Yaws re-emergence and bacterial drug resistance selection after mass administration of azithromycin: a genomic epidemiology investigation

Mathew A Beale, Marc Noguera-Julian, Charmie Godornes, Maria Casadellà, Camila González-Beiras, Mariona Parera, August Kapa Jnr, Wendy Houinei, James Wangi, Marc Corbacho-Monne, Roger Paredes, Fernando Gonzalez-Candelas, Michael Marks, Sheila A Lukehart, Nicholas R Thomson*, Oriol Mitjà*

Summary

Background In a longitudinal study assessing the WHO strategy for yaws eradication using mass azithromycin treatment, we observed resurgence of yaws cases with dominance of a single JG8 sequence type and emergence of azithromycin-resistant *Treponema pallidum* subspecies *pertenue* (*T p pertenue*). Here, we analyse genomic changes in the bacterial population using samples collected during the study.

Methods We did whole bacterial genome sequencing directly on DNA extracted from 37 skin lesion swabs collected from patients on Lihir Island, Papua New Guinea, between April 1, 2013, and Nov 1, 2016. We produced phylogenies and correlated these with spatiotemporal information to investigate the source of new cases and the emergence of five macrolide-resistant cases. We used deep amplicon sequencing of surveillance samples to assess the presence of minority macrolide-resistant populations.

Findings We recovered 20 whole *T p pertenue* genomes, and phylogenetic analysis showed that the re-emerging JG8 sequence type was composed of three bacterial sublineages characterised by distinct spatiotemporal patterns. Of five patients with resistant *T p pertenue*, all epidemiologically linked, we recovered genomes from three and found no variants. Deep sequencing showed that before treatment, the index patient had fixed macrolide-sensitive *T p pertenue*, whereas the post-treatment sample had a fixed resistant genotype, as did three of four contact cases.

Interpretation In this study, re-emergence of yaws cases was polyphyletic, indicating multiple epidemiological sources. However, given the genomic and epidemiological linkage of resistant cases and the rarity of resistance alleles in the general population, azithromycin resistance is likely to have evolved only once in this study, followed by onward dissemination.

Funding Wellcome and Provincial Deputation of Barcelona.

Introduction

Yaws, caused by the bacterium *Treponema pallidum* subspecies *pertenue* (*T p pertenue*), is a neglected tropical disease and a major cause of morbidity in regions of Africa, Asia, and the South Pacific.[1,2] Infection with *T p pertenue* is clinically apparent in the early stages, but population-level disease control is complicated by the bacterium's ability to undergo periods of asymptomatic latency that can span five to ten years, thus evading strategies targeting only clinically active cases.[1] This challenge influenced WHO in designing its Morges Strategy[3] for yaws eradication using mass drug administration (MDA) of single-dose azithromycin for all community members, regardless of infection status.

In a longitudinal study, which was part of a yaws elimination programme including more than 16 000 residents of Lihir Island, Papua New Guinea, a single MDA with azithromycin reduced the prevalence of yaws from 1·8% (238 of 16 092) to 0·1% (17 of 17 339) after 18 months,[4] but infection began to re-emerge after 24 months, with a significant increase (difference 0·3%; 95% CI 0·1–0·4; p<0·0001) in prevalence to 0·4% (51 of 18 836) at 42 months.[4] We previously reported the findings of baseline and follow-up surveys,[4] with use of a novel multilocus sequence typing (MLST) scheme[5] to analyse 239 samples, and showed a transition from a diverse population of *T p pertenue* sequence types before the commencement of MDA to the dominance of a single sequence type (JG8) 24 months after MDA.[4] We also reported the detection of five cases of *T p pertenue* with genotypic resistance to azithromycin mediated by A2059G mutations in the 23S ribosomal genes (one of two mutations, along with A2058G, known to cause resistance in *T pallidum*), all belonging to the JG8 sequence type.[4]

In this study, we validated the previous molecular typing system using whole genomes sequenced directly from clinical swab DNA,[6] and we used the increased resolution provided by genomic tools in combination with additional epidemiological data to conduct a

Biology (I2SysBio), Valencia, Spain (Prof F Gonzalez-Candelas PhD); CIBER in Epidemiology and Public Health, Valencia, Spain (Prof F Gonzalez-Candelas); Clinical Research Department (M Marks PhD) and Department of Infectious and Tropical Diseases (Prof N R Thomson), Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK; Hospital for Tropical Diseases, London, UK (M Marks); and Division of Public Health, School of Medicine and Health Sciences, University of Papua New Guinea, Port Moresby, Papua New Guinea (O Mitjà)

Correspondence to:
Dr Mathew A Beale, Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, UK
mathew.beale@sanger.ac.uk

or

Dr Oriol Mitjà, Lihir Medical Center, International SOS-Newcrest Mining, Lihir Island, Papua New Guinea
omitja@flsida.org

Research in context

Evidence before this study

We searched PubMed using the terms "yaws", "Treponema pallidum", "mass treatment", and "azithromycin", without date or language restrictions. The original search was done in August, 2019, with regular monitoring of newly published research by MAB thereafter. Mass drug administration (MDA) using azithromycin (single-dose 30 mg/kg, maximum 2 g) has been a successful strategy to reduce yaws prevalence in endemic communities, with trials done in Papua New Guinea and Ghana. However, no long-term follow-up studies have been done other than in Papua New Guinea, where a resurgence of yaws cases was observed 24 months after MDA. Multilocus sequence typing (MLST) showed a change in *Treponema pallidum* subspecies *pertenue* (*T p pertenue*) to a single dominant JG8 type, and identified resistant strains with the same MLST type, but this method lacked the resolution to determine the epidemiological mechanism leading to yaws re-emergence and the manifestation of macrolide resistance.

Added value of this study

To our knowledge, this is the first study to use whole genome sequencing (WGS) for a detailed investigation of *T p pertenue* transmission and to apply deep amplicon sequencing to genomic epidemiology investigation. We aimed to *T p pertenue* macrolide-resistance mutations. Using the increased resolution of WGS over other techniques, we showed that re-emergence of yaws after MDA was the result of at least three distinct bacterial sublineages with different spatiotemporal properties (probably due to several infected people missing treatment), rather than a single source, and that patients with macrolide-resistant yaws had genomically indistinguishable *T p pertenue*, indicating a recent common ancestor, and possibly a recent chain of transmission originating from the index case. Deep amplicon sequencing corroborates that PCR-tested macrolide-sensitive samples do not harbour resistant minority populations at detectable levels.

Implications of all the available evidence

Our data suggest that yaws re-emergence after MDA was driven by multiple sources, and therefore we recommend higher population coverage and intensive post-MDA surveillance. These data show that resistance probably evolved once during the trial, rather than being widespread. We recommend careful monitoring of affected communities during post-MDA follow-up in order to rapidly detect new emergence of azithromycin resistance, and the consideration of alternative treatments for cases detected after MDA.

genomic epidemiology investigation. We aimed to determine the epidemiological mechanisms and source of infection leading to re-emergence of yaws cases after MDA. We also investigated the epidemiological links between cases of drug resistance and assessed how macrolide resistance emerged under the selection pressure applied by single-dose MDA.

Methods

Study population

We included all patients with PCR-confirmed *T p pertenue* during a yaws elimination programme conducted between April 1, 2013, and Nov 1, 2016, in Lihir Island, Papua New Guinea, covering a population of more than 16 000. All participants, or their guardians for participants younger than 16 years, provided oral informed consent to be screened and treated during the study. Written informed consent was obtained from all patients with lesions before enrolment. The protocol was approved by the National Medical Research Advisory Committee of the Papua New Guinea Ministry of Health (MRAC number 12.36).

Data and swab collection

We matched laboratory results to case reports containing demographic data (age, sex, place of origin, and place of residence) and clinical data (disease stage and duration). We investigated possible contact with the known drug-resistant cases and recorded sociodemographic risk factors (family relations, regular place of socialising, place of religious worship, and regular travel to or visitors from another village).

Swabs of skin lesions from patients with suspected yaws were taken before MDA (April, 2013; survey round 1), and at 6-monthly surveys throughout the subsequent 42 months (to October, 2016; survey rounds 2–8), as previously described.[4] PCR performed on DNA extracted from swabs identified 239 of 777 samples to be *T p pertenue*-positive,[4,7] and from previous MLST results,[5] 193 samples were fully typed and available. Macrolide resistance single nucleotide polymorphisms (SNPs) were initially detected by restriction fragment length polymorphism (RFLP).[7,8]

Genome sequencing

Of the 193 available samples, we stratified by MLST type, geography, and surveillance round, randomly selecting samples within the strata using computer-generated random lists. We filtered these samples based on a quantitative PCR (qPCR) cycle threshold (Ct) of less than 32, as described previously.[9] Additionally, we selected six samples from five patients regardless of qPCR Ct value, because of their involvement in a suspected transmission chain (based on epidemiology and observed treatment failure).

Whole genome sequencing (WGS) was performed directly on *T p pertenue* DNA extracted from 37 samples using the pooled sequence-capture approach, as previously described (appendix 1 p 3).[6,9]

Phylogenomic analysis

We did a phylogenomic analysis of the study genomes, contextualised using 20 publicly available *T p pertenue* genomes derived from both humans and non-human primates,[6,10–14] and one *T pallidum* subspecies *endemicum* genome as an outgroup (BosniaA;[15] appendix 1 pp 4–6). We mapped sequencing reads to a common reference genome (Samoa D) to generate a multiple sequence alignment, including genomes for which we had sufficient coverage: minimum five reads per site, to call variants for at least 843 104 (74%) of 1 139 330 genomic reference positions. We screened our alignment for recombination using Gubbins v1.4.10,[16] and we used IQ-Tree v1.6.10 to infer a maximum likelihood phylogeny from the resulting recombination-masked SNP alignment.[9,17] We tested the robustness of our tree topology through Bayesian reanalysis using PhyloBayes v3.0 (appendix 1 pp 4–5).[18] To assess the phylogenetic patterns underlying clonal expansion of the JG8 sequence type, we did ancestral reconstruction of the SNPs along the phylogenetic branches within the JG8 lineage and used rPinecone[19] to define phylogenetically clustered samples, as previously described,[9] based on separation by more than four SNPs. To analyse the phylogenomic context of a sample yielding only 54% genome coverage, we subsampled the multiple sequence alignment to include only supported sites for that genome (appendix 1 p 5).

Deep sequencing analysis

To detect mixed resistance alleles and investigate how the A2059G macrolide resistance mutation first evolved, we designed discriminatory PCR primers specific to the regions flanking the individual copies of 23S rDNA (*T p pertenue* has two copies of the gene), and generated amplicons for each operon for six samples (in triplicate) from five patients with resistant strains and two samples (single replicate) with sensitive strains from epidemiologically linked participants (appendix 1 p 3). To investigate the broader prevalence of mixed 23S *T p pertenue* resistance alleles in Lihir patients, we stratified samples by survey round only, selecting 31 samples within the

strata from individuals infected with macrolide-sensitive *T p pertenue*, and used the same method to amplify the *T p pertenue* 23S operons (single replicate). Amplicons were sequenced (Illumina MiSeq, Illumina, San Diego, CA, USA), and the relative proportions of sequencing

***Figure 1:* Maximum likelihood whole genome sequencing phylogeny of 40 *T p pertenue* genomes, including 20 genomes from Lihir, Papua New Guinea**
The JG8 MLST type represents an extremely clonal lineage of *T p pertenue* found in Lihir and the Solomon Islands, with all JG8 samples separated by less than 20 single nucleotide polymorphisms. Types TD6 and SE7 are genomically distant. Three genomes carrying the A2059G mutation are shown. Branches with Ultra-Fast bootstrap values of 95% or higher are labelled. Branches are scaled by nucleotide substitutions per site. Sequence types shown comprise alleles taken from directly amplified and sequenced *T p pertenue* MLST genes or those determined in silico using the genome sequence data. New allelic signatures not described in the Godornes MLST scheme[5] are labelled white. The presence of A2059G mutations was assessed previously by restriction fragment length polymorphism if available,[4] or in silico from the genome sequence. MLST=multilocus sequence typing. *T p pertenue=Treponema pallidum* subspecies *pertenue*.
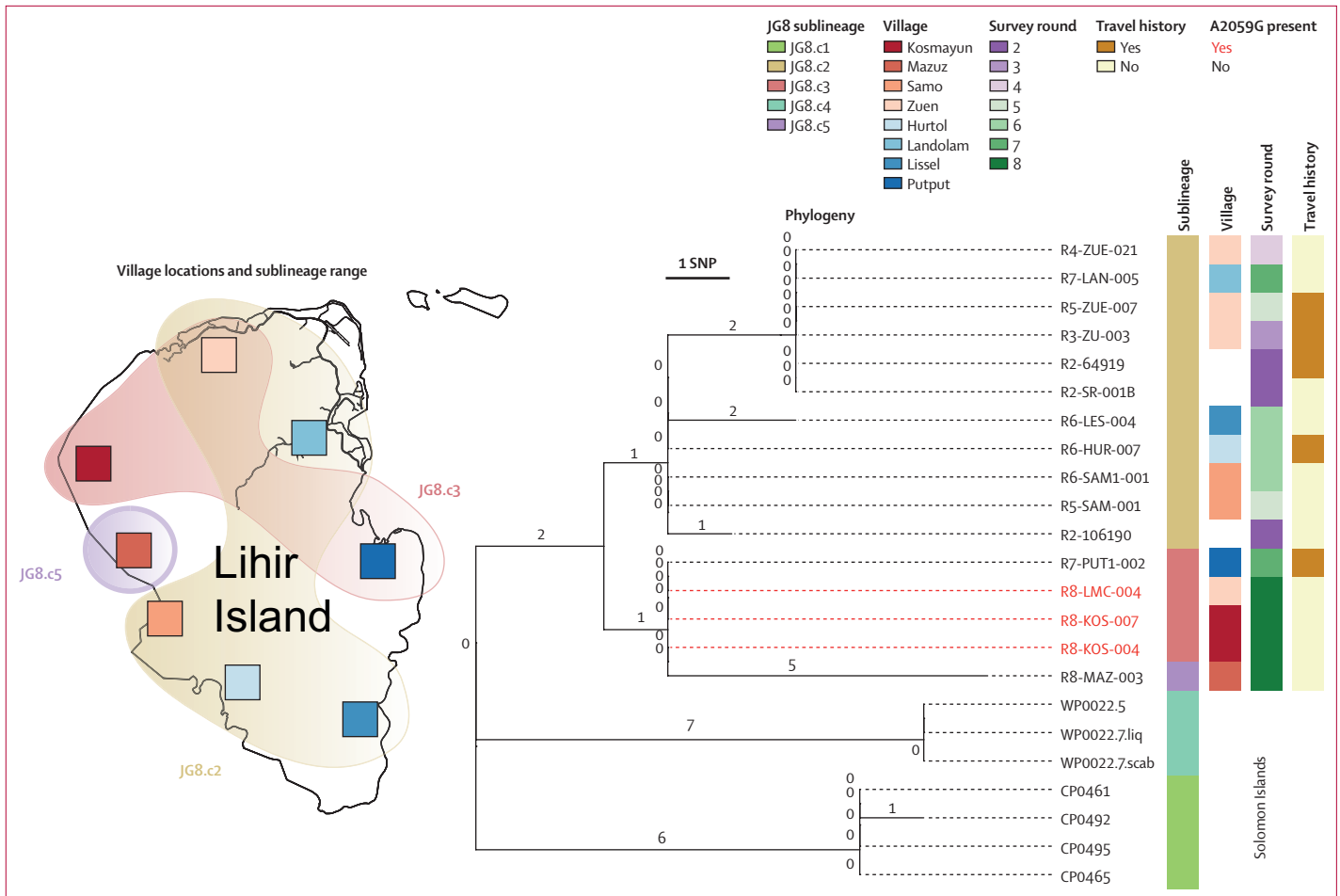
Figure 2: JG8 subtree phylogeny

Phylogeny includes 16 genomes from Lihir, Papua New Guinea, and seven from the Solomon Islands. Branches are scaled and labelled according to the absolute count of non-recombining base substitutions (relative to the parent and child nodes determined by ancestral state reconstruction). Tip labels for macrolide-resistant strains are shown in red. Coloured tracks show phylogenomic sublineage, village, survey round, and whether the patient had a recent history (previous 3 months) of travel off the island. Village data were unavailable for three patient samples (shown in white). The map shows village locations in squares, and the areas of shading show villages where sublineages were detected. SNP=single nucleotide polymorphism.

reads corresponding to each gene copy and each allele (A2058G, A2059G) were analysed, using both allele counting and read haplotype counting approaches (appendix 1 pp 3–4).

Role of the funding source

The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and MAB, NRT, and OM had final responsibility for the decision to submit for publication.

Results

Of 232 patients with *T p pertenue*-positive samples available from the Lihir trial, we selected 72 as examples of *T p pertenue* diversity, covering the full geographic and temporal range of the study period. From these, we performed WGS directly on DNA extracted from clinical swabs for 31 samples with qPCR Ct of less than 32,

generating 17 high-quality genomes. Six additional samples, selected because they were collected from five patients involved in a suspected transmission chain, yielded three high-quality genomes and one partial genome. In total, we recovered 20 near-whole *T p pertenue* genomes with more than 74% genome coverage from the Lihir trial samples (appendix 1 pp 9–11). Refining our previous observations,[9] we recovered no high-quality genomes from samples with qPCR Ct of more than 30 (appendix 1 p 9, appendix 2).

Of 20 genomes recovered, 16 were of the JG8 sequence type (figure 1). In-silico MLST of seven closely related *T p pertenue* genomes from the Solomon Islands[6] determined that these were also sequence type JG8 under the *T p pertenue* MLST scheme.[5] In addition, WGS showed a large putative recombination event, previously described in the Solomon Island *T p pertenue* strains[6] and covering ten coding sequences (spanning TPESAMD_0856a to TPESAMD_0866), was also present

See **Online** for appendix 2

in the JG8 genomes from Lihir (appendix 1 p 12). Phylogenomic analysis of all available *T p pertenue* genomes showed that the JG8 sequence type is extremely clonal (figure 1), with all JG8 genomes separated by fewer than 20 pairwise SNPs, and the topology of the JG8 subtree was built on 28 phylogenetically informative sites (appendix 1 p 13). Although low coverage in some sites might hide SNPs that could affect the JG8 subtree, Bayesian reanalysis confirmed that we had determined the most probable topology (appendix 1 pp 13, 14).

Although the majority of Lihir genomes (16 of 20) recovered were part of the clonal JG8 sequence type, one genome recovered from a sample taken early in MDA follow-up (R3-SAL-007; round 3) was of the TD6 sequence type, separated from the JG8 genomes by a long branch of 133 base substitutions. Three genomes from Lihir were sequence type SE7, forming a distinct phylogenomic clade more closely related to genomes from Indonesia or Western Samoa than to the JG8 sequence type genomes from Lihir (figure 1).

We investigated whether the spread of the JG8 sequence type after MDA could be explained by a single clonal expansion, or through multiple epidemiological routes. Using ancestral reconstruction of the 28 phylogenetically discriminatory SNPs in the JG8 sequence type, we found five distinct phylogenetic clusters (sublineages) separated by at least four SNPs, including three sublineages from Lihir (figure 2; sublineages JG8.c2, JG8.c3, JG8.c5). Sublineage JG8.c2 samples (n=11) were found in five different villages (Hurtol, Landolam, Lissel, Samo, and Zuen) and sublineage JG8.c3 samples (n=5) were found in three villages (Kosmayun, Putput, and Zuen); a sublineage JG8.c5 sample from Mazuz village was separated from the JG8.c3 genomes by five SNPs. All three JG8 sublineages from Lihir were detected in the later rounds of post-MDA follow-up, indicating they were all involved in yaws re-emergence. As figure 2 shows, rather than being the result of a single clonal expansion, the cases had distinct geographical, temporal, and phylogenomic properties, reflecting separate epidemiological and evolutionary histories. Five of eleven JG8.c2 genomes were sampled from patients with a history of recent travel outside of Lihir (in the past 3 months).

We intentionally oversampled and examined a putative transmission network of five cases with treatment failure;[4] all of the *T p pertenue* samples from these cases were of the JG8 sequence type, and all carried the A2059G mutation linked to azithromycin resistance in *T pallidum* subspecies *pallidum*.[8] We identified patient 1 (sample R7-KOS-003) as representing the earliest macrolide-resistant case (index case; June, 2016), with the patient reporting having an ulcer for at least 3 months before formal diagnosis. 7 months earlier (October, 2015), patient 1 had been diagnosed and treated with azithromycin for yaws; RFLP testing showed that this earlier infection was macrolide-sensitive (A2059). Epidemiological tracing identified

potential transmission links between all five resistant cases (figure 3); four of the children were from the same village (Kosmayun), and the index case (patient 1) was a sibling of patient 2, playmate of patients 3 and 4, and school friend of patient 5 (sample R8-LMC-004). This link between patients 1 and 5 was not described previously[4] and was discovered only during follow-up interview. Patient 5 is from Zuen village, 7 km from Kosmayun, but both patients attend a school in Sale, which is equidistant between the villages. Two additional cases from Kosmayun (patients 6 and 7) epidemiologically linked to patient 4 (sibling and playmate) were identified within the timeframe of the investigation, both of whom were infected with macrolide-sensitive (A2059) *T p pertenue*.
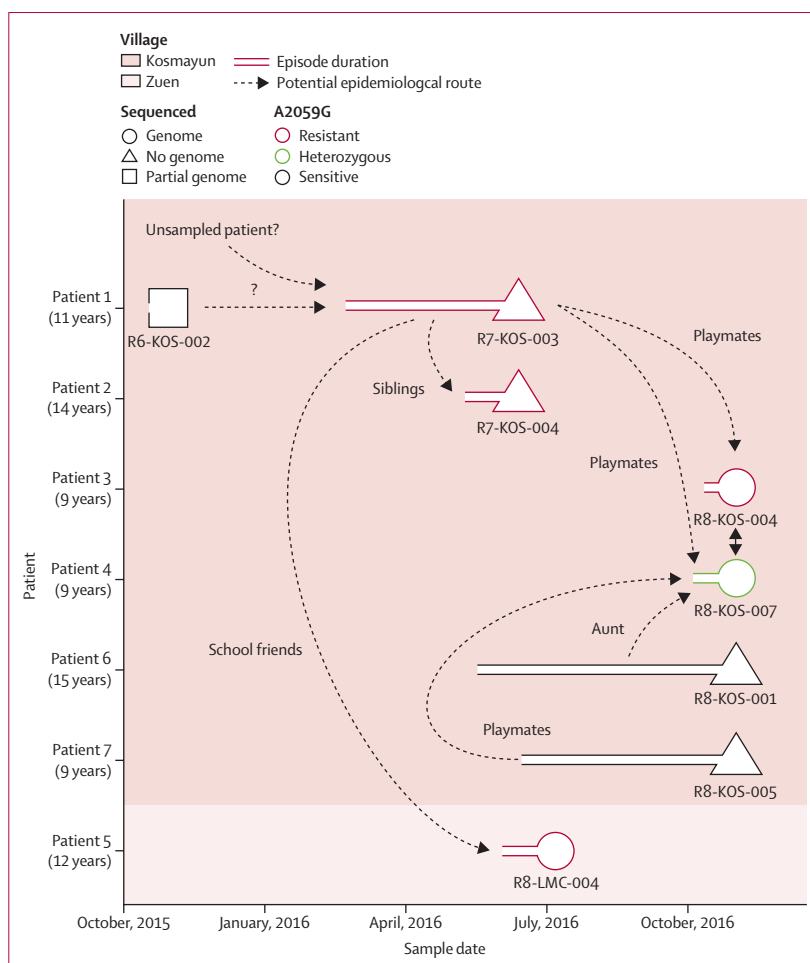

Figure 3: **Epidemiological relationships between macrolide-resistant samples**
Episode duration before sampling is based on patient reporting. Patient 1, living in Kosmayun village, was positive for wild-type *T p pertenue* in October, 2015, but follow-up in June, 2016, identified A2059G resistant *T p pertenue* in patient 1 and their sibling patient 2. A third patient (patient 5) was identified in Zuen village in July, 2016 (school friend of patient 1). Two more patients from Kosmayun (playmates of patient 1) were detected with resistant *T p pertenue* in November, 2016. All patient samples were typed as JG8, and of three near-whole genomes recovered from three patients, all were identical over the length recovered. A partial genome (54% of length covered) was obtained from the index case (patient 1, R6-KOS-002), which was identical to the others over the common regions (appendix 1 p 15). In determining the origin of patient 1's resistant sample, it was not possible to conclusively discriminate between relapsed infection and de-novo reinfection from an unsampled patient. *T p pertenue*=Treponema pallidum subspecies *pertenue*.
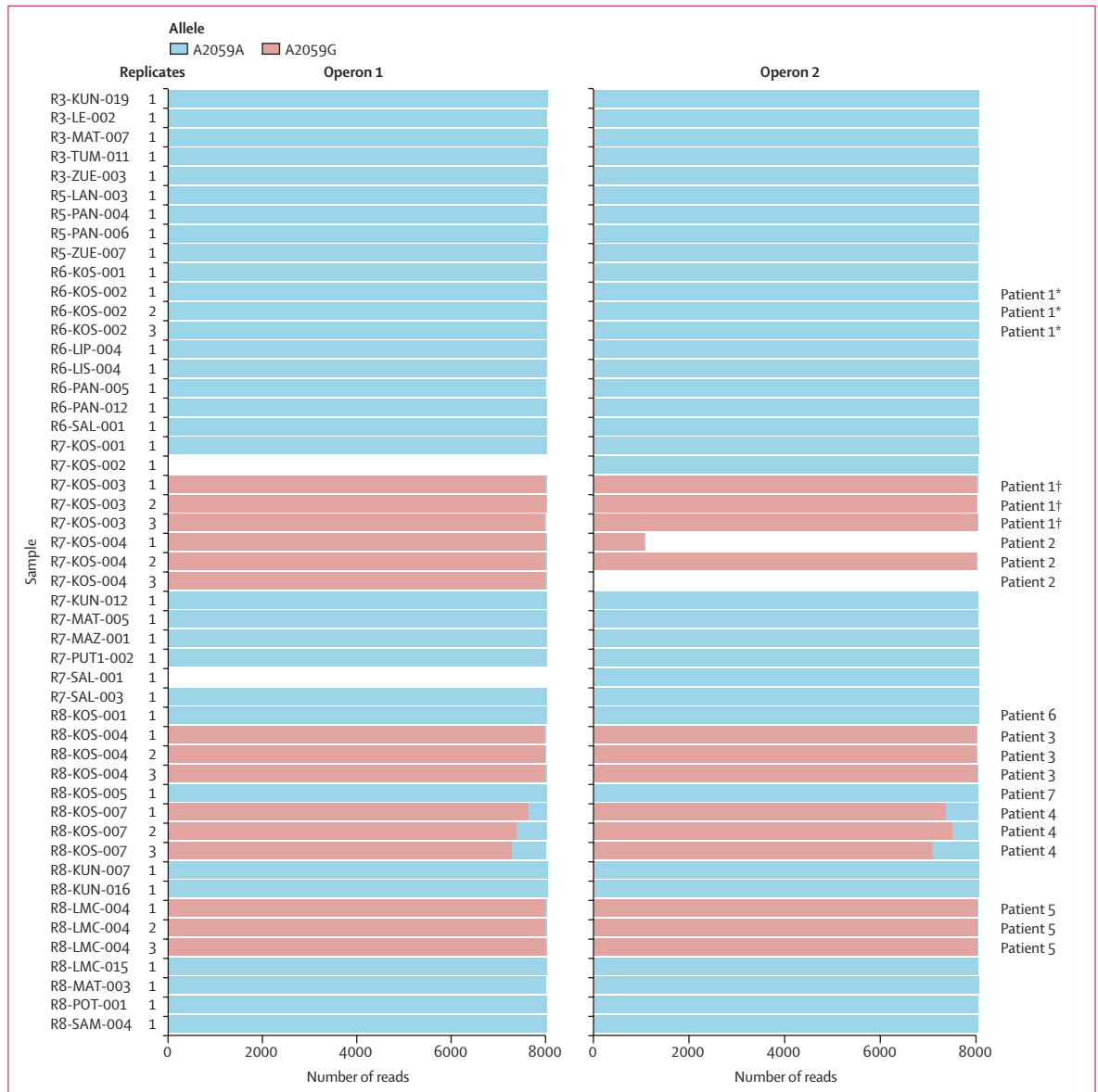
***Figure 4:* Deep amplicon sequencing and evaluation of 23S A2059 allele frequencies**
Samples within the putative transmission network are either fixed at A2059A (blue) or A2059G (red), with the exception of sample R8-KOS-007 (patient 4). There was no evidence of minority resistance alleles in any samples outside of the putative transmission network. Because there are two genomic copies of the 23S gene, all samples were tested using operon-specific PCR amplicons. Samples from the putative transmission network were tested in triplicate and are labelled by patient. Sequencing reactions yielding less than 1000 reads were excluded (shown in white). *Before treatment. †After treatment.

We were able to recover near-complete *T p pertenue* genomes from patients 3, 4, and 5. We compared the 28 discriminatory sites from these three genomes (R8-KOS-004, R8-LMC-004, and R8-KOS-007); in three-way pairwise genome comparisons we identified no SNPs. However, because of insufficient coverage, five or six sites in each pairwise comparison were excluded, and we could not rule out the possibility of a SNP being present at those excluded positions (appendix 1 p 6). The three genomes were otherwise identical. Another JG8 genome (R7-PUT1–002) was also

identical to the other three samples over the 28 sites (apart from 3–5 undetermined sites), but this sample lacked either an epidemiological link to the other cases, or the resistance allele (variants in the 23S region, including A2059G, were excluded from pairwise counts). These four genomes together comprise sublineage JG8.c3 in figure 2. We attempted to sequence *T p pertenue* from the pretreatment sample from October, 2015, for patient 1 (R6-KOS-002; lacking the macrolide resistance allele), but due to low pathogen load we were only able to recover 54% (625 633 of

1139 330 sites) of the genome; R6-KOS-002 was identical to R8-KOS-007 and R8-KOS-004 over the high-quality common recovered regions, and phylogenetic reconstruction using the 625 high-quality SNP sites from this genome replicated the sample groupings of figures 1 and 2 (appendix 1 p 15).

We used deep amplicon sequencing (figure 4, appendix 1 p 16, appendix 3) to investigate the evolution of *T p pertenue* drug-resistance variants in patient 1. For the earlier wild-type (A2059) sample (R6-KOS-002), we obtained sequencing coverage of more than 8000× for each 23S rRNA allele, with three technical replicates. A maximum of 18 reads (<0·22% A2059G) supported the A2059G mutation in any one replicate, which was well below our threshold for minority variant calling (1%) and equivalent to background sequencing error (figure 4). This result could be because the mutation was present but below our limit of detection; alternatively, the mutation could have appeared subsequent to this sample being taken, or it could represent a case of reinfection, transmitted from an unsampled individual. Deep sequencing from patient 1's post-treatment sample, as well as three of four contact cases with A2059G identified using RFLP, showed a fixed genotype, with more than 99% of sequencing reads (from a minimum of 797 reads) carrying the A2059G mutation at both 23S copies (figure 4). In patient 4 (sample R8-KOS-007), we found only 91% of reads (range 87–95% over three technical replicates, consistent for both gene copies) corresponded to A2059G. Manual examination of both the deep sequencing amplicon reads and the WGS reads (derived from a different DNA extract of the same swab) showed a mixture of correctly paired and mapped reads with both alleles at position A2059, consistent with a heterozygous SNP (ie, a mixed population of sensitive and resistant bacteria). Further analysis of read haplotypes also supported that this sample contained a mixed population of resistant and sensitive bacteria (appendix 1 p 16). We had insufficient coverage from WGS to robustly probe genome-wide minority variants and exclude the possibility of co-infection in this patient. Deep sequencing of samples from patients 6 and 7, who were close contacts of patient 4, showed that these were fixed wild-type (A2059).

To investigate the prevalence of minority *T p pertenue* drug-resistance variants more broadly, 31 samples were selected from throughout the post-MDA surveillance period, except rounds 1, 2, and 4 due to insufficient residual DNA (appendix 1 pp 10–11), from participants residing in 15 of 28 villages on Lihir Island. Samples were taken from nine female and 22 male participants (median age 9 years [IQR 7–12]), with four of seven MLST types represented. Using the same deep amplicon sequencing approach (but with a single technical replicate), we found no evidence for low-frequency macrolide resistance at either position A2058G or A2059G in any of these samples (figure 4).

Discussion

Building upon our previous work in which we resolved yaws molecular diagnostic failures through genomics,[6] in this study, we used the increased resolution afforded by WGS phylogenomic analyses to investigate post-MDA yaws re-emergence and macrolide resistance. Single-dose mass treatment with azithromycin has been a valuable tool in low-income and middle-income countries, directed towards the elimination of both trachoma[20,21] and yaws.[22,23] In the context of yaws eradication efforts, re-emergence of cases after treatment and bacterial drug resistance represent critical issues with this strategy.

See **Online** for appendix 3

Although a small proportion of cases identified during follow-up were of genetically distinct MLST types (eg, SE7), consistent with importation of strains from outside Lihir,[4] the majority of re-emergent cases were of the endemic JG8 sequence type. The MDA intervention had a lasting effect on the *T p pertenue* population structure in Lihir, with a diverse *T p pertenue* population being replaced by a clonal expansion of the JG8 sequence type, the most common lineage before the MDA.[4] Using high-resolution WGS, we showed that the re-emerging JG8 samples were split into at least three distinct phylogenetic sublineages. Contextualised by our previous estimate that the mean molecular clock rate of *T pallidum* is approximately one SNP every 4–5 years,[9] the geographical dispersal and genetic separation of samples such as R8-MAZ-003 in sublineage JG8.c5 from other samples by five SNPs is indicative of distinctly different evolutionary histories within the JG8 sequence type. These sublineages support clinical observations that MDA treatment escape is related to multiple epidemiological routes rather than a single point source (either local or imported), consistent with multiple cases of missed MDA, followed by reactivation of latent *T p pertenue*.

We note that five of 11 patients infected with sublineage JG8.c2, the largest and most geographically dispersed JG8 sublineage studied, had a history of travel away from Lihir. This observation could suggest importation of the JG8.c2 sublineage, but we had insufficient sampling to address this.

We described the genomic relatedness of *T p pertenue* among the first five people diagnosed with macrolide-resistant yaws, representing 2·4% (five of 209) of post-MDA cases. Macrolide resistance in *T pallidum* is thought to be caused by one of two mutations targeting the two genomic copies of the *T p pertenue* 23S gene: A2058G has been linked to multiple independent evolutionary events in *T p pallidum*,[9] whereas A2059G (the only variant detected in this study) is less common. In this study, all individuals with resistant *T p pertenue* were epidemiologically linked, had the same A2059G mutation, and JG8.c3 genomes sequenced from three cases were identical over the supported sites. Given the thoroughness of the surveillance and RFLP resistance testing in the post-MDA study population, it is unlikely that we failed to identify additional resistant *T p pertenue* cases. The most parsimonious

interpretation of these data is that a single, de-novo evolutionary mutation conferred macrolide resistance, followed by onward transmission. An alternative explanation for the cluster of macrolide-resistant *T p pertenue* described in this study would be independent A2059G (but not A2058G) resistance mutations arising in five separate patients infected with genomically identical *T p pertenue*, with the epidemiological linkage being coincidental.

We did not detect the A2059G mutation in the index case (patient 1) before treatment, and this patient's initial lesion resolved after treatment with azithromycin, suggesting the mutation was acquired after commencement of treatment. This acquisition could either be due to selection of de-novo resistance in the original strain, or through reinfection from an unsampled individual during the latter months of the treatment phase, when the drug levels in tissue were waning. Such a scenario of reinfection during the subtherapeutic period might explain selection of a resistant subpopulation.

Independent analyses of the sample from patient 4 indicated that this patient's *T p pertenue* was heterozygous for the macrolide resistance allele in both copies of 23S rRNA. Assuming this epidemiologically linked patient was indeed part of the same transmission network (rather than a second de-novo evolutionary event or transmission from an unsampled resistant case), we envisage two possible explanations for this heterozygosity. First, the patient could have been co-infected with a second wild-type JG8 *T p pertenue* strain; we were unable to exclude the possibility of co-infection through minority variant analysis of the WGS reads. Of note, patient 4 shared close epidemiological links with patients 6 and 7, who were both infected with wild-type JG8 sequence-type *T p pertenue*, making co-infection a plausible scenario. Second, the patient's *T p pertenue* A2059G allele could have reverted to wild-type. During this study, we identified samples with a qPCR Ct as low as 23·8 (13 003 copies/µL; R2-SR-001B; appendix 1 p 9, appendix 2), obtained from a DNA extraction of 100 µL. Ulcer swabs can therefore contain more than 1 300 000 bacterial *T p pertenue* copies, and this represents a small subsample of the bacteria present in a yaws ulcer. We therefore consider it likely that, given time, such a mutation to wild-type A2059 could have occurred in a lesion simply by chance. However, for any mutation to reach a frequency of 9% in a lesion (as detected in patient 4) implies selective pressure (eg, fitness cost). We had insufficient data to conclusively discriminate between these two scenarios, but we consider co-infection to be a more likely scenario.

Although resistance is likely to have evolved only once in this study, there is clearly a risk of further emergence. Our broader sampling and deep sequencing from patients throughout the survey period showed that, in this study, *T p pertenue* samples that were sensitive according to RFLP methods did not harbour A2058G or A2059G mutations, even at low frequency. Although

these mutations might occur naturally within a patient at low frequency, we would expect them to be selectively neutral or slightly detrimental, and to disappear within a few generations without the selective pressure of drug exposure to drive them to fixation. It will be important in future work to characterise appropriate thresholds at which a minority resistance allele is likely to evolve into a resistant case after drug exposure, as has been done in the HIV field.[24] Analysis of global syphilis lineages suggests that once macrolide resistance evolves and is selected for in a patient, it can persist in lineages,[9] without any evidence of a fitness cost of possessing such variants. Broader MDA with azithromycin could therefore promote resistance in *T p pertenue*, and consideration should therefore be given to alternative treatment options and resistance mitigation strategies.

Our study has several strengths. First, we combined detailed epidemiological analysis of spatiotemporal data and phylogenomic reconstruction of bacterial genomes, resulting in the best evidence available to understand yaws transmission and bacterial population evolution after MDA. The use of genomics enabled previously unachievable levels of strain discrimination, allowing us to determine that resurgent JG8 cases originated from multiple sources rather than a single point source. We recovered 20 near-complete *T p pertenue* genomes, substantially increasing the number of published genomes. However, the inherent limitations of high contamination and low bacterial load in yaws swabs prevented us from sampling the population more completely, such that we might have missed unsampled diversity. Furthermore, although we attempted to make our sample selection representative of the population structure in Lihir, sampling was not random and laboratory staff were not blinded to clinical or demographic data, which could have introduced bias. Second, we recovered near-complete *T p pertenue* genomes from three macrolide-resistant cases, and moving beyond MLST-level groupings, we showed that these *T p pertenue* genomes were virtually identical, indicating that they share a recent common ancestor and perhaps a single chain of transmission. Finally, we applied deep amplicon sequencing to search for minority *T p pertenue* resistance alleles, showing that RFLP-tested macrolide sensitive samples do not harbour resistant *T p pertenue* 23S minority alleles, as well as providing evidence supporting the de-novo evolution of the *T p pertenue* A2059G mutation within a single patient or lesion.

Our study provides new scientific data about the characteristics, epidemiology, and transmission dynamics of yaws. The genomic dissection renders compelling, high-resolution evidence on the factors that underpin yaws re-emergence and drug resistance after a single round of MDA. These data will support improvement of policies and strategies to eradicate yaws, and will also inform strategy for MDA campaigns in other diseases for which population data are sparse. As part of the global eradication

initiative, MDA campaigns against yaws have the potential to significantly reduce transmission of infection in many high-prevalence areas. In these settings, flexible and urgent solutions are needed for a rational approach to counteract recrudescence of infection and drug resistance. We recommend high MDA coverage to reduce the number of missed people with active or latent yaws (eg, non-cooperation, people working elsewhere), and intensive post-MDA surveillance for the detection of the residual yaws cases and to control importation and onward transmission of new infections. We also recommend careful monitoring of affected communities during MDA follow-up in order to rapidly detect and retreat individuals with clinical treatment failure using alternative treatments (eg, benzathine benzylpenicillin). If emergence of azithromycin resistance is confirmed in the lab, then we recommend implementing a comprehensive strategy to mitigate resistance dissemination, including treatment of all new yaws cases and their contacts in the affected area with alternative regimens. Genomic data can play a central role in monitoring and anticipating the dynamics of transmission of yaws to minimise the risk of re-emergence.

Contributors
MAB, MM, SAL, and OM did the literature search. MAB, FG-C, MM, SAL, NRT, and OM designed the study. CG-B, MC-M, AKJ, WH, JW, RP, MM, and OM collected the data. MAB, MN-J, CG, MC, CG-B, and MP did the laboratory work. MAB and MN-J analysed the data. MAB prepared the figures. MAB, MN-J, FG-C, SAL, NRT, and OM interpreted the data. MAB, MN-J, SAL, NRT, and OM wrote the manuscript. All authors reviewed and approved the final manuscript.

Data sharing
Raw sequencing reads from whole genome sequencing were deposited at the European Nucleotide Archive under BioProject PRJEB34799. Raw sequencing reads from deep amplicon sequencing were deposited at the Short Read Archive under BioProject PRJNA575636. All accessions used in this project are listed in appendices 2 and 3, with the appropriate sample identifier. The maximum likelihood phylogeny and metadata from this study are also available for interactive exploration in a Microreact project.

References
1 Mitjà O, Asiedu K, Mabey D. Yaws. *Lancet* 2013; **381:** 763–73.
2 Mitjà O, Marks M, Konan DJP, et al. Global epidemiology of yaws: a systematic review. *Lancet Glob Health* 2015; **3:** e324–31.
3 WHO. Eradication of yaws—the Morges strategy. *Wkly Epidemiol Rec* 2012; **87:** 189–94.
4 Mitjà O, Godornes C, Houinei W, et al. Re-emergence of yaws after single mass azithromycin treatment followed by targeted treatment: a longitudinal study. *Lancet* 2018; **391:** 1599–607.
5 Godornes C, Giacani L, Barry AE, Mitja O, Lukehart SA. Development of a Multilocus Sequence Typing (MLST) scheme for *Treponema pallidum* subsp. *pertenue*: application to yaws in Lihir Island, Papua New Guinea. *PLoS Negl Trop Dis* 2017; **11:** e0006113.
6 Marks M, Fookes M, Wagner J, et al. Diagnostics for yaws eradication: insights from direct next-generation sequencing of cutaneous strains of *Treponema pallidum. Clin Infect Dis* 2018; **66:** 818–24.
7 Lukehart SA, Godornes C, Molini BJ, et al. Macrolide resistance in *Treponema pallidum* in the United States and Ireland. *N Engl J Med* 2004; **351:** 154–58.
8 Matějková P, Flasarová M, Zákoucká H, et al. Macrolide treatment failure in a case of secondary syphilis: a novel A2059G mutation in the 23S rRNA gene of *Treponema pallidum* subsp *pallidum. J Med Microbiol* 2009; **58:** 832–36.
9 Beale MA, Marks M, Sahi SK, et al. Genomic epidemiology of syphilis reveals independent emergence of macrolide resistance across multiple circulating lineages. *Nat Commun* 2019; **10:** 3255.
10 Cejková D, Zobaníková M, Chen L, 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.
11 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.
12 Strouhal M, Mikalová L, Havlíčková P, 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.
13 Strouhal M, Mikalová L, Haviernik J, et al. Complete genome sequences of two strains of *Treponema pallidum* subsp *pertenue* from Indonesia: modular structure of several treponemal genes. *PLoS Negl Trop Dis* 2018; **12:** e0006867.
14 Knauf S, Gogarten JF, Schuenemann VJ, et al. Nonhuman primates across sub-Saharan Africa are infected with the yaws bacterium *Treponema pallidum* subsp *pertenue. Emerg Microbes Infect* 2018; **7:** 157.
15 Staudová B, Strouhal M, Zobaníková M, 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 Negl Trop Dis* 2014; **8:** e3261.
16 Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015; **43:** e15.
17 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; **32:** 268–74.
18 Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 2009; **25:** 2286–88.
19 Wailan AM, Coll F, Heinz E, et al. rPinecone: define sub-lineages of a clonal expansion via a phylogenetic tree. *Microb Genomics* 2019; **5:** e000264.
20 West ES, Munoz B, Mkocha H, et al. Mass treatment and the effect on the load of *Chlamydia trachomatis* infection in a trachoma-hyperendemic community. *Invest Ophthalmol Vis Sci* 2005; **46:** 83–87.
21 West SK, Munoz B, Mkocha H, et al. Infection with *Chlamydia trachomatis* after mass treatment of a trachoma hyperendemic community in Tanzania: a longitudinal study. *Lancet* 2005; **366:** 1296–300.
22 Mitjà O, Hays R, Ipai A, et al. Single-dose azithromycin versus benzathine benzylpenicillin for treatment of yaws in children in Papua New Guinea: an open-label, non-inferiority, randomised trial. *Lancet* 2012; **379:** 342–47.
23 Marks M, Mitjà O, Bottomley C, et al. Comparative efficacy of low-dose versus standard-dose azithromycin for patients with yaws: a randomised non-inferiority trial in Ghana and Papua New Guinea. *Lancet Glob Health* 2018; **6:** e401–10. .
24 Vandenhende M-A, Bellecave P, Recordon-Pinson P, et al. Prevalence and evolution of low frequency HIV drug resistance mutations detected by ultra deep sequencing in patients experiencing first line antiretroviral therapy failure. *PLoS One* 2014; **9:** e86771.