum clade credibility (MCC) trees, corresponding to the sampled states of the converged Markov Chain Monte Carlo (MCMC) chain with a CoV larger respectively smaller than 1, shows that extreme levels of among branch rate variation (CoV > 1) involve unrealistically high rates ($\sim 10^{-2}$ substitutions/site/year) on particular internal branches linking the subclades (Supplementary Fig. S3). This is likely consequential to the combination of the parameter-rich relaxed clock model and data with noisy and/or weak temporal signal. Because very non-clock like data impede reliably estimating substitution rates (Trovao et al. 2015; Raghwani et al. 2016), we attempted to shrink the level of among-lineage rate heterogeneity in the Campine clade by imposing a monophyly constraint that involves the Gierle-Zevendonk-Turnhout subclade (see Supplementary Fig. S3, red box). The rationale is that this subclade has 100 per cent posterior support for being monophyletic in the unconstrained analysis and requiring these taxa to co-cluster does not allow for the topology of the MCC tree that is seen for high levels of among branch rate variation (CoV > 1, Supplementary Fig. S3).

Because of these uncertainties, the tip-date randomization test was done with both types of clock models (Fig. 4). This reveals that the Ardennes PUUVs clearly represent a 'measurably evolving population' (Drummond, Pybus, and Rambaut 2003). The evolutionary rate estimates for the Ardennes clade are 1.8×10^{-4} s/s/y (95% HPD: $1.1 \times 10^{-4} - 2.5 \times 10^{-4}$) under a strict clock and 2.0×10^{-4} s/s/y (95% HPD: $1.2 \times 10^{-4} - 3.0 \times 10^{-4}$) under a relaxed clock model and are robust to the clock model specification (Figs 4 and 5). Although the monophyly constraint on the Gierle-Zevendonk-Turnhout subclade improved the model fit (Table 2) and indeed reduces the level of rate heterogeneity (Fig. 3), it is clear from the large overlap of the 95 per cent Bayesian credible interval (BCI) of the tip-date informed evolutionary rate estimate with the 95 per cent BCIs of the estimates obtained from tip-date randomized datasets, that the substitution rate estimate from the Campine dataset is not as clearly informed by the data as for the Ardennes clade (Fig. 4). Yet, that the posterior density of the tip-date informed substitution rate is shifted towards higher values when compared with the null distribution (i.e. the expectation when randomizing the sampling dates), shows that there is signal for a build-up of genetic divergences over the sampling time. Indeed, analogously to the decreasing slope (i.e. substitution rate) with decreasing clock signal in a regression of root-to-tip divergence versus sampling time, it is expected that with decreasing temporal signal the substitution rate will increasingly tend towards the null distribution, which is opposite to what is observed here. Furthermore, the rate estimates of the Campine dataset largely overlap with the Ardennes estimates, irrespective of the type of clock model (Fig. 5). Overall, these results show that the temporally annotated PUUV genetic data can be used to reliably frame this virus' recent evolution in absolute time.

2.4 The impact of landscape factors on the *Puumala* virus dispersal velocity

Combining geographic data with dated sequence evolution enables the reconstruction of virus dispersal dynamics. In turn, this allows relating molecular epidemiology with landscape ecology and testing of hypotheses about the effect of landscape configuration on PUUV dispersal. Here, we pursue this by first reconstructing the dispersal history of the Ardennes and Campine clades using a Bayesian continuous phylogeographic method (Lemey et al. 2010). In a next step, the SERAPHIM R package (Dellicour et al. 2016; Dellicour, Rose and Pybus 2016) is used to investigate the impact of three landscape features (forest areas, croplands, and primary roads) on the PUUV dispersal velocity.

From the land cover maps with superimposed geo-referenced phylogenies (Fig. 6), it is clear that PUUV migration involves both forested and non-forested (mostly croplands in these regions) areas, and that primary roads form no absolute barriers. When relating landscape features to PUUV virus dispersal velocity, we find modest Bayes factor support (BF > 3) for forest areas facilitating PUUV dispersal in the Ardennes clade, while croplands act as a resistance factor, impeding PUUV dispersal (Fig. 7). For the Campine region, however, no significant support was obtained for any of the tested landscape features. Furthermore, primary roads do not act as a significant resistance factor for PUUV dispersal velocities in either clade (Supplementary Tables S2 and S3). However, the level of statistical support does not quantify the contribution of landscape factors in explaining differences in migration rates. For this, we consider the ratio of explained heterogeneity of the environmental versus null model (R^2_{env}/R^2_{null}, Supplementary Tables S2 and S3), that shows that forest and non-forest cover have a similar added value, with an average increase of R^2 of 42.2 and 43.8 per cent above what is captured by

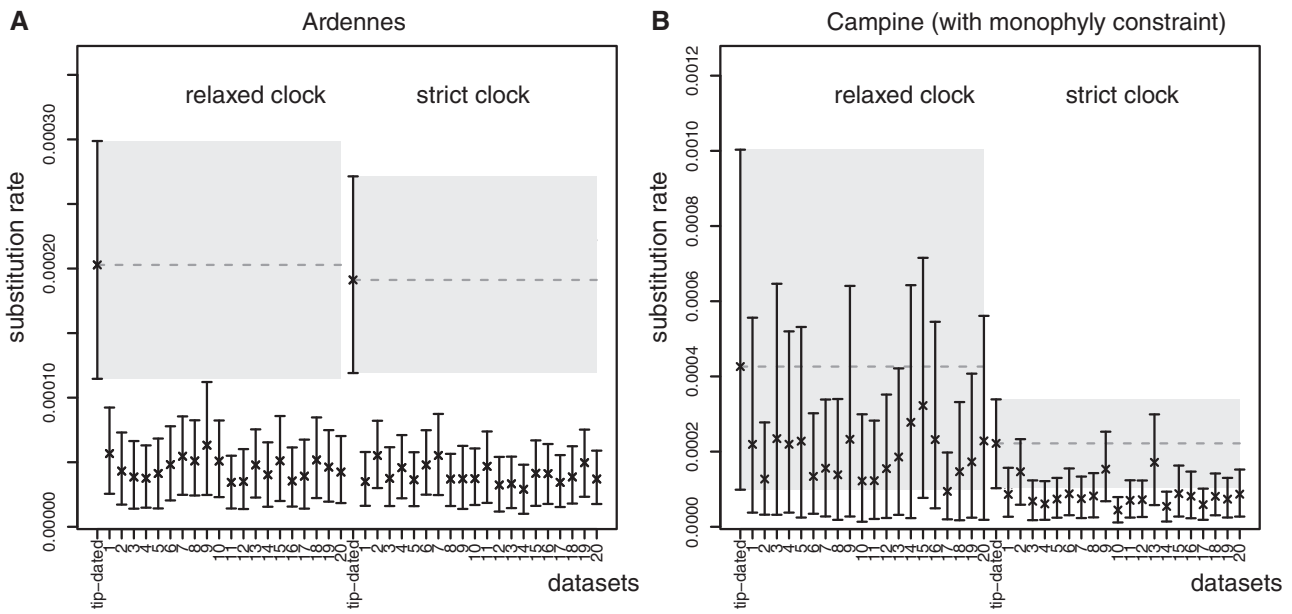

Figure 4. Tip-date randomization under a strict and relaxed clock model. Comparison of the mean and 95 per cent BCI of the evolutionary rate estimate from the actual datasets and twenty tip-date randomized datasets for the Ardennes clade (A) and Campine clade (B). The mean (crosses) and 95 per cent BCI estimates (whiskers) from the actual datasets are projected by the dashed light-gray line and background respectively, to facilitate the interpretation. Numbers on the x-axis denote the tip-date randomization replicates.

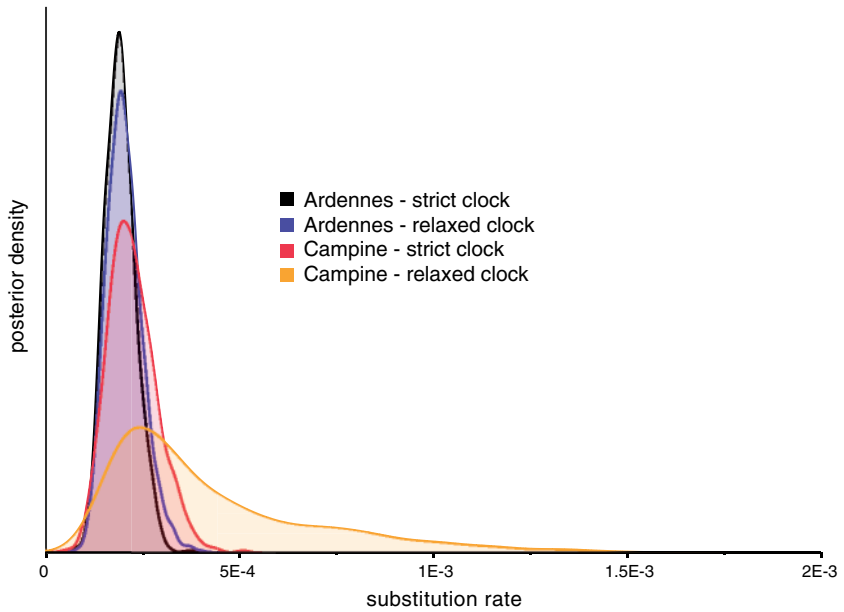

Figure 5. Evolutionary rate estimates under a strict and relaxed clock model. The evolutionary rate of the Ardennes and Campine clades were estimated under a strict and a relaxed molecular clock model. Correspondence between the posterior densities and the dataset-model combination is indicated in this figure.

geographic distances alone. Importantly, the high absolute values of R^2_{env} when treating cropland as a resistance (median 0.59, 95% CI: 0.19–0.80) and forest as a conductance factor (median 0.55, 95% CI: 0.16–0.80) testify to the importance of both types of land cover in shaping PUUV dispersal.

3. Discussion

Collecting genomic data is becoming a routine part of many hantavirus surveillance programs. Through phylogenetic analyses, this enables linking virus evolution to migration processes as well as statistically identifying correlates of epidemic spread, without having to rely on time series of incidence rates. These incidence data are often absent or scarce and can be unpredictable owing to the difficulties in their collection. To reliably bridge evolutionary information with data on epidemiological processes, however, it is crucial that the statistical models used to connect genetic information to time-stamped data—referred to as 'molecular clock models'—are accurately calibrated.

Based on a unique sampling effort covering approximately three decades, we obtained the best-informed short-term

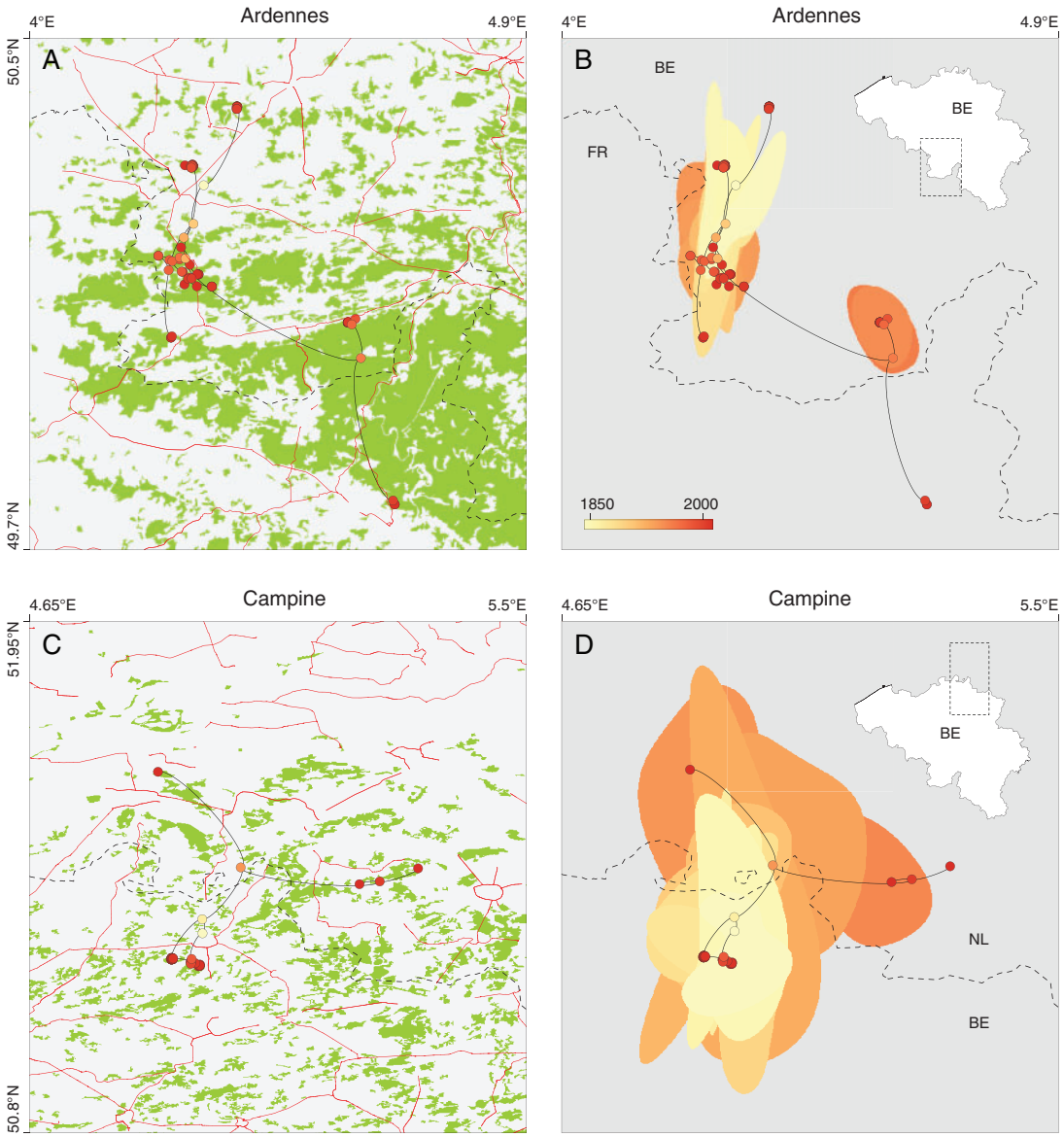

Figure 6. Reconstructed spatiotemporal diffusion for the Ardennes and Campine clades. For the Ardennes (A, B) and Campine clades (C, D), consensus trees are overlaid on the forest coverage (green, forested land; gray, other land cover) and primary roads (red lines) in panels (A) and (C). Consensus trees are also displayed alongside the 50 per cent HPD uncertainty region based on 100 trees sampled from the post burn-in posterior distribution (panels B and D). Nodes of the consensus trees are colored according to a color scale ranging from yellow (oldest) to red (most recent). Fifty per cent HPD regions were computed for successive time layers and then superimposed using the same time-reflecting color scale.

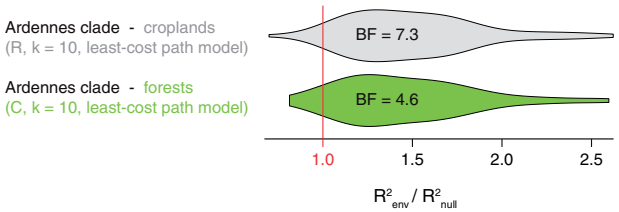

Figure 7. Croplands hamper and forests facilitate the spread of PUUV in the Ardennes. 'C' and 'R' indicate if the considered environmental raster was considered as a conductance ('C') or resistance factor ('R'), and k is the rescaling parameter used to transform the initial raster. Estimated Q distributions and related BF supports are reported for combinations of environmental factor, k value and path model for which $R > 1$ in at least 90 per cent of the measurements (see Supplementary Tables S2 and S3 for the complete results). BF supports were estimated with a randomization procedure detailed in Supplementary Appendix S1

evolutionary rate estimates of the PUUV nucleocapsid gene (S segment) to date, and exhaustively tested their reliability. Our work provides a substantial improvement on previously published PUUV substitution rate estimates in three related aspects. First, by considering PUUV evolution in isolated populations, as opposed to mixing data from populations that evolved in geographic isolation over extended periods of time (Ramsden et al. 2008), we avoid the biasing effect of the time-dependent rate decline that is expected to manifest on the long, deep branches connecting taxa from separated populations (Bennett et al. 2014). Second, our longitudinal sampling covers around three decades, meaning that the molecular clock can be calibrated more accurately in comparison to more shallow sampling schemes (Weber de Melo et al. 2015), in much the same way as simple regression analyses benefit from measurements that

cover a long range of the variable under investigation. Third, the authenticity of the temporal signal was comprehensively analyzed, which is a prerequisite for reliance on the reported estimates (Zhang and Holmes 2014).

We demonstrate that the rate estimate for the Ardennes population is well-informed by the data. Although the temporal structure in the Campine clade is less pronounced, the estimates for both populations are in close agreement (Fig. 5). Strikingly, our PUUV evolutionary rate estimates are also highly similar to substitution rates reported for S segment in other hantaviruses estimated from datasets with comparable genomic coverage and sampling time frame (Supplementary Fig. S4). This indicates that, contrary to what is often observed for other viruses (Streicker et al. 2012; Worobey, Han, and Rambaut 2014), differences in life-history traits appear to not substantially impact short-term evolutionary rates of these, and perhaps all, hantaviruses. The high transition/transversion ratios observed for the Ardennes and Campine datasets (with respective mean ratios of twelve (95% BCI: 8–16) and seventeen (95% BCI: 9–26)) testify that synonymous mutations are favored over amino acid changing mutations. Presumably, this is due to a history of virus-host co-evolution of these hantaviruses being sufficiently long to have reached a fitness plateau, resulting in strong purifying selection, as was previously reported for PUUV and other hantaviruses (Razzauti et al. 2013; Laenen et al. 2016).

Bank vole population sizes and PUUV prevalence can be highly variable (Tersago et al. 2008; Weber de Melo et al. 2015; Voutilainen et al. 2016). This, however, is not reflected in changes in the effective number of infections through time (Fig. 2), what indicates that expanding lineages are usually transient and have remained unsampled. As a matter of fact, it has been shown that the temporal distribution of sampling determines to what extent fluctuations in cyclic growth patterns are reflected in coalescent-based reconstructions of the demographic history (Stack et al. 2010). This illustrates the need for surveillance programs with an improved spacing of sampling times, preferentially generating long sequences from all three hantavirus segments to capture the multiannual population dynamics in detail. Improved sampling will also allow relating the periodic virus population growth and decline to the HFRS incidence and its potential drivers, such as host population density and climatologic factors, while avoiding the problems associated with *post hoc* methods (Gill et al. 2016; Volz and Didelot 2018). In addition, enhanced sampling is needed to assess the importance of segment reassortment for PUUV emergence dynamics (Razzauti et al. 2013; Szabo et al. 2017; Laenen et al. 2018).

Persistence of a limited set of lineages can also account for the absence of a clear imprint of variation in the rate of evolution between lineages in the virus genome. Specifically, many evolutionary dead-ends can eliminate the diversifying impact of lineage-specific differences in transmission dynamics on the substitution rate when, of the many variants that evolve at different rates at a single moment, only those with similar transmission dynamics survive in the longer-term. Importantly, both this and the constancy in population size (Fig. 2) align with the recent observation of long-term shedding of PUUV by bank voles in natural settings as a means to overcome host population bottlenecks (Voutilainen et al. 2015) and support the view that this hantavirus' long-term life strategy includes prolonged periods of replication in chronically infected hosts.

The PUUV and, more generally, the hantavirus emergence potential is sensitive to changes in environmental factors, and many modeling efforts have attempted to identify climatic, anthropogenic, and other ecological determinants of hantavirus-associated disease incidence (Linard et al. 2007; Palo 2009; Barrios et al. 2010, 2013a,b; Tersago et al. 2011; Haredasht et al. 2013a,b; Khalil et al. 2014; Voutilainen et al. 2016; Tian et al. 2017, 2018). To date, conditioning inferences on the spatial and temporal dynamics of gene flow has been unattainable for these studies. Thanks to the presence of sufficient temporal signal in our data, however, we could exploit the concordant timescales of epidemiological and evolutionary dynamics to evaluate the influence of environmental variables on PUUV dispersal velocity.

As forests are the preferred, yet not unique, habitat of bank voles (Gurnell 1985; Canova and Fasola 1991; Van Apeldoorn et al. 1992; Mazurkiewicz 1994; Wijnhoven et al. 2005; Torre and Arrizabalaga 2008), and roads can act as barriers to bank vole migration (Gerlach and Musolf 2000), we opted to investigate the relevance of forest and non-forest cover, and the presence of main roads in explaining migration rate heterogeneity. There was positive support for forest cover to facilitate PUUV dispersal and, vice versa, for non-forest cover as a resistance factor in the Ardennes clade. On the other hand, the type of land cover could not explain the variation in migration rates better than geographic distance alone in the Campine population. This is not due to a low amount of dispersal velocity heterogeneity in the Campine clade (Supplementary Fig. S5), but most likely relates to the substantial fragmentation of land cover in the Campine area, leading to less opportunity for a particular land cover type to leave an imprint on the migration rate. However, it should be noted that the environmental factors are assumed constant in time, and our approach currently cannot consider time-variable predictors. This may be particularly relevant for the Campine region, which has been subjected to substantial deforestation in the 20th century (Van der Veken, Verheyen, and Hermy 2004).

Given the impact of landscape structures, in particular forest fragmentation, on host abundance and spatial behavior (Linard et al. 2007; Guivier et al. 2011), it is likely that the extent of land cover fragmentation in the Campine area accounts for the noticeable lower effective number of infections through time (mean = 43, 95% BCI: 18–72) when compared with the Ardennes (mean = 100, 95% BCI: 58–153), where PUUV can more rapidly be reseeded following local extinction (Guivier et al. 2011). The relevance of land cover is further highlighted by historical HFRS incidence data: in the north of Belgium (cf. Campine clade), where forests are more fragmented, a lower incidence of HFRS is seen, while for the south of Belgium (cf. Ardennes clade) where forests are more connected, an endemic region with high HFRS incidence rate is present. These results further strengthen the notion that the type of land cover can be a valuable variable in predictive models of PUUV spread and infection risk (Linard et al. 2007; Guivier et al. 2011; Barrios et al. 2013b). Finally, in line with previous findings that only highways strongly impede bank vole migration (Gerlach and Musolf 2000), the presence of primary roads does not seem to affect the migration rate of PUUV in the Ardennes and Campine regions.

In summary, we demonstrated that PUUV genomic surveillance data are amenable to quantifying the rate of the PUUV dispersal dynamics and identifying its drivers. Our results will assist in identifying intervention points and improve the efficiency of public health resource allocations.

4. Materials and methods

4.1 Sample collection and sequencing

Rodent trappings were conducted on eight different locations in Belgium from 2006 to 2016. On each location, one hundred

Longworth or Sherman live traps were placed in a 10 × 10 grid at 10 m intervals at each location. Captured bank voles were sedated with isoflurane followed by a 10 µl blood collection through the retro-orbital sinus. The collected blood was tested in the field for the presence of PUUV IgG antibodies using the ReaScan® Ab-Dect *Puumala* IgG kit (Reagena). PUUV antibody-positive bank voles were sacrificed and immediately stored at −20°C. Bank vole kidney and lung tissues were dissected and preserved in RNAlater stabilization solution (Life Technologies). Kidney and/or lung tissue was homogenized with the Minilys homogenizer (Bertin Technologies) followed by total RNA extraction with the RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. Tissues were screened for the presence of hantavirus RNA as described previously (Laenen et al. 2016). In addition, PUUV isolates, obtained from low passaging on VeroE6 cells, from bank voles captured between 1984 and 1986 were included in sequencing. From the positive samples and cultures, the complete S segment was amplified with the One-Step RT-PCR kit (Qiagen) using the following conditions: 50°C for 30 min, 95°C for 15 min followed by 40 cycles of amplification (30 s at 94°C, 30 s at 53°C, 2 min at 72°C) with a final extension of 10 min at 72°C. PCR amplicons were purified using ExoSAP-IT® PCR Product Cleanup (Affymetrix) and sequenced according to the ddNTP chain termination method with the BigDye Terminator v3.1 cycle sequencing kit (Life Technologies) on an Applied Biosystems 3130xl Genetic Analyzer. Primers used to amplify and sequence the S segment are available in Supplementary Table S1.

4.2 Querying the GenBank sequence database

All available PUUV sequence data were downloaded from NCBI GenBank. S segment sequences were filtered by aligning all data to a reference S segment sequence (GenBank accession number NC_005224) with a pairwise codon-aware alignment tool (http://regatools.med.kuleuven.be/sequencetool/sequencetool.wt), allowing for at most ten frameshifts.

4.3 Bayesian inference of sequence evolution

For the analyses where no molecular clock was enforced, MrBayes v.3.2.6 (Ronquist et al. 2012) was used to estimate the posterior probabilities of trees using Bayesian inference through MCMC simulation, employing a General Time Reversible (GTR) substitution model and a discretized Γ distribution using default priors (Yang 1994).

For the specific clades where preliminary tests (see below) indicated the presence of a measurable build-up of genetic change over the sampling time span (the Ardennes and Campine clades), timed evolutionary histories were estimated using BEAST v1.8.4 (Drummond et al. 2012). Here, a strict or relaxed molecular clock model (Drummond et al. 2006) was combined with the HKY substitution model (Hasegawa, Kishino, and Yano 1985) and substitution rate heterogeneity across positions was again modeled with a discretized Γ distribution (Yang 1994). After an initial exploration of the demographic history with a flexible non-parametric coalescent model (Gill et al. 2013; see Fig. 2), we decided to assume a constant population size for both datasets when estimating clade-specific substitution rates. The fit of different models to the data was estimated using the Generalized Stepping Stone marginal likelihood estimator as implemented in BEAST (Baele et al. 2016).

4.4 Evaluating the temporal signal

The initial evaluation of the temporal signal and data quality with TempEst (Rambaut et al. 2016) was based on a maximum likelihood tree estimated with PhyML in Seaview (Gouy, Guindon, and Gascuel 2010), with GTR correction for multiple hits, Γ-distributed rate variation among sites (Yang 1994) and not assuming a molecular clock. A Mantel test (Murray et al. 2016) served to identify whether genetically similar taxa were more likely to have been sampled around the same time, and the significance of the correlation between the root-to-tip divergences and sampling times was estimated against a null distribution obtained by 1,000 times randomly reassigning the sampling dates to the taxa. The significance of the tip-date informed evolutionary rate signal was also determined, while accounting for the genetic relationships. This involved creating twenty datasets with randomly permuted tip-dates (Murray et al. 2016) and comparing the mean substitution rate estimate from the observed data with the 95 per cent BCIs estimated from the randomized datasets (Firth et al. 2010), assuming the same evolutionary models as in the tip-date informed analyses.

4.5 Testing the impact of landscape factors on the rate of geographic spread

The workflow for statistically analyzing the impact of landscape features (forest areas, croplands, and primary roads) on the rate of PUUV geographic spread is similar to those previously described by Dellicour, Rose and Pybus (2016) and Dellicour et al. (2017) and is summarized below. All required scripts are available in the SERAPHIM R package (Dellicour et al. 2016; R Core Team 2017). A more detailed exposition of this method is available in Supplementary Appendix S1.

Spatially and temporally referenced phylogenies for the Ardennes and Campine clades were inferred with the continuous phylogeographic method implemented in BEAST 1.8.4 (Lemey et al. 2010), using the same demographic and substitution models as those used for estimating the substitution rates were specified (see above). Because the evolutionary rate estimates largely overlap irrespective of the clock model (Fig. 5) and a strict clock model is not rejected by the data (Table 2), we chose the strict clock model for an optimal dating precision (Ho et al. 2005; Drummond et al. 2006). Information on the timed history of spread was extracted from a subset of 100 trees sampled at regular intervals from the post burn-in posterior distribution of trees. Specifically, all branches are treated as movement vectors that represent conditionally independent viral lineage dispersal events (Pybus et al. 2012) with known start and end locations (latitude and longitude) and start and end dates. These vectors are subsequently used to investigate the impact of landscape features on the PUUV dispersal velocity by computing 'environmental distances' that is spatial distances that are weighted according to the values of an environmental raster mirroring the spatial heterogeneity of the tested landscape features (Dellicour, Rose and Pybus 2016). Environmental distance computation was performed using two different models: 1, the least-cost path model, which uses a least-cost algorithm to obtain the migration route between the start and end points (Dijkstra 1959), and 2, the Circuitscape path model that can accommodate uncertainty in the migration route (McRae 2006; Mcrae et al. 2008). By correlating environmental distances to the time that was needed by the virus to migrate from start to end location, one can test the impact of landscape features on the dispersal velocity. Practically, this is

done by calculating the statistic $R = R^2_{env}/R^2_{null}$, where R^2_{env} is the coefficient of determination obtained by regressing branch durations against environmental distances computed on the environmental raster, and R^2_{null} is the coefficient of determination obtained by regressing branch durations against environmental distances computed on the 'null' raster. The R statistic, therefore, represents how much variation in migration rates, beyond that explained by geographic distance alone, is explained by considering the spatial heterogeneity of the environmental variable. Note that this statistic R is slightly different from the statistic Q ($= R^2_{env} - R^2_{null}$), previously used in Dellicour et al. (2017).

Because it is *a priori* unknown to what extent the geographic variation of the environmental factor affects the level at which a cell in the raster acts as a conductance/resistance factor relative to the null raster, the environmental distances were calculated from several rasters with rescaled cell values (Dellicour et al. 2017). For this, the original raster cell values were transformed according to $v_t = 1 + k^*(v_o/v_{max})$, where v_t and v_o are the transformed and original raster cell values, v_{max} is the maximum raster cell value recorded in the raster and k is 10, 100, or 1,000. Also, each landscape feature was tested as a resistance factor (i.e. impeding virus movement) and as a conductance factor (i.e. facilitating virus movement). In summary, for each combination of environmental factor, k parameter value and path model, an R distribution is derived from the 100 sampled phylogenies.

A landscape factor was considered as potentially explanatory, and BF support for the significance of an R distribution was calculated only when the distribution of regression coefficients was positive and when p ($R > 1$) was at least 90 per cent.

Data availability

Sequences generated for this study have been submitted to NCBI GenBank under accession numbers MG812385-MG812468.

Supplementary data

Supplementary data are available at Virus Evolution online.

Acknowledgements

This work was supported by KU Leuven (grant numbers IDO/07/005, CREA/11/027). B.Va, S.D., and B.Vr. were supported by the Research Foundation Flanders - Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO, Flanders, Belgium) (project number 1S28617N to B.Va.). S.D. was also supported by the Fonds National de la Recherche Scientifique (FNRS, Belgium).

We are grateful to Prof. Guido van der Groen for providing us *Puumala* virus isolates. We would like to thank all students and colleagues who assisted with bank vole trappings. We are thankful to Ankje de Vries for providing sampling information of the *Puumala* virus strains originating from the Netherlands.

Conflict of interest: None declared.

References

Baele, G., Lemey, P., and Suchard, M. A. (2016) 'Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty', *Systematic Biology*, 65: 250–64.

Barrios, J. M. et al. (2013a) 'Relating Land Cover and Spatial Distribution of Nephropathia Epidemica and Lyme Borreliosis in Belgium', *International Journal of Environmental Health Research*, 23: 132–54.

—— et al. (2013b) 'Seasonal Vegetation Variables and Their Impact on the Spatio-Temporal Patterns of Nephropathia Epidemica and Lyme Borreliosis in Belgium', *Applied Geography*, 45: 230–40.

—— et al. (2010) 'Satellite Derived Forest Phenology and Its Relation with Nephropathia Epidemica in Belgium', *International Journal of Environmental Research and Public Health*, 7: 2486–500.

Bennett, S. N. et al. (2014) 'Reconstructing the Evolutionary Origins and Phylogeography of Hantaviruses', *Trends in Microbiology*, 22: 473–82.

Biek, R., and Real, L. A. (2010) 'The Landscape Genetics of Infectious Disease Emergence and Spread', *Molecular Ecology*, 19: 3515–31.

Bloomquist, E. W., Lemey, P., and Suchard, M. A. (2010) 'Three Roads Diverged? Routes to Phylogeographic Inference', *Trends in Ecology & Evolution*, 25: 626–32.

Canova, L., and Fasola, M. (1991) 'Communities of Small Mammals in Six Biotopes of Northern Italy', *Acta Theriologica*, 36: 73–86.

Castel, G. et al. (2015) 'Complete Genome and Phylogeny of *Puumala hantavirus* Isolates Circulating in France', *Viruses*, 7: 5476–88.

Clement, J. et al. (1997) 'The Hantaviruses of Europe: From the Bedside to the Bench', *Emerging Infectious Diseases*, 3: 205–11.

de Vries, A. et al. (2016) 'Characterization of *Puumala hantavirus* in Bank Voles from Two Regions in The Netherlands Where Human Cases Occurred', *The Journal of General Virology*, 97: 1500–10.

Dellicour, S. et al. (2016) 'SERAPHIM: Studying Environmental Rasters and Phylogenetically Informed Movements', *Bioinformatics (Oxford, England)*, 32: 3204–6.

—— et al. (2017) 'Using Viral Gene Sequences to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics', *Molecular Biology and Evolution*, 34: 2563–71.

——, Rose, R., and Pybus, O. G. (2016) 'Explaining the Geographic Spread of Emerging Epidemics: A Framework for Comparing Viral Phylogenies and Environmental Landscape Data', *BMC Bioinformatics*, 17: 82.

Dijkstra, E. W. (1959) 'A Note on Two Problems in Connexion with Graphs', *Numerische Mathematik*, 1: 269–71.

Drummond, A., Pybus, O. G., and Rambaut, A. (2003) 'Inference of Viral Evolutionary Rates from Molecular Sequences', *Advances in Parasitology*, 54: 331–58.

Drummond, A. J. et al. (2006) 'Relaxed Phylogenetics and Dating with Confidence', *PLoS Biology*, 4: e88.

—— et al. (2003) 'Measurably Evolving Populations', *Trends in Ecology & Evolution*, 18: 481–8.

—— et al. (2005) 'Bayesian Coalescent Inference of past Population Dynamics from Molecular Sequences', *Molecular Biology and Evolution*, 22: 1185–92.

—— et al. (2012) 'Bayesian Phylogenetics with BEAUti and the BEAST 1.7', *Molecular Biology and Evolution*, 29: 1969–73.

Duchêne, S. et al. (2015) 'The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data', *Molecular Biology and Evolution*, 32: 1895–906.

Ecke, F. et al. (2017) 'Dampening of Population Cycles in Voles Affects Small Mammal Community Structure, Decreases Diversity, and Increases Prevalence of a Zoonotic Disease', *Ecology and Evolution*, 7: 5331–42.

Essbauer, S. S. et al. (2007) 'Nephropathia Epidemica in Metropolitan Area, Germany', *Emerging Infectious Diseases*, 13: 1271–3.

Firth, C. et al. (2010) 'Using Time-Structured Data to Estimate Evolutionary Rates of Double-Stranded DNA Viruses', *Molecular Biology and Evolution*, 27: 2038–51.

Gerlach, G., and Musolf, K. (2000) 'Fragmentation of Landscape as a Cause for Genetic Subdivision in Bank Voles', *Conservation Biology*, 14: 1066–74.

Gill, M. S. et al. (2016) 'Understanding past Population Dynamics: Bayesian Coalescent-Based Modeling with Covariates', *Systematic Biology*, 65: 1041–56.

—— et al. (2013) 'Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for Multiple Loci', *Molecular Biology and Evolution*, 30: 713–24.

Gouy, M., Guindon, S., and Gascuel, O. (2010) 'SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building', *Molecular Biology and Evolution*, 27: 221–4.

Gray, R. R. et al. (2011) 'The Mode and Tempo of Hepatitis C Virus Evolution within and among Hosts', *BMC Evolutionary Biology*, 11: 131.

Guivier, E. et al. (2011) 'Landscape Genetics Highlights the Role of Bank Vole Metapopulation Dynamics in the Epidemiology of Puumala Hantavirus', *Molecular Ecology*, 20: 3569–83.

Gurnell, J. (1985) 'Woodland Rodent Communities. Symp', *Zool. Soc. Lond*, 55: 377–411.

Haredasht, S. A. et al. (2013a) 'Ecological Niche Modelling of Bank Voles in Western Europe', *International Journal of Environmental Research and Public Health*, 10: 499–514.

—— et al. (2014) 'Modelling Seasonal and Multi-Annual Variation in Bank Vole Populations and Nephropathia Epidemica', *Biosystems Engineering*, 121: 25–37.

—— et al. (2013b) 'Model-Based Prediction of Nephropathia Epidemica Outbreaks Based on Climatological and Vegetation Data and Bank Vole Population Dynamics', *Zoonoses and Public Health*, 60: 461–77.

Hasegawa, M., Kishino, H., and Yano, T. (1985) 'Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA', *Journal of Molecular Evolution*, 22: 160–74.

Heiske, A. et al. (1999) 'A New Clethrionomys-Derived Hantavirus from Germany: Evidence for Distinct Genetic Sublineages of Puumala Viruses in Western Europe', *Virus Research*, 61: 101–12.

Heller, R., Chikhi, L., and Siegismund, H. R. (2013) 'The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History', *PLoS One*, 8: e62992.

Heyman, P., Vaheri, A., and Members, E. (2008) 'Situation of Hantavirus Infections and Haemorrhagic Fever with Renal Syndrome in European Countries as of December 2006', *Euro Surveillance*, 13:

Ho, S. Y. et al. (2005) 'Accuracy of Rate Estimation Using Relaxed-Clock Models with a Critical Focus on the Early Metazoan Radiation', *Molecular Biology and Evolution*, 22: 1355–63.

Holmes, E. C. et al. (2016) 'The Evolution of Ebola Virus: Insights from the 2013-2016 Epidemic', *Nature*, 538: 193–200.

Johansson, P. et al. (2008) 'Puumala Hantavirus Genetic Variability in an Endemic Region (Northern Sweden)', *Infection, Genetics, and Evolution*, 8: 286–96.

Kallio, E. R. et al. (2009) 'Cyclic Hantavirus Epidemics in Humans–Predicted by Rodent Host Dynamics', *Epidemics*, 1: 101–7.

—— et al. (2006) 'Prolonged Survival of *Puumala hantavirus* outside the Host: Evidence for Indirect Transmission via the Environment', *The Journal of General Virology*, 87: 2127–34.

Khalil, H. et al. (2014) 'The Importance of Bank Vole Density and Rainy Winters in Predicting Nephropathia Epidemica Incidence in Northern Sweden', *PLoS One*, 9: e111663.

Korva, M. et al. (2013) 'Phylogeographic Diversity of Pathogenic and Non-Pathogenic Hantaviruses in Slovenia', *Viruses*, 5: 3071–87.

Kruger, D. H. et al. (2015) 'Hantaviruses–Globally Emerging Pathogens', *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, 64: 128–36.

Laenen, L. et al. (2016) 'Spatio-Temporal Analysis of Nova Virus, a Divergent Hantavirus Circulating in the European Mole in Belgium', *Molecular Ecology*, 25: 5994–6008.

—— et al. (2018) 'A Novel Hantavirus of the European Mole, Bruges Virus, Is Involved in Frequent Nova Virus Coinfections', *Genome Biology and Evolution*, 10: 45–55.

Lemey, P. et al. (2010) 'Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time', *Molecular Biology and Evolution*, 27: 1877–85.

Linard, C. et al. (2007) 'Environmental Conditions and Puumala Virus Transmission in Belgium', *International Journal of Health Geographics*, 6: 55.

Luis, A. D. et al. (2015) 'Environmental Fluctuations Lead to Predictability in Sin Nombre Hantavirus Outbreaks', *Ecology*, 96: 1691–701.

Lythgoe, K. A. et al. (2017) 'Short-Sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections', *Trends in Microbiology*, 25: 336–48.

Maes, P. et al. (2004) 'Hantaviruses: Immunology, Treatment, and Prevention', *Viral Immunology*, 17: 481–97.

Mazurkiewicz, M. (1994) 'Factors Influencing the Distribution of the Bank Vole in Forest Habitats', *Acta Theriologica*, 39: 113–26.

McRae, B. H. (2006) 'Isolation by Resistance', *Evolution; International Journal of Organic Evolution*, 60: 1551–61.

Mcrae, B. H. et al. (2008) 'Using Circuit Theory to Model Connectivity in Ecology, Evolution, and Conservation', *Ecology*, 89: 2712–24.

Murray, G. G. et al. (2016) 'The Effect of Genetic Structure on Molecular Dating and Tests for Temporal Signal', *Methods in Ecology and Evolution*, 7: 80–9.

Palo, R. T. (2009) 'Time Series Analysis Performed on Nephropathia Epidemica in Humans of Northern Sweden in Relation to Bank Vole Population Dynamic and the NAO Index', *Zoonoses and Public Health*, 56: 150–6.

Pybus, O. G. et al. (2012) 'Unifying the Spatial Epidemiology and Molecular Evolution of Emerging Epidemics', *Proceedings of the National Academy of Sciences of the United States of America*, 109: 15066–71.

——, Tatem, A. J., and Lemey, P. (2015) 'Virus Evolution and Transmission in an Ever More Connected World', *Proceedings of the Royal Society B: Biological Sciences*, 282: 20142878.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raghwani, J. et al. (2016) 'Exceptional Heterogeneity in Viral Evolutionary Dynamics Characterises Chronic Hepatitis C Virus Infection', *PLoS Pathogens*, 12: e1005894.

Rambaut, A. et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.

Ramsden, C. et al. (2008) 'High Rates of Molecular Evolution in Hantaviruses', *Molecular Biology and Evolution*, 25: 1488–92.

Razzauti, M. et al. (2013) 'Microevolution of *Puumala hantavirus* during a Complete Population Cycle of Its Host, the Bank Vole (*Myodes glareolus*)', *PLoS One*, 8: e64447.

Ronquist, F. et al. (2012) 'MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space', *Systematic Biology*, 61: 539–42.

Sauvage, F. et al. (2003) 'Modelling Hantavirus in Fluctuating Populations of Bank Voles: The Role of Indirect Transmission on Virus Persistence', *Journal of Animal Ecology*, 72: 1–13.

Schilling, S. et al. (2007) 'Hantavirus Disease Outbreak in Germany: Limitations of Routine Serological Diagnostics and Clustering of Virus Sequences of Human and Rodent Origin', *Journal of Clinical Microbiology*, 45: 3008–14.

Stack, J. C. et al. (2010) 'Protocols for Sampling Viral Sequences to Study Epidemic Dynamics', *Journal of the Royal Society, Interface*, 7: 1119–27.

Streicker, D. G. et al. (2012) 'Rates of Viral Evolution Are Linked to Host Geography in Bat Rabies', *PLoS Pathogens*, 8: e1002720.

Szabo, R. et al. (2017) 'Phylogenetic Analysis of Puumala Virus Strains from Central Europe Highlights the Need for a Full-Genome Perspective on Hantavirus Evolution', *Virus Genes*, 53: 913–7.

Tersago, K. et al. (2008) 'Population, Environmental, and Community Effects on Local Bank Vole (*Myodes glareolus*) Puumala Virus Infection in an Area with Low Human Incidence', *Vector Borne and Zoonotic Diseases*, 8: 235–44.

——, Verhagen, R., and Leirs, H. (2011) 'Temporal Variation in Individual Factors Associated with Hantavirus Infection in Bank Voles during an Epizootic: Implications for Puumala Virus Transmission Dynamics', *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, 11: 715–21.

Tian, H. et al. (2018) 'Urbanization Prolongs Hantavirus Epidemics in Cities', *Proceedings of the National Academy of Sciences of the United States of America*, 115: 4707–12.

—— et al. (2017) 'Anthropogenically Driven Environmental Changes Shift the Ecological Dynamics of Hemorrhagic Fever with Renal Syndrome', *PLoS Pathogens*, 13: e1006198.

Torre, I., and Arrizabalaga, A. (2008) 'Habitat Preferences of the Bank Vole *Myodes glareolus* in a Mediterranean Mountain Range', *Acta Theriologica*, 53: 241–50.

Trovao, N. S. et al. (2015) 'Host Ecology Determines the Dispersal Patterns of a Plant Virus', *Virus Evolution*, 1: vev016.

Van Apeldoorn, R. C. et al. (1992) 'Effects of Habitat Fragmentation on the Bank Vole, Clethrionomys Glareolus, in an Agricultural Landscape', *Oikos*, 65: 265–74.

Van der Veken, S., Verheyen, K., and Hermy, M. (2004) 'Plant Species Loss in an Urban Area (Turnhout, Belgium) from 1880 to 1999 and Its Environmental Determinants', *Flora-Morphology, Distribution, Functional Ecology of Plants*, 199: 516–23.

Volz, E. M., and Didelot, X. (2018) 'Modeling the Growth and Decline of Pathogen Effective Population Size Provides Insight into Epidemic Dynamics and Drivers of Antimicrobial Resistance', *Systematic Biology*, 67: 719–28.

Voutilainen, L. et al. (2016) 'Temporal Dynamics of *Puumala hantavirus* Infection in Cyclic Populations of Bank Voles', *Scientific Reports*, 6: 21323.

—— et al. (2015) 'Life-Long Shedding of *Puumala hantavirus* in Wild Bank Voles (*Myodes glareolus*)', *The Journal of General Virology*, 96: 1238–47.

Vrancken, B. et al. (2014) 'The Genealogical Population Dynamics of HIV-1 in a Large Transmission Chain: Bridging within and among Host Evolutionary Rates', *PLoS Computational Biology*, 10: e1003505.

——, Suchard, M. A., and Lemey, P. (2017) 'Accurate Quantification of within- and between-Host HBV Evolutionary Rates Requires Explicit Transmission Chain Modelling', *Virus Evolution*, 3: vex028.

Weber de Melo, V. et al. (2015) 'Spatiotemporal Dynamics of *Puumala hantavirus* Associated with Its Rodent Host, *Myodes glareolus*', *Evolutionary Applications*, 8: 545–59.

Wijnhoven, S. et al. (2005) 'Flooding Ecology of Voles, Mice and Shrews: The Importance of Geomorphological and Vegetational Heterogeneity in River Floodplains', *Acta Theriologica*, 50: 453–72.

Worobey, M., Han, G. Z., and Rambaut, A. (2014) 'A Synchronized Global Sweep of the Internal Genes of Modern Avian Influenza Virus', *Nature*, 508: 254–7.

Yang, Z. (1994) 'Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites: Approximate Methods', *Journal of Molecular Evolution*, 39: 306–14.

Zhang, Y. Z., and Holmes, E. C. (2014) 'What Is the Time-Scale of Hantavirus Evolution? ', *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 25: 144–5.