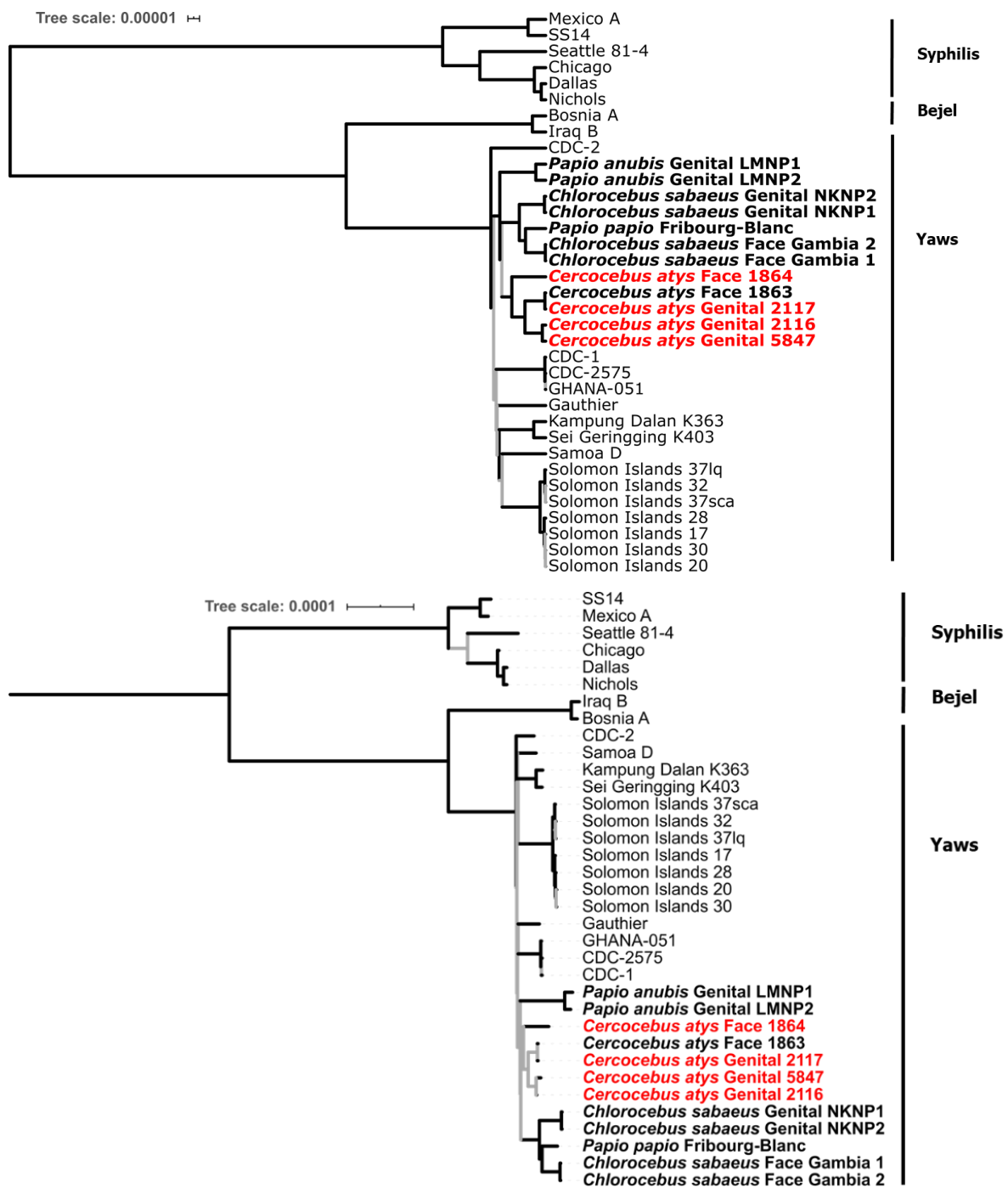
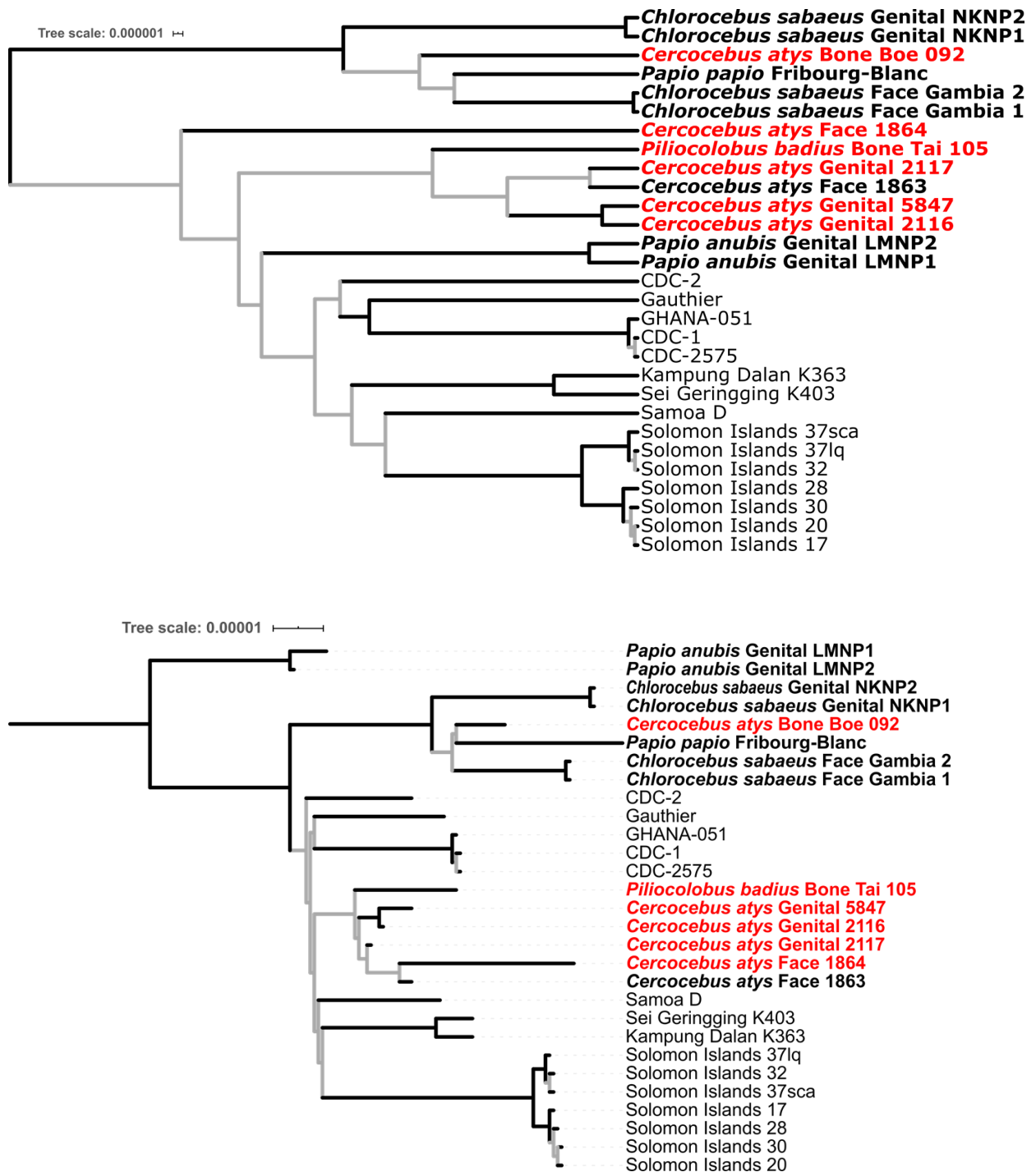


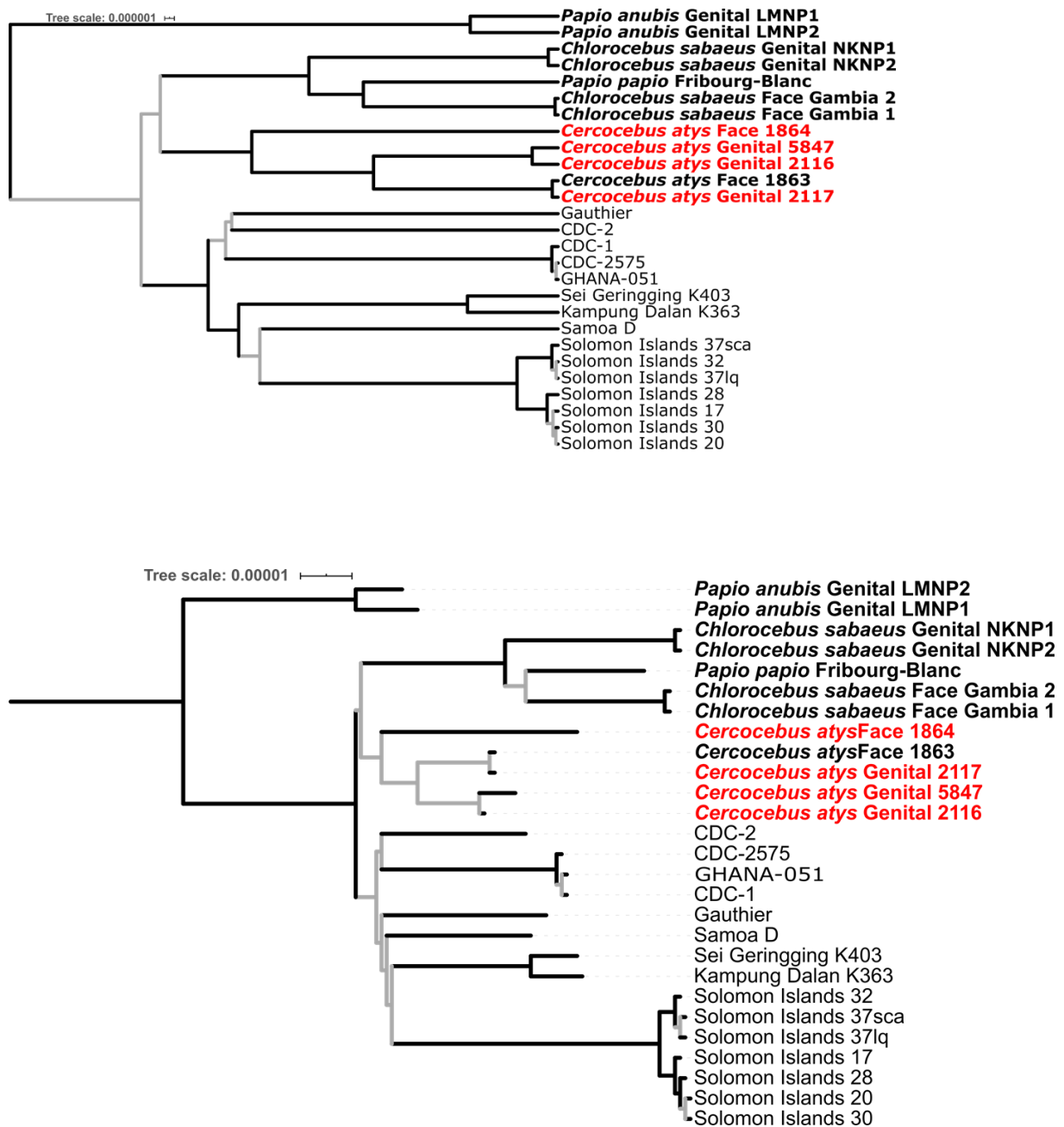
**Fig. S1. Maximum clade credibility (top: relaxed clock model assuming a birth-death process) and maximum likelihood (bottom) trees at 5X coverage and 95% threshold. All simian infecting strains are shown in bold with tip labels showing the species sampled, location of the lesion biopsied or swabbed and sample ID. Genomes generated in this study are shown in red. 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 gray. The scale shows nucleotide substitutions per site.**



**Fig. S2. Maximum clade credibility (top: relaxed clock model assuming a birth-death process) and maximum likelihood (bottom) trees at 10X coverage and 95% threshold.** All simian infecting strains are shown in bold with tip labels showing the species sampled, location of the lesion biopsied or swabbed and sample ID. Genomes generated in this study are shown in red. 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 indicated in gray. The scale shows nucleotide substitutions per site.



**Fig. S3. Maximum clade credibility (top: relaxed clock model assuming a birth-death process) and maximum likelihood (bottom) trees from ingroup analysis of TPE strains at 5X coverage and 95% threshold.** All simian infecting strains are shown in bold with tip labels showing the species sampled, location of the lesion biopsied or swabbed and sample ID. Genomes generated in this study are shown in red. 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 indicated in gray. The scale shows nucleotide substitutions per site.



**Fig. S4. Maximum clade credibility (top: relaxed clock model assuming a birth-death process) and maximum likelihood (bottom) trees from ingroup analysis of TPE strains at 10X coverage and 95% threshold.** All simian infecting strains are shown in bold with tip labels showing the species sampled, location of the lesion biopsied or swabbed and sample ID. Genomes generated in this study are shown in red. 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 indicated in gray. The scale shows nucleotide substitutions per site.

**Table S2: Primers used for the *Treponema pallidum* screening in this study**

Type of PCR	Primer pair 5'-3'	Product size (bp)	Reference
<i>polA</i> PCR	Forward_AGGATCCGGCATATGTCCAA Reverse_GTGAGCGTCTCATATTCCAAA	67	Leslie et al. 2007
Sequencing Hemi_PCR	M13-Foward_GTAAAACGACGGCCAG	105	
Sequencing Hemi_PCR	M13_Reverse_CAGGAAACAGCTATGAC	105	
Fusion PCR	Fusion_M13_Foward_GTAAAACGACGGCCAGAGGATCCGGCATATGTCCAA	105	
Fusion PCR	Fusion_M13_Reverse_CAGGAAACAGCTATGACGTGAGCGTCTCATATTCCAAA	105	
<i>cfpA</i> nested PCR	Forward_GAGTCCCAATGTGTTTCATCC Reverse_TAGGATGGCAATCTCCTTCG	352	Harper et al. 2012
	Forward_GAGCGTCTGGACGTAATGG Reverse_TAGGATGGCAATCTCCTTCG	189	
<i>polA</i> qPCR	Forward_AGGATCCGGCATATGTCCAA Reverse_GTGAGCGTCTCATATTCCAAA Probe_6FAM-ATGCACCAGCTTC+G+A	67	Leslie et al. 2007

**Table S3: Published *Treponema pallidum* genomes from humans and NHPs used in this study**

<b>Isolate ID</b>	<b>GenBank Accession ID</b>	<b>Host</b>	<b>TP spectrum</b>	<b>Country</b>	<b>Reference</b>
Bosnia A	CP007548.1	<i>Homo Sapiens</i>	Bejel	Bosnia	1
Iraq_B	CP032303.1	<i>Homo Sapiens</i>	Bejel	Iraq	2
Nichols	NC_021490.2	<i>Homo Sapiens</i>	Syphilis	USA	3
SS14	NC_021508.1	<i>Homo Sapiens</i>	Syphilis	USA	3
Chicago	NC_017268.1	<i>Homo Sapiens</i>	Syphilis	USA	4
Mexico A	NC_018722.1	<i>Homo Sapiens</i>	Syphilis	Mexico	3
Dallas	NC_016844.1	<i>Homo Sapiens</i>	Syphilis	USA	5
Seattle 81-4	CP003679.1	<i>Homo Sapiens</i>	Syphilis	USA	6
Fribourg-Blanc	NC_021179.1	<i>Papio papio</i>	Yaws	Guinea	7
Samoa D	NC_016842.1	<i>Homo Sapiens</i>	Yaws	Samoa	8
Gauthier	NC_016843.1	<i>Homo Sapiens</i>	Yaws	Republic of the Congo	8
CDC-1	NZ_CP024750.1	<i>Homo Sapiens</i>	Yaws	Ghana	9
CDC-2	NC_016848.1	<i>Homo Sapiens</i>	Yaws	Ghana	8
CDC_2575	NZ_CP020366.1	<i>Homo Sapiens</i>	Yaws	Ghana	10
Ghana-051	NZ_CP020365.1	<i>Homo Sapiens</i>	Yaws	Ghana	10
Kampung_Dalan_K363	CP024088.1	<i>Homo Sapiens</i>	Yaws	Indonesia	11
Sei_Geringging_K403	CP024089.1	<i>Homo Sapiens</i>	Yaws	Indonesia	11
Solomon Islands 03*	ERR1470343	<i>Homo Sapiens</i>	Yaws	Solomon Islands	12
Solomon Islands 17	ERR1470344	<i>Homo Sapiens</i>	Yaws	Solomon Islands	12
Solomon Islands 20	ERR1470335	<i>Homo Sapiens</i>	Yaws	Solomon Islands	12
Solomon Islands 28	ERR1470338	<i>Homo Sapiens</i>	Yaws	Solomon Islands	12
Solomon Islands 30	ERR1470334	<i>Homo Sapiens</i>	Yaws	Solomon Islands	12
Solomon Islands 32	ERR1470342	<i>Homo Sapiens</i>	Yaws	Solomon Islands	12
Solomon Islands 37 liq	ERR1470330	<i>Homo Sapiens</i>	Yaws	Solomon Islands	12

Solomon Islands 37 sca	ERR1470331	<i>Homo Sapiens</i>	Yaws	Solomon Islands	12
Gambia-1	SRR4308597	<i>Chlorocebus sabeus</i>	Yaws	Gambia	13
Gambia-2	SRR4308605	<i>Chlorocebus sabaesus</i>	Yaws	Gambia	13
Senegal NKNP-1	SRR4308606	<i>Chlorocebus sabaesus</i>	Yaws	Senegal	13
Senegal NKNP-2	SRR4308607	<i>Chlorocebus sabaesus</i>	Yaws	Senegal	13
LMNP-1	CP021113.1	<i>Papio anubis</i>	Yaws	Tanzania	13
LMNP-2_BS5	SRR4308598	<i>Papio anubis</i>	Yaws	Tanzania	13
LMNP-2_BS6	SRR4308599	<i>Papio anubis</i>	Yaws	Tanzania	13
LMNP-2_BS7	SRR4308601	<i>Papio anubis</i>	Yaws	Tanzania	13
LMNP-2_BS8	SRR4308602	<i>Papio anubis</i>	Yaws	Tanzania	13

\*while this genome was initially included, it was subsequently removed from the analyses as very few positions passed our quality filters. When it was included in an initial analysis, it was positioned on a very long branch.

## References

1. Štaudová B, Strouhal M, Zobaníková M, Čejková D, Fulton LL, Chen L, et al. Whole genome sequence of the *Treponema pallidum* subsp. *endemicum* Strain Bosnia A: The genome is related to yaws treponemes but contains few loci similar to syphilis treponemes. *PLoS Neglected Tropical Diseases*. 2014;8(11): e3261. doi:10.1371/journal.pntd.0003261
2. Mikalová L, Strouhal M, Oppelt J, Grange PA, Janier M, Benhaddou N, et al. Human *Treponema pallidum* 11q/j isolate belongs to subsp. *endemicum* but contains two loci with a sequence in TP0548 and TP0488 similar to subsp. *pertenue* and subsp. *pallidum*, respectively. *PLOS Neglected Tropical Diseases*. 2017;11(3): e0005434. doi:10.1371/journal.pntd.0005434
3. Pětrošová H, Pospíšilová P, Strouhal M, Čejková D, Zobaníková M, Mikalová L, et al. Resequencing of *Treponema pallidum* ssp.

*pallidum* strains Nichols and SS14: Correction of sequencing errors resulted in increased separation of syphilis treponeme subclusters.

*PLoS ONE*. 2013;8(9). doi:10.1371/journal.pone.0074319

4. Giacani L, Jeffrey BM, Molini BJ, Le HVT, Lukehart SA, Centurion-Lara A, et al. Complete genome sequence and annotation of the *Treponema pallidum* subsp. *pallidum* Chicago strain. *Journal of Bacteriology*. 2010;192(10): 2645–2646. doi:10.1128/JB.00159-10
5. Zobaníková M, Mikolka P, Čejková D, Pospíšilová P, Chen L, Strouhal M, et al. Complete genome sequence of *Treponema pallidum* strain DAL-1. *Standards in Genomic Sciences*. 2012;7(1): 12–21. doi:10.4056/sigs.2615838
6. Giacani L, Iverson-Cabral SL, King JCK, Molini BJ, Lukehart SA, Centurion-Lara A. Complete genome sequence of the *Treponema pallidum* subsp. *pallidum* Sea81-4 strain. *Genome Announcements*. 2014;2(2). doi:10.1128/genomeA.00333-14
7. Zobaníková M, Strouhal M, Mikalová L, Čejková D, Ambrožová L, Pospíšilová P, et al. Whole genome sequence of the *Treponema* Fribourg-Blanc: Unspecified simian isolate is highly similar to the yaws subspecies. *PLoS Neglected Tropical Diseases*. 2013;7(4): e2172. doi:10.1371/journal.pntd.0002172
8. Čejková D, Zobaníková M, Chen L, Pospíšilová P, Strouhal M, Qin X, 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 Neglected Tropical Diseases*. 2012;6(1). doi:10.1371/journal.pntd.0001471
9. Liska SL, Perine PL, Hunter EF, Crawford JA, Feeley JC. Isolation and transportation of *Treponema pertenue* in golden hamsters. *Current Microbiology*. 1982;7(1): 41–43. doi:10.1007/BF01570978



10. Strouhal M, Mikalová L, Havlíčková P, Tenti P, Čejková D, Rychlík I, 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 Neglected Tropical Diseases*. 2017;11(9): e0005894. doi:10.1371/journal.pntd.0005894
11. Strouhal M, Mikalová L, Haviernik J, Knauf S, Bruisten S, Noordhoek GT, et al. Complete genome sequences of two strains of *Treponema pallidum* subsp. *pertenue* from Indonesia: Modular structure of several treponemal genes. *PLOS Neglected Tropical Diseases*. 2018;12(10): e0006867. doi:10.1371/journal.pntd.0006867
12. Marks M, Fookes M, Wagner J, Butcher R, Ghinai R, Sokana O, et al. Diagnostics for yaws eradication: insights from direct next-generation sequencing of cutaneous strains of *Treponema pallidum*. *Clinical infectious diseases*. 2018;66(6): 818–824. doi:10.1093/cid/cix892
13. Knauf S, Gogarten JF, Schuenemann VJ, Nys HM De, Düx A, Strouhal M, et al. Nonhuman primates across sub-Saharan Africa are infected with the yaws bacterium *Treponema pallidum* subsp. *pertenue*. *Emerging microbes & infections*. 2018;7(1): 1–4. doi:10.1038/s41426-018-0156-4

**Table S4: Recombinant and/or paralogous genes removed from the phylogenetic analyses of TPE strains in this study. The gene coordinates used are according to the simian strain Fribourg-Blanc (Refseq: NC\_021179).**

<b>Gene number</b>	<b>Gene name</b>	<b>Gene coordinates interval</b>	<b>Reference</b>
TP0117	tprC	134,904 -> 136,700	1,2
TP0131	tprD	152,349 -> 154,139	2
TP0133	hypothetical protein	154,340 -> 155,572	3
TP0136	hypothetical protein	157,933 -> 159,378	4,5
TP0317	tprG	334,105 -> 336,375	1,2
TP0326	bamA gene	346,037 -> 348,535	6–8
TP0462	V-type ATP synthase subunit A	493,089 -> 494,372	1
TP0483	hypothetical protein	514,800 -> 516,011	1
TP0488	mcp-2	523,150 -> 525,687	6,7,9
TP0548	hypothetical protein	593,973 -> 595,229	9
TP0620	tprI	673,370 -> 675,199	2
TP0621	tprJ	675,258 -> 677,528	1,2
TP0856	lipoprotein	935,859 -> 937,043	3
TP0858	lipoprotein	937,145 -> 938,374	3,10
TP0865 (Arora 2016)	UPF0164	946,114 -> 947,559	4
TP1031	tprL	1,126,698 -> 1,128,368	7
TP0897	tprK	977,074 -> 978,654	3
TP0433	acidic repeat protein	462,226 -> 464,100	3
TP0470	tetratricopeptide repeat protein	499,308 -> 500,537	3
16S>23S>5S	rRNA>rRNA>rrfRNA	231, 197->236,175	3
16S>23S>5S	rRNA>rRNA>rrfRNA	279, 609->284, 577	3

**Notes:** spacers between the 16S-23S-5S were also removed from the analysis

## References

1. Grillová L, Oppelt J, Mikalová L, Nováková M, Giacani L, Niesnerová A, et al. Directly Sequenced Genomes of Contemporary Strains of Syphilis Reveal Recombination-Driven Diversity in Genes Encoding Predicted Surface-Exposed Antigens. *Frontiers in Microbiology*. 2019;10 (7): 1691. doi:10.3389/fmicb.2019.01691
2. Gray RR, Mulligan CJ, Molini BJ, Sun ES, Giacani L, Godornes C, et al. Molecular evolution of the tprC, D, I, K, G, and J genes in the pathogenic genus *Treponema*. *Molecular biology and evolution*. 2006;23(11): 2220–2233. doi:10.1093/molbev/msl092
3. Strouhal M, Mikalová L, Haviernik J, Knauf S, Bruisten S, Noordhoek GT, et al. Complete genome sequences of two strains of *Treponema pallidum* subsp. *pertenue* from Indonesia: Modular structure of several treponemal genes. Caimano MJ (ed.) *PLOS Neglected Tropical Diseases*. 2018;12(10): e0006867. doi:10.1371/journal.pntd.0006867
4. Arora N, Schuenemann VJ, Jäger G, Peltzer A, Seitz A, Herbig A, et al. Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. *Nature Microbiology*. 2016;2(1): 16245. doi:10.1038/nmicrobiol.2016.245
5. Grillová L, Bawa T, Mikalová L, Gayet-Ageron A, Nieselt K, Strouhal M, et al. Molecular characterization of *Treponema pallidum* subsp. *pallidum* in Switzerland and France with a new multilocus sequence typing scheme. Lin B (ed.) *PLOS ONE*. 2018;13(7): e0200773. doi:10.1371/journal.pone.0200773
6. Pětrošová H, Zobaníková M, Čejková D, Mikalová L, Pospíšilová P, Strouhal M, et al. Whole Genome Sequence of *Treponema pallidum* ssp. *pallidum*, Strain Mexico A, Suggests Recombination between Yaws and Syphilis Strains. Picardeau M (ed.) *PLoS Neglected Tropical Diseases*. 2012;6(9): e1832. doi:10.1371/journal.pntd.0001832
7. Štaudová B, Strouhal M, Zobaníková M, Čejková D, Fulton LL, Chen L, et al. Whole Genome Sequence of the *Treponema pallidum* subsp. *endemicum* Strain Bosnia A: The Genome Is Related to Yaws *Treponemes* but Contains Few Loci Similar to Syphilis *Treponemes*. Yang R (ed.) *PLoS Neglected Tropical Diseases*. 2014;8(11): e3261. doi:10.1371/journal.pntd.0003261
8. Harper KN, Ocampo PS, Steiner BM, George RW, Silverman MS, Bolotin S, et al. On the origin of the treponematoses: A phylogenetic approach. *PLoS Neglected Tropical Diseases*. 2008;2(1). doi:10.1371/journal.pntd.0000148
9. Mikalová L, Strouhal M, Oppelt J, Grange PA, Janier M, Benhaddou N, et al. Human *Treponema pallidum* 11q/j isolate belongs to subsp. *endemicum* but contains two loci with a sequence in TP0548 and TP0488 similar to subsp. *pertenue* and subsp. *pallidum*, respectively.

Picardeau M (ed.) *PLOS Neglected Tropical Diseases*. 2017;11(3): e0005434. doi:10.1371/journal.pntd.0005434

10. Marks M, Fookes M, Wagner J, Butcher R, Ghinai R, Sokana O, et al. Diagnostics for Yaws Eradication: Insights From Direct Next-Generation Sequencing of Cutaneous Strains of *Treponema pallidum*. *Clinical infectious diseases*. 2018;66(6): 818–824. doi:10.1093/cid/cix8