Geographically structured genomic diversity of non-human primate-infecting *Treponema pallidum* subsp. *pertenue*

Benjamin Mubemba[1,2], Jan F. Gogarten[1,3], Verena J. Schuenemann[4,5], Ariane Düx[1], Alexander Lang[1], Kathrin Nowak[1], Kamilla Pléh[1], Ella Reiter[5], Markus Ulrich[1], Anthony Agbor[6], Gregory Brazzola[6], Tobias Deschner[6], Paula Dieguez[6], Anne-Céline Granjon[6], Sorrel Jones[6], Jessica Junker[6], Erin Wessling[7], Mimi Arandjelovic[6], Hjalmar Kuehl[6,8], Roman M. Wittig[6], Fabian H. Leendertz[1,*,†] and Sébastien Calvignac-Spencer[1,3,*,†]

Abstract

Many non-human primate species in sub-Saharan Africa are infected with *Treponema pallidum* subsp. *pertenue*, the bacterium causing yaws in humans. In humans, yaws is often characterized by lesions of the extremities and face, while *T. pallidum* subsp. *pallidum* causes venereal syphilis and is typically characterized by primary lesions on the genital, anal or oral mucosae. It remains unclear whether other *T. pallidum* subspecies found in humans also occur in non-human primates and how the genomic diversity of non-human primate *T. pallidum* subsp. *pertenue* lineages is distributed across hosts and space. We observed orofacial and genital lesions in sooty mangabeys (*Cercocebus atys*) in Taï National Park, Côte d'Ivoire and collected swabs and biopsies from symptomatic animals. We also collected non-human primate bones from 8 species in Taï National Park and 16 species from 11 other sites across sub-Saharan Africa. Samples were screened for *T. pallidum* DNA using polymerase chain reactions (PCRs) and we used in-solution hybridization capture to sequence *T. pallidum* genomes. We generated three nearly complete *T. pallidum* genomes from biopsies and swabs and detected treponemal DNA in bones of six non-human primate species in five countries, allowing us to reconstruct three partial genomes. Phylogenomic analyses revealed that both orofacial and genital lesions in sooty mangabeys from Taï National Park were caused by *T. pallidum* subsp. *pertenue*. We showed that *T. pallidum* subsp. *pertenue* has infected non-human primates in Taï National Park for at least 28 years and has been present in two non-human primate species that had not been described as *T. pallidum* subsp. *pertenue* hosts in this ecosystem, western chimpanzees (*Pan troglodytes verus*) and western red colobus (*Piliocolobus badius*), complementing clinical evidence that started accumulating in Taï National Park in 2014. More broadly, simian *T. pallidum* subsp. *pertenue* strains did not form monophyletic clades based on host species or the symptoms caused, but rather clustered based on geography. Geographical clustering of *T. pallidum* subsp. *pertenue* genomes might be compatible with cross-species transmission of *T. pallidum* subsp. *pertenue* within ecosystems or environmental exposure, leading to the acquisition of closely related strains. Finally, we found no evidence for mutations that confer antimicrobial resistance.

DATA SUMMARY

INTRODUCTION

Spirochete bacteria belonging to the species *Treponema pallidum* (TP) have affected humankind since at least the late 15th century [1] and cause a large global disease burden in humans [2, 3]. Three pathogenic subspecies are currently recognized that are morphologically similar, but genetically and epidemiologically distinguishable [4]. Clinically, pathogenic TP subspecies cause three distinct disease syndromes; yaws (subsp. *pertenue*; TPE), venereal syphilis (subsp. *pallidum*; TPA) and bejel (subsp. *endemicum*; TEN) [5, 6]. Though usually treatable with antibiotics, these treponematoses remain major public health threats across the globe [2–4]. For example, in 2012, nearly 8 million new cases of venereal syphilis were reported globally [7], and in 2018, an estimated 80000 new yaws cases occurred in the 15 countries where the disease remains endemic [8]. Exact estimates of the number of bejel cases from the Sahel region and Arabian Peninsula where the disease is endemic are not available [9]. Efforts to reduce the prevalence of these treponematoses are underway, particularly for yaws, which an ongoing campaign aims to eradicate globally by 2030 [10]. Eradication efforts face a number of challenges, such as issues concerning the availability of diagnostics, the distribution of treatment options and resistance to available antibiotics.

One of the key questions for the potential success of eradication efforts and for understanding TPE evolution is the degree to which other animals are infected with these pathogens and whether cross-species transmission occurs. To date, only TPE has been shown to infect non-human primates (NHPs), as suggested by the finding of TPE DNA in lesion samples from a number of different NHP species [11–13]. The TPE strains infecting humans and NHPs are extremely similar. For example, the first genome sequenced from a TPE infecting a wild NHP – the strain Fribourg-Blanc obtained from a Guinea baboon (*Papio papio*: RefSeq ID: NC_021179.1) – exhibits a 99.97% sequence identity with human-infecting strains [13]. There is no clear evidence for phylogenetic separation of NHP-infecting and human-infecting TPE strains, as strains do not form well-supported reciprocally monophyletic groups [12, 13]. Although this pattern might be suggestive of past transmission events between NHPs and humans, whether such events really happened remains uncertain due to a paucity of genomic data from both humans and wildlife. However, in an experimental setting, the Fribourg-Blanc strain induced classical yaws symptoms in humans [14] and human-infecting TPE strains were reported to elicit yaws-like symptoms in NHPs [15]. This suggests that molecular compatibility barriers to cross-species transmission of TPE

Impact Statement

We were able to reconstruct three new *Treponema pallidum* subsp. *pertenue* genomes from swabs/biopsies and three partial genomes from bone samples, adding to the limited number of *T. pallidum* subsp. *pertenue* genomes existing today. Phylogenomic analyses revealed that simian *T. pallidum* subsp. *pertenue* strains did not form monophyletic clades based on the host species or the types of clinical symptoms manifested, but rather clustered by geographical origin. This is compatible with cross-species transmission of *T. pallidum* subsp. *pertenue* within ecosystems where this disease occurs. Additionally, this study joins a growing body of evidence that human and wildlife bones are useful for tracing the history of treponemal diseases. Using bones, we confirmed that *T. pallidum* subsp. *pertenue* infected previously undescribed non-human primate hosts and has been present in Taï National Park for at least 28 years.

are low, though other barriers to spillover might exist [16]. In fact, recent multilocus sequence typing (MLST) of genes unique to TPE showed that TPE strains infecting wild NHPs cluster according to geography and not by host species in Tanzania, which is compatible with local cross-species transmission of TPE in natural systems [17].

Here, we aimed to further explore TPE genomic diversity across space, hosts and clinical manifestations. For this, we first focused our analyses on a single group of sooty mangabeys (*Cercocebus atys*) in Taï National Park (TNP), Côte d'Ivoire, whose affected individuals showed orofacial or genital lesions. We then drew on recent findings that TP can be detected in NHP bones [18] to screen NHP bones from TNP and examine whether closely related strains of TPE circulate in this ecosystem, as seen in Tanzania [17]. In addition, we also analysed non-symptomatic bones from 11 species sampled across 8 sub-Saharan countries to further explore the range of hosts that this pathogen infects and be in a position to detect potential broad-scale spatial patterns. Finally, we examined the TPE genomes we generated for evidence of mutations associated with antibiotic resistance, as ongoing eradication efforts are based on antibiotic therapy.

METHODS

Study sites and samples

In January 2014, sooty mangabeys from a habituated social group in TNP were observed with orofacial lesions and lesions of their distal extremities; samples were collected from two symptomatic animals and through whole-genome sequencing, TPE was determined to be the cause of the infection [12]. Over the next 2 years, other individuals started showing genital ulcerations and necrotizing dermatitis on the inner parts of the thighs and abdomen, often with visible

Fig. 1. Lesions due to TPE infection in sooty mangabeys. (a) Necrotizing dermatitis of inner parts of the thighs and ventral abdomen with yellowish crusts. (b) Necrotic orofacial lesions. (c) Genital lesions in females. (d) Genital lesions in males.

yellow crusts. Orofacial and genital lesions were still observed on other animals in the group during the study period (Fig. 1). Three more individuals with visible lesions were chemically immobilized using a combination of xylazine (1 mg kg^{-1}) and ketamine (10 mg kg^{-1}) administered by blowpipe. Biopsy and swab samples were collected from orofacial and genital lesions

(Table 1). Samples were preserved in RNAlater Thermo Fisher Scientific, Waltham, MA, USA) and shipped to the Robert Koch Institute, Berlin, Germany for molecular analysis.

To obtain insights into NHP treponematoses in TNP over the last three decades and to obtain an understanding of the

Table 1. Lesion and sample types observed at TNP and PCR screening results for each animal

Animal ID	Species	Lesions observed	Sample type	*PolA* PCR	*cfpA* PCR
5847	*C. atys*	Anogenital lesions	Lesion biopsy	Negative	Negative
		Anogenital lesions	Lesion swab	Positive	Positive
		Facial lesions	Lesion biopsy	Negative	Negative
		Body ventral lesions	Lesion biopsy	Negative	Negative
2116	*C. atys*	Genital lesions	Lesion swab	Positive	Positive
2117	*C. atys*	Genital lesions	Lesion swab	Positive	Positive
1864	*C. atys*	Orofacial lesions	Lesion biopsy	Positive	Positive

Animal ID 1864 was previously reported in Knauf *et al.* [12] and in this study we prepared a new library and resequenced it to improve genome coverage.

host range of this pathogen, we screened non-symptomatic bones collected opportunistically at TNP (*n*=67; Table S1, available in the online version of this article). Previous studies have shown that informative TP sequences can be obtained from such skeletal remains [18]. We also screened NHP bones collected from 11 additional field sites in sub-Saharan Africa (*n*=83; Table S1) through the support of the Pan African Programme: The Cultured Chimpanzee (PanAf; http://panafrican.eva.mpg.de/). NHP bones were assigned to particular species using molecular methods targeting the 16S ribosomal RNA gene as described previously [19], coupled with morphological assignment by experts in the field.

DNA extraction

DNA was extracted from skin biopsies (*n*=4) and swabs (*n*=3) using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's instructions (Table 1). DNA was extracted from bones using a silica-based method. Briefly, non-lesioned bones were drilled using a fine drill bit at slow speed to produce ~150 mg of bone powder. The drilling was performed in a designated sealed glove box, both to prevent any contamination of the bones and to prevent exposure of researchers to pathogens that might be present in the bones [e.g. in TNP, cases of sylvatic anthrax (*Bacillus cereus* biovar *anthracis*) are frequent, and this pathogen can be cultured from bones] [20]. The box was UV-sterilized and surfaces were bleached following the drilling of each bone and extraction. Drill bits, the metal pieces that actually come into contact with the bones, were changed for each bone to prevent cross-contamination. DNA extraction from bone powder was performed following a protocol that was modified from one that was used previously [21, 22] and is described in detail elsewhere [23]. Extracted DNA from all samples was quantified with a Qubit fluorometer using a double-stranded DNA high-sensitivity assay kit (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. DNA was subsequently stored at −20 °C.

Screening for *T. pallidum*

To screen for TP DNA in swabs and biopsy samples, we performed an end-point polymerase chain reaction (PCR) assay targeting the 67 bp of the *polA* gene fragment, using primers previously developed for screening human clinical specimens (Table S2) [24]. PCR reactions were performed in 25 µl reactions; up to 200 ng of DNA was amplified using 1.25 U of high-fidelity Platinum *Taq* polymerase (Thermo Fisher Scientific, Waltham, MA, USA), 10× PCR buffer (Thermo Fisher Scientific, Waltham, MA, USA), 200 µM dNUTPs, 4 mM MgCl2 and 200 nM of both forward and reverse primers. The thermal cycling profile was as follows; denaturation at 95 °C for 5 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 30 s and 72 °C for 1 min, with a final elongation step at 72 °C for 10 min. Known TPE-positive DNA extracts and negative controls were included.

The 67 bp amplified product is too short for direct Sanger sequencing, so to confirm the results of this initial screening positive samples were further tested with a semi-nested assay

targeting the cytoplasmic filament protein gene (*cfpA*) [25]. These primer pairs yield a 352 bp outer product in the primary PCR and a 189 bp inner fragment in the second round PCR (Table S2). In both primary and second round PCR assays, the reactions were performed as follows: ~200 ng of DNA was amplified in a 25 µl reaction using 1.25 U of high-fidelity Platinum *Taq* polymerase, 10× PCR buffer, 200 µM dNUTPs, 4 mM MgCl2 and 200 nM of both forward and reverse primers. For the second round PCR, 2 µl of a 1:20 dilution of the first round PCR product was used as input template. The thermal cycling profiles for both rounds of the nested PCRs were the same as in the initial screening assay described above.

PCR products were visualized on a 1.5% agarose gel stained with GelRed (Thermo Fisher Scientific, Waltham, MA, USA). Positive bands were purified using the PureLink Quick Gel Extraction kit (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's protocol. Purified products were stored at −20 °C until they were sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific, Waltham, MA, USA) and sequences were compared to publicly available sequences in EMBL through blast [26]. All samples that tested positive in the confirmatory *cfpA* PCR assay were selected for whole-genome in-solution hybridization capture and high-throughput sequencing.

Potential DNA degradation in bone samples precluded the use of the *cfpA* PCR assay (demonstrated in [18]). To select the most promising samples, we estimated copy numbers in the bones using a real-time qPCR also targeting the 67 bp fragment of the *polA* gene [24]. Samples were tested in duplicate. Briefly, 5 µl of total DNA was amplified in a 25 µl qPCR reaction containing 10× PCR buffer, 200 µM dNUTPs, 4 mM MgCl2, 300 nM of both forward and reverse primers, 100 nM of a specific probe and 0.5U of high-fidelity Platinum *Taq* polymerase. The thermal cycling profile was set as follows DNA denaturation at 95 °C for 10 min followed by 45 cycles at 95 °C for 15 s and 60 °C for 34 s. All bone samples that had detectable TP DNA in duplicate reactions were selected for in-solution hybridization capture and high-throughput sequencing.

Library preparation, genome-wide capture and high-throughput sequencing

For biopsy and swab sample extracts, we sheared 1000 ng of DNA per sample to 400 bp fragments using the Covaris S2 ultrasonicator (intensity: 4; duty cycle: 10%; cycles per burst: 200; treatment time: 55 s; and temperature: 4–5 °C). Bone samples were not sheared due to the potentially fragmented nature of DNA in these older specimens. Two library preparation methods were applied (Table 2). Single-indexed Illumina libraries had previously been built and sequenced from samples 1864 and 22_52 using the Accel NGS ds DNA Library Preparation kit (Swift Biosciences, Ann Arbor, MI, USA) [12, 18]. In an effort to improve genome coverage for these samples, we generated new dual-indexed libraries using the NEBNext Ultra II DNA Library Prep kit (New England

Table 2. Summary table for library preparation, capture and sequencing methods used for each sample analysed in this study

Sample ID	NHPs species	Sample type	Library preparation method		Site/country
			NEBNext	**Accel NGS**	
Boe_092	*C. atys*	Bone	Yes	No	Boe/GB
2117	*C. atys*	Swab – genital lesion	Yes	No	TNP/CI
5847	*C. atys*	Swab – genital lesion	Yes	No	TNP/CI
2116	*C. atys*	Swab – genital lesion	No	Yes	TNP/CI
1864	*C. atys*	Biopsy – face lesion	Yes	Yes	TNP/CI
Tai_105	*P. badius*	Bone	Yes	No	TNP/CI
22_52	*P. badius*	Bone	Yes	Yes	TNP/CI
Control	Chicken DNA	Tissue	Yes	No	NA

Samples 1864 and 22_52 had been sequenced previously, but in this study we prepared new libraries and resequenced them to improve genome coverage [12, 18].

NA, not applicable.

Biolabs, Ipswich, MA, USA). The different library preparation methods used reflect the adoption of a new method in the laboratory because of higher library conversion rates with the NEBNext Ultra II DNA Library Prep kit; we have no reason to think that these different methods will systematically affect the genomes generated. For all other samples, DNA was converted into dual-indexed Illumina libraries using the NEBNext Ultra II DNA Library Prep kit (New England Biolabs, Ipswich, MA, USA; samples: Tai_105, Boe_092, 2116, 2117, 5847). All generated libraries were quantified using the KAPA library quantification kit (KAPA Biosystems, Wilmington, MA, USA) following the manufacturer's instructions. In all library preparations, chicken DNA was included as a control, as no subspecies of TP is known to infect birds.

Libraries were enriched for *T. pallidum* sequences using in-solution hybridization capture with biotinylated RNA baits following the manufacturer's protocol (myBaits, Arbor Biosciences, Ann Arbor, MI, USA). The baits spanned the simian-derived Fribourg-Blanc reference genome (RefSeq ID: NC_021179.1) with a two-fold tiling. In-solution hybridization capture was performed for two rounds of 48 h each. After each round of capture, a post-capture amplification step was performed using the KAPA HiFi HotStart library amplification kit (KAPA Biosystems) with P5 and P7 Illumina primers to generate ~200 ng of enriched DNA per sample. The post-capture amplification thermal profile was as follows: initial hot start at 98 °C for 2 min followed by 12 to 16 cycles at 98 °C for 20 s, 65 °C for 30 s and 72 °C for 45 s, with a final elongation step at 72 °C for 5 min. Enriched libraries were quantified using the KAPA library quantification kit (KAPA Biosystems, Wilmington, MA, USA). Prior to sequencing, libraries were diluted to 4 nM and pooled for sequencing on an Illumina MiSeq (Illumina, San Diego, CA, USA) with 300 bp paired end reads (V3 chemistry; Table 2) at the Robert Koch Institute (Berlin, Germany).

Bioinformatics analysis

The paired-end reads generated here, along with published SRAs from prior TP sequencing efforts (Table S3), were trimmed using Trimmomatic v0.38, removing the leading and trailing reads below a quality score of 30, clipping any part of the read where the average base quality across 4 bp was less than 30, and removing reads less than 30 bp in length [27]. The surviving read pairs were merged using Clip and Merge version 1.7.8 with default settings [28]. Merged reads and surviving single-end reads were combined and mapped to TPE Fribourg-Blanc (RefSeq ID: NC_021179.1) using BWA-MEM [29] with a minimum seed length of 29. Mapped reads were sorted using Picard's SortSam and subsequently deduplicated with Picard's MarkDuplicates (https://broadinstitute.github.io/picard/index.html). Alignments with a mapping quality score smaller than 30 and a mapping length lower than 30 were also removed using SAMtools [30]. We called two consensus sequences for each genome using Geneious v11.1.5 [31]: (i) the first consensus required a minimum of 10 unique reads per position (i.e. 10× coverage) with at least 95% identity for a base to be unambiguously called; (ii) the second consensus required 5× coverage and at least 95% identity. Positions that did not meet the coverage requirement were subsequently treated as missing data. In the following, we mostly report on the results obtained with the 5×, 95% dataset, which includes all the (partial) genomes generated in this study. To confirm these findings, we repeated the analyses using the 10×, 95% dataset that included only the higher-quality genomes available, thereby excluding two partial genomes of lower quality that we assembled from bone material (Tai 105 and Boe 092).

Each set of whole-genome consensus sequences was aligned using the multiple sequence alignment programme MAFFT [32]. We then removed all previously described paralogous and putative recombinant genes (Table S4) and selected

conserved blocks using the Gblocks tool [33] in SeaView v4 [34]. Phylogenetic inference was performed on the resulting alignments of informative positions after stripping all identical sites and ambiguities in the final datasets: 5×, 95% (1737 positions: 37 sequences including the two TPE genomes from bones) and 10×, 95% (1739 positions: 35 sequences). To account for the effects of potential long branch attraction, we ran analyses on versions of these alignments that either included or excluded non-TPE sequences.

Final alignments were uploaded to the online ATGC PhyML-SMS tool (http://www.atgc-montpellier.fr/phyml-sms/) for construction of a maximum-likelihood (ML) phylogeny using smart model selection [35] with the Bayesian information criterion and subtree pruning and regrafting (SPR) for tree improvement, but otherwise using default settings. Branch robustness was estimated using the Shimodaira–Hasegawa approximate likelihood ratio test (SH-like aLRT) [36]. We also used RAxML Next Generation [37], which allowed for an accounting of the number of invariant sites that had been stripped from the alignment to correct for acquisition bias, using 200 bootstrap replicates and the model selected by PhyML-SMS [38, 39]. Both ML analyses retrieved very similar trees; here we report the ML trees from this analysis that accounted for the number of invariant sites. The ML trees were then rooted using TempEst (version 1.5.1), which estimated the best-fitting root of these phylogenies using the heuristic residual mean squared function, which minimizes the variance of root-to-tip distances [40]. Evolutionary pairwise distance between TP strains were extracted from the X5, 95% ml phylogeny using the Patristic program [41].

To explore the robustness of our phylogenetic analysis, we also ran Bayesian phylogenetic analyses with BEAST (version 1.10.4) using four different models. These models all used the general time-reversible (GTR) nucleotide substitution model identified in the PhyML-SMS analysis, but combined either a strict clock model or an uncorrelated lognormal relaxed clock model with demographic models assuming a coalescent process and a constant population size or a birth–death process. These models covered a plausible range of clock and tree priors, similar to those explored in another recent publication on syphilis genomics [42]. We incorporated the number of invariant sites into these models to account for sampling and ascertainment bias. For all models, we examined the output of multiple runs for convergence and appropriate sampling of the posterior using Tracer (version 1.7.1) [43] before merging runs using Log Combiner (version 1.10.4) [44]. The maximum clade credibility (MCC) tree was then identified from the posterior set of trees and annotated with Tree Annotator (version 1.10.4: distributed with BEAST). The resulting ML and MCC tree files were edited using iTOL (https://itol.embl.de/) [45].

For those bone samples for which the aforementioned phylogenetic pipeline was unable to resolve the position of their partial genomes in the phylogeny, we performed phylogenetic read placement using the evolutionary placement algorithm tool EPA-ng to determine the position of individual reads on the TP phylogeny [46]. We performed read placement on a red colobus bone sample (22-52; this particular sample exhibited a suspiciously long branch in initial ML and Bayesian analyses, which we considered to be potentially indicative of an assembly problem, e.g. the inclusion of reads from TPE and an environmental spirochete) and two previously sequenced chimpanzee bone samples (11786 and 15028) that were collected in TNP, for which the presence of TP was confirmed but no subspecies assignment had been performed [18]. Briefly, we selected filtered merged and single-end reads that mapped to TPE and were 50 bp or longer. The surviving reads were then aligned to the whole-genome alignment using the parsimony-based phylogeny-aware read alignment program (PaPaRa) [47]. The resulting PaPaRa alignment was then split into the query reads and the original alignment using the split function available in the EPA-ng toolkit [47]. To estimate the best fitting evolutionary model for the phylogenetic placement of the query reads, both the reference tree and reference alignment were evaluated using the RaxML-ng toolkit [48]. The best fitting model was then used to place query reads on the reference tree with the EPA-ng tool. The resulting EPA-ng jplace tree files were visualized as heat trees depicting the percentage of reads placed on each branch using the gappa toolkit [49]. For tree visualization, every query read was treated as a point mass concentrated on the highest-weight placement and the multiplicity of each query read was set to 1.

RESULTS

Screening PCR and whole-genome capture

All symptomatic animals sampled by biopsy or swab tested positive for *T. pallidum* in at least one of the sample types collected (Table 1). Sequences generated from the respective assays were all identical (a representative sequence was uploaded to Zenodo: doi.org/10.5281/zenodo.3540499). In addition, based on the *polA* gene qPCR assay, we detected TP DNA in NHP bones from the following species/subspecies and field sites: 14 western chimpanzees (*Pan troglodytes verus*), 5 western red colobus (*Piliocolobus badius*), 2 Diana monkeys (*Cercopithecus diana*), and 1 sooty mangabey from TNP in Côte d'Ivoire; 5 eastern chimpanzees (*Pan troglodytes schweinfurthii*) from Bili-Uere in the Democratic Republic of the Congo; 2 western chimpanzees from Loango National Park, Gabon; 1 sooty mangabey from Boe in Guinea-Bissau; 1 western chimpanzee from East Nimba, Liberia; 1 western chimpanzee and one lesser spot-nosed monkey (*Cercopithecus petaurista*) from the nationwide survey in Liberia; 2 western chimpanzees from Sapo National Park, Liberia; 1 eastern black-and-white colobus (*Colobus guereza*) from Gashaka Gumti National Park, Nigeria; 1 blue monkey (*Cercopithecus mitis*) and 1 eastern black-and-white colobus from Budongo in Uganda (Table S1).

From the *C. atys* study group in TNP, we were able to sequence three new TPE genomes from biopsy and swab samples with 5× coverage of 47.5–97.5% of the genome (Table 3). We reanalysed a previously sequenced sample from a *C. atys*

Table 3. Mapping results for non-human primate *T. pallidum* subsp. *pertenue* strains from TNP and BNP determined in this study and those sequenced previously

Sample ID	NHPs species	Sample type	Deduplicated reads mapped	Positions covered [1×]	Positions covered [5×]	% genome coverage [1×]	% genome coverage [5×]
Boe 092	*C. atys*	Bone	36407	793392	247484	69.6	21.7
2117	*C. atys*	Swab – genital lesion	44143	1009366	541728	88.5	47.5
5847	*C. atys*	Swab – genital lesion	457063	1117566	1111969	98.0	97.5
2116	*C. atys*	Swab – genital lesion	131550	1119023	1066350	98.1	93.5
1864	*C. atys*	Biopsy – face lesion	202416	1116210	1089159	97.9	95.5
Tai 105	*P. badius*	Bone	42719	867318	448209	76.1	39.3
22_52	*P. badius*	Bone	43373	711608	306789	62.4	26.9
Control	Chicken	Tissue	1	52	0	0.005	0.0

with facial lesions (1864) [12], and improved the 5× genome coverage from 82.4 to 95.5%. Further, we recovered partial TPE genomes from two *P. badius* bones (Tai 105 and 22_52) from TNP, Côte d'Ivoire, as well as one *C. atys* (Boe 092) bone from Boe, Guinea Bissau, with 76.1, 62.4 and 69.6%, respectively at 1× genome coverage and 39.3, 26.9 and 21.7% at 5× genome coverage, respectively. In the control library (chicken DNA), only a single TP read survived quality control. All genomes were distinct, including those obtained from individuals belonging to a single social group of *C. atys*. None of the genomes generated in this study had the A2058G and A2059G mutations in the 23S ribosomal RNA gene, which have been demonstrated to confer antimicrobial resistance to macrolide antibiotics [50].

Phylogenetic analyses

Phylogenetic analysis of genomes generated as part of this study, all other TPE and TEN genomes, and a representative selection of TPA genomes from GenBank (Table S3) yielded tree topologies largely consistent in both the RAxML next-generation and BEAST-based approaches and the two minimum coverage thresholds used. The ML and MCC tree topologies resolved into distinct reciprocally monophyletic groups representing the TP subspecies (TPA, TPE and TEN). The TPE clade included both human- and all NHP-infecting strains, while TPA and TEN clades consisted only of human strains (Fig. 2). These analyses clearly showed that all NHP-derived strains were TPE, but the relationships within the TPE clade were poorly resolved, potentially due to the long branches separating these clades. Long branches have the potential to complicate phylogenetic analyses [51]; thus we focused our examination of relationships of strains with the TPE clade on the ingroup phylogenetic analyses presented below.

Our ingroup phylogenetic analysis revealed that neither the ML nor the MCC tree topologies showed monophyly of TPE strains infecting NHPs based on their clinical manifestations, regardless of the minimum coverage thresholds used. In other words, TNP strains causing different clinical manifestations did not form statistically supported, reciprocally monophyletic groups based on clinical manifestations sampled (Figs 2 and 3, Table 4).

Rather, it appeared that simian strains formed reciprocally monophyletic groups based on geography, as strains from the same location generally seemed to cluster together. We observed the most consistent support for a clade of NHP-infecting TPE strains from Senegal, The Gambia and Guinea-Bissau, with all analyses and models showing this geographical clustering, in most cases with strong statistical support (Figs 3 and S1, Table 4). This clade consistently included a sooty mangabey bone-derived TPE (Boe_092) collected in Guinea-Bissau and included the Fribourg-Blanc strain isolated from a baboon in Guinea and green monkeys (*Chlorocebus sabaeus*) from neighbouring Senegal and The Gambia (Fig. 3, Table 4). We found less consistent support for a clade including NHP-infecting TPE strains from TNP, specifically TPE-infecting western red colobus and sooty mangabeys (Figs 3 and S1, Table 4). For the 5×, 95% dataset, statistical support for this TNP clade was relatively weak, predominantly due to inconsistent placement of the low-coverage 1864 genome; support for a monophyletic TNP clade was much stronger in the 10×, 95% dataset (Fig. S1, Table 4). In addition, we observed that human-derived TPE strains formed a monophyletic group with relatively strong statistical support in the 10×, 95% dataset (Fig. S1, Table 4).

The analysis of bone-derived TPE strains further supported a geographical structuring of NHP TPE genomic diversity. Read placement of bone-derived TPE reads from one western red colobus specimen (22_52) and two western chimpanzees specimens (11786 and 15028) from TNP showed that the majority of reads fell within the TNP clade (Fig. 4).

Within these geographical clusters, for TPE-infecting particular species at each field site, there appeared to be differences in average genetic distances between strains. The sooty mangabey-infecting strains at TNP had an average patristic distance of 4.23333E-05%, compared to 0.000152, 0.000002 and 0.000002% for strains infecting *Papio anubis* in
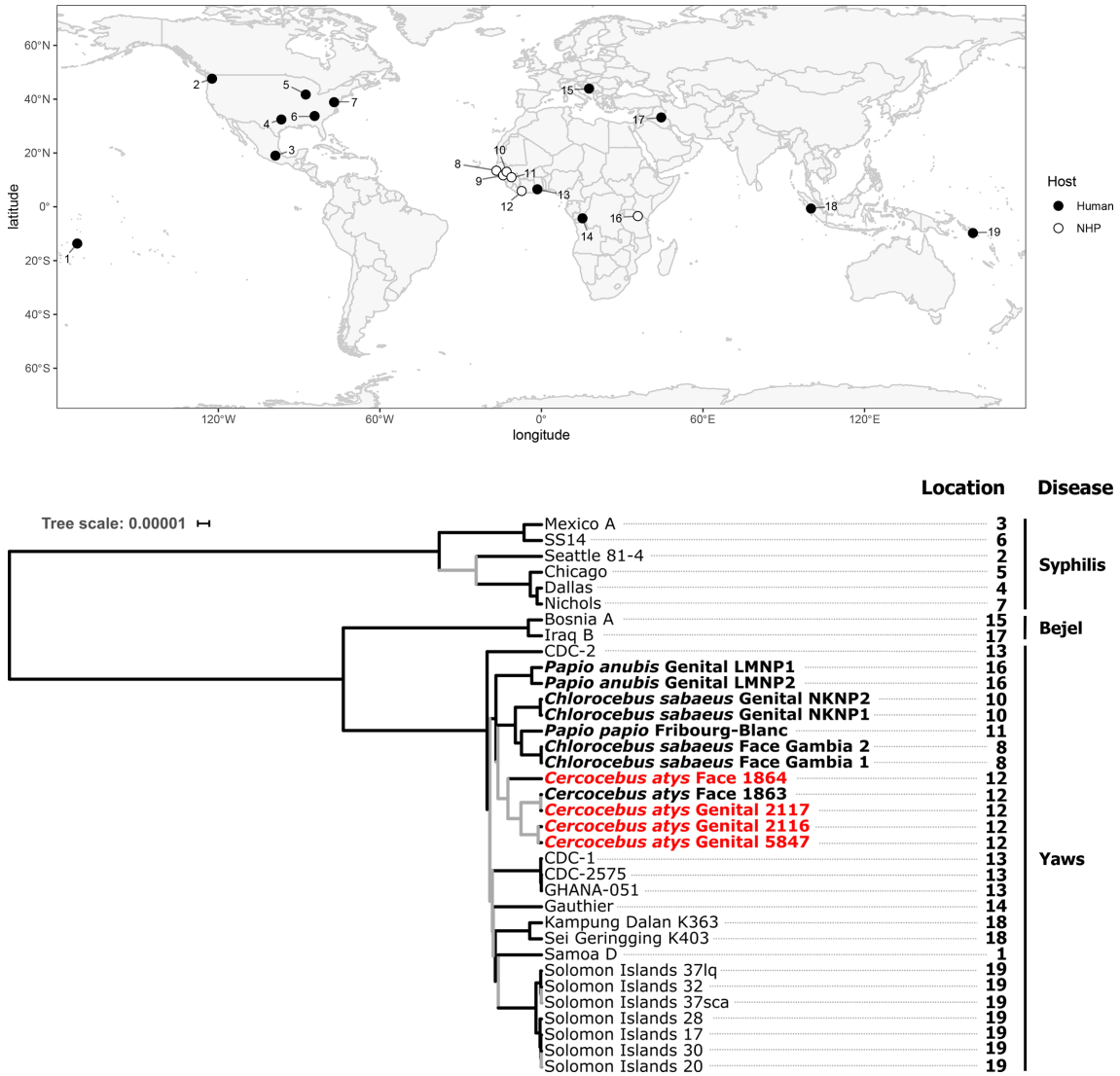
Fig. 2. Maximum clade credibility tree of *T. pallidum* strains (relaxed clock model assuming a birth–death process) and map showing the sampling locations for all TP genomes that appear in the tree. All simian-infecting strains are shown in bold with tip labels showing the host species, location of the sampled lesion and sample ID. Genomes generated in this study are shown in red with a minimum coverage of 10× to call a base, and a threshold of 95% identity for a base to be called. Branches supported by SH-like aLRT values <0.90 in the maximum-likelihood tree and posterior probabilities <0.95 in the maximum clade credibility tree are indicated in grey. The scale shows nucleotide substitutions per variable site.

Tanzania and *Chlorocebus sabaeus* monkeys in Senegal and The Gambia, respectively (Table S5), although sample sizes were small, precluding a formal statistical analysis.

DISCUSSION

In agreement with previous studies [11, 13, 17], our phylogenetic analyses demonstrated that the NHP-infecting TP strains all belonged to the TPE clade. We confirmed the presence of TPE in both orofacial and genital lesions, adding to a growing body of evidence that TPE causes a diversity of symptoms in NHPs [11, 25, 52–55]. We found no evidence that TPE strains causing these different symptoms formed

separate monophyletic groups, suggesting that TPE pathology may not be determined by bacterial properties alone, but also by host factors (e.g. individual immune status) or the route of exposure.

While TPE genomes did not cluster by symptoms, we found that simian TPE strains from different ecosystems in sub-Saharan Africa formed monophyletic groups largely based on their geographical origin. These results support and extend the findings of Chuma *et al.* [17], who used multilocus sequence typing (MLST) data to reveal that TPE sequences from NHPs in Tanzania did not form monophyletic clades based on host species, but rather clustered based on geography
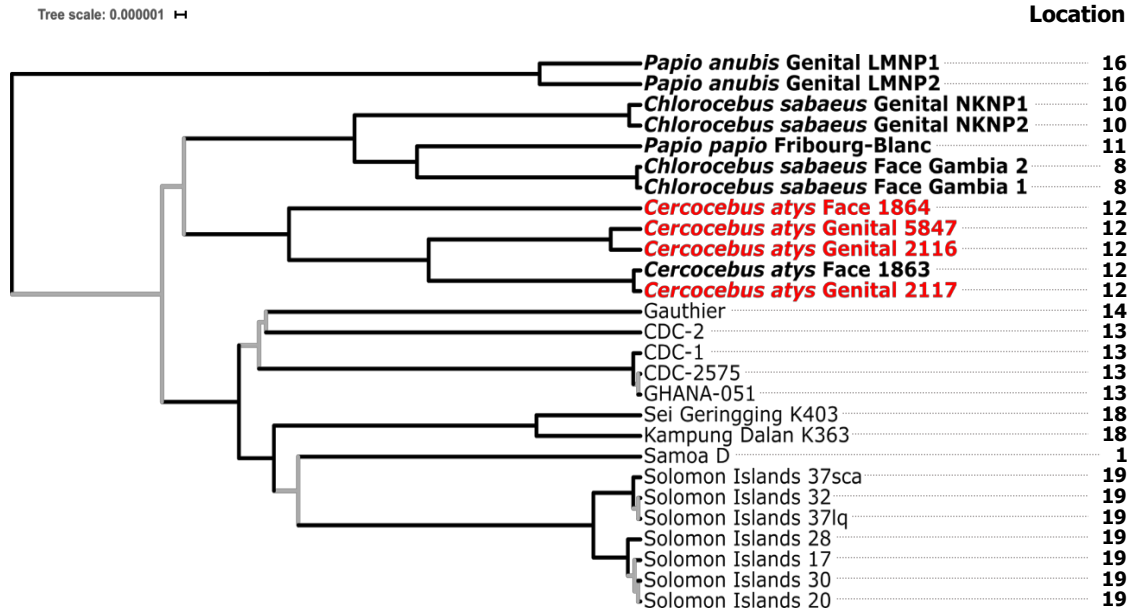
Fig. 3. Maximum clade credibility tree from the ingroup analysis of TPE strains (relaxed clock model assuming a birth–death (speciation) process). All simian-infecting strains are shown in bold with tip labels showing the host species, location of the sampled lesion and sample ID. Genomes generated in this study are shown in red with a minimum coverage of 10× to call a base, and a threshold of 95% identity for a base to be called. Branches supported by SH-like aLRT values <0.90 in the maximum-likelihood tree and posterior probabilities <0.95 in the maximum clade credibility tree are indicated in grey. The sampling locations that appear on this tree are those shown on the map in Fig. 2. The scale shows nucleotide substitutions per variable site.

Table 4. Summary table of TPE ingroup phylogenetic analyses. The support for monophyletic groups of simian and human strains based on the different probabilistic methods, molecular clocks and tree priors is shown

Dataset	Probabilistic method	Clock	Tree prior	Based on clinical manifestations	Senegal–The Gambia–Guinea-Bissau	Tai National Park	Human-infecting strains
5×, 95%	ML	NA	NA	No	Yes	Yes (weak)	No
	BMCMC	Strict	Coalescent	No	Yes	No	No
	BMCMC	Strict	Speciation	No	Yes	No	No
	BMCMC	Relaxed	Coalescent	No	Yes	No	Yes (weak)
	BMCMC	Relaxed	Speciation	No	Yes	No	Yes (weak)
10×, 95%	ML	NA	NA	No	Yes	Yes (weak)	Yes (weak)
	BMCMC	Strict	Coalescent	No	Yes	Yes	Yes
	BMCMC	Strict	Speciation	No	Yes	Yes	Yes
	BMCMC	Relaxed	Coalescent	No	Yes	Yes	Yes
	BMCMC	Relaxed	Speciation	No	Yes	Yes	Yes

NA, not applicable.
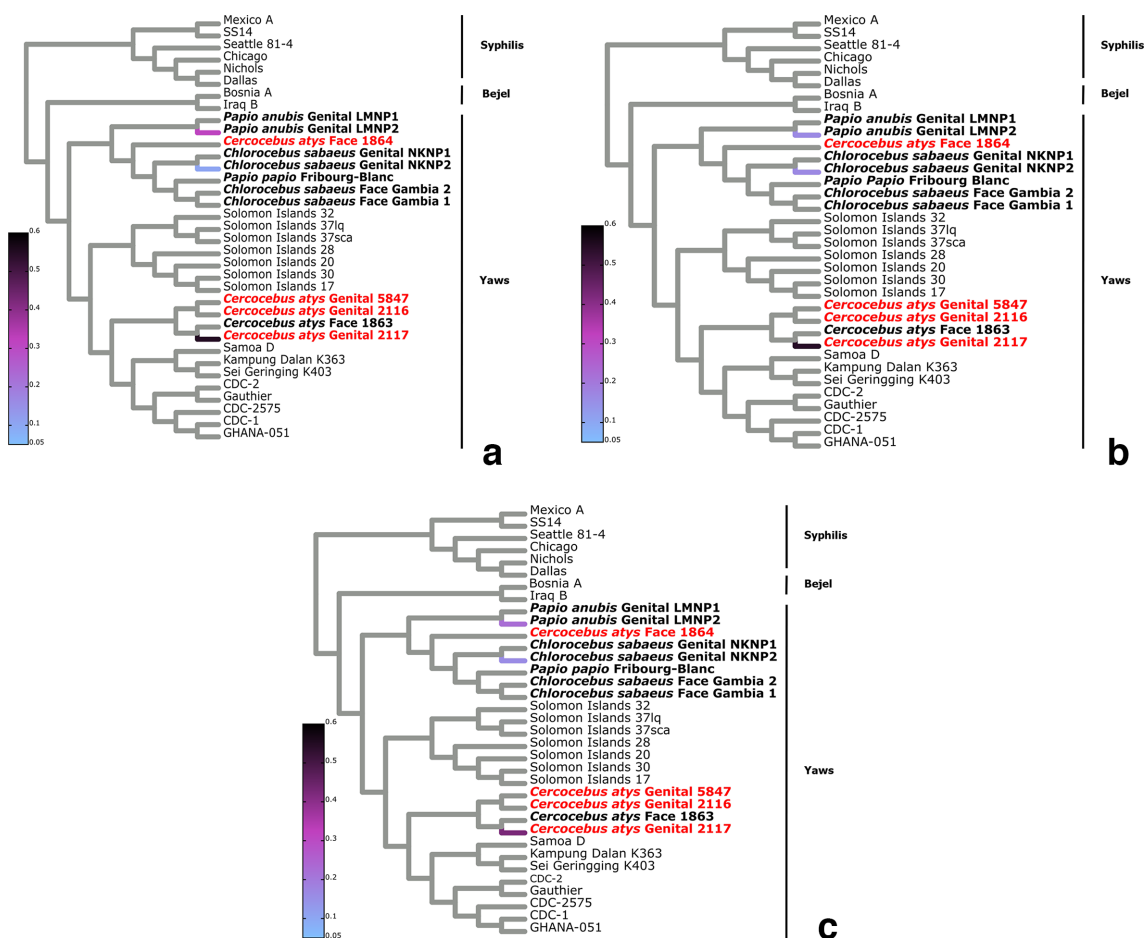
Fig. 4. Phylogenetic read placement of bone samples. Heat tree visualization of phylogenetic placement of TPE mapped reads from bone samples on to the TP MCC (10× coverage and 95% threshold) reference tree using the evolutionary placement algorithm (EPA-ng). The approximate percentage of reads placed on to a particular branch of the *T. pallidum* cladogram is shown as a linearly scaled colour density. Genomes generated in this study are shown in red. (a) Sample 11786 (*P. troglodytes verus*: total number of reads, 517), (b) sample 15028 (*P. troglodytes verus*: total number of reads, 3581) and (c) sample 22_52 (*P. badius*: total number of reads, 19389).

[17]. Here, considering a larger geographical scale, we find that TPE strains from TNP cluster separately from strains infecting NHPs in Guinea, Guinea-Bissau, Senegal and The Gambia (all countries in close proximity to one another). This geographical signal was supported by phylogenetic read placement of TPE reads derived from the bones of two NHP species collected in TNP, which fell predominantly in the TNP clade. This observed incongruence between the host phylogeny and the phylogeny of their TPE infections could be indicative of cross-species transmission or infection from some common unknown source within a habitat [56, 57].

We observed high TPE diversity in some ecosystems, which could lend support to the hypothesis of exposure of NHPs to a diversity of strains, potentially from other species sharing an ecosystem. We observed the highest level of diversity within one region in the TPE strains sampled from a single sooty mangabey group in the TNP ecosystem. Indeed, the average number of single-nucleotide polymorphisms (SNPs) separating any two TNP sooty mangabey-infecting TPE genomes

was 97 (5×, 95% identity threshold). While one pair differed at only 3 positions, potentially suggesting an epidemiological link, the most divergent pair had 264 SNPs. Considering that the TPE mutation rate in humans has been estimated to be ~1.21×10^{-7} per nucleotide site per year (or lower [58]), equalling fewer than 0.138 mutations per year across the 1.14 MB TPE genome, the 264 SNPs between the 2 most divergent TNP strains would likely have required hundreds to thousands of years to accumulate (Table S6). High levels of divergence precluding a direct epidemiological link were also observed for two TPE strains sampled from the Lake Manyara National Park ecosystem in Tanzania [17]. In other ecosystems, TPE genomes from the same species were very similar, compatible with an epidemiological link (green monkeys in Bijilo Forest Park, The Gambia and Niokolo National Park, Senegal [12]). The factors driving the high diversity of TPE in some ecosystems and not others are unknown, but between-species transmission in diverse primate ecosystems, such as TNP, could play a role.

Interspecies interactions that could facilitate transmission via direct contact between NHP species inhabiting TNP are well documented; these include a strong predator–prey relationship between chimpanzees and red colobus, as well as direct contact between monkeys that spend large amounts of time in mixed-species associations (e.g. grooming, fighting, play, mating [59, 60]). Another transmission mode that has long been suggested for TPE is vectorial transmission; under experimental conditions viable TP spirochetes were transmitted by flies between different host species causing clinical disease [61, 62]. Knauf *et al.* amplified TP DNA from flies in ecosystems where TP infections in NHPs are common [63] and Gogarten *et al.* showed that flies carrying TP DNA formed high-density persistent associations with NHP social groups in TNP [64]. Primate-associated flies were observed to move between groups of different species, suggesting that they could be involved in transmitting the yaws pathogen between species, even when NHP are not found in mixed-species associations [64]. Further work is needed to confirm whether flies or other arthropods actually transmit TPE in the wild [65] and to understand the routes and rates of transmission of TPE between NHP species.

Wide spatio-temporal differences in sampling of human- and simian-derived TPE strains precludes a robust assessment of whether zoonotic transmission occurs. The sampling of NHP TPE genetic diversity has improved over the last several years, with 14 of the 29 TPE genomes included in the current analyses originating from NHPs in sub-Saharan Africa. Unfortunately, in countries where NHP-infecting TPE strains have been sequenced, no genomic data from human infections are available. Given the geographical signal observed for simian isolates [17], future studies may benefit from investigating human yaws infections in these regions. Such data will help determine whether zoonotic inter-species transmission of TPE between humans and NHPs actually occurs. Our analysis found some support for the monophyly of human-infecting strains; if larger datasets support this finding, it will suggest that zoonotic spillover does not occur frequently (or at all). If zoonotic spillover of TPE from NHPs to humans does occur, it is noteworthy that all NHP-derived TPE strains characterized to date [17], including those generated in this study, have no mutations in the 23S ribosomal RNA gene, which is known to confer antimicrobial resistance to macrolide antibiotics [50]. Therefore, the available antibiotic treatments are expected to be effective against NHP-derived strains. However, we caution against overinterpreting this finding, as sample sizes are small, meaning that low-frequency mutations would likely remain undetected.

This study joins a growing body of evidence that human and wildlife bones are a useful resource for generating sufficient TPE reads to inform phylogenetic analyses that can both extend the known host range of TPE and push back the date of emergence for TPE in particular populations [18, 66]. Despite a lack of clinical evidence from western chimpanzees and western red colobus in TNP, the phylogenetic placement of bone-derived TPE reads into the TNP clade confirms that sooty mangabeys are not the only species affected by TPE in this ecosystem. The oldest bone from which we were able to phylogenetically assign reads to the TPE clade was collected in 1992, complementing clinical evidence from TNP that only started accumulating from 2014 onward [12]. Our results suggest that, together with archaeological collections [67–69], natural history specimens can provide important genomic information regarding TPE and thereby further our understanding of its ecology and evolution.

CONCLUSIONS

We found that genomically diverse TPE circulate in NHP, even in a single social group [17]. On a larger spatial scale, whole-genome sequences and bone-derived TPE sequences allowed phylogenetic analyses that revealed that the genomic diversity of TPE strains derived from NHPs is geographically structured. This pattern is compatible with cross-species transmission of TPE within ecosystems, although how often and by what means this transmission occurs remains an important area of future research.

Wild Chimpanzee Foundation, Guinea; Foundation Chimbo (Boe); Wild Chimpanzee Foundation, Liberia; Gashaka Primate Project, Nigeria; Wildlife Conservation Society (WCS) Nigeria, Nigeria; Fongoli Savanna Chimpanzee Project, Senegal; Jane Goodall Insitute Spain (Dindefelo), Senegal; Budongo Conservation Field Station, Uganda; Ngogo Chimpanzee Project, Uganda.

References

1. Armelagos GJ, Zuckerman MK, Harper KN. The science behind pre-Columbian evidence of syphilis in Europe: research by documentary. *Evol Anthropol* 2012;21:50–57.

2. Lithgow KV, Hof R, Wetherell C, Phillips D, Houston S *et al*. A defined syphilis vaccine candidate inhibits pallidum. *Nat Commun* 2017;8:1–10.

3. Giacani L, Lukehart SA. The endemic treponematoses. *Clin Microbiol Rev* 2014;27:89–115.

4. Marks M, Solomon AW, Mabey DC. Endemic treponemal diseases. *Trans R Soc Trop Med Hyg* 2014;108:601–607.

5. Čejková D, Zobaníková M, Chen L, Pospíšilová P, Strouhal M *et al*. 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 Negl Trop Dis* 2012;6:e1471.

6. Centurion-Lara A, Molini BJ, Godornes C, Sun E, Hevner K *et al*. Molecular differentiation of *Treponema pallidum* subspecies. *J Clin Microbiol* 2006;44:3377–3380.

7. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M *et al*. Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. *PLoS One* 2015;10:e0143304.

8. WHO. Yaws. Available: https://www.who.int/news-room/factsheets/detail/yaws 2019.

9. WHO. First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases. 2010.

10. Dyson L, Mooring EQ, Holmes A, Tildesley MJ, Marks M. Insights from quantitative and mathematical modelling on the proposed 2030 goals for yaws. *Gates Open Res* 2019;3:1576.

11. Chuma IS, Batamuzi EK, Collins DA, Fyumagwa RD, Hallmaier-Wacker LK *et al*. Widespread *Treponema pallidum* Infection in Nonhuman Primates, Tanzania. *Emerg Infect Dis* 2018;24:1002–1009.

12. Knauf S, Gogarten JF, Schuenemann VJ, De Nys HM, Düx A *et al*. Nonhuman primates across sub-Saharan Africa are infected with the yaws bacterium *Treponema pallidum* subsp. *pertenue*. *Emerg Microbes Infect* 2018;7:1–4.

13. Zobaníková M, Strouhal M, Mikalová L, Cejková D, Ambrožová L *et al*. Whole genome sequence of the *Treponema* Fribourg-Blanc: unspecified simian isolate is highly similar to the yaws subspecies. *PLoS Negl Trop Dis* 2013;7:e2172.

14. Smith JL, David NJ, Indgin S, Israel CW, Levine BM *et al*. Neuro-ophthalmological study of late yaws and pinta. II. The Caracas project. *Sex Transm Infect* 1971;47:226–251.

15. Lawton Smith J, Israel CW. Recovery of spirochaetes in the monkey by passive transfer from human late sero-negative syphilis. *Br J Vener Dis* 1968;44:109–115.

16. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI *et al*. Pathways to zoonotic spillover. *Nat Rev Microbiol* 2017;15:502–510.

17. Chuma IS, Roos C, Atickem A, Bohm T, Anthony Collins D *et al*. Strain diversity of *Treponema pallidum* subsp. *pertenue* suggests rare interspecies transmission in African nonhuman primates. *Sci Rep* 2019;9:14243.

18. Gogarten JF, Düx A, Schuenemann VJ, Nowak K, Boesch C *et al*. Tools for opening new chapters in the book of *Treponema pallidum* evolutionary history. *Clin Microbiol Infect* 2016;22:916–921.

19. Calvignac-Spencer S, Merkel K, Kutzner N, Kühl H, Boesch C *et al*. Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Mol Ecol* 2013;22:915–924.

20. Hoffmann C, Zimmermann F, Biek R, Kuehl H, Nowak K *et al*. Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. *Nature* 2017;548:82–86.

21. Rohland N, Hofreiter M. Ancient DNA extraction from bones and teeth. *Nat Protoc* 2007;2:1756–1762.

22. Gamba C, Hanghøj K, Gaunitz C, Alfarhan AH, Alquraishi SA *et al*. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol Ecol Resour* 2016;16:459-69.

23. Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M *et al*. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Res* 2011;21:1705–1719.

24. Leslie DE, Azzato F, Karapanagiotidis T, Leydon J, Fyfe J. Development of a real-time PCR assay to detect *Treponema pallidum* in clinical specimens and assessment of the assay's performance by comparison with serological testing. *J Clin Microbiol* 2007;45:93–96.

25. Harper KN, Fyumagwa RD, Hoare R, Wambura PN, Coppenhaver DH *et al*. *Treponema pallidum* infection in the wild baboons of East Africa: distribution and genetic characterization of the strains responsible. *PLoS One* 2012;7:e50882.

26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.

27. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 2014;30:2114–2120.

28. Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C *et al*. EAGER: efficient ancient genome reconstruction. *Genome Biol* 2016;17:60.

29. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* 2013:arXiv:1303.3997.

30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.

31. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M *et al*. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;28:1647–1649.

32. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.

33. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;56:564–577.

34. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010;27:221–224.

35. Lefort V, Longueville J-E, Gascuel O. SMS: smart model selection in PhyML. *Mol Biol Evol* 2017;34:2422–2424.

36. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.

37. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35:4453–4455.

38. Leaché AD, Banbury BL, Felsenstein J, de Oca AN-M, Stamatakis A. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Syst Biol* 2015;64:1032–1047.

39. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol* 2001;50:913–925.

40. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2016;2:vew007.

41. Fourment M, Gibbs MJ. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol Biol* 2006;6:1.

42. Arora N, Schuenemann VJ, Jäger G, Peltzer A, Seitz A *et al*. Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. *Nat Microbiol* 2016;2:16245.

43. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol* 2018;67:901–904.

44. Drummond AJ, Rambaut A. Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7:214.

45. Letunic I, Bork P. Interactive tree of life (iTOL) V4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.

46. Barbera P, Kozlov AM, Czech L, Morel B, Darriba D *et al*. EPA-ng: massively parallel evolutionary placement of genetic sequences. *Syst Biol* 2019;68:365–369.

47. Berger SA, Stamatakis A. Aligning short reads to reference alignments and trees. *Bioinformatics* 2011;27:2068–2075.

48. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35:4453–4455.

49. Czech L, Barbera P, Stamatakis A. Genesis and Gappa: library and toolkit for working with phylogenetic (placement) data. *bioRxiv* 2019;647958.

50. Stamm LV, Bergen HL. A point mutation associated with bacterial macrolide resistance is present in both 23S rRNA genes of an erythromycin-resistant *Treponema pallidum* clinical isolate. *Antimicrob Agents Chemother* 2000;44:806–807.

51. Bergsten J. A review of long-branch attraction. *Cladistics* 2005;21:163–193.

52. Knauf S, Batamuzi EK, Mlengeya T, Kilewo M, Lejora IAV *et al*. *Treponema* infection associated with genital ulceration in wild baboons. *Vet Pathol* 2012;49:292–303.

53. Baylet R, Thivolet J, Sepetjian M, Nouhouay Y, Baylet M. [Natural open treponematosis in the *Papio papio* baboon in Casamance]. *Bull Soc Pathol Exot Filiales* 1971;64:842–846

54. Fribourg-Blanc A, Mollaret HH. Natural treponematosis of the African primate. *Primates Med* 1969;3:113–121

55. Mubemba B, Chanove E, Mätz-Rensing K, Gogarten JF, Düx A *et al*. Yaws disease caused by *Treponema pallidum* subspecies *pertenue* in wild chimpanzee, guinea, 2019. *Emerg Infect Dis* 2020;26:1283–1286.

56. Davies TJ, Pedersen AB. Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proc Biol Sci* 2008;275:1695–1701.

57. Pedersen AB, Davies TJ. Cross-species pathogen transmission and disease emergence in primates. *Ecohealth* 2009;6:496–508.

58. Strouhal M, Mikalová L, Havlíčková P, Tenti P, Čejková D *et al*. Complete genome sequences of two strains of *Treponema pallidum* subsp. *pertenue* from Ghana, Africa: Identical genome sequences in samples isolated more than 7 years apart. *PLoS Negl Trop Dis* 2017;11:e0005894.

59. Boesch C, Boesch H. Hunting behavior of wild chimpanzees in the Taï National Park. *Am J Phys Anthropol* 1989;78:547–573.

60. Gogarten JF, Akoua-Koffi C, Calvignac-Spencer S, Leendertz SAJ, Weiss S *et al*. The ecology of primate retroviruses – an assessment of 12 years of retroviral studies in the Taï National Park area, Côte d'Ivoire. *Virology* 2014:460–461.

61. Kumm HW, Turner TB. The transmission of yaws from man to rabbits by an insect vector, *Hippelates pallipes* Loew 1. *Am J Trop Med Hyg* 1936;s1-16:245–271.

62. Satchell GH, Harrison RA. Experimental observations on the possibility of transmission of yaws by wound-feeding Diptera, in Western Samoa. *Trans R Soc Trop Med Hyg* 1953;47:148–153.

63. Knauf S, Raphael J, Mitjà O, Lejora IAV, Chuma IS *et al*. Isolation of *Treponema* DNA from Necrophagous flies in a natural ecosystem. *EBioMedicine* 2016;11:85–90.

64. Gogarten JF, Düx A, Mubemba B, Pléh K, Hoffmann C *et al*. Tropical rainforest flies carrying pathogens form stable associations with social nonhuman primates. *Mol Ecol* 2019;28:4242–4258.

65. Stamm LV. Flies and yaws: molecular studies provide new insight. *EBioMedicine* 2016;11:9–10.

66. Fabricius T, Winther C, Ewertsen C, Kemp M, Nielsen SD. Osteitis in the dens of axis caused by *Treponema pallidum*. *BMC Infect Dis* 2013;13:347.

67. Majander K, Pfrengle S, Neukamm J, Kocher A, du PL *et al*. Ancient bacterial genomes reveal a formerly unknown diversity of *Treponema pallidum* strains in early modern Europe. *bioRxiv* 2020.

68. Schuenemann VJ, Kumar Lankapalli A, Barquera R, Nelson EA, Iraíz Hernández D *et al*. Historic *Treponema pallidum* genomes from Colonial Mexico retrieved from archaeological remains. *PLoS Negl Trop Dis* 2018;12:e0006447.

69. Barquera R, Lamnidis TC, Lankapalli AK, Kocher A, Hernández-Zaragoza DI *et al*. Origin and health status of first-generation Africans from early colonial Mexico. *Current Biology* 2020;30:2078-2091.e11:2078–2091.