Nonhuman primates across sub-Saharan Africa are infected with the yaws bacterium *Treponema pallidum* subsp. *pertenue*

Running title: Yaws bacterium in wild nonhuman primates

Sascha Knauf (sknauf@dpz.eu)[1¶], Jan F. Gogarten (jan.gogarten@gmail.com)[2,3,4¶], Verena J. Schuenemann (verena.schuenemann@uni-tuebingen.de)[5¶], Hélène M. De Nys (denyshelene@yahoo.com)[2,#¶], Ariane Düx (duexa@rki.de)[2,6], Michal Strouhal (mstrouhal@med.muni.cz)[7], Lenka Mikalová (lmikal@med.muni.cz)[7], Kirsten I. Bos (bos@shh.mpg.de)[8], Roy Armstrong (roy.armstrong@cumbria.ac.uk)[9], Emmanuel K. Batamuzi (emmanuelbatamuzi@gmail.com)[10,§], Idrissa S. Chuma (chumaidr@gmail.com)[11], Bernard Davoust (bernard.davoust@gmail.com)[12], Georges Diatta (georges.diatta@ird.fr)[13], Robert D. Fyumagwa (rfyumagwa@yahoo.com)[14], Reuben R. Kazwala (kazwala@suanet.ac.tz)[15], Julius D. Keyyu (julius.keyyu@gmail.com)[14], Inyasi A. V. Lejora (inyasi.lejora@tanzaniaparks.go.tz)[11], Anthony Levasseur (anthony.levasseur@univ-amu.fr)[12], Hsi Liu (hcl6@cdc.gov)[16], Michael A. Mayhew (mic1mayhew@hotmail.com)[9], Oleg Mediannikov (olegusss1@gmail.com)[12], Didier Raoult (Didier.raoult@gmail.com)[12,13], Roman M. Wittig (wittig@eva.mpg.de)[3,17], Christian Roos (CRoos@dpz.eu)[18], Fabian H. Leendertz (leendertzf@rki.de)[2], David Šmajs (dsmajs@med.muni.cz)[7&*], Kay Nieselt (kay.nieselt@uni-tuebingen.de)[19&], Johannes Krause (krause@shh.mpg.de)[5,8&*], Sébastien Calvignac-Spencer (CalvignacS@rki.de)[2,6&*]

[1]Work Group Neglected Tropical Diseases, Pathology Unit, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Göttingen, Germany.

[2]Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Seestraße 10, 13353 Berlin, Germany.

[3]Primatology Department, Max Planck Institute for Evolutionary Anthropology, Deutscher Pl. 6, 04103 Leipzig, Germany.

[4]Department of Biology, McGill University, 1205 Avenue du Docteur-Penfield, Montréal, QC H3A 1B1, Quebec, Canada.

[5]Institute for Archaeological Sciences, University of Tübingen, Ruemelinstrasse 23, 72070 Tübingen, Germany.

[6]Viral Evolution, Robert Koch Institute, Seestraße 10, 13353 Berlin, Germany.

[7]Department of Biology, Faculty of Medicine, Masaryk University, Kamenice 5/A6, 625 00 Brno, Czech Republic.

[8]Max Planck Institute for the Science of Human History, Kahlaische Strasse 10, 7745 Jena, Germany.

[9]Department of Science, Natural Resources and Outdoor Studies, University of Cumbria, Fusehill Street, Carlisle, Cumbria, CA12HH, UK.

[10]Department of Veterinary Surgery and Theriogenology, College of Veterinary and Medical Sciences, Sokoine University of Agriculture, Morogoro, Tanzania.

[11]Ecological Monitoring Department, Tanzania National Parks (TANAPA), P.O.Box 3134, Arusha, Tanzania.

[12]Research Unit of Emerging Infectious and Tropical Diseases (URMITE) UM63, CNRS 7278, IRD 198, INSERM 1095, Aix-Marseille University, 27, Bd Jean Moulin, 13385 Marseille cedex 05, Marseille, France.

[13]Research Unit of Emerging Infectious and Tropical Diseases (URMITE) IRD 198, Campus IRD/UCAD, Hann Les Maristes, Dakar, Senegal.

[14]Tanzania Wildlife Research Institute (TAWIRI), P.O. Box 661, Arusha, Tanzania.

[15]Department of Veterinary Medicine and Public Health, College of Veterinary and Medical Sciences, Sokoine University of Agriculture, Morogoro, Tanzania.

[16]National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Diseases Control and Prevention, Atlanta, GA, USA.

[17]Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques, BP 1303, Abidjan 01, Côte d'Ivoire.

[18]Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Göttingen, Germany.

[19]Center for Bioinformatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany.

[#]Current address: Unité Mixte Internationale 233, Institut de Recherche pour le Développement, INSERM U1175, and University of Montpellier, Montpellier, France.

[*]Correspondence to: Dr. Sébastien Calvignac-Spencer, Robert Koch Institute, Seestrasse 10, 13353 Berlin, Germany, tel.: +49187542501, calvignacs@rki.de; Prof. David Šmajs, Faculty of Medicine, Masaryk University, 625 00 Brno, Czech Republic, tel.: +420 549 497 496, dsmajs@med.muni.cz; Prof. Dr. rer. nat. Johannes Krause, Max Planck Institute for the Science of Human History, Jena, Germany, tel.: +49 3641 686600, krause@shh.mpg.de.

[¶]These authors contributed equally to this work.

[§]Deceased.

[&]These authors also contributed equally to this work.

Dear Editor,

The bacterium *Treponema pallidum* (*TP*) causes human syphilis (subsp. *pallidum*; *TPA*), bejel (subsp. *endemicum*; *TEN*) and yaws (subsp. *pertenue*; *TPE*)[1]. While syphilis reached a world-wide distribution[2], bejel and yaws are endemic diseases. Bejel is found in dry areas in Sahelian Africa and Saudi Arabia, whereas yaws is present in the humid tropics[1]. Yaws is currently reported endemic in 14 countries and an additional 84 countries have a known history of yaws but lack recent epidemiological data[3,4]. The disease was subject to global eradication efforts in the mid-20th century but reemerged in West Africa, Southern Asia, and the Pacific region[5]. New large-scale treatment options triggered the ongoing second eradication campaign, which aims to eradicate yaws globally by 2020[5].

TPE is usually considered as a strictly human pathogen. This perception may however partly result from the lack of detailed data on nonhuman primate (NHP)-infecting treponemes. Indeed, a number of African NHPs show skin ulcerations suggestive of treponemal infection and antibodies against *TP* have been detected in wild NHP populations[6,7]. While genetic studies confirmed monkeys and great apes are infected with *TP* strains[8-10], most of these analyses only determined short DNA sequences. The small number of polymorphic sites examined largely precludes assignment of these strains to a particular *TP* subspecies[9], especially considering that sporadic recombination events between subspecies have been reported[11]. The only simian strain whose whole genome was sequenced - Fribourg-Blanc, isolated from a Guinea baboon (*Papio papio*) in 1966[7] - unambiguously clustered with human-infecting *TPE* strains[12].

A fundamental question with regard to yaws evolution and possibly yaws eradication is whether humans and NHPs are commonly infected with the same pathogen, *TPE*, and whether transmission between NHPs and humans occurs. To determine which pathogen causes treponematoses in NHPs across sub-Saharan Africa, we collected samples from symptomatic wild individuals belonging to three NHP species (*Cercocebus atys*, *Chlorocebus sabaeus*, and *Papio anubis*) from four independent populations in West and East Africa (**Fig 1**, **Supplementary Table S1, Supplementary Materials**). Samples were collected at Taï National Park (TaïNP; Côte d'Ivoire), Bijilo Forest Park (BFP, the Gambia), Niokolo-Koba National Park (NKNP, Senegal), and Lake Manyara National Park (LMNP,

Tanzania). Monkeys presented yaws-like orofacial and limb lesions (TaïNP, BFP) or ulcerative anogenital skin lesions (BFP, NKNP, LMNP)[9].

Using PCR, we showed the presence of *TP* in skin lesion biopsies or swabs from NHPs at TaïNP (*C. atys*), BFP, and NKNP (*C. sabaeus*). *TP* infection in olive baboons (*P. anubis*) at LMNP had previously been confirmed[6]. Two samples per NHP population were selected for whole genome sequencing based on high *TP* copy number or the ability to amplify long PCR fragments (**Supplementary Table S2**). To overcome the background of host genomic DNA, we used targeted DNA capture coupled with NGS to reconstruct whole *TP* genomes[2,8]. Following quality filtering, removal of PCR duplicates, merging of different sequencing runs from the same sample, and mapping against the *TPE* strain Fribourg-Blanc reference genome, we obtained a range of 22,886-470,303 DNA sequencing reads per sample. All samples showed at least 80% coverage of the reference genome with depth coverage of three or higher; average genome coverage depth was between 6.1 and 121.0-fold (**Supplementary Table S3**).

We generated maximum likelihood, Bayesian and maximum parsimony trees based on the genomes reconstructed in our study and all available reference genomes (total sequence length: 1,133,379 nucleotides). In all trees, *TPE* and *TPA* strains formed reciprocally monophyletic groups, with a mean *TPE*/*TPA* strain divergence of 0.099%. NHP-infecting *TP* strains all clustered with human-infecting *TPE* strains (**Fig 1**; **Supplementary Figure S1**). The *TPE* clade exhibited a star-like branching pattern whereby all basal branches were very short and received low statistical support. Importantly, this pattern does not support a clear reciprocal monophyly of the *TPE* strains infecting humans and NHPs. In line with this, the minimum divergence between strains infecting humans and NHPs was lower than the maximum divergence amongst human and NHP-infecting strains (0.011% versus 0.015% and 0.024%). Human-infecting *TPE* strains Samoa D, CDC-2, CDC-2575, Ghana-051, and Gauthier, which span a broad geographic and temporal range (at least four decades), were less divergent from each other than the two strains infecting sooty mangabeys from a single social group at TaïNP (0.011% versus 0.017% sequence divergence, respectively). While intra-group strain divergence was low for the two African green monkey populations and the olive baboons (0.0003% and 0.0017%, respectively), intra-species strain divergence for African green monkeys was relatively high when

compared to the divergence observed between the two most divergent human strains (0.0094% versus 0.015%).

For the sample LMNP-1, we determined the complete genome sequence and structure (average depth of coverage: 169x; GenBank: CP021113; **Supplementary Tables S5-6**)[12]. The LMNP-1 genome showed the same structure as published complete genomes of human-infecting *TPE* strains and the simian strain Fribourg-Blanc. It was more similar to the human-infecting *TPE* Gauthier strain than the simian isolate Fribourg-Blanc, showing differences at 266 and 325 chromosomal positions, respectively. Most differences were single nucleotide substitutions or small indels (**Supplementary Table S7**). The LMNP-1 and Gauthier strains exhibited the same number of the 24-bp repeats in the *TP_0470* gene (n=25) and Gauthier had only one 60-bp repeat more than LMNP-1 strain in the *arp* gene (LMNP-1 n=9 vs. Gauthier n=10). All 60-bp repeats in the *arp* gene of LMNP-1 were of Type II and were identical to other *TPE* strains[13]. The *tprK* gene of LMNP-1 only had three variable regions, V5-V7, when compared to other *TPE* strains. In addition to differences in *TP_0433*, *TP_0470*, and *tprK* genes, relatively large indels were determined in *TPEGAU_0136* (33-nt long deletion; specific for strains Gauthier and Samoa D), in *TPFB_0548* (42-nt long deletion; specific for strain Fribourg-Blanc), in *TPEGAU_0858* (79-nt long deletion; specific for strain Gauthier), in the intergenic region (IGR) between *TPEGAU_0628* and *TPEGAU_0629* (302-nt long deletion; specific for strain Gauthier), and in IGR between *TPFB_0696* and *TPFB_0697* (430-nt long insertion; specific for strain Fribourg-Blanc); the length of other sequence differences ranged between 1-15 nts. RNA operons structure of the LMNP-1 genome (coordinates 231,180-236,139; 279,584-284,533; according to *TPE* strain Gauthier: NC_016843.1) was similar to strains Gauthier, CDC-2, and Fribourg-Blanc, but different to the strains Samoa D, Samoa F, and CDC-1. The LMNP-1 16S-5S-23S was identical in both operons and 23S rRNA sequences were identical to other *TPE* strains except for strain Fribourg-Blanc (having a single nucleotide difference at position 458). We did not find any mutations associated with macrolide resistance (e.g. A2058G, A2059G)[14]. When the two NHP-infecting TPE strains, Fribourg-Blanc and LMNP-1, were compared to the closest human-pathogenic TPE strains CDC-2 and Gauthier, respectively, only 7.2% and 9.1% of all coding sequences (77 and 97 coding sequences out of 1065)

contained amino acid substitutions, suggesting limited functional divergence (**Supplementary Tables S7-9**).

Our findings unambiguously indicate that at least three African NHP species (representing four populations) from West and East Africa currently suffer from treponematosis caused by *TPE*. Taking into account the isolation of the Fribourg–Blanc strain from Guinea baboons in 1966 and its recent sequencing and identification as a member of the *TPE* clade[12], this represents four African NHP species and five populations whose symptoms can be explained by *TPE* infections. Coupled with a growing number of clinical and serological observations[6,7,9,10], this suggests infection of NHPs with *TPE* is common throughout sub-Saharan Africa. Humans are not the exclusive host for the yaws bacterium and NHPs are infected with the same bacterial agent.

TPE strains in NHPs exhibit considerable genetic diversity, which at least equals that found among published human-infecting *TPE* strains. Importantly, we found no evidence for a clear sub-differentiation of NHP- and human-infecting *TPE* strains, i.e. these strains did not form well-supported reciprocally monophyletic groups. Rather, the star-like topology of our phylogenomic tree suggests a rapid initial radiation of the ancestor of *TPE* which may have involved transmission across primate species barriers in a relatively distant past (with respect to the *TPE* clade depth). These results neither support, nor allow us to exclude, a possible recent transmission of *TPE* between NHPs and humans, especially due to the large geographic and temporal separation between the two groups of samples being compared. A major hurdle in identifying such potential transmission events is the availability of enough bacterial genomes. Despite large numbers of human cases, very few genomes have been determined from human-infecting *TPE* strains and only from a very limited geographic range. Generating additional human-infecting *TPE* genomes represents an important area of research that, coupled with the genomes of *TPE* strains infecting NHPs presented here, could now enable the detection of recent zoonotic transmission events, would any exist.

Since yaws has not been reported for several decades in humans in countries where we find NHPs to be infected with *TPE*, we expect that if transmission happens, it is only at very low frequency (as is the case for many zoonotic diseases). Of course, such low frequency zoonotic transmission does not explain the reemergence of yaws, which is the consequence of continued human-to-human

transmission. However, now that eradication of yaws seems within reach[15], the finding that *TPE* strains circulate in NHPs certainly calls for more research into their diversity and zoonotic potential.

Data availability

All raw read files have been deposited in NCBI as part of the BioProject PRJNA343706.

Competing interests

The authors declare that they have no competing interests.

Supporting information

Supplementary Materials. This document comprises supporting methods and additional results.

Supplementary Figure S1. Phylogenetic trees of *TP* whole genome sequences.

Supplementary Table S1. Nonhuman primates anesthetized for this study.

Supplementary Table S2. Molecular analyses (PCR and sequencing) performed on blood samples, skin tissue samples, and lesion swabs.

Supplementary Table S3. Read mapping and genotyping results.

Supplementary Table S4. Published genomes used for phylogenetic analyses.

Supplementary Table S5. List of primers used for long-range PCR amplification of *TP* intervals of the East African baboon genome (strain LMNP-1).

Supplementary Table S6. Summary of the PSGS sequencing results of four genomic DNA (gDNA) pools of the East African baboon genome (strain LMNP-1).

Supplementary Table S7. Number of nucleotide differences (i.e. indels and SNVs) of various lengths between the genome of the baboon (strain LMNP-1) and the published *TPE* genome of strains Gauthier and Fribourg-Blanc.

Supplementary Table S8. Proteins encoded by the *TPE* strain Fribourg-Blanc genome with 1 and more amino acid changes when compared to the *TPE* strain CDC-2 proteome.

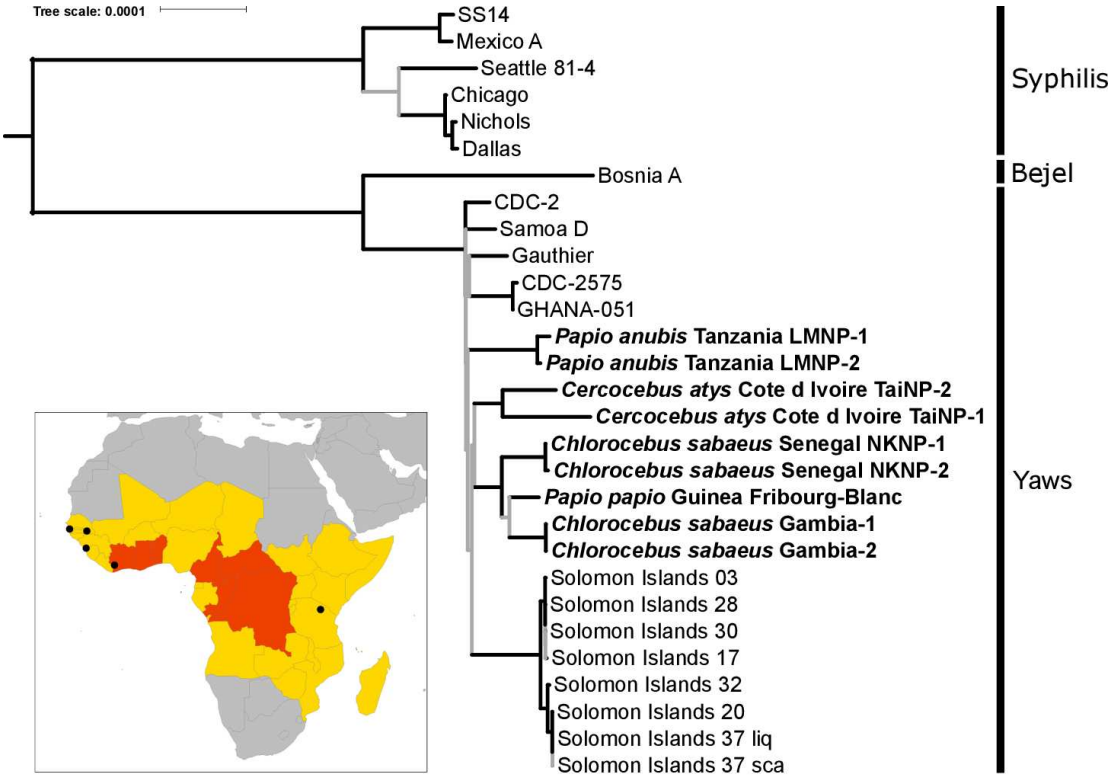
205 **Supplementary Table S9**. Proteins encoded by the LMNP-1 baboon strain with one and more amino
206 acid changes when compared to the *TPE* strain Gauthier proteome.
207

Figure Legend

Fig 1. Phylogenomic analysis of NHP- and human-infecting *Treponema pallidum* strains. In this maximum likelihood tree, nodes that had less than 95% ultrafast bootstrap approximation support are indicated with grey lines. Tip labels indicate the NHP species sampled, the country of origin, and the sample ID. The scale is in nucleotide substitution per site. The inset is a map of Africa where sites of origin of the NHP samples from which a *TP* genome was determined are indicated with black circles. A country's 2013 yaws status based on the World Health Organization's Global Health Observatory (http://www.who.int/gho/en/) is indicated by its color: grey indicates no previous history of yaws infections in humans, yellow indicates a country previously endemic for yaws though the current status is unknown, and countries in red indicate countries which are currently considered endemic for yaws.

Figures

Fig 1



Acknowledgments

This article represents a chapter in the doctoral dissertation of JFG, and has thus benefited greatly from the input of his supervisor committee and advisors Jonathan Davies, David Marcogliese, Charles Nunn, and Louis Lefebvre. Simone Lüert and Ulla Thiesen are thanked for discussions and technical support. We thank the Ivorian Ministry of Environment and Forests, the Ministry of Research, the directorship of the TaïNP, the Office Ivoirien des Parcs et Réserves, the Centre Suisse de Recherche Scientifique, the Taï Chimpanzee Project and the Taï Monkey project and their teams of field assistants for their support. We thank Jonathan Müller-Tiburtius, Therese Löhrich, Sylvain Lemoine, Simon T. Kannieu, Daniel Gnimion, Richard Peho, and Martina Magris for field assistance at TNP. For assistance and support in BFP we thank the Department of Parks and Wildlife Management and the Forestry Department. For assistance and support with sequencing and read processing at the RKI we are grateful to Wojciech Dabrowski, Julia Hinzmann, Andreas Nitsche and Julia Tesch. We would like to thank the Senegalese authorities, in particular the officials of NKNP, who helped our work in Senegal immensely. We also thank Annick Abeille, who performed the initial PCR screening and sequencing of the Senegalese green monkey samples. For the work on East African simian strains, we would like to thank TAWIRI, TANAPA, LMNP staff, especially the chief park warden, ecology monitoring unit and park rangers, as well as Tanzania Commission for Science and Technology (COSTECH).

References

1 Giacani, L. & Lukehart, S. A. The endemic treponematoses. *Clinical microbiology reviews* **27**, 89-115, doi:10.1128/CMR.00070-13 (2014).

2 Arora, N. *et al.* Origin of modern syphilis and emergence of a pandemic Treponema pallidum cluster. *Nat Microbiol* **2**, 16245, doi:10.1038/nmicrobiol.2016.245 (2016).

3 Marks, M. Yaws: towards the WHO eradication target. *Trans R Soc Trop Med Hyg*, doi:10.1093/trstmh/trw032 (2016).

4 World Health Organization. Eradication of yaws: procedures for verification and certification of interruption of transmission., (Geneva, 2018).

5 Asiedu, K., Fitzpatrick, C. & Jannin, J. Eradication of yaws: historical efforts and achieving WHO's 2020 target. *PLoS neglected tropical diseases* **8**, e3016, doi:10.1371/journal.pntd.0003016 (2014).

6 Felsenfeld, O. & Wolf, R. H. Serological reactions with treponemal antigens in nonhuman primates and the natural history of treponematosis in man. *Folia Primatol (Basel)* **16**, 294-305 (1971).

7 Fribourg-Blanc, A. & Mollaret, H. H. Natural treponematosis of the African primate. *Primates Med* **3**, 113-121 (1969).

8 Gogarten, J. F. *et al.* Tools for opening new chapters in the book of Treponema pallidum evolutionary history. *Clin Microbiol Infect* **22**, 916-921, doi:10.1016/j.cmi.2016.07.027 (2016).

9 Knauf, S. *et al.* Treponema infection associated with genital ulceration in wild baboons. *Veterinary pathology* **49**, 292-303, doi:10.1177/0300985811402839 (2012).

10 Chuma IS, B. E., Collins DA, Fyumagwa RD, Hallmaier-Wacker LK, Kazwala RR, et al. Widespread *Treponema pallidum* Infection in Nonhuman Primates, Tanzania. *Emerg Infect Dis*. **24**, 1002-1009, doi: https://doi.org/10.3201/eid2406.180037 (2018).

11 Petrosova, H. *et al.* Whole genome sequence of Treponema pallidum ssp. pallidum, strain Mexico A, suggests recombination between yaws and syphilis strains. *PLoS neglected tropical diseases* **6**, e1832, doi:10.1371/journal.pntd.0001832 (2012).

12 Zobanikova, M. *et al.* Whole genome sequence of the Treponema Fribourg-Blanc: unspecified simian isolate is highly similar to the yaws subspecies. *PLoS neglected tropical diseases* **7**, e2172, doi:10.1371/journal.pntd.0002172 (2013).

13 Harper, K. N. *et al.* The sequence of the acidic repeat protein (arp) gene differentiates venereal from nonvenereal *Treponema pallidum* subspecies, and the gene has evolved under strong positive selection in the subspecies that causes syphilis. *FEMS immunology and medical microbiology* **53**, 322-332, doi:10.1111/j.1574-695X.2008.00427.x (2008).

14 Lukehart, S. A. *et al.* Macrolide resistance in *Treponema pallidum* in the United States and Ireland. *The New England journal of medicine* **351**, 154-158, doi:10.1056/NEJMoa040216 (2004).

15 Marks, M. *et al.* Mathematical Modeling of Programmatic Requirements for Yaws Eradication. *Emerg Infect Dis* **23**, 22-28, doi:10.3201/eid2301.160487 (2017).