

Genome-wide support for incipient Tula hantavirus species within a single rodent host lineage

Anton Labutin and Gerald Heckel*

Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, Bern 3012, Switzerland
<https://orcid.org/0000-0002-0162-323X>

*Corresponding author: E-mail: gerald.heckel@unibe.ch

Abstract

Evolutionary divergence of viruses is most commonly driven by co-divergence with their hosts or through isolation of transmission after host shifts. It remains mostly unknown, however, whether divergent phylogenetic clades within named virus species represent functionally equivalent byproducts of high evolutionary rates or rather incipient virus species. Here, we test these alternatives with genomic data from two widespread phylogenetic clades in *Tula orthohantavirus* (TULV) within a single evolutionary lineage of their natural rodent host, the common vole *Microtus arvalis*. We examined voles from forty-two locations in the contact region between clades for TULV infection by reverse transcription (RT)-PCR. Sequencing yielded twenty-three TULV Central North and twenty-one TULV Central South genomes, which differed by 14.9–18.5 per cent at the nucleotide and 2.2–3.7 per cent at the amino acid (AA) level without evidence of recombination or reassortment between clades. Geographic cline analyses demonstrated an abrupt (<1 km wide) transition between the parapatric TULV clades in continuous landscape. This transition was located within the Central mitochondrial lineage of *M. arvalis*, and genomic single nucleotide polymorphisms showed gradual mixing of host populations across it. Genomic differentiation of hosts was much weaker across the TULV Central North to South transition than across the nearby hybrid zone between two evolutionary lineages in the host. We suggest that these parapatric TULV clades represent functionally distinct, incipient species, which are likely differently affected by genetic polymorphisms in the host. This highlights the potential of natural viral contact zones as systems for investigating the genetic and evolutionary factors enabling or restricting the transmission of RNA viruses.

Keywords: host-pathogen co-evolution; *Microtus arvalis*; common vole; virus speciation; genomic admixture; RNA virus

Introduction

The evolution of parasites is usually tightly linked to their hosts, driven to a large degree by the parasites' total loss of fitness when they fail to infect their host species (Vale and Little 2009; Penczykowski, Laine, and Koskella 2016; Simmonds, Aiewsakun, and Katzourakis 2019; Ebert and Fields 2020). Similarity in physiological features, such as in the immune system, affects the range of potential host species, and thus, there is often a strong phylogenetic component in host-parasite relationships (Clark et al. 2018; Mollentze et al. 2020). Acellular parasites (viruses and bacteriophages) are particularly affected by incompatibilities with their hosts, as they lack independent metabolisms and require tight interaction with the host's cellular machinery for successful reproduction and transmission (Simmonds, Aiewsakun, and Katzourakis 2019). As a result, functional diversification and speciation in viruses depend on the genetic environment encountered in their hosts, with host-virus co-divergence (Switzer et al. 2005; Rector et al. 2007) and host switches (Parrish et al. 2008; Lin et al. 2012; Mélade et al. 2016) being the most common drivers of viral speciation.

Similar to the variety of definitions of species in their eukaryotic hosts (e.g. Galtier 2019; Kollár, Pouličková, and Dvořák 2022), there is no universal consensus on the definition of virus species, and thus, their diversity is difficult to estimate. The most commonly accepted set of definitions of virus species is provided by the International Committee on Taxonomy of Viruses (ICTV). However, these definitions vary for each virus family and distinguish virus species based on a combination of factors, such as 'natural and experimental host range, cell and tissue tropism, pathogenicity, vector specificity, antigenicity, and the degree of relatedness of their genomes or genes' (International Committee on Taxonomy of Viruses (ICTV) 2021). Depending on the particular case, these criteria are subject to discussion and carry a level of ambiguity between many viral families (Adams et al. 2013; Van Regenmortel et al. 2013; Shapiro and Polz 2015). The criteria of the ICTV are not uniformly applied across all virus families, and as a result, viral species designations tend to lack a universal biological foundation (Simmonds et al. 2017).

The definitions of the ICTV are well suited for identifying deeply separated virus species, but were not intended to classify

virus taxa that have not reached the threshold levels of evolutionary divergence but show already functionally relevant differences acquired in the process of speciation analogous to incipient eukaryotic species. A simplified species concept for acellular organisms has been proposed recently based on rates of gene flow (Bobay and Ochman 2018). This concept establishes species boundaries between virus populations in contact with one another that show no signs of gene exchange, i.e. reassortment and recombination. In this study, we use this definition to distinguish incipient virus species, although it has some limitations. Viruses with only a single segment are unable to reassort, which restricts their potential for genomic exchange to recombination. Furthermore, suitable conditions for potential co-infections by different virus clades need to exist, i.e. there should be no ecological barrier to reassortment and recombination.

Many named virus species are subdivided into deeply diverged phylogenetic clades, which can be spatially associated with the distribution of closely related host species (e.g. de Bellocq et al. 2015) or evolutionary lineages within the host (Saxenhofer et al. 2019). These clades could indicate speciation events within the virus cryptic to current taxonomy. Polymorphisms at relatively few host loci may be decisive for functional virus divergence (Meyer et al. 2016), and even single host genes have been shown to restrict the range of species that a virus can infect (Stremlau et al. 2004; van Doremalen et al. 2014; Long et al. 2016). Thus, viral clades that are confined to parapatric distribution ranges within a single evolutionary host lineage could actually represent distinct evolutionary units. These may result from cryptic or incipient speciation events on the virus side related to few genetic host polymorphisms (Saxenhofer et al. 2022). Parapatric distribution ranges have been detected at many levels of virus evolution, but further inferences are often limited because the geographic scale analysed is typically very coarse compared to the dispersal ability of the natural hosts and genetic resolution is often lacking (e.g. Drewes et al. 2017; Jeske et al. 2019).

Here, we investigated potential incipient speciation between two clades of Tula hantavirus (TULV) within its reservoir host, the common vole (*Microtus arvalis*). The genus *Orthohantavirus* contains currently thirty-eight virus species recognized by ICTV (Walker et al. 2020), which have often been linked to a single reservoir host species each. Both long-term co-speciation and occasional host shifts have been shown to play an extensive role in the deeper evolutionary history of hantaviruses (Guo et al. 2013). Recently, the ICTV proposed a new definition for classifying hantavirus species, setting the cut-off at a pairwise-evolutionary distance (PED) of 0.1 in DEmARC analysis for separate species (Laenen et al. 2019; Kuhn et al. 2023), which is likely to affect the species number once implemented.

TULV, like all hantaviruses, is a three-segmented, negative-strand RNA virus (Vaheri et al. 2013). It is horizontally transmitted in its rodent reservoir hosts causing chronic, asymptomatic infections (Forbes, Sironen, and Plyusnin 2018). The distribution of deep phylogenetic clades in TULV in Europe is partly associated with morphologically cryptic evolutionary lineages in its reservoir host, which suggests the potential importance of co-divergence in this system (Heckel et al. 2005; Saxenhofer et al. 2019; Schmidt et al. 2016; see later). Detailed analyses of the contact region between the TULV clades Central South (TULV-CEN.S) and Eastern South (TULV-EST.S) revealed an extremely tight association with the Central and Eastern evolutionary lineages in *M. arvalis* (Saxenhofer et al. 2019). These TULV clades occupy non-overlapping distribution ranges despite frequent dispersal and local gene flow between their host lineages. Their adaptive divergence and functional

separation likely evolved only after a spillover event across the host hybrid zone (Saxenhofer et al. 2019). Further analyses identified a relatively small number of host genes that may contribute to limiting the effective transmission of these two TULV clades across the hybrid zone between host lineages (Saxenhofer et al. 2022).

In this study, we use this virus–host system to test whether two deeply diverged phylogenetic clades of TULV within a single evolutionary lineage of *M. arvalis* show spatial and genomic properties consistent with distinct ‘biological species’ *sensu* Bobay and Ochman (2018). We build on phylogeographic data of partial TULV sequences, which indicated that the TULV clades Central North (TULV-CEN.N) and Central South (TULV-CEN.S) have parapatric distribution ranges within the central mitochondrial lineage of *M. arvalis*, which occupies large areas of Central Europe (Fig. 1; Schmidt et al. 2016). We combined fine-scale sampling of the potential contact region with genome-wide sequence data of the virus and the host populations in order to determine the extent of evolutionary divergence across the sampling area. If the level of functional divergence between the clades matched or exceeded TULV-CEN.S and TULV-EST.S from Saxenhofer et al. (2019), we would expect to find a sharp transition without genetic exchange between the clades despite their genetically similar hosts.

Materials and Methods

Sampling

We sampled common voles (*M. arvalis*) at forty-five trapping locations in Eastern Germany (Fig. 1) where we expected the contact between TULV-CEN.N to the north and TULV-CEN.S to the south based on Schmidt et al. (2016) and Saxenhofer et al. (2019). We termed this sampling area the ‘Central transect’ because the TULV contact was expected within the Central common vole lineage. Southernmost sampling locations from the Central transect partially overlapped with the Porcelain transect from Saxenhofer et al. (2019; 2022). Rodent trapping was performed after ethical evaluation and approval by the Bernese cantonal commission on animal experimentation under permits BE-33/14 and BE-86/17. Common voles were trapped using snap traps and stored at -20°C immediately after collection. An overview over all samples analysed in this study can be found in Supplementary Table S1.

TULV screening, whole-genome sequencing, and phylogeny

We screened 247 adult common voles from forty-two sampling sites for TULV infections. Screening was only performed for adult voles with a body weight of at least 20 g because infection rates of lighter animals are extremely low (<0.5 per cent; Schmidt et al. 2021; unpublished data) and juveniles of infected mothers are protected by maternal antibodies (Kallio et al. 2006). We extracted RNA from lung tissue with a modified QIAzol protocol (Schmidt et al. 2016). TULV infections were detected by PCR amplification of a fragment of the nucleocapsid gene in the S-segment of TULV using the RT-PCR assay detailed in Essbauer et al. (2006) and gel electrophoresis. Library preparation, sequencing, and genome assembly followed the hybrid sequence capture protocol in Hiltbrunner and Heckel (2020). We prepared RNA libraries for all TULV-infected vole hosts, as well as five additional TULV samples from sites in the western end of the Porcelain transect from Saxenhofer et al. (2019). These samples constituted four novel genomes and one replicate TULV genome (MarDHg01) originally obtained with shotgun sequencing and comparatively low read depth in

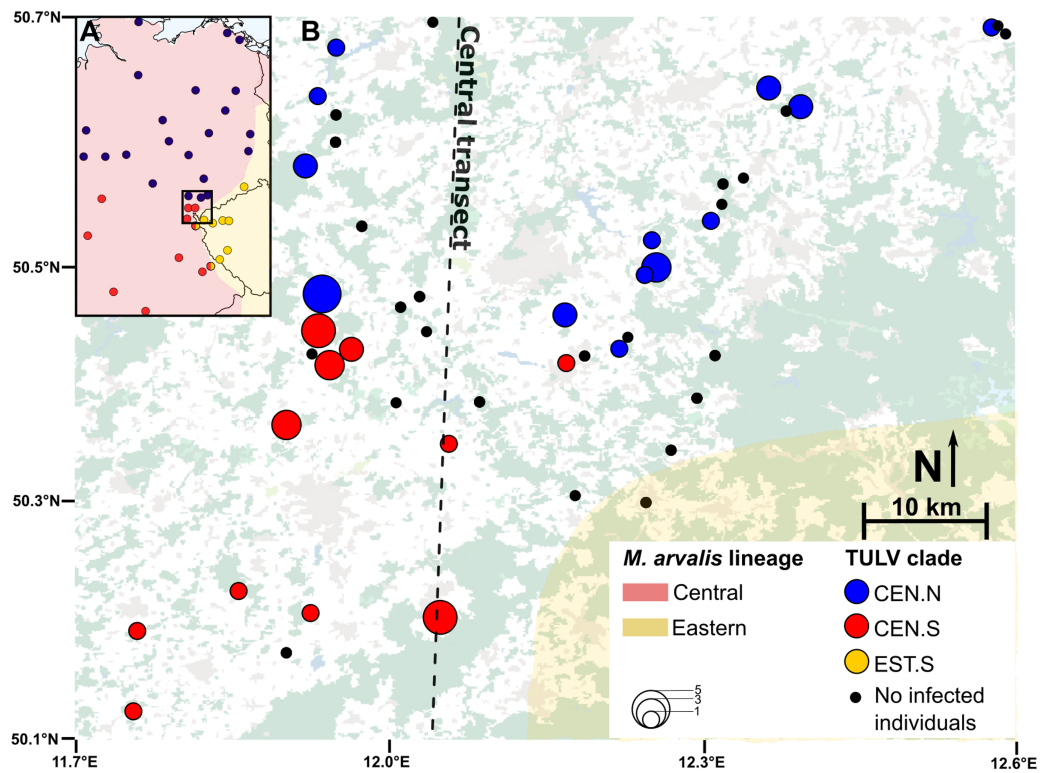


Figure 1. Contact area between Central European TULV clades in their natural rodent host. (A) Overview of our study area (square) in eastern Germany and the western Czech Republic with the evolutionary lineages Central (red shaded) and Eastern (yellow shaded) in the common vole (*M. arvalis*). The distribution of the TULV clades TULV-CEN.N (blue), TULV-CEN.S (red), and TULV-EST.S (yellow) is shown across Central Europe (based on Schmidt et al. 2016; Saxenhofer et al. 2019). (B) Sampling sites of common voles across the Central transect. Coloring of populations with infected individuals corresponds to (A). Circle sizes correspond to the frequency of infected individuals. Black dots indicate sampling sites where no infected individuals were detected. The area shaded in yellow indicates the extent of the Eastern *M. arvalis* lineage, as also indicated in (A) (Saxenhofer et al. 2019; 2022). The dashed line indicates the axis of the Central transect. The background map shows bodies of water in blue, settlements in grey, forests in green, and fields in white.

Saxenhofer et al. (2019). Two novel samples failed library quality controls and were not included in further downstream processing.

Custom baits from Hiltbrunner and Heckel (2020) were used to capture and enrich viral sequences in libraries, which were then sequenced on a MiSeq (Illumina, San Diego, CA, USA) with 2× 300 cycles on the Next-Generation Sequencing Platform of the University of Bern. We used the Iterative Virus Assembler (Hunt et al. 2015) for *de novo* assembly of TULV genomes, which were then reference mapped against the viral consensus genomes for quality control and inference of mapping statistics (see details in Hiltbrunner and Heckel 2020). For two TULV genomes, missing information for nucleotides in the L-segment (four positions in MarDOK02 and thirty-one in MarDNk29) was imputed based on the sequence of the closest genetic relative from the same sampling site (Saxenhofer et al. 2019). For each genome, we calculated the proportion of sites with a read depth of at least 3× and 20×, respectively, as well as the average genomic coverage using R.

TULV phylogenetic analysis

Phylogenetic analysis and clade assignment of the TULV genomes were based on the coding nucleotide sequences (CDSs) and the derived AA sequences of the viral genomes. For the CDS, both the concatenated sequence of all segments and individual segments were analysed. We used published TULV genomes with a complete CDS (excluding genomes with missing sites or sections of the CDS) from Central Europe (Kukkonen, Vaheri, and Plyusnin

1998; Saxenhofer et al. 2019; Hiltbrunner and Heckel 2020) as the reference. In addition, four published TULV genomes from *Microtus obscurus* from China (Chen et al. 2019) and two published *Puumala orthohantavirus* genomes (Vapalahti et al. 1992; Piiparinen et al. 1997; Ali et al. 2015) were included as outgroups (Supplementary Table S2).

The phylogenetic analysis of nucleotide sequences was performed using MrBayes version 3.2.7a (Ronquist et al. 2012) on the CIPRES platform (Miller, Pfeiffer, and Schwartz 2010). We performed Markov chain Monte Carlo (MCMC) sampling for up to 10⁸ generations in four independent runs comprising four chains, implementing reversible-jump sampling over the entire general time-reversible substitution model space (Huelsenbeck, Larget, and Alfaro 2004). After discarding a burn-in fraction of 25 per cent, samples were recorded every 10³ generations. Chains converged after 155,000 generations. The phylogenetic analysis for AA sequences was performed in MEGA X (Kumar et al. 2018). Tree topology was inferred using the maximum-likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model (Jones, Taylor, and Thornton 1992) with 1,000 bootstraps. The final tree was obtained by applying neighbour-joining and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model and then selecting the topology with superior log-likelihood value. Phylogenetic trees were drawn and edited using the online platform iTOL v5 (Letunic and Bork 2019). Maps visualizing the viral clade distribution were created in R using the geosphere package (Hijmans et al. 2017), and

topographic backgrounds are based on Globeland30 (Jun, Ban, and Li 2014).

TULV sequence diversity and signatures of selection

We used DnaSP version 5 (Rozas et al. 2003) to estimate genome-wide nucleotide diversity and between-clade divergence and perform sliding-window analyses (window size = 30 and step size = 10) of d_N/d_S ratios (ratio of non-synonymous to synonymous substitutions) and D_{XY} (average number of nucleotide substitutions per site) across the CDS of all three TULV segments. AA divergence was estimated in MEGA X by calculating the mean p -distance within and between TULV clades. Additionally, PEDs were also calculated using Tree-Puzzle version 5.2, using a maximum-likelihood approach with a Whelan and Goldman substitution model, analogous to the DEmARC analysis in Laenen et al. (2019). We tested for signatures of selection using CodeML, which is part of the PAML package version 4.9 (Yang 2007), implementing both the branch-site model (Zhang, Nielsen, and Yang 2005) and clade model C (Bielawski and Yang 2004). Phylogenetic trees for model fitting were created using the RAxML software version 8.2.12 on the CIPRES platform. Both branch-site and clade model C likelihoods were compared to respective null hypotheses using likelihood ratio tests (LRTs) and χ^2 distributions. Bayes Empirical Bayes (Yang, Wong, and Nielsen 2005) inference was used to detect sites under positive selection. To test for rate variation at synonymous sites, we performed additional scans for selection in FUBAR (Murrell et al. 2013) and MEME (Murrell et al. 2012) in HYPHY (Pond, Frost, and Muse 2005) on the Datamonkey webserver (Delpont et al. 2010). Posterior probabilities >0.85 or P values <0.1 were considered as evidence of positive selection for sites. RDP4 (Martin et al. 2015) was used with the concatenated genomes of all available TULV (Fig. 2) in order to detect reassortments and recombination. We used all methods available within the software to detect potential recombination events and only retained those which were detected across all methods to minimize uncertainty in recombination and breakpoint identification (Martin et al. 2017).

Sequencing of host mitochondrial DNA

Common vole DNA was extracted according to a standard phenol-chloroform protocol. We used mtDNA for an initial assessment of the evolutionary lineages of voles across the Central transect. We sequenced at least 288 basepairs (bp) of the cytochrome *b* gene following Fink, Excoffier, and Heckel (2004), which allows unambiguous differentiation between mtDNA lineages (Sutter, Beysard, and Heckel 2013). We sequenced a total of 132 individuals, consisting of 119 new individuals from this study and thirteen additional ones from Saxenhofer et al. (2019) in the Porcelain transect, which have not been previously sequenced (Supplementary Table S1). We analysed at least two individuals per population whenever possible to obtain a general overview of host lineages across the entire Central transect. Mitochondrial lineages were assigned based on reference sequences from Braaker and Heckel (2009) (Supplementary Table S2). Phylogenetic analysis was performed in the same way as for the CDS of the TULV genomes (see earlier).

Genotyping of host nuclear DNA

Genome-wide nuclear DNA (nucDNA) was used to infer the genetic structure of hosts via Genotyping by Sequencing (GBS) (Elshire et al. 2011). We sequenced at least five individuals per sampling site across the Central transect whenever possible, for a total of 200 individuals (Supplementary Table S1). Sequencing was carried out by LGC Biosearch Technologies (Berlin, Germany) using

Illumina NovaSeq 6000 and PstI/MspI as restriction enzymes. We combined our dataset with GBS data of 216 additional individuals from the Porcelain transect (Saxenhofer et al. 2022) processed under the same conditions. This separate dataset consisted of voles from the Central and Eastern lineages, as well as admixed individuals, and served as a reference for the assignment of the newly genotyped 200 individuals to the evolutionary lineages.

SNP calling was performed for all 416 individuals together using the GBS v2 pipeline (Tassel 5) (Glaubitz et al. 2014) with the *M. arvalis* genome (BioProject ID: PRJNA737461, Gouy et al., submitted) as reference. We utilized default parameters, except requiring a minimum of five reads to identify a unique tag. We only retained bi-allelic SNPs and called genotypes if individuals had a read depth of at least five at the locus. After SNP calling, we removed all loci with complex indels, a minor allele frequency of less than 5 per cent, more than 20 per cent missing data or observed heterozygosity greater than 50 per cent, which may indicate loci that contain paralogues merged together (White et al. 2013). Individuals with more than 50 per cent missing data were also removed (seven individuals, all from Saxenhofer et al. (2022)). We performed a linkage disequilibrium k -nearest neighbors imputation in TASSEL 5 (Glaubitz et al. 2014) for remaining missing data based on the most common state of the allele across the ten closest genetic neighbours, calculated across the thirty SNPs with the highest LD towards the missing site and keeping Ns in the case of ties. A total of 12.8 per cent of data were missing in the dataset of 409 individuals, of those 99.93 per cent were imputed. Sites which still contained missing data after imputation were discarded. To address potential batch effects of combining two independent GBS datasets, six of the 200 sequenced individuals were replicates of samples from Saxenhofer et al. (2022). One of the replicates was among the seven samples which failed quality control, leaving a total of five effective replicates. We performed all analyses of host population structures with the original dataset before imputation, a second dataset after imputation of missing data, and a third, very stringently filtered dataset in which we removed any loci from our analysis at which SNPs differed between the originals and replicates. We found minor quantitative differences between the three datasets but identical qualitative patterns across all analyses and show only the results for the stringently filtered dataset.

Host population structure

The genetic structure of common voles in the Central and Porcelain transect was analysed using the ADMIXTURE 1.3 software (Alexander, Novembre, and Lange 2009). Cross-validation (CV) was performed for cluster numbers from one through five, and CV error rates were used to determine the optimal number of clusters. ADMIXTURE was re-run for the optimal cluster number with 1,000 bootstrap replicates to establish cluster membership of all individuals. In addition, we performed a principal component analysis (PCA) using the SNPRelate package (Zheng et al. 2012) in R. For display, individuals from the Porcelain transect were assigned to evolutionary lineages based on cluster membership from the three-cluster model in admixture: Central lineage: $q_{\text{cluster1}} + q_{\text{cluster2}} \geq 0.9$, Central-Eastern admixed: $q_{\text{cluster1}} + q_{\text{cluster2}} \geq 0.1$ and $q_{\text{cluster3}} \geq 0.1$, and Eastern lineage: $q_{\text{cluster3}} \geq 0.9$.

We analysed genetic distances between host populations in order to compare the extent of population structures within the Central transect to the hybrid zone in the Porcelain transect. We first calculated pairwise F_{ST} (Weir and Cockerham 1984) between populations of four or more individuals within each transect

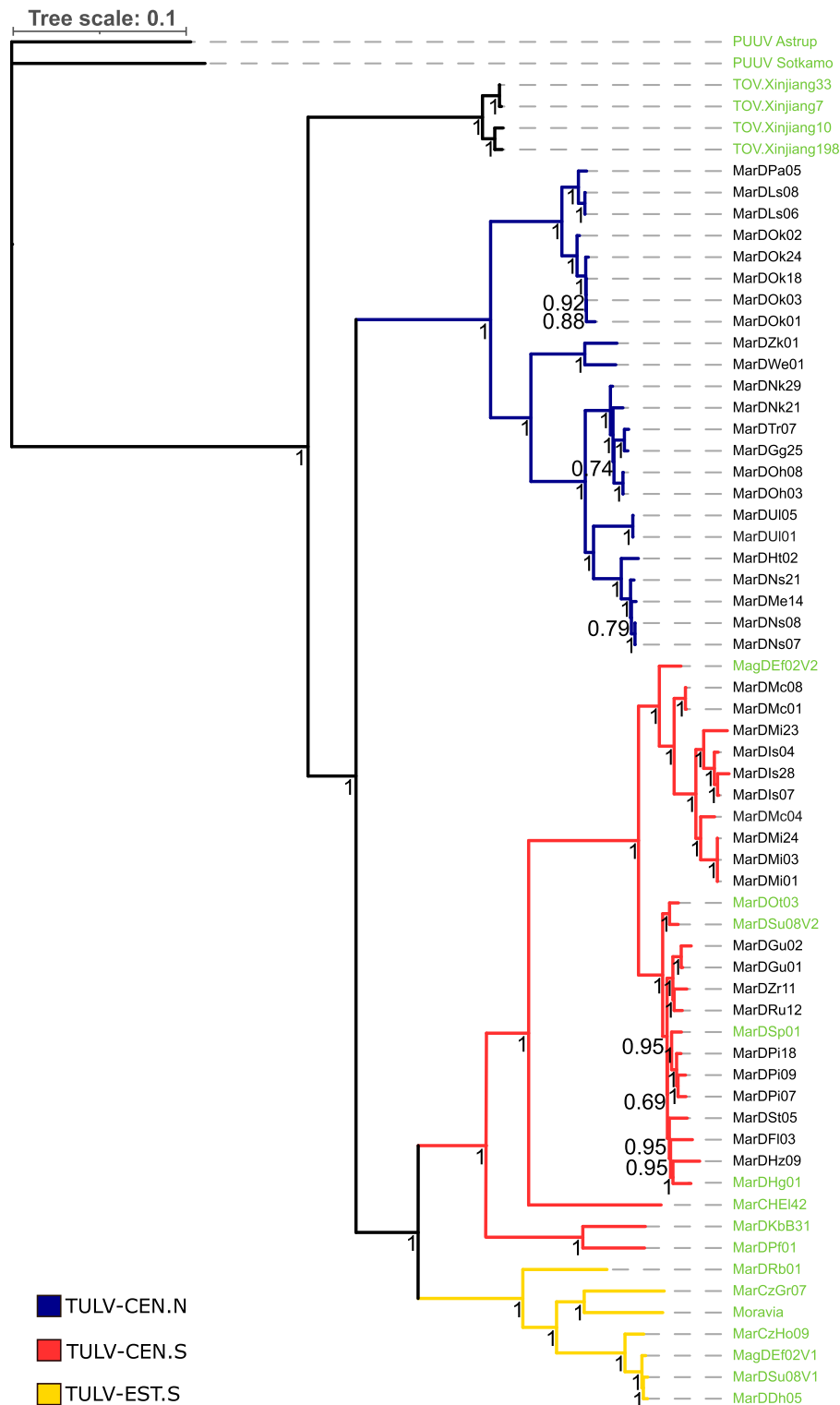


Figure 2. Phylogenetic relationships of TULV genomes from Central Europe with complete coding sequences. Phylogenetic analysis was based on the concatenated complete coding regions of TULV with samples from Xinjiang (China) and *Puumala orthohantavirus* as outgroups. Names in black show new TULV genome sequences from this study, while green represents reference and outgroup sequences. Bayesian posterior probabilities are included for all nodes. The scale bar on top shows evolutionary distance in substitutions per nucleotide.

with *vcftools* (Danecek et al. 2011). We then tested for isolation-by-distance (IBD) relationships between genetic and geographic distances among the populations within each transect via Mantel tests using the *ecodist* package (Goslee and Urban 2007) in R.

We fitted linear models for each transect to test for differences in the slopes of regressions. We compared models with different slopes for each transect with a model with a single slope for both transects using a LRT in R.

Cline analyses

In order to quantify the width and centre of the TULV clade contact zone and compare it to the genomic transition of host DNA in the Central transect, we performed an analysis of geographic clines using the HZAR package in R (Derryberry et al. 2014). We fitted a one-dimensional axis along the Central transect, which minimized the geographic distance between the TULV-CEN.N and TULV-CEN.S clades (Fig. 1). Sampling locations were projected onto the transect axis, and distances are given between the projection points (Beysard and Heckel 2014; Saxenhofer et al. 2019). For geographic cline fitting, we categorized each population based on its proportion of evolutionary lineages for mtDNA of the hosts, cluster memberships for nucDNA of the hosts, and the clade membership for TULV. For each of the different data types, four cline models were applied with increasing complexity of parameterization: null model (no cline within the sampling region), Model 1 (cline boundaries set to minimum and maximum observed frequency, free cline centre, and width), Model 2 (minimum and maximum frequency is free), and Model 3 (additional free parameters for independent exponential tails). We compared likelihood scores of all four cline models for each dataset and estimated cline parameters for the model with the highest likelihood, performing 10^5 generations of MCMC sampling in three independent chains and with a burn-in period of 10^4 iterations. Concordance of cline centres and widths was tested with a LRT in R. The test statistic was calculated as two times the difference between the log-likelihood of the alternative model of individual cline widths and centres for both datasets and a null model predicting a concordant cline through two combined datasets. Significance was determined based on a χ^2 distribution with two degrees of freedom.

Results

TULV clade distribution and divergence

Screening of 247 adult voles for TULV RNA detected the S-segment fragment in forty-one individuals. We combined these with five TULV-positive samples from Saxenhofer et al. (2019), for which complete genomes are not yet available, for whole-genome sequencing. Forty-four of these samples passed library quality control, and hybrid sequence capture yielded a total of 2,409,747 TULV sequence reads (range: 4,435–222,961 per sample) that could be *de novo* assembled and backmapped. Read depth was very large across most of the forty-four sequenced genomes with an average of 960 \times (maximum: 3,103 \times , minimum: 31 \times ; Supplementary Table S3). All genomes covered between 99.2 per cent and 100 per cent of the TULV ‘Moravia’ reference genome (Kukkonen, Vaheeri, and Plyusnin 1998), and 99.25 per cent of all sites were covered by at least three reads and 96.4 per cent by at least twenty reads. Re-sequencing of a TULV sample (MarDHg01) from

Saxenhofer et al. (2019) showed an identical sequence, albeit with a much larger depth (203 \times with sequence capture vs 31 \times with earlier shotgun sequencing).

Phylogenetic analysis assigned the forty-four new samples into twenty-three TULV-CEN.N and twenty-one TULV-CEN.S genomes both at the nucleotide and AA level (Fig. 2, Supplementary Fig. S1). Phylogenetic assignment of viral genomes to the two clades was consistent for all genomic segments. Our analyses with RDP4 indicated a potential reassortment event within the TULV-CEN.S clade (Supplementary Fig. S2) but no consistent evidence of recombination between the clades in any TULV genome segment. The newly sequenced TULV-CEN.N samples were found exclusively in the northern part of the Central transect and TULV-CEN.S in the southern part, without any discernible physical barriers (e.g. rivers, forests) to host dispersal or indications of lower host density between them (Fig. 1, Fig. 4B).

We included a published TULV-CEN.S genome (MarDsp01) from Saxenhofer et al. (2019) sampled in close proximity (6 km) to the Central transect in downstream analyses of the differences between the virus clades at the local geographic scale. Nucleotide divergence was 18 per cent between clades, while within-clade diversity was 6.9 per cent and 4 per cent for TULV-CEN.N and TULV-CEN.S, respectively (Table 1). Divergence in AA sequence was 3.25 per cent between clades and 0.57 per cent and 0.49 per cent within TULV-CEN.N and TULV-CEN.S, respectively. DEmARC analysis of the full AA sequence calculated PEDs at <0.04 and <0.01 between and within clades, respectively. Overall d_N/d_S ratios were very similar between clades (0.011) and within TULV CEN.N (0.009) and TULV CEN.S (0.012). A sliding-window analysis showed that d_N/d_S between the TULV-CEN.N and TULV-CEN.S clades was consistently low along the genome except for a d_N/d_S spike at the beginning of the M-segment (Supplementary Fig. S3). Statistical support for positive selection was only provided by FUBAR for Codon 18 in this region of the M-segment (Supplementary Tables S4 and S5). All other statistical tests supported purifying selection along most of the TULV genome (Supplementary Tables S4 and S6). Our assessment of selective constraints across the phylogeny with clade Model C of CodeML confirmed purifying selection as the main selective force affecting 88.5–96 per cent of the viral CDS (Supplementary Table S5). However, the analysis also indicated divergent selective constraints between TULV-CEN.N and TULV-CEN.S particularly for the M-segment, with a significantly elevated proportion of sites under non-purifying/neutral selection compared to the null model (TULV-CEN.N: $P=0.023$, TULV-CEN.S: $P<0.001$; Supplementary Table S5). Additional positions of interest may include codons 644 and 764, both of which encode AAs that were fixed between the two clades and fell into different sidechain polarity groups, which may alter their binding and surface reactivity properties.

Table 1. Genome-wide sequence divergence of phylogenetic clades in TULV. The table shows the within-clade diversity of TULV-CEN.N ($n=23$) and TULV-CEN.S ($n=22$) and net divergence between the two clades across the Central transect. For each genome segment, the estimates are given for the CDS, the AA sequence, and d_N/d_S ratios.

Segment	Length	TULV-CEN.N ($n=23$)			TULV-CEN.S ($n=22$)			Between clades		
		CDS (%)	AA (%)	d_N/d_S	CDS (%)	AA (%)	d_N/d_S	CDS (%)	AA (%)	d_N/d_S
L	6,459	8.82	0.76	0.009	3.12	0.40	0.012	18.54	3.67	0.007
M	3,423	4.65	0.41	0.009	6.10	0.79	0.014	18.12	2.29	0.009
S	1,287	3.91	0.04	0.004	2.62	0.15	0.005	14.88	3.67	0.023
All	11,169	6.92	0.57	0.009	3.98	0.49	0.012	17.99	3.25	0.011

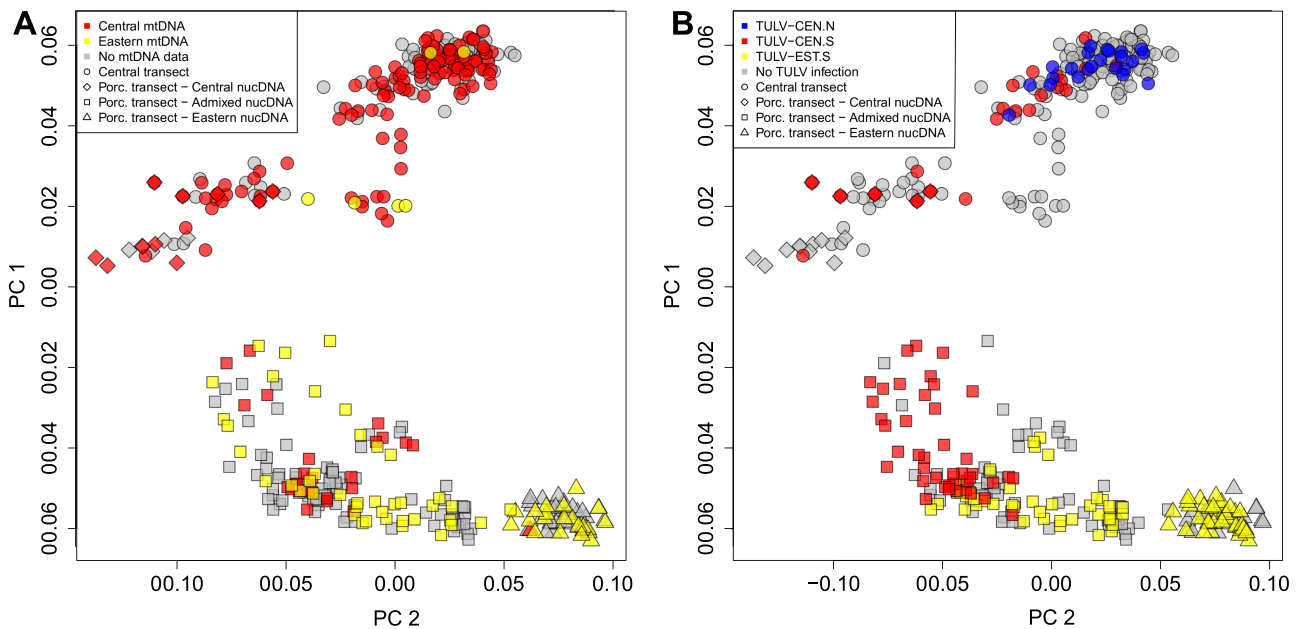


Figure 3. Associations of nuclear genomic variation in *M. arvalis* with (A) mitochondrial DNA and (B) TULV clades in the larger contact region between TULV clades. Both plots show the result of a PCA on nuclear SNPs in all 404 common voles. (A) Distribution of the Central (red) and Eastern (yellow) mitochondrial lineages in vole hosts. Grey: not sequenced for mitochondrial DNA. (B) Distribution of infections by TULV-CEN.N (blue), TULV-CEN.S (red) and TULV-EST.S (yellow) across all infected individuals. Grey: no TULV RNA detected. Principal component (PC)1 and PC2 explained 6.02 per cent and 1.53 per cent of total variance, respectively. Symbol shape indicates the transect of origin and lineage membership based on Fig. 3: circle: Central transect; rhombus: Porcelain transect, Central lineage; square: Porcelain transect, admixed between Central and Eastern lineage; and triangle: Eastern lineage.

Microtus arvalis lineage distribution and population structure

Our survey of host mtDNA at forty-four sampling locations confirmed that the Central transect contained 114 Central and only five Eastern lineage sequences in the east in voles close to the hybrid zone (Fig. 1, Supplementary Fig. S4). For the nuclear genomes, the most stringently filtered final dataset consisted of 6,947 SNPs typed in 404 vole individuals (excluding technical replicates) across the Central and Porcelain transects. The PCA showed a population structure in the region, which resembles the actual geographical distribution of the samples and the presence of the Central and Eastern evolutionary lineages (Fig. 3). Individuals from the Central transect and the western part of the Porcelain transect were separated from the Eastern lineage part of the Porcelain transect, largely reflecting mtDNA lineage distributions (Fig. 3A). Continuous genomic transitions in host nucDNA in both transects were associated with largely discrete patterns of infection by the three TULV clades (Fig. 3B).

Genetic clustering of nucDNA with ADMIXTURE revealed very similar support for three (CV error $K=3$: 0.50142) and four genetic clusters (CV error $K=4$: 0.50130) in our vole dataset. Partitioning into $K=3$ showed a genomic and spatial transition between Cluster 1 in the northern part of the Central transect and Cluster 2 in the southern part (Fig. 4A). The large majority of voles infected with TULV-CEN.N or TULV-CEN.S in the Central transect contained mostly Cluster 1 ancestry. Cluster 2 showed further a gradual transition into Cluster 3 that was largely composed of individuals in the Porcelain transect assigned as Central–Eastern hybrids and Eastern lineage common voles in Saxenhofer et al. (2022). The gradual shift of ancestry toward the east within the Porcelain transect reflects the admixture zone between the host lineages and the abrupt transition between TULV-CEN.S and TULV-EST.S (Fig. 4B; Saxenhofer et al. 2019; Saxenhofer et al. 2022). Admixture

analyses with $K=4$ produced highly similar genomic and spatial patterns and assigned the fourth genetic cluster mostly to voles in the immediate zone of hybridization in the Porcelain transect (Supplementary Fig. S5).

Direct comparisons of the spatial and genetic transitions between hosts and virus clades were performed using geographical cline analyses along the Central transect. For mtDNA, the cline null model was favoured consistent with no cline along the transect axis despite a few vole populations with introgressed Eastern lineage mtDNA (Fig. 5A; Supplementary Table S7). For nucDNA, combining admixture components from Clusters 1 and 2 vs. Cluster 3 also favoured the cline null model (Fig. 5B). The alternative combination of Cluster 1 vs. Clusters 2 and 3 supported a gradual transition along the Central transect (Fig. 5C) with the cline centre at 54.2 km (46.9–64.4 km 95 per cent confidence interval (CI)) and a width of 36.5 km (20.5–62.7 km 95 per cent CI). In contrast to the hosts, the clinal transition from the TULV-CEN.N to the TULV-CEN.S clade was extremely steep (Fig. 5D) with an estimated cline width of 0.0048 km (0.0003–1.3013 km 95 per cent CI) and the cline centre at 38.4 km (38.2–38.9 km 95 per cent CI). The TULV and host clines differed significantly in their centres and widths (LRT: $P < 0.0001$) (Supplementary Table S7).

Given largely gradual genomic transitions in common voles, we compared the extent of host population structures within the Central transect to that of the hybrid zone in the Porcelain transect. We found significant associations of genetic and spatial distances between populations (IBD) in the Central transect (Mantel test R^2 : 0.238, $P < 0.0001$) and in the Porcelain transect (Mantel test R^2 : 0.728, $P < 0.0001$; Fig. 6). However, the slope of the IBD relationship was significantly steeper (LRT: $P < 0.0001$) for populations across the hybrid zone between the Central and Eastern lineages (slope: 0.00101, Fig. 6B) than for vole populations in the Central transect (slope: 0.000414, Fig. 6A). This supported that the

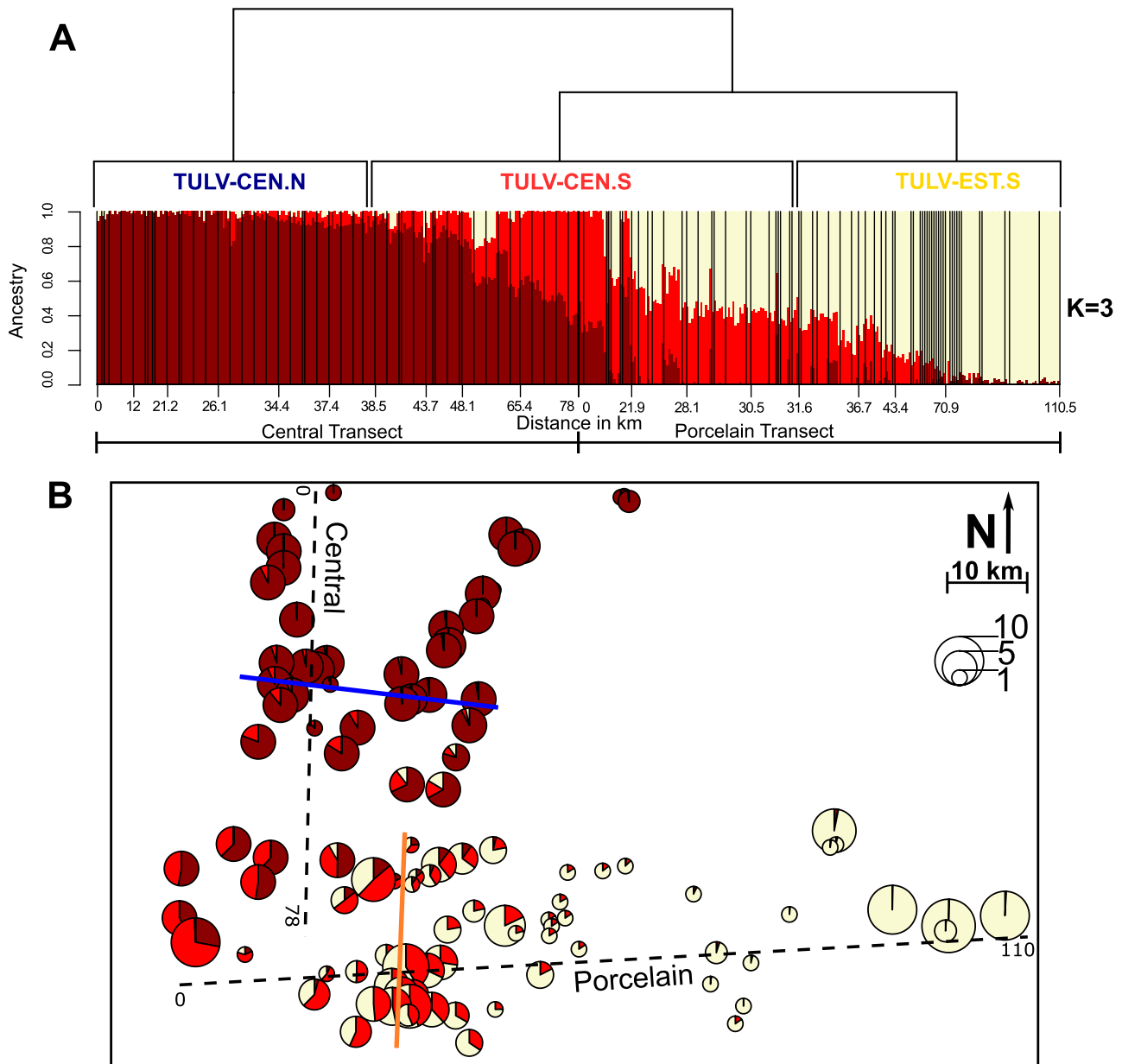


Figure 4. Genetic admixture of *M. arvalis* hosts in the larger contact region of TULV-CEN.N, TULV-CEN.S and TULV-EST.S. (A) Admixture analysis of nucDNA of 404 samples for $K=3$. Each vertical bar represents the assignment of an individual to the genetic clusters (Cluster 1: dark red, Cluster 2: light red, and Cluster 3: light yellow). Geographical distances are given as the distance of the sampling site in kilometres from the respective transect start. Black vertical lines separate individuals from different sampling sites. The cladogram above the barplot indicates the phylogenetic relationships of the TULV-CEN.N, TULV-CEN.S and TULV-EST.S clades and their distribution along the two transects. (B) Spatial overview of cluster membership of voles for the three-cluster model ($K=3$) from Admixture. Circle sizes are proportional to the number of individuals analysed per sampling site. The solid blue and orange lines show the location of contact between the TULV-CEN.N and TULV-CEN.S clades and TULV-CEN.S and TULV-EST.S clades, respectively.

gradual genomic transition in the Central transect likely reflected the genetic structure between host populations of the same lineage.

Discussion

Our study demonstrates genome-wide separation of two hantavirus clades occurring parapatrically in the common vole *M. arvalis*. The narrow viral clade transition without physical barriers to vole dispersal and thus also to virus transmission and the level of TULV divergence observed in our study are highly similar to a TULV contact zone, which is associated with two distinct

evolutionary lineages in the vole host (Saxenhofer et al. 2019). These results suggest that viral divergence much below the level of officially named virus species can be ecologically and functionally highly relevant in natural populations.

The virus clade transition and host population structure

Our analysis of common vole population structure showed a gradual genetic transition from north to south within the Central transect and low levels of admixture from the Eastern evolutionary lineage (Fig. 4). The small spatial scale of our study coupled

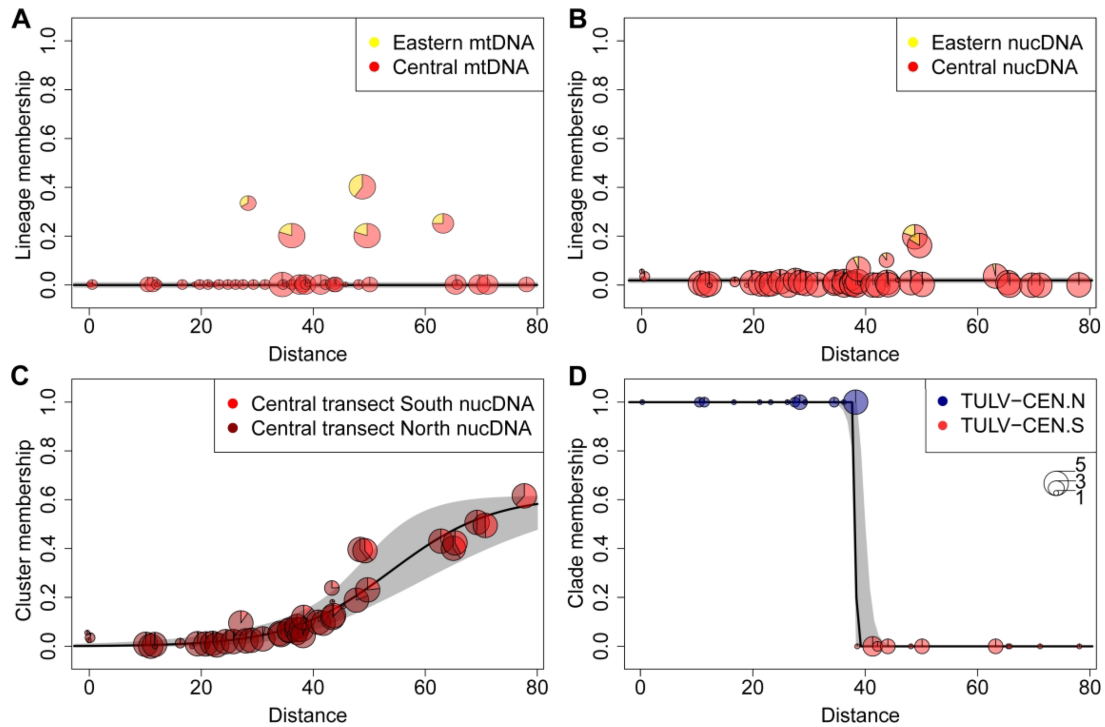


Figure 5. Geographical clines along the Central transect crossing the contact zone between the TULV-CEN.N and TULV-CEN.S clades. The transect was characterized based on the mtDNA of 119 voles (A), nucDNA of 200 voles (B and C), and clade assignment of forty-five TULV-infected voles (D). The y-axis shows the average membership towards the respective host lineages, genetic clusters, or TULV clades for each sampling site. In (A) and (B), the null model (no cline) had the lowest Akaike information criterion c (AICc). (C) shows an alternative membership assignment for nucDNA with a cline centre at 54.1 km and a cline width of 36.5 km. (D) The cline transition from TULV-CEN.N (blue) to TULV-CEN.S (red) occurred at 38.4 km with an estimated cline width of 0.007 km. 95 per cent credible cline regions are shown in grey. Distances are given relative to the northern end of the transect. Circle sizes correspond to sample sizes per site.

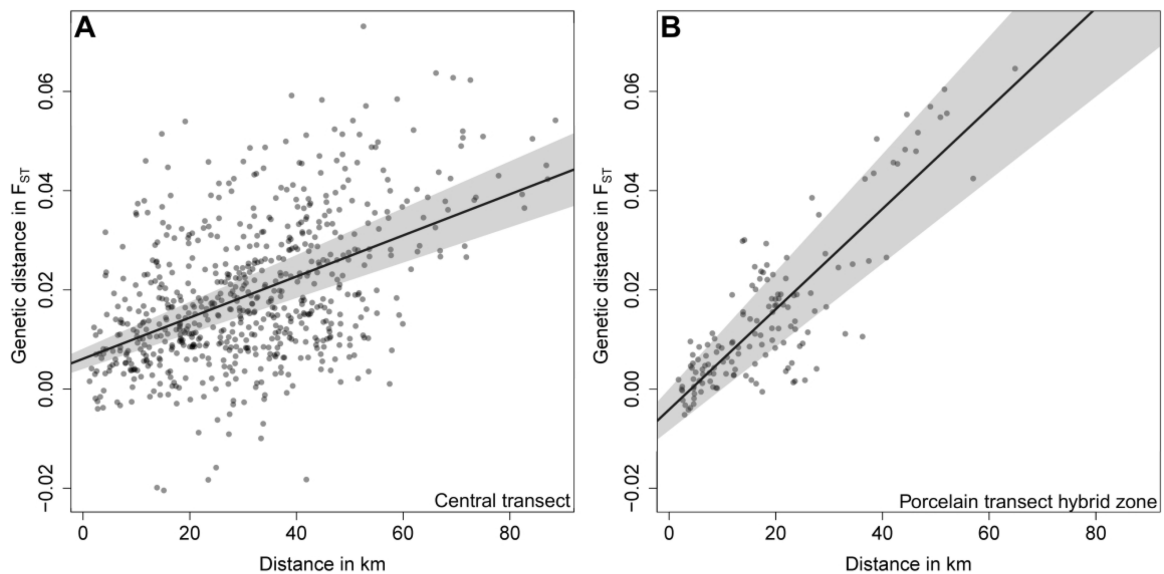


Figure 6. Genetic differentiation between common vole sampling locations in the Central and Porcelain transects. (A) Pairwise comparisons between locations along the Central transect. (B) Pairwise comparisons between locations along the Porcelain transect crossing the hybrid zone between the Central and Eastern evolutionary lineages in *M. arvalis*. (A) features a slope estimate of 0.000414, which is significantly lower (LRT: $P < 0.0001$) than the slope estimate of (B) at 0.00101. Pairwise F_{ST} was estimated between sampling sites comprising four or more individuals. 95 per cent confidence intervals around linear regression lines are plotted in grey.

with the absence of potential barriers to vole dispersal makes it unlikely that the population structure is strongly impacted by climatic differences (Gloria-Soria et al. 2017), landscape connectivity (Bastos-Silveira et al. 2012; Gryseels et al. 2017; Somoano et al.

2022), or other extrinsic factors. The genetic patterns in the Central transect are best explained by the population structure and IBD within a single evolutionary host lineage rather than by the presence of two host lineages (Fig. 5A and B).

IBD patterns have been documented in *M. arvalis* on large geographical scales (Heckel et al. 2005) as a consequence of the species' limitation to short distances of dispersal and gene flow (Schweizer, Excoffier, and Heckel 2007; Hahne et al. 2011). Clustering algorithms like in the ADMIXTURE software sometimes tend to interpret IBD patterns as a transition between genetic clusters (Meirmans 2012) such as those inferred along our Central transect. Saxenhofer et al. (2019), (2022) classified the common voles in the area of Cluster 2 (Fig. 4B) as Central lineage, consistent with our mtDNA data. We cannot, however, fully exclude that Cluster 2 in our analysis may be associated with admixture between the Central and the Eastern evolutionary lineages. It is thus possible that introgression of specific alleles from the Eastern host lineage with barrier effects towards TULV-CEN.N could locally contribute to limiting the range of the clade (see later; Saxenhofer et al. 2022). However, the contact zone between TULV-CEN.N and TULV-CEN.S extends 500–600 km westward to the Netherlands entirely through the Central host lineage (Heyman et al. 2002; Heckel et al. 2005; Lischer, Excoffier, and Heckel 2014; Schmidt et al. 2016, 2021; Maas et al. 2017; Wang, Peischl, and Heckel 2023) (Fig. 1). It is thus unlikely that genetic variation specific to the Eastern lineage has an effect on the distribution of the TULV-CEN.N clade beyond host populations in close proximity to the Central–Eastern hybrid zone.

An alternative explanation for the transition of TULV clades along the Central transect would be their evolutionary association with an undetected sublineage within the Central host lineage. This scenario would require within the Central lineage a complete, large-scale replacement of mitochondrial DNA of one of the two sublineages and a shift of the TULV clade contact in our study area (Fig. 4A, Supplementary Table S7). Differences between the distribution of mtDNA lineages and nucDNA patterns have been observed in hybrid zones of several *Microtus* species including *M. arvalis* (Braaker and Heckel 2009; Bastos-Silveira et al. 2012; Beysard et al. 2012; Sutter, Beysard, and Heckel 2013; Beysard and Heckel 2014; Beysard, Krebs-Wheaton, and Heckel 2015). However, dense geographical coverage of genetic studies has provided no support for potential sublineages in the Central host lineage (Fink, Excoffier, and Heckel 2004; Heckel et al. 2005; Braaker and Heckel 2009; Martínková et al. 2013; Beysard and Heckel 2014; Fischer et al. 2014; Schmidt et al. 2016; Wang, Peischl, and Heckel 2023), making this evolutionary scenario unlikely.

Evolutionary history of TULV clades and *M. arvalis* in Central Europe

The co-location of many borders between TULV clades and *M. arvalis* lineages (Fig. 1; Saxenhofer et al. 2019; Schmidt et al. 2016) suggests a major role of co-divergence processes in the evolutionary history of this system, similar to other hantaviruses and their hosts (Guo et al. 2013). A recent study suggested a potential origin of TULV in the Black Sea region based on partial S-segment sequence data (Cirkovic et al. 2023). The vast distribution range of TULV is only very sparsely sampled in the east (Cirkovic et al. 2023), but this would indicate that the Western and Central European TULV clades originated from a single ancestral strain from Eastern Europe similar to the deeper evolutionary history of *M. arvalis* (Heckel et al. 2005). The divergence of evolutionary lineages in *M. arvalis* is associated with the separation of populations in multiple refugia across Europe during the last glacial maximum (LGM) (Heckel et al. 2005; Lischer, Excoffier, and Heckel 2014), which probably also contributed to the separation of phylogeographic clades in TULV. The sister clades TULV-CEN.S and TULV-EST.S are likely the result of a host switch from Central to Eastern lineage hosts following the secondary contact of the two

vole lineages in the hybrid zone much after the LGM (Beysard and Heckel 2014; Saxenhofer et al. 2019). The divergence between these clades has then been accumulated over time after being isolated in their respective host lineages.

The history of divergence is less clear for the more distant relationship of TULV-CEN.N and TULV-CEN.S (Fig. 2) given the lack of genomes from other European clades, e.g. TULV Eastern North (TULV-EST.N) in the northern part of the Eastern host lineage. Partial genome sequence data exist but provide no robust support for the more basal nodes in phylogenetic reconstructions of TULV (e.g. Saxenhofer et al. 2017, 2019; Schmidt et al. 2021; Cirkovic et al. 2023). The wider application of sequence capture methods such as the one used here has the potential to provide much more resolution of (co-)evolutionary processes in TULV and other pathogens in the future (e.g. Hiltbrunner and Heckel 2020; Jeske et al. 2021).

Genome-wide isolation and incipient species in TULV

The absence of detectable recombination or reassortment between TULV-CEN.N and TULV-CEN.S in our study indicates distinct species according to the concept for acellular organisms of Bobay and Ochman (2018). However, due to their low divergence at the AA level and several other criteria, they do not meet the ICTV thresholds for distinct virus species. Rather, they can be classified as incipient species, given that barriers to genetic exchange are apparently already in effect. In general, reassortment events in hantaviruses have been observed only infrequently and all documented cases occurred within ICTV-recognized species (Klempa 2018). Co-infection with two different TULV clades, as a necessary prerequisite for reassortment or recombination to occur, has been detected only in a few individuals at the contact between TULV-CEN.S and TULV-EST.S (Saxenhofer et al. 2019; Hiltbrunner and Heckel 2020), but not in the present study. TULV prevalence generally varies between 0 and 58 per cent across populations, with an average of around 15–20 per cent in adult voles (Schmidt-Chanasit et al. 2010; Schmidt et al. 2016, 2021; Maas et al. 2017; Jeske et al. 2021). Moderate rates of infection among both dispersing and local vole individuals are therefore likely to limit the potential for double infections with two viral clades. Potential traces of older recombination events within TULV clades were detected in a few sequences, but no evidence of recombination between clades has been found (Saxenhofer et al. 2019; Hiltbrunner and Heckel 2020).

Reassortments between TULV-CEN.N and TULV-CEN.S may occasionally happen, because two potentially reassorted samples were detected in a population about 360 km to the west of our study area (Schmidt et al. 2021). The inference suggesting likely reassortment was based only on partial TULV genome sequences, making the final distinction from recombination or co-infection impossible. Given the relatively dense sampling of TULV diversity in the studies of clade contact zones here and in Saxenhofer et al. (2019), we suggest that the products of reassortment or recombination between clades are probably less fit than their ancestral counterparts and get purged from populations (see e.g. McDonald et al. 2016).

Adaptive interactions between TULV and vole genomes

Genome-wide data have suggested a particular adaptive role for clade-specific differences in a set of five codons at the beginning of the TULV M-segment in this study for TULV-CEN.N and TULV-CEN.S and Saxenhofer et al. (2019) for TULV-CEN.S and TULV-EST.S. This region corresponds to the N-terminus of the glycoprotein and may represent a signal peptide or the beginning of the N-terminal

ectodomain (Vaheiri et al. 2013; Ganaie and Mir 2014; Li et al. 2016). Interactions between this region and specific host genes are hypothesized to be drivers of TULV diversification and speciation (Saxenhofer et al. 2019, 2022). In particular, Codon 17 of the M-segment of TULV-CEN.S and TULV-EST.S was found to be under positive selection (Saxenhofer et al. 2019) and our analyses here showed indications for Codon 18. The more than five-fold increase in d_N/d_S at the beginning of the M-segment compared to the rest of the genome (Supplementary Fig. S3) suggests differences in the adaptive regime. However, high rates of synonymous variation and purifying selection on linked sites may prevent the detection of an even stronger signature of positive selection.

On the host side, several candidate genes with a potential effect as a barrier to the transmission of non-adapted TULV-CEN.S or TULV-EST.S clades have been identified in a genome-wide association study (Saxenhofer et al. 2022). These genes may be targets for interactions with TULV in general and the region at the beginning of the M-segment in particular. However, most of the SNP alleles significantly associated with infection by TULV-CEN.S or TULV-EST.S in the hybrid zone between evolutionary lineages were not present in the Central transect dataset. The sharp parapatric distribution of the TULV-CEN.N and TULV-CEN.S clades is nevertheless likely to be associated with yet undefined genomic barriers in the vole host. It is possible that the same genome region in different TULV clades may interact differently with multiple genomic regions in the common vole hosts depending on their deeper evolutionary background. An analogous system was described for the 3' untranslated region in the Dengue virus and its mosquito and human hosts (Villordo et al. 2015). The viral region has specific adaptations to each host, potentially suppressing Interferon- α/β activities in human cells and RNAi pathways in mosquito cells (Villordo et al. 2015). Extended genomic and functional analyses will be necessary in the future to determine the general and specific interactions between TULV clades and their hosts.

Evolving virus species in single host species

The generation of new species in parasites in general and viruses in particular is most commonly studied in the context of host-parasite co-divergence and host shifts (Switzer et al. 2005; Rector et al. 2007; Parrish et al. 2008; Sharp and Simmonds 2011; Lin et al. 2012; Longdon et al. 2014; Mélade et al. 2016). Speciation or co-existence of sibling parasite species within single host species has only been documented in a few cases and requires specific conditions, e.g. within species variation of key host genes (Pérez-Tris et al. 2007; Meyer et al. 2016; Martinů, Hypša, and Štefka 2018; Saxenhofer et al. 2019; Martinů et al. 2020; Chaikeratisak et al. 2021). Viral contact zones within a single reservoir host can provide important insights into the genetic environment that can facilitate speciation and the extent of functional divergence between incipient species. For example, the contacts between parapatric arenaviruses within morphologically cryptic evolutionary lineages of their rodent host show similarities with the TULV case (Gryseels et al. 2017; Cuypers et al. 2020) even though lower spatial and genetic resolution limits the conclusions. In comparison, murine cytomegalovirus shows multiple viral clades that are associated with *Mus musculus musculus* or *M. m. domesticus* in the European hybrid zone, but lacks the sharp contact (de Bellocq et al. 2015). The murine cytomegalovirus genome has been shaped by extensive recombination (Smith et al. 2013), which indicates that its clades belong to the same virus species despite association with separate host species. These findings emphasize that host divergence is not necessarily a direct indicator for virus divergence in closely related species and highlight the importance of

high-resolution studies for the discovery and characterization of barriers to viral transmission in genetically similar hosts.

Conclusions

The diversification of viruses within and between host species involves complex processes, which are vital for our understanding of the evolution and management of viruses around the globe. The combined analysis of virus and host genomes allowed us to determine that parapatric TULV clades behave from an evolutionary point of view like incipient species within a highly similar genetic environment. A particular feature of this system is that the multiple levels of divergence in both the virus and the host enable further studies on hierarchical phylogenetic levels and thus have the potential to provide much deeper insights than systems with one level only. Further studies support the idea that small genetic differences not only between viruses but also between hosts can be crucial for limiting the infection ranges of viruses (Stremlau et al. 2004; van Doremalen et al. 2014; Long et al. 2016). The partial decoupling of virus evolution from relatively simple co-divergence processes stresses the importance of combining information on both agents in host-parasite systems. We expect that in-depth characterization and phylogenomic analyses of pathogen populations together with their reservoir hosts will contribute to uncovering the full spectrum of factors that constitute fundamental barriers to viral transmission in genetically highly similar hosts and ultimately drive viral speciation.

Data availability

Tula Hantavirus S-segment and common vole mtDNA sequences: National Center for Biotechnology Information GenBank accession numbers OP173222-OP173485. Genotype by Sequencing data: Sequence read archive accession number: PRJNA869681. Keyfile for Genotype by Sequencing data: Dryad repository DOI: 10.5061/dryad.w0vt4b905.

Supplementary data

Supplementary data is available at *VEVOLU Journal* online.

Acknowledgements

We thank Susanne Tellenbach for laboratory support; Melanie Hiltbrunner, Moritz Saxenhofer, Guy Schmidrig, Joana Rechsteiner, Tobias Böni, Carla Anliker, Mauro Gwerder, Till Meyer zu Westram, Marco Felber, Dario Neokleous, and Alexandra Zurlinden for assistance with sample collection; Stephan Peischl for advice on statistical analysis; and Kimberly Gilbert for comments on the manuscript. We thank the Next-Generation Sequencing Platform of the University of Bern for the sequencing services. We thank the anonymous reviewers for insightful suggestions that helped to improve the manuscript. This study was supported by grant 31003A_176209 from the Swiss National Science Foundation to Gerald Heckel.

Conflict of interest: None declared.

References

Adams, M. J. et al. (2013) 'Recently Agreed Changes to the International Code of Virus Classification and Nomenclature', *Archives of Virology*, 158: 2633–9.

- Alexander, D. H., Novembre, J., and Lange, K. (2009) 'Fast Model-based Estimation of Ancestry in Unrelated Individuals', *Genome Research*, 19: 1655–64.
- Ali, H. S. et al. (2015) 'Complete Genome of a Puumala Virus Strain from Central Europe', *Virus Genes*, 50: 292–8.
- Bastos-Silveira, C. et al. (2012) 'Deep Mitochondrial Introgression and Hybridization among Ecologically Divergent Vole Species', *Molecular Ecology*, 21: 5309–23.
- Beysard, M. et al. (2012) 'Asymmetric and Differential Gene Introgression at a Contact Zone between Two Highly Divergent Lineages of Field Voles (*Microtus Agrestis*)', *Journal of Evolutionary Biology*, 25: 400–8.
- Beysard, M., and Heckel, G. (2014) 'Structure and Dynamics of Hybrid Zones at Different Stages of Speciation in the Common Vole (*Microtus Arvalis*)', *Molecular Ecology*, 23: 673–87.
- Beysard, M., Krebs-Wheaton, R., and Heckel, G. (2015) 'Tracing Reinforcement through Asymmetrical Partner Preference in the European Common Vole *Microtus Arvalis*', *BMC Evolutionary Biology*, 15: 170.
- Bielawski, J. P., and Yang, Z. (2004) 'A Maximum Likelihood Method for Detecting Functional Divergence at Individual Codon Sites, with Application to Gene Family Evolution', *Journal of Molecular Evolution*, 59: 121–32.
- Bobay, L.-M., and Ochman, H. (2018) 'Biological Species in the Viral World', *Proceedings of the National Academy of Sciences*, 115: 6040–5.
- Braaker, S., and Heckel, G. (2009) 'Transalpine Colonisation and Partial Phylogeographic Erosion by Dispersal in the Common Vole (*Microtus Arvalis*)', *Molecular Ecology*, 18: 2518–31.
- Chaikeeratisak, V. et al. (2021) 'Viral Speciation through Subcellular Genetic Isolation and Virogenesis Incompatibility', *Nature Communications*, 12: 1–9.
- Chen, J.-T. et al. (2019) 'Identification and Characterization of a Novel Subtype of Tula Virus in *Microtus Arvalis Obscurus* Voles Sampled from Xinjiang, China', *Infection Genetics & Evolution*, 75: 104012.
- Cirkovic, V. et al. (2023) 'Phylogeographic Analysis of Tula Hantavirus Highlights a Single Introduction to Central Europe', *Virus Evolution*, 8: veac112.
- Clark, N. J. et al. (2018) 'Climate, Host Phylogeny and the Connectivity of Host Communities Govern Regional Parasite Assembly', *Diversity and Distributions*, 24: 13–23.
- Cuypers, L. N. et al. (2020) 'Three Arenaviruses in Three Subspecific Natal Multimammate Mouse Taxa in Tanzania: Same Host Specificity, but Different Spatial Genetic Structure?', *Virus Evolution*, 6: veaa039.
- Danecek, P. et al. (2011) 'The Variant Call Format and VCFtools', *Bioinformatics*, 27: 2156–8.
- de Bellocq, J. G. et al. (2015) 'Murine Cytomegalovirus Is Not Restricted to the House Mouse *Mus Musculus Domesticus*: Prevalence and Genetic Diversity in the European House Mouse Hybrid Zone', *Journal of Virology*, 89: 406–14.
- Delpont, W. et al. (2010) 'Datamonkey 2010: A Suite of Phylogenetic Analysis Tools for Evolutionary Biology', *Bioinformatics*, 26: 2455–7.
- Derryberry, E. P. et al. (2014) 'HZAR: Hybrid Zone Analysis Using an R Software Package', *Molecular Ecology Resources*, 14: 652–63.
- Drewes, S. et al. (2017) 'Host-associated Absence of Human Puumala Virus Infections in Northern and Eastern Germany', *Emerging Infectious Diseases*, 23: 83–6.
- Ebert, D., and Fields, P. D. (2020) 'Host-parasite Co-evolution and Its Genomic Signature', *Nature Reviews Genetics*, 21: 754–68.
- Elschire, R. J. et al. (2011) 'A Robust, Simple Genotyping-by-sequencing (GBS) Approach for High Diversity Species', *PLoS One*, 6: e19379.
- Essbauer, S. et al. (2006) 'A New Puumala Hantavirus Subtype in Rodents Associated with an Outbreak of Nephropathia Epidemica in South-East Germany in 2004', *Epidemiology & Infection*, 134: 1333–44.
- Fink, S., Excoffier, L., and Heckel, G. (2004) 'Mitochondrial Gene Diversity in the Common Vole *Microtus Arvalis* Shaped by Historical Divergence and Local Adaptations', *Molecular Ecology*, 13: 3501–14.
- Fischer, M. C. et al. (2014) 'Continental-scale Footprint of Balancing and Positive Selection in a Small Rodent (*Microtus Arvalis*)', *PLoS One*, 9: e112332.
- Forbes, K. M., Sironen, T., and Plyusnin, A. (2018) 'Hantavirus Maintenance and Transmission in Reservoir Host Populations', *Current Opinion in Virology*, 28: 1–6.
- Galtier, N. (2019) 'Delineating Species in the Speciation Continuum: A Proposal', *Evolutionary Applications*, 12: 657–63.
- Ganaie, S. S., and Mir, M. A. (2014) 'The Role of Viral Genomic RNA and Nucleocapsid Protein in the Autophagic Clearance of Hantavirus Glycoprotein Gn', *Virus Research*, 187: 72–6.
- Glaubitz, J. C. et al. (2014) 'TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline', *PLoS One*, 9: e90346.
- Gloria-Soria, A. et al. (2017) 'Infection Rate of *Aedes Aegypti* Mosquitoes with Dengue Virus Depends on the Interaction between Temperature and Mosquito Genotype', *Proceedings of the Royal Society B: Biological Sciences*, 284: 20171506.
- Goslee, S. C., and Urban, D. L. (2007) 'The Ecodist Package for Dissimilarity-based Analysis of Ecological Data', *Journal of Statistical Software*, 22: 1–19.
- Gryseels, S. et al. (2017) 'When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure in the Spatial Spread of Arenaviruses', *PLoS Pathogens*, 13: 1–22.
- Guo, W.-P. et al. (2013) 'Phylogeny and Origins of Hantaviruses Harbored by Bats, Insectivores, and Rodents', *PLoS Pathogens*, 9: e1003159.
- Hahne, J. et al. (2011) 'Establishment Success and Resulting Fitness Consequences for Vole Dispersers', *Oikos*, 120: 95–105.
- Heckel, G. et al. (2005) 'Genetic Structure and Colonization Processes in European Populations of the Common Vole, *Microtus Arvalis*', *Evolution*, 59: 2231–42.
- Heyman, P. et al. (2002) 'Tula Hantavirus in Belgium', *Epidemiology & Infection*, 128: 251–6.
- Hijmans, R. J. et al. (2017) 'Package Geosphere', *Spherical Trigonometry*, 1: 1–45.
- Hiltbrunner, M., and Heckel, G. (2020) 'Assessing Genome-Wide Diversity in European Hantaviruses through Sequence Capture from Natural Host Samples', *Viruses*, 12: 749.
- Huelsensbeck, J. P., Larget, B., and Alfaro, M. E. (2004) 'Bayesian Phylogenetic Model Selection Using Reversible Jump Markov Chain Monte Carlo', *Molecular Biology and Evolution*, 21: 1123–33.
- Hunt, M. et al. (2015) 'IVA: Accurate de Novo Assembly of RNA Virus Genomes', *Bioinformatics*, 31: 2374–6.
- International Committee on Taxonomy of Viruses (ICTV). (2021) <<https://ictv.global/taxonomy/>> accessed 15 Jul 2021.
- et al. (2019) 'Field Vole-associated Traemmersee Hantavirus from Germany Represents a Novel Hantavirus Species', *Virus Genes*, 55: 848–53.
- Jeske, K. et al. (2021) 'Hantavirus-*Leptospira* Coinfections in Small Mammals from Central Germany', *Epidemiology & Infection*, 149: 97.
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) 'The Rapid Generation of Mutation Data Matrices from Protein Sequences', *Bioinformatics*, 8: 275–82.

- Jun, C., Ban, Y., and Li, S. (2014) 'Open Access to Earth Land-cover Map', *Nature*, 514: 434.
- Kallio, E. R. et al. (2006) 'Maternal Antibodies Postpone Hantavirus Infection and Enhance Individual Breeding Success', *Proceedings of the Royal Society B: Biological Sciences*, 273: 2771–6.
- Klempa, B. (2018) 'Reassortment Events in the Evolution of Hantaviruses', *Virus Genes*, 54: 638–46.
- Kollár, J., Pouličková, A., and Dvořák, P. (2022) 'On the Relativity of Species, or the Probabilistic Solution to the Species Problem', *Molecular Ecology*, 31: 411–8.
- Kuhn, J. H. et al. (2023) 'Pending Reorganization of Hantaviridae to Include Only Completely Sequenced Viruses: A Call to Action', *Viruses*, 15: 660.
- Kukkonen, S. K., Vaheri, A., and Plyusnin, A. (1998) 'Completion of the Tula Hantavirus Genome Sequence: Properties of the L Segment and Heterogeneity Found in the 3' termini of S and L Genome RNAs', *Journal of General Virology*, 79: 2615–22.
- Kumar, S. et al. (2018) 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms', *Molecular Biology and Evolution*, 35: 1547–9.
- Laenen, L. et al. (2019) 'Hantaviridae: Current Classification and Future Perspectives', *Viruses*, 11: 788.
- Letunic, I., and Bork, P. (2019) 'Interactive Tree of Life (ItoL) V4: Recent Updates and New Developments', *Nucleic Acids Research*, 47: W256–9.
- Li, S. et al. (2016) 'A Molecular-level Account of the Antigenic Hantaviral Surface', *Cell Reports*, 15: 959–67.
- Lin, X.-D. et al. (2012) 'Cross-species Transmission in the Speciation of the Currently Known Murinae-associated Hantaviruses', *Journal of Virology*, 86: 11171–82.
- Lischer, H. E. L., Excoffier, L., and Heckel, G. (2014) 'Ignoring Heterozygous Sites Biases Phylogenomic Estimates of Divergence Times: Implications for the Evolutionary History of *Microtus Voles*', *Molecular Biology and Evolution*, 31: 817–31.
- Long, J. S. et al. (2016) 'Species Difference in ANP32A Underlies Influenza A Virus Polymerase Host Restriction', *Nature*, 529: 101–4.
- Longdon, B. et al. (2014) 'The Evolution and Genetics of Virus Host Shifts', *PLoS Pathogens*, 10: e1004395.
- Maas, M. et al. (2017) 'High Prevalence of Tula Hantavirus in Common Voles in the Netherlands', *Vector-Borne and Zoonotic Diseases*, 17: 200–5.
- Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.
- et al. (2017) 'Detecting and Analyzing Genetic Recombination Using RDP4', in J. M. Keith (ed) *Bioinformatics*, pp. 433–60. Springer.
- Martínková, N. et al. (2013) 'Divergent Evolutionary Processes Associated with Colonization of Offshore Islands', *Molecular Ecology*, 22: 5205–20.
- Martinů, J. et al. (2020) "'Parasite Turnover Zone" at Secondary Contact: A New Pattern in Host–parasite Population Genetics', *Molecular Ecology*, 29: 4653–64.
- Martinů, J., Hypša, V., and Štefka, J. (2018) 'Host Specificity Driving Genetic Structure and Diversity in Ectoparasite Populations: Coevolutionary Patterns in *Apodemus* Mice and Their Lice', *Ecology and Evolution*, 8: 10008–22.
- McDonald, S. M. et al. (2016) 'Reassortment in Segmented RNA Viruses: Mechanisms and Outcomes', *Nature Reviews, Microbiology*, 14: 448–60.
- Meirmans, P. G. (2012) 'The Trouble with Isolation by Distance', *Molecular Ecology*, 21: 2839–46.
- Mélade, J. et al. (2016) 'An Eco-epidemiological Study of Morbilli-related Paramyxovirus Infection in Madagascar Bats Reveals Host-switching as the Dominant Macro-evolutionary Mechanism', *Scientific Reports*, 6: 1–12.
- Meyer, J. R. et al. (2016) 'Ecological Speciation of Bacteriophage Lambda in Allopatry and Sympatry', *Science*, 354: 1301–4.
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010) 'Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees'. 2010 Gateway Computing Environments Workshop (GCE), Institute of Electrical and Electronics Engineers New Orleans, LA, USA, 14 November 2010. pp. 1–8.
- Mollentze, N. et al. (2020) 'Virulence Mismatches in Index Hosts Shape the Outcomes of Cross-species Transmission', *Proceedings of the National Academy of Sciences*, 117: 28859–66.
- et al. (2012) 'Detecting Individual Sites Subject to Episodic Diversifying Selection', *PLoS Genet*, 8: e1002764.
- Murrell, B. et al. (2013) 'FUBAR: A Fast, Unconstrained Bayesian Approximation for Inferring Selection', *Molecular Biology and Evolution*, 30: 1196–205.
- Parrish, C. R. et al. (2008) 'Cross-species Virus Transmission and the Emergence of New Epidemic Diseases', *Microbiology and Molecular Biology Reviews*, 72: 457–70.
- Penczykowski, R. M., Laine, A.-L., and Koskella, B. (2016) 'Understanding the Ecology and Evolution of Host-parasite Interactions across Scales', *Evolutionary Applications*, 9: 37–52.
- Pérez-Tris, J. et al. (2007) 'Within-host Speciation of Malaria Parasites', *PLoS One*, 2: e235.
- Piiparinen, H. et al. (1997) 'Sequence Analysis of the Puumala Hantavirus Sotkamo Strain L Segment', *Virus Research*, 51: 1–7.
- Pond, S. L. K., Frost, S.D.W., and Muse, S. V. (2005) 'HyPhy: hypothesis testing using phylogenies', in *Bioinformatics*, Vol 21 pp. 676–9.
- Rector, A. et al. (2007) 'Ancient Papillomavirus-host Co-speciation in Felidae', *Genome Biology*, 8: 1–12.
- 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.
- Rozas, J. et al. (2003) 'DnaSP, DNA Polymorphism Analyses by the Coalescent and Other Methods', *Bioinformatics*, 19: 2496–7.
- Saxenhofer, M. et al. (2017) 'Revised Time Scales of RNA Virus Evolution Based on Spatial Information', *Proceedings of the Royal Society B: Biological Sciences*, 284: 20170857.
- et al. (2019) 'Secondary Contact between Diverged Host Lineages Entails Ecological Speciation in a European Hantavirus', *PLoS Biology*, 17: e3000142.
- et al. (2022) 'Host Genetic Factors Associated with the Range Limit of a European Hantavirus', *Molecular Ecology*, 31: 252–65.
- et al. (2016) 'High Genetic Structuring of Tula Hantavirus', *Archives of Virology*, 161: 1135–49.
- Schmidt, S. et al. (2021) 'Spatial and Temporal Dynamics and Molecular Evolution of Tula Orthohantavirus in German Vole Populations', *Viruses*, 13: 1132.
- Schmidt-Chanasit, J. et al. (2010) 'Extensive Host Sharing of Central European Tula Virus', *Journal of Virology*, 84: 459–74.
- Schweizer, M., Excoffier, L., and Heckel, G. (2007) 'Fine-scale Genetic Structure and Dispersal in the Common Vole (*Microtus Arvalis*)', *Molecular Ecology*, 16: 2463–73.
- Shapiro, B. J., and Polz, M. F. (2015) 'Microbial Speciation', *Cold Spring Harbor Perspectives in Biology*, 7: a018143.
- Sharp, P. M., and Simmonds, P. (2011) 'Evaluating the Evidence for Virus/host Co-evolution', *Current Opinion in Virology*, 1: 436–41.
- Simmonds, P. et al. (2017) 'Virus Taxonomy in the Age of Metagenomics', *Nature Reviews, Microbiology*, 15: 161–8.
- Simmonds, P., Aiewsakun, P., and Katzourakis, A. (2019) 'Prisoners of War – Host Adaptation and Its Constraints on Virus Evolution', *Nature Reviews, Microbiology*, 17: 321–8.

- Smith, L. M. et al. (2013) 'The Genome of Murine Cytomegalovirus Is Shaped by Purifying Selection and Extensive Recombination', *Virology*, 435: 258–68.
- Somoano, A. et al. (2022) 'A Borage Landscape Restricts the Gene Flow of Pest Vole Populations', *Life*, 12: 800.
- Stremlau, M. et al. (2004) 'The Cytoplasmic Body Component TRIM5 α Restricts HIV-1 Infection in Old World Monkeys', *Nature*, 427: 848–53.
- Sutter, A., Beysard, M., and Heckel, G. (2013) 'Sex-specific Clines Support Incipient Speciation in a Common European Mammal', *Heredity*, 110: 398–404.
- Switzer, W. M. et al. (2005) 'Ancient Co-speciation of Simian Foamy Viruses and Primates', *Nature*, 434: 376–80.
- Vaheri, A. et al. (2013) 'Uncovering the Mysteries of Hantavirus Infections', *Nature Reviews, Microbiology*, 11: 539–50.
- Vale, P. F., and Little, T. J. (2009) 'Measuring Parasite Fitness under Genetic and Thermal Variation', *Heredity*, 103: 102–9.
- van Doremalen, N. et al. (2014) 'Host Species Restriction of Middle East Respiratory Syndrome Coronavirus through Its Receptor, Dipeptidyl Peptidase 4', *Journal of Virology*, 88: 9220–32.
- Van Regenmortel, M. H. V. et al. (2013) 'Virus Species Polemics: 14 Senior Virologists Oppose a Proposed Change to the ICTV Definition of Virus Species', *Archives of Virology*, 158: 1115–9.
- Vapalahti, O. et al. (1992) 'Cloning and Sequencing of Puumala Virus Sotkamo Strain S and M RNA Segments: Evidence for Strain Variation in Hantaviruses and Expression of the Nucleocapsid Protein', *Journal of General Virology*, 73: 829–38.
- Villordo, S. M. et al. (2015) 'Dengue Virus RNA Structure Specialization Facilitates Host Adaptation', *PLoS Pathogens*, 11: e1004604.
- Walker, P. J. et al. (2020) 'Changes to Virus Taxonomy and the Statutes Ratified by the International Committee on Taxonomy of Viruses (2020)', *Archives of Virology*, 165: 2737–48.
- Wang, X., Peischl, S., and Heckel, G. (2023) 'Demographic History and Genomic Consequences of 10,000 Generations of Isolation in a Wild Mammal', *Current Biology*, 33: 2051–62.
- Weir, B. S., and Cockerham, C. C. (1984) 'Estimating F-statistics for the Analysis of Population Structure', *Evolution*, 38: 1358–70.
- White, T. A. et al. (2013) 'Adaptive Evolution during an Ongoing Range Expansion: The Invasive Bank Vole (*Myodes glareolus*) in Ireland', *Molecular Ecology*, 22: 2971–85.
- Yang, Z. (2007) 'PAML 4: Phylogenetic Analysis by Maximum Likelihood', *Molecular Biology and Evolution*, 24: 1586–91.
- Yang, Z., Wong, W. S. W., and Nielsen, R. (2005) 'Bayes Empirical Bayes Inference of Amino Acid Sites under Positive Selection', *Molecular Biology and Evolution*, 22: 1107–18.
- Zhang, J., Nielsen, R., and Yang, Z. (2005) 'Evaluation of an Improved Branch-site Likelihood Method for Detecting Positive Selection at the Molecular Level', *Molecular Biology and Evolution*, 22: 2472–9.
- Zheng, X. et al. (2012) 'A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data', *Bioinformatics*, 28: 3326–8.

Virus Evolution, 2024, **10**(1), 1–14

DOI: <https://doi.org/10.1093/ve/veae002>

Advance Access Publication 5 January 2024

Research Article

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.