

Genetic diversity of *Treponema paraluisleporidarum* isolates in European lagomorphs

Linda Hisgen¹, Erik Agren², Alex Barlow³, Marcus Fährndrich⁴, Ulrich Voigt⁴, Luisa Fischer⁵, Linda Grillová⁶, Luisa Hallmaier-Wacker⁷, Maria Kik⁸, Jana Klink⁴, Jitka Krenova⁹, Antonio Lavazza¹⁰, Simone Lueert¹, Markéta Nováková¹¹, Darina Čejková⁶, Carlo Pacioni¹², Christian Roos¹³, Tiziana Trogu¹⁰, David Smajs⁶, and Sascha Knauf¹

¹Friedrich-Loeffler-Institut Bundesforschungsinstitut für Tiergesundheit

²National Veterinary Institute

³University of Bristol School of Veterinary Sciences

⁴Tierärztliche Hochschule Hannover

⁵State Agency for Nature, Environment and Consumer Protection North Rhine-Westphalia

⁶Brno University of Technology

⁷Deutsches Primatenzentrum GmbH - Leibniz-Institut für Primatenforschung

⁸Utrecht University

⁹Masaryk University

¹⁰Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna Bruno Ubertini

¹¹University of Veterinary and Pharmaceutical Sciences Brno

¹²Arthur Rylah Institute for Environmental Research

¹³German Primate Center

April 15, 2023

Abstract

The bacterium *Treponema paraluisleporidarum* causes syphilis in Lagomorphs. In a set of 1,095 samples from four species – European brown hare (*Lepus europaeus*), mountain hare (*Lepus timidus*), Corsican hare (*Lepus corsicanus*) and European rabbit (*Oryctolagus cuniculus*) – we genotyped the strains that infect wild lagomorphs. Samples originate from Sweden, the Netherlands, the United Kingdom, Germany, the Czech Republic and Italy. The phylogenetic analyses of two informative gene targets (tp0488 and tp0548) showed high genetic diversity among the lagomorph-infecting treponemes. More specifically, we found a high number of nucleotide variants and various short repeat units in the tp0548 locus which have not been described for human syphilis and primate yaws causing *Treponema pallidum*. While the functional aspect of these short repeat units remains subject to ongoing investigations, it likely enables the pathogen to better survive in its lagomorph host. Our data did not support any geographic clustering which is equally reflected in the host population genetics as shown by mitochondrial genome data corresponding to the sampled lagomorph populations. This is unexpected and in contrast with what has been shown for nonhuman primate infection with *T. pallidum*. In the future, the combination of multi-locus sequence typing and WGS from modern and ancient samples from a wide geographic range and multiple lagomorph species will contribute to a better understanding of the epidemiology and evolutionary path of lagomorph-infecting treponemes. In conclusion, our current study demonstrates a high genetic variation of the syphilis-causing pathogen in a higher number of positively PCR-tested European lagomorphs (n=496/1095).

Introduction

Syphilis and yaws are caused by *Treponema pallidum* (*TP*) ssp. *pallidum* (syphilis; *TPA*) and ssp. *pertenue* (yaws; *TPE*), respectively. While syphilis is a human disease with several hundred-year long history, yaws is a disease of humans and nonhuman primates (Majander et al., 2020; Beale et al., 2021). Studies on modern *TPA* lineages controversially discuss a common ancestor of all *TPA* strains in the 1700s (Arora et al., 2016). However, the evolution of *TPA*, its closest relative *TPE* and the genetically closely related lagomorph syphilis-causing bacterium *Treponema paraluisleporidarum* ecovar Lepus (*TP* eL) in hares and *T. paraluisleporidarum* ecovar Cuniculus (*TP* eC) in rabbits remains elusive. Only a single complete genome of a rabbit-infecting laboratory-maintained strain from the USA has been whole-genome sequenced until today (Šmajš et al., 2011). Comparative studies showed that this strain, *TP* eC strain Cuniculi A, has a 98.1% whole-genome identity to the human syphilis-causing *TPA* strain Nichols (Šmajš et al., 2011; Pětrošová et al., 2013). Although its genome is slightly smaller compared to that of the human infecting syphilis bacterium there is a general genome synteny (Šmajš et al., 2011). Yet, all information on genomic deletions, insertions or changes that likely code for the lagomorph host specificity is derived from the single published *TP* eC strain Cuniculi A genome (Strouhal et al., 2007; Šmajš et al., 2011). Consequently, the basis for human-pathogenicity cannot be identified unless more *Treponema* genomes of lagomorph origin are analysed.

Our previous studies have demonstrated anti-*Treponema* antibodies and the presence of the bacterium in several European brown hare (*Lepus europaeus*; EBH) populations (Nováková et al., 2019; Hisgen et al., 2020, 2021). Serology and quantitative PCR were, however, unable to distinguish infection with *TP* eL, *TP* eC or the genetically closely related *TP*. We note here that the latter is not known to infect wild lagomorphs, but rabbits are traditionally used to cultivate human treponemes in vivo which highlights the ability for *TP* to adapt and survive in the lagomorph host (Lukehart and Marra, 2007). In this study and prior to whole-genome sequencing, we were interested in the molecular epidemiology and the genetic diversity of treponemes infecting European lagomorphs. In the light of the previously reported wide-spread infection in the wild lagomorph-host, which argues for a well-established disease in European lagomorphs, we hypothesised that hare infecting strains have a strain diversity as high as the one seen in a comparative wild host-pathogen system – African nonhuman primates infected with the sister-bacterium *TPE* (Chuma et al., 2019). We predicted a geographic clustering of strains isolated from structurally connected hare populations. We used the naturally occurring *Treponema* infection in nonhuman primates as a direct comparison since the infection in wild lagomorphs is equally not under selection pressure from antibiotic treatment (Knauf et al., 2018) and has multiple host species involved. Due to the high genetic similarity of human and nonhuman primate infecting *TP* and hare and rabbit infecting *TP* eL/C (98.1% genome identity; (Šmajš et al., 2011)), we applied a multi-locus sequence typing system that was originally designed for *TPE* infection in nonhuman primates (Chuma et al., 2019).

Materials and methods

Study design, sampling locations and animals

We tested a total of 1,095 samples from legally hunted or otherwise deceased (e.g., found dead) EBHs (*L. europaeus*, n=1,042), mountain hares (*Lepus timidus*, n=5), Corsican hares (*Lepus corsicanus*, n=2), and European rabbits (*Oryctolagus cuniculus*, n=39) that were opportunistically and randomly collected between 2016 and 2021 in six European countries (Germany (n=938), Sweden (n=4), England (n=25), Italy (n=81), the Netherlands (n=32), the Czech Republic (n=15); Figure 1). The 39 rabbit samples included six euthanised pet rabbits (*O. cuniculus*) and one sample from a live domestic rabbit which was presented with syphilitic lesions in a veterinary clinic. In the latter, swabs were taken purposely for diagnostic reasons.

Add Figure 1 about here

Sampling

We swabbed the vagina in female lagomorphs using sterile polyester tipped swabs (AF.022, abf diagnostics, Kranzberg, Germany) and removed the corpus penis using a sterile scalpel blade in males. In animals with crusty lesions, we sampled the affected skin area using sterile scalpel blades. All samples were stored in 2 ml safe seal tubes (Sarstedt, Nürnberg, Germany) containing 500 μ l of sterile filtered custom-made lysis buffer (10mM Tris-HCl, pH 8.0; 0.1M EDTA, pH 8.0; 0.5% SDS). Samples were frozen at -20 °C until further processing.

DNA extraction

DNA was extracted from swab and tissue material using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's guidance with some minor modifications. Briefly, we extracted DNA from swabs and tissue samples according to the protocols published by Chuma et al. (2018) and Hisgen et al. (2021), respectively. Subsequently, glycogen precipitation was performed to clean and concentrate the DNA. The method followed the procedure described by Knauf et al. (2016). We measured the DNA yield using a NanoDrop photometer (ThermoFisher Scientific, Darmstadt, Germany).

Polymerase chain reactions

Host genus confirmation

Samples that were received from third parties (e.g., hunters) were subject to host genus confirmation. For this reason, we amplified a 1,486 bp long region of the immunoglobulin heavy chain gene using two independent PCRs (IGHGCH2 and IGHG hinge regions) to differentiate lagomorphs on the genus level (Lavazza et al., 2015). Briefly, the 50 μ l reaction contained 25 μ l 2x Phanta Max Master Mix (Vazyme Biotech Co. Ltd., Nanjing, China), 19 μ l RNase free water, 2 μ l of the respective 10 μ M primer and template DNA. Cycling conditions were identical for both PCRs: 3 min initial denaturation at 95°C followed by 40 amplification cycles of 15 sec at 95°C, 15 sec at 60°C and an elongation phase at 72°C for 30 sec, followed by a post-extension step at 72°C for 5 min.

Treponema multi-locus sequence typing

In our previous work, we identified two variable gene loci for strain typing of *TPE* (*tp0488* and *tp0548*) in nonhuman primates (Chuma et al., 2019), which we adapted for the use in our lagomorph samples. We note here the close genetic relationship between *TPE* and *TP* eC, which is over 98% based on the whole genome (Šmajš et al., 2011; Čejková et al., 2012), 97.8% for the *tp0488* gene and 90.3% for the *tp0548* gene with all amplification primer binding sites being 100% conserved (Šmajš et al., 2011).

tp0488 : The PCR amplifies an ~830 bp region of the methyl-accepting chemotaxis protein 2 gene (*mcp* 2). We checked the previously published primers (Chuma et al., 2019) for compatibility to the published *TP* eC strain Cuniculi A (GenBank CP002103.1) and amplified the gene target as described previously (Chuma et al., 2019). Briefly, the 51 μ l reaction volume comprised 45 μ l Platinum PCR Super Mix High Fidelity (Thermo Fisher Scientific, Darmstadt, Germany), 2 μ l of each 10 μ M primer and template DNA, respectively. The amplification was performed using a SensoQuest Thermocycler (SensoQuest, Goettingen, Germany) applying the following conditions: two min pre-denaturation at 94°C followed by 80 cycles of 15 sec at 94°C, 15 sec at 59°C and 60 sec at 68°C.

tp0548 : We used a nested PCR to amplify a ~1,070 bp region of the *tp0548* gene (encoding for a predicted outer membrane protein), using primers and cycling conditions as published elsewhere (Chuma et al., 2019) with the only exception of using a different polymerase. Briefly, the 50 μ l reaction solution contained 25 μ l 2x Phanta Max Master Mix (Vazyme Biotech Co. Ltd., Nanjing, China), 19 μ l RNase free water, 2 μ l of the respective 10 μ M primer and template DNA. For the nested PCR, we used 2 μ l of the first PCR reaction. Cycling conditions for the first and nested PCR were 3 min initial denaturation at 95°C followed by 35 amplification cycles of 15 sec at 95°C, 15 sec at 48°C and an elongation phase at 72°C for 90 sec (first PCR) and 60 sec (nested PCR), respectively. Each PCR run ended with a post-extension step at 72°C for 5 min.

Gel electrophoresis, DNA purification and Sanger sequencing

All amplified samples were run on a 1.5% agarose gel and DNA products of correct size were extracted using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Subsequently, extracted DNA was sent for Sanger sequencing using the Microsynth Laboratory service (Microsynth, Göttingen, Germany). For the *tp0488* gene product, and for the IGHGCH2 and the IGHG hinge region amplicons, we utilized the respective forward primer for sequencing. The *tp0548* amplicons, however, were sequenced bidirectionally using the internal sequencing primers published elsewhere (Matějková et al., 2009).

Sanger sequencing data analysis

Sanger sequence data were evaluated, edited and aligned using Geneious Prime 2021.2.2 (Biomatters Limited, Auckland, New Zealand) and 4Peaks sequence viewer (Nucleobytes B.V., Aalsmeer, the Netherlands). We compared sequence data to respective orthologs available in GenBank using BLAST search (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). *Treponema* sequence data were analysed for positive gene selection following the tools and algorithm described by Maděránková et al (2019). Briefly, positively selected sites were determined from sequence alignments using: (i) a codon-based Site model implemented in EasyCodeML package (Gao et al., 2019) and/or (ii) a mixed effects model of evolution (MEME) using hypothesis testing approach via the Datamonkey webserver (Murrell et al., 2012; Weaver et al., 2018). For CodeML analysis, the phylogenetic trees were constructed using RAXML-NG tool (Kozlov et al., 2019). Phylogenetic trees and networks were constructed with IQ-TREE 2.0.7 (Minh et al., 2020), Mr. Bayes 3.2.7 (Ronquist et al., 2012) and the minimum spanning trees were inferred using MSTree V2 algorithm within GrapeTree program (Zhou et al., 2018).

Maximum-likelihood trees in IQ-TREE were constructed with 1,000 ultrafast bootstrap replicates (Hoang et al., 2018) and the best-fit model as obtained by IQ-Tree's ModelFinder (Kalyaanamoorthy et al., 2017) according to the Bayesian Information Criterion (BIC). Tree reconstructions based on Bayesian inference in MrBayes were conducted with 1,000,000 generations with sampling every 100 generations and a burn-in of 25%. To check for convergence of all parameters and adequacy of the burn-in, we investigated the uncorrected potential scale reduction factor (PSRF) (Gelman and Rubin, 1992) as calculated by MrBayes. We used *T. pallidum* subsp. *endemicum* strain Iraq B (GenBank CP032303.1) as an outgroup to root the tree.

Mitochondrial DNA amplification and high-throughput sequencing

We amplified mitochondrial (mt)-genomes of 95 randomly selected European brown hares equally distributed across all sampling sites (n= 1 to 5 per site, n=36 sites). Randomisation was performed using Research Randomizer 4.0 (<http://www.randomizer.org/>). In an initial step, we performed two independent long range PCRs using *Lepus europaeus* -specific primers covering the mt-genome range (KY211031) 19-9,464 (mtF1_Lepus_S_Roos19 5'-AAA GCA AAG CAC TGA AAA TGC T and mtF1_Lepus_AS_Roos19 5'-CCA AAA CTA ACT GAT TGG AAG T) and 8,500-484 (mtF2_Lepus_S_Roos19 5'-ATT AGT CCA ACA ACA GCC CTA and mtF2_Lepus_AS_Roos19 5'-CTT AGC TAT CGT GAG TTC GAA). Primers were designed based on available mt-genome data in GenBank using Geneious Prime 2021.2.2 software. PCR reactions were adjusted to 50 µl and included the following components: 10 µl 5x PrimeSTAR GXL buffer (Takara Bio Europe SAS/Clontech Labs, Saint-Germain-en-Laye, France), 1 µl PrimeSTAR GXL DNA Polymerase (Takara Bio Europe SAS, Saint-Germain-en-Laye, France), 4 µl dNTP mixture (2.5 mM each), 28.5 µl RNase free water, 2 µl of each 10 µM primer and ~250 ng DNA (1-3.5 µl depending on DNA concentration). Cycling conditions were as follows: 35 cycles at 98°C for 10 sec, 59°C for 15 sec and 68°C for 12 min. Subsequently DNA amplicons were purified using SPRIselect magnetic beads (Beckman Coulter, Inc., Krefeld, Germany) followed by the determination of DNA concentration utilizing the Qubit 4.0 fluorometer (ThermoFisher Scientific, Darmstadt, Germany). Next, we pooled 250 ng of each PCR product and adjusted the final volume to 26 µl that were subsequently used for next-generation library preparation. Libraries were prepared using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, Frankfurt am Main, Germany) according to the manufacturer's instructions in combination with the NEBNext Multiplex Oligos for Illumina

(NEBE6440, 96 Unique Dual Index Primers plate) to ensure that all samples can be sequenced at once. DNA was enzymatically fragmented to an average size of 300-700bps. Following the manufacturer's guidance and prior to sequencing, library quantification was performed with the NEBNext Library Quant Kit for Illumina (New England Biolabs) run on a StepOnePlus RealTime PCR System (Thermo Fisher Scientific, Darmstadt, Germany). In a final step, we normalized the samples to a concentration of 10 nM each using the previously generated quantification data. Samples were then pooled and sent to the NGS Integrative Genomics Core Unit (NIG, University Medical Center, Goettingen, Germany) for Illumina MiSeq 250 bp paired-end sequencing.

High-throughput sequencing data analysis

Mt-genome assembly was conducted with the Geneious 11.1.3 package (<https://www.geneious.com/>). First, demultiplexed raw sequence reads were trimmed and quality-filtered with BBDuk 37.64 of the BBTools package (<https://jgi.doe.gov/data-and-tools/bbtools/>) and duplicate reads were removed with Dedupe 37.64 (BBTools package); for both steps we applied standard settings. For mt-genome assembly, cleaned reads were mapped onto the reference mt-genome of *L. europaeus* (GenBank: NC_004028.1) using the Geneious assembler with standard settings. Newly produced mt-genomes were manually checked and then annotated with Geneious. For phylogenetic tree reconstruction, we added additional mt-genome sequences from EBHs from Sweden, Poland, Greece, Cyprus and Turkey available in GenBank and aligned them as described above. Tree reconstructions based on the maximum-likelihood algorithm and Bayesian inference were performed as described above using IQ-TREE and MrBayes software.

Results

Host genus confirmation

We tested a total of 230 MLST-positive samples for their host genus using amplification of two chromosomal loci. Of these, we retrieved only the IGHGCH2 target gene sequence from one sample and for 13 samples we were left with qualitative useful sequences from the IGHG hinge region only. The BLAST search of all 230 sequences including the 13 sequences where we had only one of the two PCRs positive resulted in a mean query coverage of $96.94 \pm 0.99\%$ (mean \pm SD). The organism outcome of the respective BLAST result was compared to the host genus that was assigned by the sample provider. We note here that only those samples that were not taken by our research team were subject to host genus confirmation. The originally declared host genus (*Lepus* versus *Oryctolagus*) was confirmed in all but one specimen. The sample (62BYMX121119) was originally labelled as of European rabbit origin but generated a sequence (joint IGHGCH2 and IGHG hinge gene target region) that is identical to the immunoglobulin gamma heavy chain constant region of EBH.

Treponema infection and clinical manifestations

Not all lagomorphs could be clinically inspected by our research team. Of those clinically examined (n=531/1,095) only 20 animals (3.8%) had crusts and ulcerations in the face (n=6) or genital region (n=14). The latter includes the domestic rabbit presented in a veterinary clinic. All samples from animals with documented facial or genital lesions – *L. europaeus* = 13, *L. timidus* = 4, *L. corsicanus* = 2 and *O. cuniculus* = 1 – were PCR positive for *T. paraluisleporidarum* (Table S1). Table 1 provides an overview about the PCR results in all lagomorphs.

Add Table 1 about here

Treponema paraluisleporidarum multi-locus sequence typing

We obtained positive PCR results from 405 samples from the *tp0488* gene (Germany: n=343/938; Sweden: n=4/4; England: n=12/25; Italy: n=14/81; the Netherlands: n=18/32; Czech Republic: n=14/15) and

439 sequences from the *tp0548* gene (Germany: n=380/938; Sweden: n=4/4; England: n=13/25; Italy: n=14/81; the Netherlands: n=13/32; Czech Republic: n=15/15; Table S1). In 57 samples we could only amplify the *tp0488* gene and in 91 samples we were only able to amplify the *tp0548* gene target. A total of 349 sequences (*tp0488* = 195 and *tp0548* = 154) had to be excluded from analysis due to low sequence quality or high background noise due to superimposed sequences. This resulted in 212, 296 and 194 sequences for *tp0488*, *tp0548* and the concatenated gene target sequences, respectively, including the reference sequence from the *TP* eC strain Cuniculi A (CP002103.1; locus tags *TPCCA_RS02365* and *TPCCA_RS02685*) and *T. pallidum* ssp. *endemicum* strain Iraq B (CP032303.1; locus tags TENDIB_0488 and TENDIB_0548). We generated a maximum-likelihood tree based on the concatenated sequences of the *tp0488* and *tp0548* genes and added the geographic origin as attributes, with samples grouped into Northern (Schleswig Holstein and northern part of Lower Saxony), Central and Western (southern part of Lower Saxony, Hesse and North Rhine-Westphalia) and Southern Germany (Bavaria and Baden-Wuerttemberg) as well as the Netherlands, Italy, Czech Republic and the United Kingdom. Overall, there is no clustering of the samples according to their geographic origin (Figure 2). While bootstrap support for nodes in the phylogenetic tree is generally low, there are some significantly supported distinctive features that are noteworthy. The tree exhibits an initial split into two clades of which one contains sequences obtained from EBHs sampled in Baden-Wuerttemberg and Bavaria (Southern Germany) and one EBH from North Rhine-Westphalia (Western Germany). In addition, the clade contains the *TPe C* reference strain Cuniculi A and three strains of mountain hares from Sweden. A sample from the fourth Swedish mountain hare (V1313_03_L1) is found in the second clade and clusters together with all other EBH samples as well as a strain that was found in a pet rabbit in Hesse (Central Germany). The latter is identical to a strain obtained from a EBH from Lower Saxony, approximately 140 km from the pet rabbit sampling location. Within both main clusters, a number of statistically supported subclades of geographically related samples e.g., from the Czech Republic or Northern Germany, were found.

In our analysis we identified positively selected sites (codons) in each of the target genes including *tp0488* (n = 29) and *tp0548* (n = 54) (Table S1). Those sites were removed from the alignments and only non-positively selected parsimony-informative sites as well as singletons were used for network constructions. The minimum spanning network that resulted from non-positively selected single nucleotide variants within the *tp0548* locus (Figure S1) does not change the overall topology of the maximum-likelihood tree shown in Figure 2 and equally lacks the overall geographic clustering of the samples. Maximum-likelihood trees for individual loci can be found in the Supplemental Material (Figure S2 and S3).

Add Figure 2 about here

Genetic diversity within *tp0548*

While the *tp0488* gene in lagomorph infecting *TPe C/L* strains shows no defined sequence variability site at a chosen minimum variant frequency of 0.25, the *tp0548* gene in our analysed samples had two hypervariable regions (V1-2). These regions range from 589,242-287 (V1) and 589,558-647 (V2) on the *TP* eC strain Cuniculi A reference genome (CP002103.1; Figure 3A) and were characterised by an aggregation of polymorphic sites, deletions and repeat-patterns. Briefly, V1 is characterized through indels and a dominating arginine, serine and glycine-coding composition. The V2 region is longer and includes various types of repetitions that are illustrated in Figure 3. Most strains (n = 203/287) present with type I repetitions that code for a KGGG amino acid motif. The median number of repetitions of this dominating type I repeat is three with a range of one to seven (Figure 3C). Besides the 228 strains that showed only one repeat type, 56 strains presented with a mosaic of two or three different repeat types, and three samples had no repeat sequence at all (Figure S4).

Add Figure 3 about here

Hare population genetics

We compared the identified diversity of *TP* eC/L with the diversity of the EBH host population based on mt-genome data and reconstructed phylogenetic trees using the maximum-likelihood algorithm and Bayesian inference (Figures S5). We obtained three major clades, referring to populations in Turkey/Cyprus, Greece and central/western Europe with the latter clade showing no or only little geographical structuring.

Discussion

We show further evidence of *TP* eC/L infection in wild lagomorphs in Europe based on DNA isolated from swab and tissue material. This complements our previous serological and qPCR findings that already suggested a widespread infection in European lagomorphs (Nováková et al., 2019; Hisgen et al., 2020, 2021). Moreover, we confirm infection in three previously reported host species (EBH, mountain hare and rabbits) and a newly identified host, the Corsican hare. Unfortunately, in this study, we cannot make a statement about prevalence rates which is particularly the case for the two Corsican hares samples that are included into our sample set. Reports about hybridization in hare species (Pierpaoli et al., 1999; Marques, Farelo et al., 2017; Seixas et al., 2018; Pohjoismäki et al., 2021) highlight feasible transmission pathways for the interspecies spread of *Treponema*. In this context, the demonstrated infection of a pet rabbit and a EBH from Central Germany with an identical strain (01GIF22900120 and 02DNTWF1100120, Figure 1) is more difficult to explain since a direct interaction of the pet rabbit and wild hares can be excluded. Whole genome sequencing of the two strains is needed to prove whether the two *TP* eC/L isolates are truly identical. To our knowledge, sexual interactions of rabbits and hares – even under natural conditions – has not been reported. In primate infection (including humans) with the related bacterium *T. pallidum* subsp. *Pertenuis*, vector transmission through flies has been discussed (Knauf et al., 2016; Houinei et al., 2017; Gogarten et al., 2019). In lagomorphs vector transmission of *Tp* eC/L has not been investigated.

Our current data add novel insight into the genetic diversity of lagomorph infecting *Treponema* from Italy, Czech Republic, Germany, Sweden, the Netherlands and the United Kingdom (Figure 1). We used our established MLST system for nonhuman primate yaws infection (Chuma et al., 2019) to equally characterise lagomorph infecting strains. This was done under the assumption that the lagomorph infecting *TP* eC/L and *T. pallidum* are closely related as shown on the basis of the single published whole genome of *TP* eC (Strouhal et al., 2007; Šmajš et al., 2011). The results presented in Figure 2, demonstrate an unexpectedly greater diversity compared to what we have seen in non-treated naturally yaws infected nonhuman primates in sub-Saharan Africa. In the case of lagomorph *tp0548* sequences, 242 variants – including differences in the length of repetitions – out of 295 obtained sequences were found, demonstrating the enormously high degree of genetic variability. Yet, most variable nucleotides in both loci, *tp0488* and *tp0548*, were found under positive selection, partly explaining the huge observed genetic diversity and the lack of geographical clustering. Interestingly, a human syphilis multilocus sequence typing system also uses partial analysis of the *tp0548* gene for molecular typing of clinical isolates. Until now, 77 different alleles of *tp0548* in *TP* have been identified from a total of 944 investigated clinical isolates (Grillová et al. 2019), suggesting that similar evolutionary forces operate on the *tp0548* locus in human and lagomorph infecting *Treponema*. The higher observed genetic diversity of *tp0548* in lagomorphs could be explained by prolonged infection in hares than in humans who are treated with antibiotics. It is open to debate whether the higher genetic diversity of lagomorph infecting *TP* eC/L mimics treponemal evolution in an untreated human population.

The strain diversity, geographic range of infection and the involvement of multiple lagomorph species are all indicators for the endemic character of the disease in European lagomorphs. In contrast to nonhuman primates (Chuma et al., 2019) and human infections (Beale et al., 2021) with the sister bacterium *T. pallidum*, our current data from lagomorphs showed only weak geographic clustering. This is unexpected, since the biology of hares – a mostly philopatric species that shows only limited dispersal activity (Avril et al., 2011) – would likely contribute to the long-term circulation of regional (dominant) strains in the different hare

metapopulations. It is open to debate whether this is the result of the positive selection of variants in the gene targets that we used for molecular typing or an effect of the anthropogenic influence on the population through trans- or relocation (Masetti and Marinis, 2008; Sokos et al., 2015; Sánchez-García et al., 2021). Until today, the management of overexploited EBH populations is based on annual restocking (Canu et al., 2013). In combination with the mt-genome data (Figure S5), which indicate a panmictic EBH population, it is most likely that EBH dispersal and associated with this, the dispersal of *TP* eL strains, is dominated by human influence.

In *TP* eC/L, the *tp0548* locus shows not only a higher number of nucleotide variants compared to nonhuman primate and human infection with *TPE* and *TPA*, respectively, but also various types of short repeat units that have not been described in primate treponematoses (Figure 3). While the functional aspect of these *tp0548* short repeat units remains subject to ongoing investigations, it is likely that it enables the pathogen to better survive in its lagomorph host. From an evolutionary perspective, short, repeated nucleotide sequences in bacteria are frequently associated with higher replication error rates caused by slipped-strand mispairing (Deitsch et al., 2009; Castillo-Lizardo et al., 2014). It seems obvious, that a provoked slipped-strand mispairing in structurally non-essential parts of antigenic outer-membrane proteins provides an advantage over spontaneous mutations in terms of immune escape.

Apparently, the current selected loci – *tp0488* and in particular *tp0548* – are not well suited for the molecular typing of treponemes of lagomorph origin. This is reflected by the high number of haplotypes. Once a reasonable number of whole genome sequences of *TP* eC/L becomes available, an in-depth revision of the current typing system is necessary to include more decently variable loci that are suitable for the epidemiological monitoring of transmission chains. Moreover, whole genome sequencing of modern and ancient samples from a wide geographic range and from multiple lagomorph species, including those that are not yet investigated (broom hare (*L. castroviejoi*) in northern Spain, the cape hare (*L. capensis*) in Sardinia and Cyprus, the Iberian hare (*L. granatensis*) on the Iberian peninsula) will help backtrack the evolutionary path of the pathogen and its relationship to modern syphilis in humans

We have demonstrated the presence of *TP* eC/L using nucleic amplification assays and subsequent Sanger sequencing. These methods prevent us from making a final statement on the viability of the treponemes. Yet, in the authors' view, the consistence of infection in lagomorphs sampled across Europe and the high copy numbers detected in samples of the internal genital of EBHs (Hisgen et al., 2021) makes an active infection likely. In human syphilis, with exception of latent syphilis, active infection is associated with clinical lesions (Lukehart, 2008), which was rarely seen in the clinically inspected lagomorphs, of which 20/532 (3,8%) had typical skin lesions. Unfortunately, many samples included into this study originated from collaborating hunting parties which limited the clinical expectation of the integument, the oral-cavity and the genital tract during the sampling procedure. In these samples, it cannot be excluded that lesions have been overlooked for example when the ulcer was located in the urethra as it is described for humans (Chambers et al., 2018).

Conclusion

In our current study, we show a high proportion of wild European lagomorphs infected with *TP* eC/L, based on the detection of the pathogen's DNA in genital swab and tissue materials (Table 1). Sequencing of the targeted gene loci revealed an unexpectedly high genetic diversity. The various types of repetitions in one of the two hypervariable regions at the *tp0548* locus have not been described in the sister bacterium *T. pallidum* subsp. *pallidum*, the causative agent of human syphilis. This warrants further research on the functional aspects of repetitive units in the genome of *TP* eC/L. A revision of the MLST system is recommended once a substantial number of lagomorph infecting treponemes has been whole genome sequences.

Acknowledgments

Our study was funded by the Deutsche Forschungsgemeinschaft (DFG KN1097/7-1 to SK) and partly the Czech Science Foundation (GC18-23521J to DŠ). We thank all contributors for their support. This included the employees of the Chemical and Veterinary Investigation Offices in North Rhine-Westphalia, namely Maja Eydner, Henning Petersen (CVUA-OWL), Annette Kuczka (CVUA-RRW), Maren Kummerfeld (CVUA-

MEL), Martin Peters and Sabine Merbach (CVUA-WFL) and the state agencies for health and food safety in Erlangen, Germany and Oberschleissheim, Germany for the sampling of lagomorphs during necropsy. Wildlife pathologists at SVA are thanked for sampling hares in the general wildlife disease surveillance programme. Moreover, we are grateful that Axel Wehrend (Clinic for Veterinary Obstetrics, Gynaecology and Andrology, Giessen, Germany) and Katja Wehrend (Fachtierarztpraxis im Kleebach) contributed an important swab sample of a domestic rabbit. All hunters are thanked for their generous help and assistance in the field. They all have contributed to the success of this study and, therefore, helped to investigate the health of wild lagomorphs.

References

- Arora, N., Schuenemann, V. J., Jäger, G., Peltzer, A., Seitz, A., Herbig, A., . . . Bagheri, H.C. (2016). Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. *Nat Microbiol*, 2, 16245. <https://doi.org/10.5167/uzh-134132>
- Avril, A., Léonard, Y., Letty, J., Péroux, R., Guitton, J.-S., & Pontier, D. (2011). Natal dispersal of European hare in a high-density population. *Mamm Biol*, 76, 148–156. <https://doi.org/10.1016/j.mambio.2010.07.001>
- Beale, M. A., Marks, M., Cole, M. J., Lee, M.-K., Pitt, R., Ruis, . . . Thomson, N.R. (2021). Global phylogeny of *Treponema pallidum* lineages reveals recent expansion and spread of contemporary syphilis. *Nat Microbiol*, 6, 1549–1560. <https://doi.org/10.1038/s41564-021-01000-z>
- Canu, A., Scandura, M., Luchetti, S., Cossu, A., Iacolina, L., Bazzanti, M., & Apollonio, M. (2013). Influence of management regime and population history on genetic diversity and population structure of brown hares (*Lepus europaeus*) in an Italian province. *Eur J Wildlife Res*, 59, 783–793. <https://doi.org/10.1007/s10344-013-0731-x>
- Castillo-Lizardo, M., Henneke, G., & Viguera, E. (2014). Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota *Pyrococcus abyssi*. *Front Microbiol*, 5, 403. <https://doi.org/10.3389/fmicb.2014.00403>
- Čejková, D., M. Zobaníková, L. Chen, P. Pospíšilová, M. Strouhal, X. Qin, L. Mikalová, S.J. Norris, D.M. Muzny, R.A. Gibbs, L.L. Fulton, E. Sodergren, G.M. Weinstock, and D. Šmajš, 2012: Whole genome sequences of three *Treponema pallidum* ssp. *pertenue* strains: yaws and syphilis treponemes differ in less than 0.2% of the genome sequence. *Plos Neglect Trop D* 6, e1471, <https://doi.org/10.1371/journal.pntd.0001471>
- Chambers, L. C., Srinivasan, S., Lukehart, S. A., Ocbamichael, N., Morgan, J. L., Lowens, M. S., . . . Manhart, L. E. (2018). Primary Syphilis in the Male Urethra: A Case Report. *Clin Infect Dis*, 68, 1231–1234. <https://doi.org/10.1093/cid/ciy771>
- Chuma, I. S., Batamuzi, E. K., Collins, D. A., Fyumagwa, R. D., Hallmaier-Wacker, L. K., Kazwala, R. R., . . . Knauf, S. (2018). Widespread *Treponema pallidum* Infection in Nonhuman Primates, Tanzania. *Emerg Infect Dis*, 24,1002–1009. <https://doi.org/10.3201/eid2406.180037>
- Chuma, I. S., Roos, C., Atickem, A., Bohm, T., Collins, D. A., Grillová, L., . . . Knauf, S. (2019). Strain diversity of *Treponema pallidum* subsp. *pertenue* suggests rare interspecies transmission in African nonhuman primates. *Sci Rep*, 9, 14243. <https://doi.org/10.1038/s41598-019-50779-9>
- Deitsch, K. W., Lukehart, S. A., & Stringer, J. R. (2009). Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol*, 7, 493–503. <https://doi.org/10.1038/nrmicro2145>
- Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., & Ho, S. Y. W. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol Evol*, 9, 3891–3898. <https://doi.org/10.1002/ece3.5015>
- Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statist Sci*, 7, 457–472. <https://doi.org/10.1214/ss/1177011136>

- Gogarten, J. F., Düx, A., Mubemba, B., Pléh, K., Hoffmann, C., Mielke, A., ... Leendertz, F. H. (2019). Tropical rainforest flies carrying pathogens form stable associations with social nonhuman primates. *Mol Ecol*, 28, 4242–4258. <https://doi.org/10.1111/mec.15145>
- Grillova, L., K. Jolley, D. Šmajš, and M. Picardeau, 2019: A public database for the new MLST scheme for *Treponema pallidum* subsp. *pallidum* : surveillance and epidemiology of the causative agent of syphilis. *PeerJ* 6, e6182, <https://doi.org/10.7717/peerj.6182>.
- Hisgen, L., Abel, L., Hallmaier-Wacker, L., Lüert, S., Lavazza, A., Trogu, T., ... Knauf, S. (2021). The distribution of lagomorph syphilis caused by *Treponema paraluisleporidarum* in Europe. *Eur J Wildlife Res*, 67, 92. <https://doi.org/10.1007/s10344-021-01535-w>
- Hisgen, L., Abel, L., Hallmaier-Wacker, L. K., Lueert, S., Siebert, U., Faehndrich, M., ... Knauf, S. (2020). High syphilis seropositivity in European brown hares (*Lepus europaeus*), Lower Saxony, Germany. *Transbound Emerg Dis*, 67, 2240–2244. <https://doi.org/10.1111/tbed.13551>
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol*, 35, 518–522. <https://doi.org/10.1093/molbev/msx281>
- Houinei, W., Godornes, C., Kapa, A., Knauf, S., Mooring, E. Q., Gonzalez-Beiras, C., ... Mitja, O. (2017). *Haemophilus ducreyi* DNA is detectable on the skin of asymptomatic children, flies and fomites in villages of Papua New Guinea. *Plos Negl Trop Dis*, 11, e0004958. <https://doi.org/10.1371/journal.pntd.0004958>
- Kalyanamoothy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermin, L. S. (2017). ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat Methods*, 14, 587–589. <https://doi.org/10.1038/nmeth.4285>
- Knauf, S., Gogarten, J. F., Schuenemann, V. J., Nys, H. M. D., Dux, A., Strouhal, M., ... Calvignac-Spencer, S. (2018). Nonhuman primates across sub-Saharan Africa are infected with the yaws bacterium *Treponema pallidum* subsp. *pertenuis*. *Emerg Microbes Infec*, 7, 157. <https://doi.org/10.1038/s41426-018-0156-4>
- Knauf, S., Raphael, J., Mitja, O., Lejora, I. A. V., Chuma, I. S., Batamuzi, E. K., ... Lukehart, S. A. (2016). Isolation of *Treponema* DNA from Necrophagous Flies in a Natural Ecosystem. *EBioMedicine*, 11, 85–90. <https://doi.org/10.1016/j.ebiom.2016.07.033>
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35, 4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>
- Lavazza, A., Cavadini, P., Barbieri, I., Tizzani, P., Pinheiro, A., Abrantes, J., ... Capucci, L. (2015). Field and experimental data indicate that the eastern cottontail (*Sylvilagus floridanus*) is susceptible to infection with European brown hare syndrome (EBHS) virus and not with rabbit haemorrhagic disease (RHD) virus. *Vet Res*, 46, 13. <https://doi.org/10.1186/s13567-015-0149-4>
- Lukehart, S. A. (2008). Biology of Treponemes. In: Holmes, K. K., Sparling, P. F., Stamm, W. E., Piot, P., Wasserheit, J. N., Corey, L., ... Watts, D. H. (Eds), *Sexually Transmitted Diseases* (pp. 647–659). New York: Mc Graw Hill Medical.
- Lukehart, S. A., & Marra, C. M. (2007). Isolation and Laboratory Maintenance of *Treponema pallidum*. *Curr Protoc Microbiol*, 7, 12A.1.1-12A.1.18. <https://doi.org/10.1002/9780471729259.mc12a01s7>
- Maděránková, D., Mikalová, L., Strouhal, M., Vadják, Š., Kuklová, I., Pospíšilová, P., ... Šmajš, D. (2019). Identification of positively selected genes in human pathogenic treponemes: syphilis-, yaws-, and bejel-causing strains differ in sets of genes showing adaptive evolution. *Plos Neglect Trop Dis*, 13, e0007463. <https://doi.org/10.1371/journal.pntd.0007463>

- Majander, K., Pfrengle, S., Kocher, A., Neukamm, J., du Plessis, L., Pla-Díaz, M., ... Schuenemann, V. J. (2020). Ancient Bacterial Genomes Reveal a High Diversity of *Treponema pallidum* Strains in Early Modern Europe. *Curr Biol*, 30, 3788-3803. <https://doi.org/10.1016/j.cub.2020.07.058>
- Marques, J. P., Farelo, L., Vilela, J., Vanderpool, D., Alves, P. C., Good, J. M., ... Melo-Ferreira, J. (2017). Range expansion underlies historical introgressive hybridization in the Iberian hare. *Sci Rep-uk* 7, 40788, <https://doi.org/10.1038/srep40788>
- Marques, J. P., Ferreira, M. S., Farelo, L., Callahan, C. M., Hackländer, K., Jenny, H., ... Melo-Ferreira, J. (2017). Mountain hare transcriptome and diagnostic markers as resources to monitor hybridization with European hares. *Sci Data* 4, 170178, <https://doi.org/10.1038/sdata.2017.178>
- Masseti, M., & Marinis, A. M. D. (2008). Prehistoric and Historic Artificial Dispersal of Lagomorphs on the Mediterranean Islands. In: Alves, P. C., Ferrand, N., & Hackländer, K. (Eds.), *Lagomorph Biology: Evolution, Ecology, and Conservation* (pp. 13–25). Heidelberg: Springer
- Matějková, P., Flasarová, M., Zákoucká, H., Bořek, M., Křemenová, S., Arenberger, P., ... Šmajš, D. (2009). Macrolide treatment failure in a case of secondary syphilis: a novel A2059G mutation in the 23S rRNA gene of *Treponema pallidum* subsp. *pallidum*. *J Med Microbiol*, 58, 832–836. <https://doi.org/10.1099/jmm.0.007542-0>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*, 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Pond, S. L. K. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *Plos Genet*, 8, e1002764. <https://doi.org/10.1371/journal.pgen.1002764>
- Nováková, M., Najt, D., Mikalová, L., Kostková, M., Vrbová, E., Strouhal, M., ... Šmajš, D. (2019). First report of hare treponematoses seroprevalence of European brown hares (*Lepus europaeus*) in the Czech Republic: seroprevalence negatively correlates with altitude of sampling areas. *BMC Vet Res*, 15, 350. <https://doi.org/10.1186/s12917-019-2086-3>
- Pětrošová, H., Pospíšilová, P., Strouhal, M., Čejková, D., Zobaníková, M., Mikalová, L., ... Šmajš, D. (2013). Resequencing of *Treponema pallidum* ssp. *pallidum* strains Nichols and SS14: correction of sequencing errors resulted in increased separation of syphilis treponeme subclusters. *Plos One*, 8, e74319. <https://doi.org/10.1371/journal.pone.0074319>
- Pierpaoli, M., Riga, F., Trocchi, V., & Randi, E. (1999). Species distinction and evolutionary relationships of the Italian hare (*Lepus corsicanus*) as described by mitochondrial DNA sequencing. *Mol Ecol*, 8, 1805–1817. <https://doi.org/10.1046/j.1365-294x.1999.00766.x>
- Pohjoismäki, J. L. O., Michell, C., Levänen, R., & Smith, S. (2021). The best of both worlds: Shortcutting evolution through adaptive hybridization in hares. PREPRINT (Version 1) available at Research Square. <https://doi.org/10.21203/rs.3.rs-329700/v1>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biol*, 61, 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Sánchez-García, C., Pérez, J. A., Armenteros, J. A., Gaudioso, V. R., & Tizado, E. J. (2021). Survival, spatial behaviour and resting place selection of translocated Iberian hares *Lepus granatensis* in Northwestern Spain. *Eur J Wildl Res*, 67, 22. <https://doi.org/10.1007/s10344-021-01464-8>
- Seixas, F. A., Boursot, P., & Melo-Ferreira, J. (2018). The genomic impact of historical hybridization with massive mitochondrial DNA introgression. *Genome Biol*, 19, 91. <https://doi.org/10.1186/s13059-018-1471-8>

Šmajš, D., Zobaníková, M., Strouhal, M., Čejková, D., Dugan-Rocha, S., Pospíšilová, P., ... Weinstock, G. M. (2011). Complete Genome Sequence of *Treponema paraluis-cuniculi*, Strain Cuniculi A: The Loss of Infectivity to Humans Is Associated with Genome Decay. *Plos One*, 6, e20415. <https://doi.org/10.1371/journal.pone.0020415>

Sokos, C., Birtsas, P., Papaspyropoulos, K. G., Giannakopoulos, A., Athanasiou, L. V., Manolakou, K., ... Billinis, C. (2015). Conservation Considerations for a Management Measure: An Integrated Approach to Hare Rearing and Release. *Environ Manage*, 55, 19–30. <https://doi.org/10.1007/s00267-014-0388-6>

Strouhal, M., Šmajš, D., Matejková, P., Sodergren, E., Amin, A., Howell, J., ... Weinstock, G. M. (2007). Genome differences between *Treponema pallidum* subsp. *pallidum* strain Nichols and *T. paraluis-cuniculi* strain Cuniculi A. *Infect Immun*, 75, 5859–5866. <https://doi.org/10.1128/IAI.00709-07>

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Pond, S. L. K. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol*, 35, 773–777. <https://doi.org/10.1093/molbev/msx335>

Zhou, Z., Alikhan, N. F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., ... Achtmann, M. (2018). GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome research*, 28, 1395–1404. <https://doi.org/10.1101/gr.232397.117>

Data Accessibility and Benefit-Sharing

Genetic data:

Treponema paraluisleporidarum haplotype sequences obtained from Sanger Sequencing are available under GenBank accession numbers OM939693-OM939724 (*tp0488*) and OM990854-OM991094 (*tp0548*). Likewise, the mt-genome sequences can be found using the GenBank accession numbers OM993354-OM991094. Immunoglobulin heavy chain gene sequence data are available under the GenBank accession numbers ON089362-ON089577. **Sample metadata:**

Data are summarized in Table S1 (TPEC/L related data) and S2 (mt-genome data).

Benefit-Sharing Statement:

The use of the samples in this study is in accordance with the legislative, administrative or policy measures of the respective party to the Nagoya Protocol. Benefits from this research accrue from the sharing of our data and results on public databases as described above.

Author Contributions

The study was conceptualised by DŠ and SK; field and laboratory work investigations were performed by LH, EOÁ, AMB, MF, LF, MJLK, JCK, AL, SL, MN, DC, CP, TT and SK. Data curation, formal analysis was done by LH, LG, LKH, CR, DŠ and SK. Funding acquisition, project administration and supervision was performed by DŠ and SK. LH, DŠ and SK prepared the original draft manuscript, all authors reviewed and edited the script.

Ethics statement

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required as the samples in this study originated from dead (hunted or otherwise deceased) animals or in one case were taken for diagnostic purposes.

Conflict of Interest

The authors declare that they have no conflict of interest.

Supplemental Information

Supplemental Information with Figures S1-5 as described in the text.

Table S1. GenBank accession numbers for the mt-genomes.

Table S2. Metadata and GenBank accession numbers for the samples included into this study.

Tables

Table 1. Overview on the infection status of the sampled lagomorphs.

Host species	N sampled	Confirmed <i>TPeC/L</i> PCR positive ⁺	<i>TPeC/L</i> PCR inconclus
<i>Lepus europaeus</i>	1,042	294	191
<i>Lepus timidus</i>	5	4	0
<i>Lepus corsicanus</i>	2	2	0
<i>Oryctolagus cuniculus</i>	39	0	2
<i>Oryctolagus cuniculus</i> f. <i>domestica</i>	7	1	0
Total	1,095	301	193

⁺at least one gene (*tp0488* or *tp0548*) positive

⁺⁺PCR product of correct size visible on agarose gel and low sequence quality or high background noise due to superimposed sequences.

Figures

Figure 1. Map of Europe showing the geographic origin and the infection status of our samples based on the host species: (A) European rabbit (*Oryctolagus cuniculus*), (B) European brown hare (*Lepus europaeus*), (C) mountain hare (*Lepus timidus*), (D) Corsican hare (*Lepus corsicanus*). Red squares indicate positive and black triangles represent negative tested animals. Coloured geographic range overlays provide information on the distribution of the different lagomorph species according to the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (<https://www.iucnredlist.org>). The distribution charts were modified to include the UK to the species distribution of *L. europaeus* (Marques, Ferreira et al., 2017). Green = mountain hare, red = European brown hare, yellow = Iberian hare (*Lepus granatensis*). The map was created in QGIS v. 3.20 using Bing Aerial (<http://ecn.t3.tiles.virtualearth.net/tiles/a{q}.jpeg?g=1>) as a map source with TM World Borders 0.3 overlay (https://thematicmapping.org/downloads/world_borders.php).

Figure 2. Maximum-likelihood tree based on the concatenated alignment (*tp0488* and *tp0548*). The tree was constructed with IQ-TREE with the best-fit model (GTR+F+R3) based on the Bayesian Information Criterion and 1,000 bootstrap replicates. We included 194 sequences, containing 156 parsimony-informative sites and 48 singletons. Only bootstrap values > 90% are shown. Scale bar represents substitutions per nucleotide site. The circle colour indicates the geographic origin of the sample: magenta = USA, yellow = Sweden, cyan = Northern Germany, black = Southern Germany, green = Central and Western Germany, orange = The Netherlands, wine-red = Italy, pink = Czech Republic and red = United Kingdom. Text colour indicates the host species: black = *Lepus europaeus*, yellow = *Lepus timidus*, blue = *Lepus corsicanus* and red = *Oryctolagus cuniculus*. The tree was rooted using *T. pallidum* subsp. *endemicum* strain Iraq B (GenBank CP032303.1).

Figure 3. Illustration of the *tp0548* gene variable regions and repeat types. **A.** Overview about the two identified variable regions (V1-2) at a minimum variant frequency of 0.25. The colours in V2 are coding for the flanking regions of the repeats upstream (blue) and downstream (red). The grey area indicates the location of the repeats. **B.** Sequence information of the identified repeat types I-XIV. Changes from the Type I repeat are highlighted in bold. **C.** Truncated violine plot of the frequency distribution of the five most abundant repeat types identified in our set of sequences. In red, the bold dashed lines indicate the median, the smaller dotted lines represent the quartiles. Individual data are superimposed with a grey cross. The total number of strains with the respective repeat type is shown on the right. Only repeat types occurring

> three and which contain only a single repeat type are shown. Figure S4 provides further information on other repeat patterns. Violin plots were created in GraphPad PRISM 9.3.1.



