Clinical Microbiology and Infection 22 (2016) 916–921

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Review

# Tools for opening new chapters in the book of *Treponema pallidum* evolutionary history

J.F. Gogarten [1, 3, 4], A. Düx [1, 2], V.J. Schuenemann [5], K. Nowak [1], C. Boesch [3], R.M. Wittig [3, 6], J. Krause [5, 7], S. Calvignac-Spencer [1, 2, *, 8], F.H. Leendertz [1, *, 8]

[1] *Epidemiology of Highly Pathogenic Microorganisms, Germany*
[2] *Viral Evolution, Robert Koch Institute, Berlin, Germany*
[3] *Primatology Department, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*
[4] *Department of Biology, McGill University, Montreal, Quebec, Canada*
[5] *Institute for Archeological Sciences, University of Tübingen, Tübingen, Germany*
[6] *Taï Chimpanzee Project, CSRS, Abidjan, Cote d'Ivoire*
[7] *Max Planck Institute for the Science of Human History, Jena, Germany*

ARTICLE INFO

*Article history:*
Available online 4 August 2016

Editor: M. Drancourt

*Keywords:*
Ancient DNA
Nonhuman primate reservoirs
Syphilis
*Treponema pallidum*
Yaws

ABSTRACT

*Treponema pallidum* infections causing yaws disease and venereal syphilis are globally widespread in human populations, infecting hundreds of thousands and millions annually respectively; endemic syphilis is much less common, and pinta has not been observed in decades. We discuss controversy surrounding the origin, evolution and history of these pathogens in light of available molecular and anthropological evidence. These bacteria (or close relatives) seem to affect many wild African nonhuman primate (NHP) species, though to date only a single NHP *Treponema pallidum* genome has been published, hindering detection of spillover events and our understanding of potential wildlife reservoirs. Similarly, only ten genomes of *Treponema pallidum* infecting humans have been published, impeding a full understanding of their diversity and evolutionary history. Research efforts have been hampered by the difficulty of culturing and propagating *Treponema pallidum*. Here we highlight avenues of research recently opened by the coupling of hybridization capture and next-generation sequencing. We present data generated with such an approach suggesting that asymptomatic bones from NHP occasionally contain enough treponemal DNA to recover large fractions of their genomes. We expect that these methods, which naturally can be applied to modern biopsy samples and ancient human bones, will soon considerably improve our understanding of these enigmatic pathogens and lay rest to old yet unresolved controversies. **J.F. Gogarten, CMI 2016;22:916**


## The Genus *Treponema*

Organisms in the genus *Treponema* (phylum *Spirochaetes*, order *Spirochetales*, family *Spirochaetaceae*) are obligate parasites distributed across a broad range of animal hosts, though a few basal species may represent free-living organisms [1]. Treponemes have been detected across many of their host's body compartments, including the oral cavity [2], hindgut [3], skin, cartilage and bone [1]. *Treponema* species can be pathogenic or nonpathogenic, with suggestions they may play a symbiotic role in some hosts. For example *Treponema* species found in termite guts perform $H_2$-$CO_2$ acetogenesis and nitrogen fixation, processes releasing carbon and energy or providing nitrogen for their host [4]. Many *Treponema* species are abundant in gut microbiomes of humans and nonhuman primates (NHP), though their functional role in these communities remains unknown [3,5]. The *Treponema* genus is however best known for its pathogenic members, which are responsible for a large current and historic human global disease burden.

* Corresponding authors. S. Calvignac-Spencer and F.H. Leendertz, P3—Epidemiology of Highly Pathogenic Microorganisms Robert Koch Institute, Seestraße 10, Berlin 13353, Germany.
*E-mail addresses:* CalvignacS@rki.de (S. Calvignac-Spencer), LeendertzF@rki.de (F.H. Leendertz).
[8] The last two authors contributed equally to this article, and both should be considered senior author.

http://dx.doi.org/10.1016/j.cmi.2016.07.027
1198-743X/

## Pathogenic *Treponema pallidum*

Subspecies of the spirochete bacterium *T. pallidum* are responsible for yaws (*T. p. pertenue*), bejel or endemic syphilis (*T. p. endemicum*) and venereal syphilis (*T. p. pallidum*). To date, the causative agent of pinta (currently classified as *T. carateum*) has not been cultured or isolated; this has precluded a determination of whether it represents a further member of the *T. pallidum* species or is divergent enough to warrant designation as its own species [6]. Pinta disease is not discussed here in detail.

The clinical presentations associated with treponemal diseases share similarities but are still distinctive. Yaws is characterized by a primary cutaneous lesion, most often presenting on lower extremities, with secondary lesions developing across the body [6]. In contrast, primary lesions of endemic syphilis infections are rarely observed, and when they are observed, they are typically found in oral mucosa [6]. Similarly, primary lesions of venereal syphilis are typically present on the genital, anal or oral mucosae. All *T. pallidum* infections can become latent, hampering treatment efforts. Left untreated, endemic syphilis and yaws can lead to destructive osteitis of the nose, palate and nasal septum, with yaws often causing lesions on the feet [6]. In contrast, late-stage venereal syphilis is more systemic, causing major neurologic and cardiovascular problems and growth of granuloma on many organs [1]. Venereal syphilis can cause adverse outcomes for pregnant women including stillbirth, early foetal death, low birth weight, preterm delivery, neonatal death and infection of the newborn [7]. It was long thought that only venereal syphilis is involved with the central nervous system, though some evidence suggests this also occurs rarely with yaws; yaws may also be transmitted vertically in some rare cases [6,8]. All *T. pallidum* are predominantly transmitted by direct contact with infectious lesions. Yaws is transmitted mostly by skin-to-skin contact, while endemic syphilis is also transmitted by contact with mucous membranes [6]. In contrast, venereal syphilis is predominantly transmitted by sexual contact, though nonsexual transmission has been described [9] and vertical transmission is a major concern [10]. Consistent with their different transmission modes, yaws and endemic syphilis predominantly infect children 2 to 15 years of age, while venereal syphilis infects mainly adults and infants [6].

In 2009 there were 10.6 million new cases of venereal syphilis in adults, with more than 36.4 million adults thought to be infected [7]; estimates for 2008 suggested more than half a million births were affected despite massive antenatal care efforts [10]. While venereal syphilis is globally distributed, its prevalence varies by region, peaking in sub-Saharan Africa [7]. Endemic syphilis is less well documented, though several cases have been reported across several countries in Africa and the Middle East [6], most recently in Iran in 2010 [11]. Yaws has been the target of major past and present eradication campaigns, which has reduced its confirmed distribution to 12 countries in Africa, Asia and the South Pacific, where it still caused more than 300 000 new cases between 2008 and 2012 [12]. Poor documentation and detection of latent infections combined with underreporting and misdiagnosis suggest that the magnitude of the disease burden for endemic syphilis and yaws is vastly underestimated. Though symptoms and progression of these infections varies, debate remains as to whether these are due to the pathogens themselves or are rather caused by differences between hosts, environmental variables or mode of infection [13,14].

## NHP Infections

NHP are also susceptible to *T. pallidum* infections, and descriptions of symptoms [15–17] combined with serologic and morphologic evidence suggest both syphilis- and yawslike infections occur in a number of wild NHP populations across Africa [16,18,19]. The presence of *T. pallidum* in inflamed tissues has however only been demonstrated in baboons [16,17,20,21]; conclusive evidence of infections in other species is still lacking. The recognition of *T. pallidum* infections in NHP led to the hypothesis that human treponematoses are zoonotic in origin [19], but whether human *T. pallidum* is the result of a single transmission event or of continuous transmission, or whether these treponemes codiverged with their primate hosts is a source of debate [22,23]. It has been suggested that eradication initiatives for yaws might be hampered by continued spillover from a NHP reservoir [19], though evidence confirming transmission events, or even cross-NHP species transmission events in the wild, are lacking.

## Molecular Diagnosis and Phylogenetic Relationships Among Subspecies

Such conclusive evidence of zoonotic transmission is lacking in large part because, despite major advances in culture techniques, *T. pallidum* remains one of the last as-yet uncultured human pathogens [24]. The three recognized subspecies are morphologically indistinguishable and antigenically cross-reactive [6]. Subspecies delineations in humans are therefore nearly always based on clinical and epidemiologic data rather than distinguishing molecular evidence [14]. Most molecular diagnosis tools are PCR based, but because *T. pallidum* is slowly evolving, the short regions amplified with these approaches fail to capture sufficient information to determine evolutionary history. A number of single nucleotide polymorphisms distinguishing subspecies have been proposed for diagnostics and untangling the evolutionary history of this species [23], though the utility and validity of this approach has been questioned [14]. The accumulation of sequences has cast doubt on a number of formerly recognized diagnostic single nucleotide polymorphisms which may be explained by mere sampling artefacts and/or recombination events [14].

Full genomes allow a resolution of evolutionary relationships that for the most part distinguishes subspecies in much the same manner as delineations based on symptoms, though misclassification has occurred when symptoms were used alone [25] (Fig. 1). From extant human strains, molecular differences across genomes were described for only a few cases; ten full genomes are published and these are >99.6% identical [26]. The whole genome of a treponeme infecting West African baboons (*Papio cynocephalus*) isolated in 1966 was recently sequenced; phylogenetic analysis suggested it is extremely similar to human *T. p. pertenue,* though more genomes of treponemes infecting NHP are required to test the hypothesis that humans acquired this infection from NHP or *vice versa* [27]. The diversity of nonpathogenic treponemes in humans and wildlife is even less well described, though preliminary insights suggest it is high [3–5], complicating efforts to design pathogen-specific primers and presenting challenges for the hybridization capture approaches described below. This is particularly true for noninvasive sampling of NHP populations using faeces, where treponeme diversity is high [3].

## Origins and History of *T. pallidum* in Human Populations

Molecular tools and the availability of full genome data promises to shed light on the origins and evolutionary history of *T. pallidum,* though many questions remain unresolved with the currently available data. The origins and spread of venereal syphilis particularly has been the focus of much of the debate and controversy. The origins and spread of yaws and endemic syphilis are equally enigmatic but have mostly been discussed in relation to the

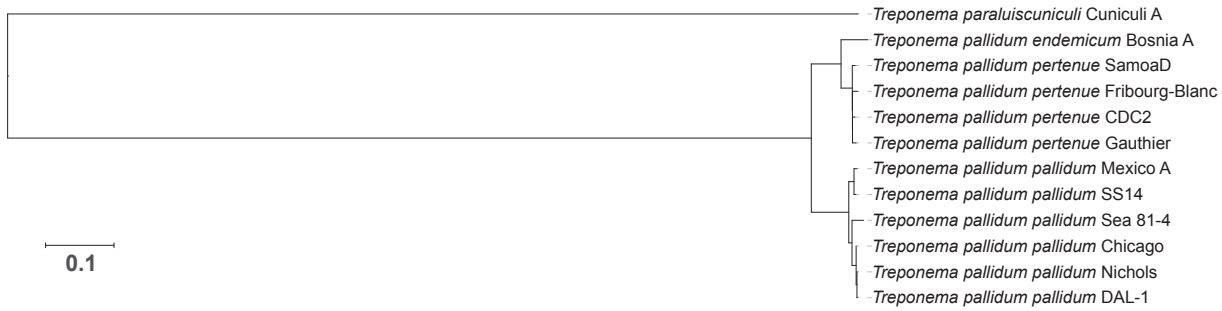


**Fig. 1.** Maximum likelihood phylogeny of *Treponema pallidum* subspecies and closely related *Treponema paraluiscuniculi* generated from full genome sequences. Highly variable *TPR* genes were removed before phylogenetic analysis, as these are under strong positive selection and often recombine, suggesting that these genes may be inappropriate for inferring phylogenetic history. Scale is in substitutions per variable site. Support values were calculated using Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-like aLRT), and all branches received values above 0.9, with the exception of short branches within *Treponema pallidum pertenue* clade between CDC2 and Gauthier genomes and between these two genomes and *T. p. pertenue* Fribourg-Blanc, which received values of 0.422 and 0.098 respectively. Details on how phylogeny was generated are provided in the Supplementary Materials.

question of the origin of venereal syphilis. More specifically, controversy has surrounded the origin of venereal syphilis in European populations; the disease seemingly appeared at the end of the 15th century and quickly turned into a major epidemic that swept across the continent. Contemporaries began questioning the origins of the disease, and the issue remains largely unresolved. Three main hypotheses have been proposed.

The Columbian hypothesis posits that when Columbus and his crew returned from the New World in 1493, they brought with them not only tobacco and corn but also a new infectious agent. Proponents of this theory recognize widespread syphilitic infections across a large temporal and spatial scale in the New World and argue that nothing comparable can be found in the European archaeological record before Columbus's return. They also consider the very rapid rise of syphilis as an indication that a novel infectious disease was spreading in a naive population [28].

Others have hypothesized that syphilis was present in the Old World long before Columbus's return and that it became more virulent around the time of Columbus or increased in prevalence and spread as a result of other social and geopolitical factors [29]. Proponents of the pre-Columbian hypothesis suggest that the progression of the disease makes it unlikely that the crew would have been exhibiting infective lesions after their return to Europe and that the time interval between the sailors return from the New World and the start of a widespread outbreak across Europe is unrealistic; these proponents draw on historic documents suggesting syphilis was present in Europe before Columbus's return [29]. Archaeological remains have been used to argue that pre-Columbian Old World skeletons show evidence of syphilitic infections [30] and that New World pre-Columbian skeletons show evidence of lesions present in young individuals, which might be more consistent with yawslike infections than venereal syphilis [31]. Much of the debate surrounds whether bone lesions in different archaeological records in the Old and New World represent *T. pallidum,* and if so, which subspecies.

Another hypothesis presented is that yaws, venereal syphilis and endemic syphilis actually represent the same pathogen, and that environmental and social conditions determine the outcome of the infection [28]; proponents of this hypothesis draw on the low diversity and difficulties in identifying genetic differences between these subspecies. The full genome evidence discussed above suggests genetic differences do exist between these different pathogens, though small sample sizes might have missed a continuum of diversity in this species, and species concepts are notoriously difficult to apply to bacterial lineages.

## Ancient Pathogen DNA

Nucleic acids from archaeological and palaeontologic remains have proven a powerful tool for examining the phylogenetic relationships between historic and modern organisms and have the potential to resolve the aforementioned controversies [32–34]. While the treasure trove of information in ancient remains is appealing, the minute amounts of degraded genetic material calls for caution; contaminations have literally plagued the ancient DNA (aDNA) field. Great care is therefore required to ensure the replicability and authenticity of any findings; guidelines have been proposed that include the need for replication by independent labs and the use of clean rooms [33]. The fragmented nature of aDNA means that only short reads can be generated with each PCR, precluding in-depth phylogenetic analyses for slowly evolving pathogens such as *T. pallidum.* The rise of next-generation sequencing (NGS) technologies provided a first means to generate large amounts of data from small amounts of starting material, circumventing many limitations of PCR-based approaches. Interestingly, data from NGS can be leveraged to validate their own authenticity, e.g. by evidencing damage motifs characteristic of ancient samples [35–37]. Ancient samples are metagenomic, containing a mix of host and environmental DNA and often only a low percentage of endogenous pathogen DNA. In some rare cases, shotgun NGS approaches have been powerful enough to generate complete or close-to-complete ancient bacterial genomes (e.g. *Tannerella forsythia* [38] and *Mycobacterium leprae* [34]), though the combination of NGS with hybridization capture was the fundamental technical leap that revolutionized this field of microbiology. Enrichment experiments have thus far succeeded in shedding light on the history and evolution of a number of human bacterial pathogens including *Mycobacterium tuberculosis* [39], *Yersinia pestis* [40], *Vibrio cholerae* [41] and *Helicobacter pylori* [42]. These enrichment approaches also allowed researchers to generate sequence information for the pathogen causing leprosy (*Mycobacterium leprae*) from both ancient samples and modern samples obtained from patient biopsy samples [34]. As is the case for *T. pallidum,* propagation by culture is difficult, which has prohibited large-scale sequencing of this pathogen; NGS approaches coupled with hybridization capture have greatly expanded our understanding of modern *M. leprae* diversity and how this relates to ancient infections [34].

Debate surrounding interpretation of archaeological evidence for *T. pallidum* infections would be clarified considerably if genome sequence information became available, particularly from pre-Columbian samples from the Old and New World. A short

treponeme PCR fragment from a 200-year-old mummy [43] and from post-Columbian foetuses in Europe [44] suggested aDNA approaches might prove effective, though in addition to being too modern to help inform the debate, the short fragments amplified precluded informative phylogenetic analyses. Many researchers have subsequently tried to use PCR-based approaches to study ancient samples with evidence of *T. pallidum* lesions, with very limited success, which has lead many to question the feasibility of using bones to study ancient *T. pallidum* infections [45]. This may partially be a product of the progression of venereal syphilis; in modern cases, the highest pathogen load is usually found in stage 1 when no bone lesions have developed; in adults bone deformations usually occur in stage 3 when it is difficult to detect the pathogen itself [44]. The availability of hybridization capture coupled with NGS clearly represents a promising way forward; this approach should outperform PCR-based approaches for screening samples with low concentrations and highly fragmented pathogen DNA, ultimately paving the way for the generation of genomewide information from both ancient and modern samples.

## Demonstrating Feasibility of Hybridization Capture Enrichment for Bone Samples

We assessed the feasibility of such an approach using a set of samples that intuitively seemed unpromising: asymptomatic bones from NHP. The ability to recover treponemal DNA from bones, and more particularly from diagnosed bones, has been a matter of controversy [45]. On the other hand, as mentioned above, many NHP populations present clinical manifestations suggestive of treponemal infections, in some groups even at very high prevalence [19].

For this study, we generated DNA extracts from contemporary nonlesioned NHP bones ($n = 51$, six species from Taï National Park, Côte d'Ivoire) and first screened them with three independent PCR systems specific to *T. pallidum* (Supplementary Materials). No bone was positive for all three PCR systems (12 tested positive for the shortest *PolA* sequence, and three were also positive for the longer *GDP* or *cfpA* fragments), suggesting *T. pallidum* DNA is highly fragmented or at low concentrations. On the basis of this screening, we selected three candidate extracts (i.e. those also positive for a longer fragment; Table 1) for which we conducted library preparation, enrichment and sequencing in two separate laboratories using standard procedures to avoid contamination (Supplementary Fig. 1). At the University of Tübingen we used a DNA microarray-based approach, with probes spanning the *T. p. pallidum* genome, and postcapture pathogen-enriched DNA was sequenced on an Illumina HiSeq 2500 [46] (Illumina, San Diego, CA, USA). At the Robert Koch Institute, we used an in-solution capture approach, with baits spanning the *T. p. pertenue* Fribourg-Blanc genome and postcapture pathogen-enriched DNA was sequenced on an Illumina MiSeq. Hybridization capture approaches have been used to enrich DNA up to 42% dissimilar to baits [47]; *T. pallidum* is characterized by extremely low diversity, suggesting these different bait sets would capture DNA from any subspecies of *T. pallidum.* Reads generated at both institutions were trimmed using Trimmomatic, mapped to *T. p. pertenue* Fribourg-Blanc using BWA-MEM [48] and deduplicated using Picard's MarkDuplicates. To ensure reads were not contributed by nonpathogenic treponemes we filtered reads; each mapped read was BLASTed against a local database of treponeme genomes, and only reads which were a hit to every published *T. pallidum* genome and where the lowest bit score for a *T. pallidum* genomes was greater than the highest bit score for published non–*T. pallidum* treponeme genomes were kept. For comparing bit scores we used R version 3.2.3 [49] with the package 'data.table' (https://cran.r-project.org/web/packages/data.table/data.table.pdf). Surviving reads were mapped to a closely related outgroup of *T. pallidum* that infects rabbits, *T. paraluiscuniculi* [50], along with *T. p. pallidum* and *T. p. endemicum* that infect humans.

Both the array-based and in-solution-based approaches generated sequences spanning the *T. p. pertenue* Fribourg-Blanc genome (Table 1 and Fig. 2), though coverage was low for the two chimpanzee samples (range, 0.036–1.8%) and moderate for the red colobus bone (7.2 and 13.9%; Table 1). Both laboratories, using unique indices and distinct library preparation methods and capture protocols, converged on similar findings for each of the bones. Combining data from the approaches resulted in 19.8% genome coverage from the red colobus bone (99.5% identical sites) and for the best chimpanzee sample 3.4% coverage (98.9% identical sites). Sequence similarity to the most closely related outgroup, *T. paraluiscuniculi,* was lower than for all three *T. pallidum* subspecies, suggesting chimpanzees and red colobus are infected with *T. pallidum* in the wild. That this approach worked for nonlesioned bones suggests it should work efficiently for tissue samples or swabs from humans or wildlife, providing a cost-effective, culture-free means of generating whole-genome data. It also suggests aDNA studies might benefit from screening nonlesioned remains. The amount of sequence information generated represents a substantial increase compared to what has been feasible with PCR-based analyses, particularly from bone samples [43,44]. Higher coverage could still be achieved by intensifying the sampling of DNA fragments, either through deeper sequencing of the same libraries and/or the generation and sequencing of further libraries. Capture experiments need not target entire genomes of *T. pallidum,* and other bait designs, e.g. targeting unique but variable regions, might increase the power of such an approach for screening samples; positive samples could subsequently be enriched with a

**Table 1**
Screening results of bone samples from Taï National Park, Côte d'Ivoire, selected for *Treponema pallidum* enrichment[a]

| Species | Year of death | *PolA* seq. | *GDP* seq. | *cfpA* seq. | Array capture approach: genome coverage (%) (% identical sites) | | | | In-solution capture: genome coverage (%) (% identical sites) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *T. paraluiscuniculi* | *T. p. endemiccum* | *T. p. pallidum* | *T. p. pertenue FB* | *T. paraluiscuniculi* | *T. p. endemiccum* | *T. p. pallidum* | *T. p. pertenue FB* |
| *Procolobus badius* | 2002 | + | + | – | 13.4 (97.8) | 13.8 (99.5) | 13.8 (99.4) | 13.9 (99.5) | 7.2 (98.6) | 7.2 (99.6) | 7.2 (99.5) | 7.2 (99.6) |
| *P. t. verus* | 1992 | + | – | + | 0.036 (97.3) | 0.036 (97.5) | 0.036 (97.5) | 0.036 (97.6) | 0.2 (95.6) | 0.2 (96.3) | 0.2 (96.3) | 0.2 (96.3) |
| *P. t. verus* | 1998 | + | + | – | 1.7 (97.5) | 1.7 (98.9) | 1.7 (98.9) | 1.8 (99.0) | 1.6 (98.1) | 1.7 (98.7) | 1.6 (98.6) | 1.7 (98.6) |

*P. t. verus, Pan troglodytes verus; FB, Treponema pallidum pertenue* Fribourg-Blanc (NC_021179) isolated from baboon (European Nucleotide Archive (ENA) study accession number PRJEB13855; http://www.ebi.ac.uk/ena/data/view/PRJEB13855).
+, Sanger sequence generated; –, no sequence generated.
[a] Sample accession numbers for raw data generated at Tübingen University using a microarray-based approach: *P. badius,* 2002 = ERS1138345; *P. t. verus,* 1992 = ERS1138347; *P. t. verus,* 1998 = ERS1138349. Sample accession numbers for raw data generated at the Robert Koch Institute using in-solution-based capture approach: *P. badius,* 2002 = ERS1138344; *P. t. verus,* 1992 = ERS1138346; *P. t. verus,* 1998 = ERS1138348 and ERS1138350 (large number of reads split into two sets of paired reads). The following reference genomes were used for mapping and comparison of percentage identical sites: *T. paraluiscuniculi* (NC_015714), *T. p. endemiccum* (NZ_CP007548), *T. p. pallidum* (NC_021490) and *T. p. pertenue FB* (NC_021179).

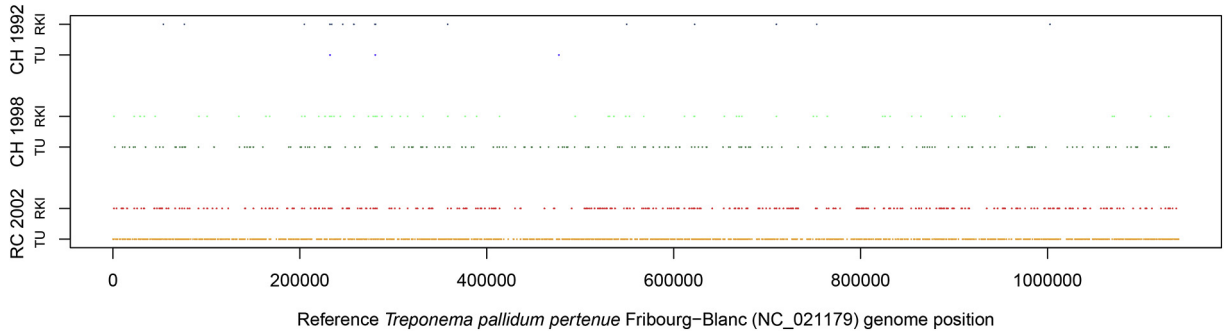


**Fig. 2.** Coverage of *Treponema pallidum pertenue* Fribourg-Blanc genome. Each point represents coverage at that location on *T. p. pertenue* Fribourg-Blanc genome for red colobus bone (RC) and two chimpanzee bones (CH) using two different enrichment approaches at Robert Koch Institute (RKI) and Tübingen University (TU). Coverage appears randomly distributed across genome from both approaches and from three samples.

genomewide approach to enable rigorous phylogenetic analysis. In any case, these experiments demonstrate the feasibility of capture enrichment in combination with NGS for studying the enigmatic pathogen, *T. pallidum*.

While field reports have suggested *T. pallidum* infections might be occurring in NHP species beyond baboons [19], our results molecularly confirm these observations and expand the NHP reservoir to include our closest NHP relatives, chimpanzees. The strongest evidence for *T. pallidum* infection came from the preferred prey species of these chimpanzees, red colobus; transmission of microorganisms in this hunter–prey relationship has been documented [51], though further studies are needed to understand the ecology and between-species transmission of *T. pallidum* in wildlife communities and to test the hypothesis of a NHP reservoir that is continuously spilling over into human populations.

## Conclusion

Despite the availability of effective antibiotic treatments, *T. pallidum* infections causing venereal syphilis and yaws are still globally widespread, while endemic syphilis continues to reappear despite widespread eradication efforts. The paucity of genome data from *T. pallidum* infecting humans or wildlife has hampered a deep understanding of the ecology and evolution of this pathogen. This lack of genomic data is notably explained by the difficulty of culturing *T. pallidum*. We show the feasibility of *T. pallidum* DNA enrichment through hybridization capture and demonstrate that in combination with NGS technologies these approaches sometimes allow for the recovery of substantial parts of treponemal genomes. These approaches should be applicable to biopsy samples from symptomatic NHP and humans and from ancient bone specimens of humans and NHP. We expect they will provide a major contribution to our understanding of these enigmatic pathogens and help resolve long-standing controversy surrounding their ecology and evolution.

## Acknowledgements

For assistance and support with sequencing at the RKI we are grateful to the Central Sequencing Laboratory, particularly J. Hinzmann, A. Nitsche and J. Tesch. For assistance with the processing of MiSeq reads and help accessing computing power at the RKI we would like to thank W. Dabrowski. For assistance with processing and accessing the Tübingen HiSeq reads we thank K. Nieselt. For the microarray design of the *T. p. pallidum* array we thank K. Bos. For assistance in implementing the bioinformatics approach for filtering the reads generated in this project and generating the phylogeny from published genomes we thank P. Gogarten. For administrative assistance we are grateful for the help of M. Kovacev-Wegener, and for assistance in the laboratory at the RKI we thank U. Thiesen, K. Merkel and A. Sachse, along with the entire P3 Project Group. For comments on an earlier version of this article, we thank A. Goedmakers.

## Transparency Declaration

JFG was supported by an NSF Graduate Research Fellowship (DGE-1142336); the Canadian Institutes of Health Research's Strategic Training Initiative in Health Research's Systems Biology Training Program; an NSERC Vanier Canada Graduate Scholarship (CGS); the Quebec Centre for Biodiversity Science's (Centre de la Science de la Biodiversité du Québec) Excellence Awards Program; McGill University, particularly through Research Travel Awards, the Graduate Research Mobility Award Program, the Dean of Science's Research Fund and the Excellence Awards Program; and a long-term Research Grant from the German Academic Exchange Service (DAAD-91525837-57048249). The bones tested in this study stem from Taï National Park, Côte d'Ivoire and were collected through the long-term research conducted with permission from the Ministère de la Recherche Scientifique, the Ministère de l'Environnement et des Eaux et Forets, the Direction Générale de la Recherche Scientifique et de l'Innovation Technique and the Office Ivoirien des Parcs et Reserves; support from the Max Planck Society was integral for the long-term field presence of the Taï Chimpanzee Project, which enabled such an extensive collection and benefited from the assistance of all students and field assistants of the project. All authors report no conflicts of interest relevant to this article.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cmi.2016.07.027.

## References

[1] Norris SJ, Paster BJ, Moter A, Göbel UB. The genus *Treponema*. The prokaryotes. New York: Springer; 2006. p. 211–34.
[2] Griffen AL, Beall CJ, Campbell JH, Firestone ND, Kumar PS, Yang ZK, et al. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. ISME J 2012;6:1176–85.
[3] Bittar F, Keita MB, Lagier JC, Peeters M, Delaporte E, Raoult D. *Gorilla gorilla gorilla* gut: a potential reservoir of pathogenic bacteria as revealed using culturomics and molecular tools. Sci Rep 2014;4:1–5.
[4] Graber JR, Leadbetter JR, Breznak JA. Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. Appl Environ Microbiol 2004;70:1315–20.
[5] Schnorr SL, Candela M, Rampelli S, et al. Gut microbiome of the Hadza hunter-gatherers. Nat Commun 2014;5:3654.

[6] Giacani L, Lukehart SA. The endemic treponematoses. Clin Microbiol Rev 2014;27:89–115.
[7] World Health Organization. Global incidence and prevalence of selected curable sexually transmitted infections, 2008. Geneva: World Health Organization; 2012.
[8] Román GC, Román LN. Occurrence of congenital, cardiovascular, visceral, neurologic, and neuro-ophthalmologic complications in late yaws: a theme for future research. Rev Infect Dis 1986;8:760–70.
[9] Krivatkin S, Krivatkina E. Secondary syphilis transmitted through non-sexual contact. Three case descriptions. J Eur Acad Dermatol Venereol 1997;9:S224.
[10] Newman L, Kamb M, Hawkes S, et al. Global estimates of syphilis in pregnancy and associated adverse outcomes: analysis of multinational antenatal surveillance data. PLoS Med 2013;10:e1001396.
[11] Abdolrasouli A, Croucher A, Hemmati Y, Mabey D. A case of endemic syphilis, Iran. Emerging Infect Dis 2013;19:162.
[12] Kazadi WM, Asiedu KB, Agana N, Mitjà O, Guinea PN. Epidemiology of yaws: an update. Clin Epidemiol 2014;6:119–28.
[13] Thornburg RW, Baseman JB. Comparison of major protein antigens and protein profiles of *Treponema pallidum* and *Treponema pertenue*. Infect Immun 1983;42:623–7.
[14] Mulligan CJ, Norris SJ, Lukehart SA. Molecular studies in *Treponema pallidum* evolution: toward clarity? PLoS Negl Trop Dis 2008;2:e184.
[15] Levréro F, Gatti S, Gautier-Hion A, Ménard N. Yaws disease in a wild gorilla population and its impact on the reproductive status of males. Am J Phys Anthropol 2007;132:568–75.
[16] 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 1970;64:842–6.
[17] Knauf S, Batamuzi EK, Mlengeya T, et al. *Treponema* infection associated with genital ulceration in wild baboons. Vet Pathol 2012;49:292–303.
[18] Lovell N, Jurmain R, Kilgore L. Skeletal evidence of probable treponemal infection in free-ranging African apes. Primates 2000;41:275–90.
[19] Knauf S, Liu H, Harper KN. *Treponema*l infection in nonhuman primates as possible reservoir for human yaws. Emerging Infect Dis 2013;19:2058.
[20] Fribourg-Blanc A, Niel G, Mollaret H. Confirmation serologique et microscopique de la treponemose du cynocephale de Guinee. Bull Soc Pathol Exot Filiales 1966;59:54–9.
[21] Harper KN, Fyumagwa RD, Hoare R, 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.
[22] Hackett CJ. On the origin of the human treponematoses (pinta, yaws, endemic syphilis and venereal syphilis). Bull World Health Organ 1963;29:7–41.
[23] Harper KN, Ocampo PS, Steiner BM, et al. On the origin of the treponematoses: a phylogenetic approach. PLoS Negl Trop Dis 2008;2:e148.
[24] Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015;28:208–36.
[25] Cameron CE, Lukehart SA, Castro C, Molini B, Godornes C, Van Voorhis WC. Opsonic potential, protective capacity, and sequence conservation of the *Treponema pallidum* subspecies *pallidum* Tp92. J Infect Dis 2000;181:1401–13.
[26] Šmajs D, Norris SJ, Weinstock GM. Genetic diversity in *Treponema pallidum*: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. Infect Genet Evol 2012;12:191–202.
[27] Zobaníková M, Strouhal M, Mikalová 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.
[28] Hudson EH. Treponematosis and man's social evolution. Am Anthropol 1965;67:885–901.
[29] Luger A. The origin of syphilis: clinical and epidemiologic considerations on the Columbian theory. Sex Transmitted Dis 1993;20:110–7.
[30] von Hunnius TE, Roberts CA, Boylston A, Saunders SR. Histological identification of syphilis in pre-Columbian England. Am J Phys Anthropol 2006;129:559–66.
[31] Baker B, Armelagos G. The origin and antiquity of syphilis: paleopathological diagnosis and interpretation. Curr Anthropol 1988;29:703.
[32] Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. Ancient DNA. Nat Rev Genet 2001;2:353–9.
[33] Poinar HN, Cooper A. Ancient DNA: do it right or not at all. Science 2000;5482:1139.
[34] Schuenemann VJ, Singh P, Mendum TA, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science 2013;341:179–83.
[35] Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 2013:btt193.
[36] Stoneking M, Krause J. Learning about human population history from ancient and modern genomes. Nat Rev Genet 2011;12:603–14.
[37] Briggs AW, Stenzel U, Johnson PL, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A 2007;104:14616–21.
[38] Warinner C, Rodrigues JFM, Vyas R, et al. Pathogens and host immunity in the ancient human oral cavity. Nat Genet 2014;46:336–44.
[39] Bouwman AS, Kennedy SL, Müller R, et al. Genotype of a historic strain of *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 2012;109:18511–6.
[40] Bos KI, Schuenemann VJ, Golding GB, et al. A draft genome of *Yersinia pestis* from victims of the Black Death. Nature 2011;478:506–10.
[41] Devault AM, Golding GB, Waglechner N, et al. Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. N Engl J Med 2014;370:334–40.
[42] Maixner F, Krause-Kyora B, Turaev D, et al. The 5300-year-old *Helicobacter pylori* genome of the Iceman. Science 2016;351:162–5.
[43] Kolman CJ, Centurion-Lara A, Lukehart SA, Owsley DW, Tuross N. Identification of *Treponema pallidum* subspecies *pallidum* in a 200-year-old skeletal specimen. J Infect Dis 1999;180:2060–3.
[44] Montiel R, Solórzano E, Díaz N, et al. Neonate human remains: a window of opportunity to the molecular study of ancient syphilis. PLoS One 2012;7:e36371.
[45] von Hunnius TE, Yang D, Eng B, Waye JS, Saunders SR. Digging deeper into the limits of ancient DNA research on syphilis. J Archaeol Sci 2007;34:2091–100.
[46] Arora N, Schuenemann VJJ, Jäger G, et al. Origin of modern syphilis and emergence of a contemporary pandemic cluster. bioRxiv 2016. http://dx.doi.org/10.1101/051037 [in press].
[47] Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced virome sequencing using targeted sequence capture. Genome Res 2015;25:1910–20.
[48] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].
[49] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
[50] Šmajs D, Zobaníková M, Strouhal M, et al. Complete genome sequence of *Treponema paraluiscuniculi*, strain Cuniculi A: the loss of infectivity to humans is associated with genome decay. PLoS One 2011;6:e20415.
[51] Gogarten JF, Akoua-Koffi C, Calvignac-Spencer 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;460:147–53.