University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Taouk, ML;Taiaroa, G;Pasricha, S;Herman, S;Chow, EPF;Azzatto, F;Zhang, B;Sia, CM;Duchene, S;Lee, A;Higgins, N;Prestedge, J;Lee, YW;Thomson, NR;Graves, B;Meumann, E;Gunathilake, M;Hocking, JS;Bradshaw, CS;Beale, MA;Howden, BP;Chen, MY;Fairley, CK;Ingle, DJ;Williamson, DA

Title:
Characterisation of Treponema pallidum lineages within the contemporary syphilis outbreak in Australia: a genomic epidemiological analysis

Date:
2022-06-01

Citation:
Taouk, M. L., Taiaroa, G., Pasricha, S., Herman, S., Chow, E. P. F., Azzatto, F., Zhang, B., Sia, C. M., Duchene, S., Lee, A., Higgins, N., Prestedge, J., Lee, Y. W., Thomson, N. R., Graves, B., Meumann, E., Gunathilake, M., Hocking, J. S., Bradshaw, C. S. ,... Williamson, D. A. (2022). Characterisation of Treponema pallidum lineages within the contemporary syphilis outbreak in Australia: a genomic epidemiological analysis. Lancet Microbe, 3 (6), pp.e417-e426. https://doi.org/10.1016/S2666-5247(22)00035-0.

Persistent Link:
https://hdl.handle.net/11343/309256

License:
CC BY-NC-ND

Characterisation of *Treponema pallidum* lineages within the contemporary syphilis outbreak in Australia: a genomic epidemiological analysis

Mona L Taouk, George Taiaroa*, Shivani Pasricha*, Sophie Herman*, Eric P F Chow, Francesca Azzatto, Bowen Zhang, Cheryll M Sia, Sebastian Duchene, Alvin Lee, Nasra Higgins, Jacqueline Prestedge, Yi Wei Lee, Nicholas R Thomson, Bianca Graves, Ella Meumann, Manoji Gunathilake, Jane S Hocking, Catriona S Bradshaw, Mathew A Beale, Benjamin P Howden, Marcus Y Chen, Christopher K Fairley, Danielle J Ingle, Deborah A Williamson*

Summary

Background The incidence of syphilis has increased markedly in the past decade in high-income countries, including Australia. To date, however, genomic studies of *Treponema pallidum* have focused mainly on the northern hemisphere. Here, we aimed to characterise the lineages of *T pallidum* driving the current syphilis epidemic in Australia.

Methods In this genomic epidemiological analysis, using phylogenomic and phylodynamic analyses, we analysed 456 high-quality *T pallidum* genomes collected from clinical samples in Australia between Oct 19, 2005, and Dec 31, 2020, and contextualised this information with publicly available sequence data. We also performed detailed genomic characterisation of putative antimicrobial resistance determinants, in addition to correlating single-locus typing of the TP0548 allele with the *T pallidum* phylogeny.

Findings Phylogenomic analyses identified four major sublineages circulating in Australia and globally, two belonging to the SS14 lineage, and two belonging to the Nichols lineage. Australian sublineages were further delineated into twelve subgroups, with five of the six largest subgroups associated with men who have sex with men, and the sixth lineage was predominantly associated with heterosexual people. Most Australian *T pallidum* genomes (398 [87%] of 456) were genotypically macrolide resistant, and TP0548 typing correlated significantly with *T pallidum* genomic subgroups.

Interpretation These findings show that the current syphilis epidemic in Australia is driven by multiple lineages of *T pallidum*, rather than one distinct outbreak. Major subgroups of *T pallidum* in Australia have emerged within the past 30 years, are closely related to global lineages, and circulate across different sexual networks. In conjunction with improved testing and treatment, these data could better inform the control of syphilis in Australia.

Funding National Health and Medical Research Council, Australian Research Council.

Copyright © 2022 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Introduction

Syphilis is a major public health problem globally, with WHO estimating 6 million new infections worldwide in 2016.[1] Untreated, syphilis can result in serious morbidity and mortality including ocular syphilis, neurosyphilis, and congenital infection.[2] The causative agent of syphilis, *Treponema pallidum* subspecies *pallidum* (TPA) cannot be cultured routinely in vitro, making large-scale molecular epidemiological studies technically challenging, and limiting our understanding of the current syphilis epidemic. Recently, targeted whole-genome sequencing has been applied to improving our understanding of the emergence and evolution of TPA.[3–5] These studies have demonstrated co-circulation of two major TPA lineages (Nichols and SS14) across multiple continents, with a large population expansion of TPA lineages during the early 2000s, consistent with the re-emergence of syphilis globally.[3,4]

In Australia, as in other high-income countries, the incidence of syphilis has increased dramatically over the past decade, with a 358% increase in new syphilis diagnoses in Australia between 2009 and 2019.[6] The syphilis epidemic in Australia has occurred disproportionately in two major populations, first in young Aboriginal and Torres Strait Islander people, and second in urban areas among men who have sex with men (MSM).[7,8] Since 2015, there has also been an increase in new syphilis diagnoses in females and heterosexual males in Australia, with cases of congenital syphilis re-emerging as a public health issue.[9] To date, there are scarce data from Australia on the circulating lineages of TPA driving the contemporary syphilis epidemic. It is unknown whether there are shared lineages between the syphilis epidemic in remote northern Australia and the concurrent epidemic in urban centres. Knowledge of factors that promote the emergence and transmission

Lancet Microbe 2022; 3: e417–26

Published Online April 20, 2022 https://doi.org/10.1016/S2666-5247(22)00035-0

*Contributed equally

Department of Microbiology and Immunology (M L Taouk BBiomed [Hons], G Taiaroa BSc [Hons], S Pasricha PhD, S Herman MSc, F Azzatto BSc [Hons], B Zhang MSc, C M Sia MSc, S Duchene PhD, J Prestedge BSc [Hons], Y W Lee BSc [Hons], Prof B P Howden PhD, D J Ingle PhD), **Department of Infectious Diseases** (M L Taouk, G Taiaroa, S Pasricha, S Herman, B Zhang, J Prestedge, Prof D A Williamson PhD), **and Microbiological Diagnostic Unit Public Health Laboratory** (Prof B P Howden), **The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia; Melbourne Sexual Health Centre, Alfred Health, Melbourne, VIC, Australia** (E P F Chow PhD, Prof J S Hocking PhD, Prof C S Bradshaw PhD, Prof M Y Chen PhD, Prof C K Fairley PhD); **Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC, Australia** (E P F Chow, Prof C S Bradshaw, Prof M Y Chen, Prof C K Fairley); **Centre for Epidemiology and Biostatistics** (E P F Chow), **Melbourne School of Population and Global Health** (Prof J S Hocking), **The University of Melbourne, Melbourne, VIC, Australia; Victorian Infectious Diseases Reference Laboratory, Royal Melbourne Hospital, Melbourne, VIC, Australia** (F Azzatto, Prof D A Williamson);

Victorian Department of Health, Melbourne, VIC, Australia (A Lee MPH, N Higgins MEpi)**; Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK** (Prof N R Thomson PhD, M A Beale PhD)**; London School of Hygiene & Tropical Medicine, London, UK** (Prof N R Thomson)**; Unit of Infectious Diseases, Royal Brisbane and Women's Hospital, Herston, QLD, Australia** (B Graves MBBS)**; Territory Pathology, Royal Darwin Hospital, Darwin, NT, Australia** (E Meumann MBBS FRACP)**; Global and Tropical Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia** (E Meumann)**; Sexual Health and Blood Borne Virus Unit, Centre for Disease Control, Northern Territory Department of Health, NT, Australia** (M Gunathilake MD)**; The Kirby Institute, University of New South Wales, Kensington, NSW, Australia** (M Gunathilake)

Correspondence to: Prof Deborah A Williamson, Department of Infectious Diseases, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne 3000, VIC, Australia
deborah.williamson@unimelb.edu.au

Research in context

Evidence before this study

To date, phylogenomic studies exploring the epidemiology and evolutionary dynamics of *Treponema pallidum* have been limited due to the difficulty in culturing the bacterium in vitro and the low levels of pathogen load in clinical samples making whole-genome sequencing technically challenging. A search of primary research articles published in PubMed up until July 27, 2021, using the search terms "Syphilis" or "*Treponema pallidum*" and "genomic" or "genome(s)" or "sequencing" returned 15 studies describing whole-genomic sequencing of *Treponema pallidum,* subspecies *pallidum*. Most studies were from North America and Europe and only two studies integrated epidemiological data on sexual networks with genomic data.

Added value of this study

In this observational genomic epidemiology study we identified and characterised the main lineages of *T pallidum* circulating in Australia. We identified two major sublineages, both of which have undergone population expansion in the past decade. Further, we showed that within these sublineages, there are several distinct putative transmission networks associated with different sexual exposure profiles, with possible bridging across different sexual networks. The addition of 456 *T pallidum* genomes from Australia represents a substantial addition to the global genomic data available for this important sexually transmitted infection. Moreover, the integration of genomic and epidemiological data provides insights into the emergence of syphilis in Australia, with globally prevalent lineages circulating in both urban and remote areas in Australia.

Implications of all the available evidence

Our findings reinforce previous work that the contemporary syphilis epidemic is mainly due to lineages that have emerged since the latter part of the 20th century. We showed the feasibility of integrating genomic and epidemiological data for investigating the dissemination of *T pallidum*, although real-time assessment of transmission is largely precluded by the slow evolutionary rate of *T pallidum*. To enhance genomic studies of *T pallidum*, more clinical samples are required, which could be achieved by improving access to PCR testing.

of particular lineages might enable improved targeting of public health interventions towards those populations at highest risk of disease. Research conducted in 2019 in our setting (contemporary, urban Australia, specifically Victoria) has combined genomic analysis and sexual behavioural data to demonstrate distinct transmission networks of other sexually-transmitted pathogens such as *Neisseria gonorrhoeae* and *Shigella* spp.[10,11] To gain insights into the syphilis epidemic in Australia, we aimed to apply targeted whole-genome sequencing to clinical samples containing TPA over a 15 year time span, covering the contemporary re-emergence of syphilis in Australia.

Methods

Setting and data sources

In this genomic epidemiological analysis, we used routine clinical samples and derived nucleic acids from individuals who had a positive TPA PCR (appendix 1 p 3) performed at the Victorian Infectious Diseases Reference Laboratory between Jan 1, 2015, and Dec 31, 2020. All exclusion and inclusion criteria for samples are described in appendix 1 (p 3). For historical context, we included ten randomly collected clinical samples of TPA from Victoria each year between 2005 and 2014. Deidentified demographic information was provided by the Victorian Department of Health (appendix 1 p 3). Additionally, we included available clinical samples collected from individuals in the Northern Territory, Australia, between 2016 and 2019. For phylogeographic context, we included sequence reads from all other published genomic studies of TPA globally (appendix 2).[3–5,12–16] Data were collected in accordance with the Victorian Public Health and Wellbeing Act 2008.[17] Ethical approval was obtained from the Alfred Hospital Ethics Committee (project 728/18) and the Human Research Ethics Committee of the Northern Territory Department of Health and Menzies School of Health Research (HREC-2019-3487).

For more on **minimap2 (v2.1)** see https://github.com/lh3/minimap2

See **Online** for appendix 1

For more on **IQ-TREE (v2.0.3)** see https://github.com/iqtree/iqtree2

For more on **rhierbaps (v1.0.1)** see https://github.com/gtonkinhill/rhierbaps

See **Online** for appendix 2

Whole-genome sequencing

Whole-genome sequencing was performed directly on TPA DNA extracted from 700 samples using a hybridisation capture approach, based on 120 nucleotide RNA baits (SureSelect Target enrichment system, Agilent, Santa Clara, CA, USA) used previously (appendix 1 pp 3–4).[4,18] Samples generated for this study were prefiltered to select reads that mapped to the TPA SS14 reference genome (accession CP004011.1) using minimap2 (v2.1) with default parameters (appendix 1 pp 11–12).[19] Quality control steps are described in appendix 1 (pp 4–5).

Phylogenetic and phylodynamic analyses

A total of 980 TPA genomes (appendix 2) were aligned to the SS14 reference genome (CP004011.1).[20] Additionally, a separate alignment of all high-quality Australian genomes was made using the same SS14 reference genome (appendix 1 pp 4–5). Maximum likelihood phylogenetic trees were inferred using IQ-TREE (v2.0.3), allowing for the built-in ModelFinder to infer the optimal model, followed by 10 000 UltraFast bootstraps (v2.0.3).[21,22] To identify lineages, rhierbaps (v1.0.1) was used, with the masked alignments as input (appendix 1 p 5).

The final alignments for phylodynamic analyses (for the global dataset and the Australian dataset) were analysed in BEAST (v1.10.4).[23] The highest supported

model was a GTR+Γ substitution model, relaxed lognormal clock with a coalescent exponential tree prior. Formal testing for temporal signal was assessed using a Bayesian evaluation of temporal signal.[24] Phylodynamic methods are described further in appendix 1 (pp 6–7).

In silico epidemiological typing and putative antimicrobial resistance determinants

Macrolide resistance-conferring variants were defined using a mapping-based approach, with sample reads aligned to the 23S rRNA gene sequence from the Nichols reference genome (CP004010.2). Additional characterisation was performed to identify non-synonymous mutations, and assess their functional significance, in penicillin-binding protein genes (*pbp1, pbp2, mrcA*), a putative β-lactamase *Tp47*, and a putative membrane protein (TP0548) using gene sequences corresponding to the Nichols genome (CP004010.2). Allele designations for TP0548 were made relative to alphabetic subtypes as described elsewhere.[25] TP0548 subtypes were correlated against the global phylogeny. Full details are described in appendix 1 (p 7).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Of the 814 samples collected from patients between Oct 19, 2005, and Dec 31, 2020, after quality control processing, 456 genomes from Australia were included in the subsequent analyses (appendix 1 p 13; appendix 2). Of the sequences associated with demographic data, 410 (95%) of 431 samples were from men and 21 (5%) of 431 were from women. Of the men who reported the sex of their sex partner (384 [94%] of 410), the majority were MSM only (327 [85%] of 384; table). The median age of individuals with included samples was 33 years (IQR 28–43); most individuals were in the 30–39-year age group. Most samples (375 [82%] of 456) were from sexual health clinics (table). Overall, the demographic characteristics of the dataset represented the broader epidemiological characteristics of syphilis in Victoria, Australia (appendix 1 p 24).

In addition to the 456 genomes from Australia, 748 genomes from previously published studies were downloaded; of these, 524 high-quality genomes were included in further analysis (appendix 1 p 13; appendix 2). Following removal of repeat and recombinant regions, the processed alignment was used to infer a maximum likelihood phylogeny (appendix 1 pp 14–15). Similar to previous work, bootstrap values supported two divergent major lineages, the SS14 lineage (previously described as SS14Ω) and the Nichols lineage (appendix 1 pp 14–15). The SS14 and Nichols lineages were separated by 111 pairwise single nucleotide polymorphisms (SNPs).[4,5]

	Individuals with *Treponema pallidum* samples (n=456)
Location	
Victoria	431 (95%)
Northern Territory	25 (5%)
Sex	
Male	410 (90%)
Female	21 (5%)
Unknown	25 (5%)
Age at diagnosis, years	
All	33 (28–43)
Male	34 (28–44)
Female	27 (23–35)
Referring clinic	
Sexual health clinic	375 (82%)
Primary care	26 (6%)
Unknown	55 (12%)
Sexual exposure, self-reported	
MSM	335 (73%)
MSMO	327 (98%)
MSMW	8 (2%)
MSW	49 (11%)
Women	21 (5%)
Not reported or unknown	51 (11%)
HIV status	
HIV positive	80 (18%)
HIV negative	302 (66%)
Not reported or unknown	74 (16%)

Data are n (%) or median (IQR). MSM=men who have sex with men. MSMO=men who have sex with men only. MSMW=men who have sex with men and women. MSW=men who have sex with women only.

Table: **Characteristics of individuals with *Treponema pallidum* samples from Australia included in this study**

Of the 980 genomes, 759 (77%) belonged to the globally dominant SS14 lineage and 221 (23%) belonged to the Nichols lineage. Of the 456 Australian genomes, 339 (74%) belonged to the SS14 lineage, and 117 (26%) belonged to the Nichols lineage, consistent with the proportional distribution of these lineages in other settings (appendix 1 pp 14–15).[3] Using hierarchical Bayesian analysis of population structure (BAPS) classification, four major BAPS groups were identified, which we designated Bayesian population (BP) 1–4 (figure 1; appendix 1 pp 14–15). Genomes from Australia were distributed throughout the phylogeny and were present in all four major BAPS groups (appendix 1 pp 9, 14–15).

Consistent with other findings,[3] the SS14-lineage was more genetically homogenous than the Nichols lineage (median pairwise SNP distances of 6 and 37, respectively, $p<0·0010$; figure 1) and comprised two main groups: BP1 (528 [70%] of 759) and BP2 (231 [30%] of 759; appendix 1 p 9). Similar to the SS14 lineage, the Nichols lineage comprised two major BAPS groups,

BP3 (116 [53%] of 221) and BP4 (69 [31%] 221; appendix 1 pp 14–15). The dominant BP3 group consisted mainly of genomes from Australia (101 [87%] of 116), although this group also contained genomes from the UK, USA, Canada, France, Ireland, and the Netherlands (appendix 2) that were previously identified as sublineage 8 in the 2019 study by Beale and colleagues.[3,4] Among the Australian genomes in the BP3 lineage, 97 were collected in Melbourne, three were from the Northern Territory, and one was from Brisbane (appendix 2). Only one BP3 sequence from Australia was identified in our random sampling from 2004–14, suggesting this lineage emerged relatively recently in our setting, consistent with findings from our temporal phylogenetic reconstruction (see below).

To investigate the emergence and evolution of TPA in the Australian context, we undertook Bayesian phylodynamic analysis using a dataset of 456 Australian genomes and 492 previously published sequences (appendix 2). We estimated a median molecular clock rate of $2{\cdot}14\times10^{-7}$ substitutions per site per year (95% highest posterior density [HPD] $1{\cdot}78\times10^{-7}$ to $2{\cdot}56\times10^{-7}$), and its date of emergence (time to the most recent common ancestor [tMRCA] of the crown node) of the SS14 and Nichols lineages at a median of 1748 (HPD 1427·7–1899·9), broadly consistent with previous studies.[3–5] Formal testing of temporal signal in the data using Bayesian evaluation of temporal signal[24] found the model with correct dates had higher support than the same Australian and global models with no dates, with a log Bayes factor of 878 for the Australian model and 968 for the global model. Further, our analysis suggested that the two main BAPS lineages co-circulating in Australia (BP1 and BP3) over the study period both became established in Australia in the latter part of the 20th century (figure 1). The median tMRCA for BP1 was 1993 (HPD 1983·7–1999·9), and for BP3 was 1981 (HPD 1955·5–1997·3; figure 1). Bayesian skyline analysis of the two largest Australian sublineages showed differences in the effective population size of the two populations (figure 2). The effective population size of BP1 suggested a sharp increase in the population around 2010, with a subsequent plateau. By contrast, the effective population size of BP3 suggested no major change in the population size over time; the slight increase in the observed trajectory had wide credible intervals (figure 2).

Among the Australian genomes, there were four major BAPS groups (BP1–4), consistent with the global distribution of BAPS groups, and twelve BAPS subgroups (appendix 1 p 16). Subgroup BP5·1 consisted of one singleton strain from the Northern Territory which did not cluster with any major BAPS groups and was distinct from other genomes in the global phylogeny (appendix 1 p 16). We identified concurrent circulation of TPA in Melbourne and the Northern Territory in three major BAPS groups, BP1, BP2, and BP3 (appendix 1 p 16), highlighting the multi-lineage nature of the current syphilis epidemic in Australia.

To examine sexual risk groups within BAPS subgroups, we linked genomic data and sexual behaviour. We focused on six dominant BAPS subgroups that contained more than 20 individuals and were supported by bootstrap replicates of 95% or more (figure 3). Phylodynamic analysis revealed that four subgroups were likely to have emerged in the 21st century, with the two largest subgroups (BP1·1 and BP3·1) emerging in the late 20th century (figure 3). Subgroups were defined as predominantly MSM-associated or heterosexual-associated depending on whether MSM or heterosexuals were associated with 50% or more of sequences in those subgroups (figure 3). Five subgroups were MSM-associated, except for BP1·5 (SS14 lineage), which was mainly heterosexual-associated (figure 3). All major subgroups contained at least one woman, and men who identified as having sex with men and women were present in three of the major subgroups (figure 3). Four of the major subgroups contained individuals with HIV, with proportions ranging from 20 (14%) of 144 (sublineage BP1·1) to 15 (48%) of 31 (sublineage BP1·3; figure 3).

All genomes were screened for mutations in the 23S rRNA gene, in addition to mutations in three previously described penicillin-binding protein genes (*pbp1, pbp,* and *mrcA*) and one putative β-lactamase gene (*Tp47*). Across the global phylogeny most sequences were genotypically macrolide-resistant (628 [83%] of 759 for SS14 and 167 [76%] of 221 for Nichols; figure 1). Of the Australian sequences, 398 (87%) of 456 were genotypically macrolide resistant (296 [88%] of 338 for SS14 and 102 [86%] of 118 for Nichols).

Similar to previous work, we observed sequence variation in penicillin-binding protein genes (*pbp1, pbp2,* and *mrcA*) and the putative β-lactamase gene, *Tp47* (appendix 1 pp 9–10, 17–18; appendix 2). The functional significance of these amino acid changes was inferred (appendix 1 pp 9, 25).[26–28] Four amino acid changes were identified as being of potential functional significance by both PolyPhen-2 and PROVEAN (SIFT predicted all neutral changes), three of which were in *mrcA* (Pro44Leu, Ala506Val, Gly708Ser) and one in *pbp1* (Pro564Leu; appendix 1 pp 9–10). There was high concordance between major BAPS lineages and TP0548 designations, with subtypes A–D representing Nichols samples and subtypes F–M, along with novel types representing SS14 samples (appendix 1 p 19; appendix 2). Further, there was significant association between BP subgroups and TP0548 subtypes ($p<0{\cdot}0010$; appendix 1 p 19). Within the SS14 lineage, the majority of BP1 samples were designated as TP0548 subtype G (421 [96%] of 440) and the majority of BP2 samples were designated as subtype F (186 [91%] of 204; appendix 1 p 19). Within the Nichols lineage, all 56 samples in BP3 were designated as subtype D (appendix 1 p 19). Moreover, subtype D was

only observed in a single sample outside BP3, this being a sample in the polyphyletic BP5 lineage. Within BP4 all but one sample was designated as subtype C (59 [98%] of 60; appendix 1 p 19).

Discussion

Over the past decade, syphilis infections have occurred disproportionately in two major risk groups in Australia, namely MSM in urban centres, and Aboriginal and Torres

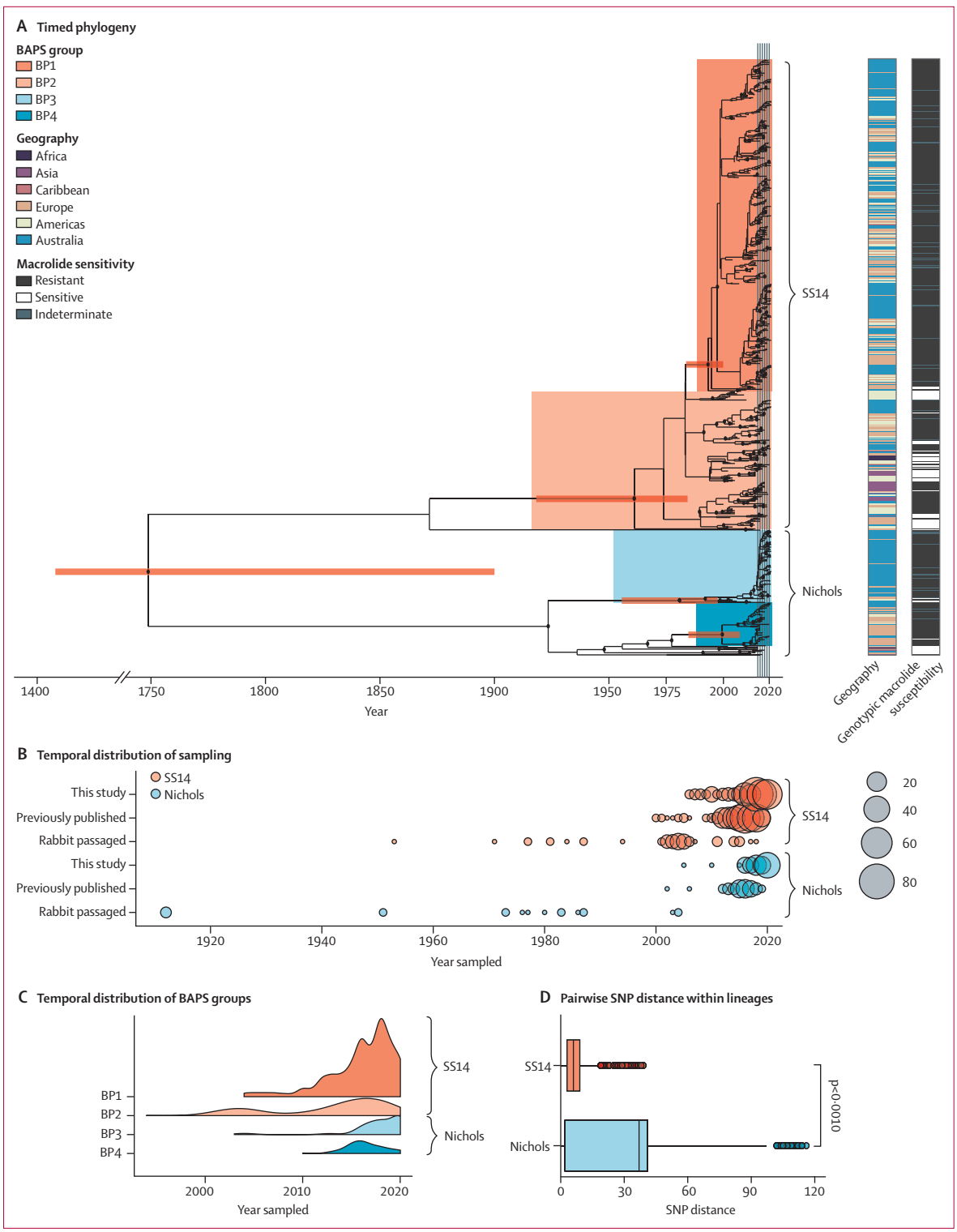


Figure 1: **Inferred global population dynamics of *Treponema pallidum***
(A) Maximum clade credibility tree inferred from 811 SNPs with 948 genomes. BAPS level 1 groups are highlighted, and continent of sample origin and genotypic macrolide susceptibility are shown in a heatmap. The interior nodes for each of the four BAPS groups, as well as the root of the phylogeny, are labelled with red bars to indicate the highest posterior density intervals. A black dot is shown on nodes with posterior support of 0·95 or more. (B) The temporal distribution of sampling for the major lineages (SS14 and Nichols) by samples collected in this study, previously published samples and rabbit passaged samples, whereby the size of the bubble is proportionate to the number of samples collected. (C) The temporal distribution of samples for each BAPS group between 1995 and 2020 (inclusive), whereby height is proportional to the number of samples collected. (D) The pairwise SNP distance between pairs of samples within each major lineage (SS14 and Nichols), whereby box plots indicate median IQR, with the whiskers representing the highest and lowest values within 1·5 × IQR of the upper and lower quartiles, and the dots representing outlier values. A Wilcoxon test was performed and found that the two groups were statistically significant (p<0·0010). BAPS=Bayesian analysis of population structure. BP=Bayesian population. SNP=single nucleotide polymorphism.

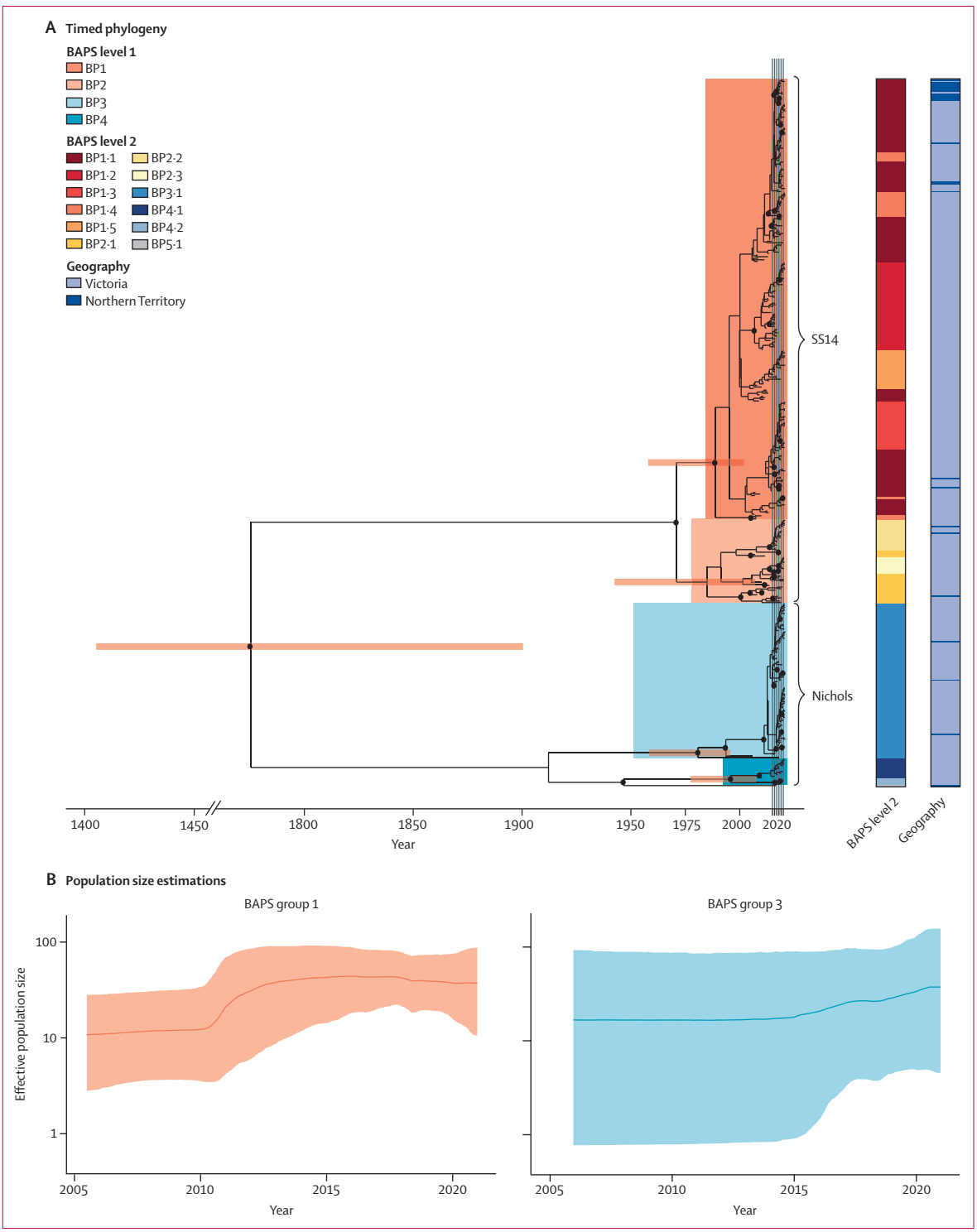


Figure 2: **Inferred population dynamics of Australian *Treponema pallidum***
(A) Time-scaled phylogeny inferred from 383 single nucleotide polymorphisms (SNPs) with 456 genomes. BAPS level 1 groups are highlighted, and BAPS level 2 groups and the Australian state of the sample collection (Victoria or the Northern Territory) are shown as a heatmap. The internal nodes as well as the root of the phylogeny are labelled with red bars to indicate the highest posterior density (HPD) intervals. A black dot is shown on nodes with posterior support of 0·95 or more. (B) Bayesian skyline plots showing the estimated effective population size through time for the two largest level 1 BAPS groups with median effective population size are shown as the black line and the 95% HPD intervals are represented by white shading. BAPS=Bayesian analysis of population structure. BP=Bayesian population.

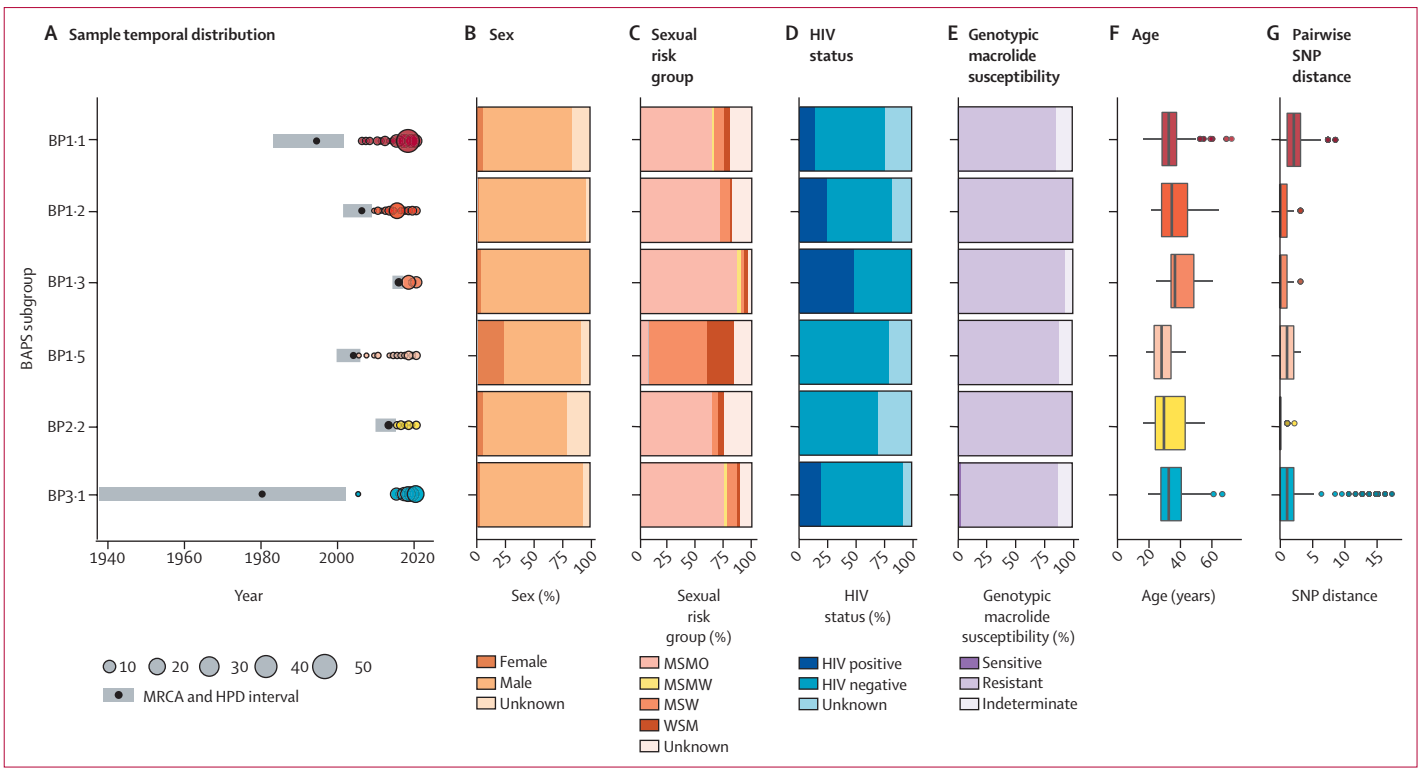


Figure 3: **Associations of major clusters of *Treponema pallidum* with population groups and antimicrobial susceptibility profiles**
(A) The temporal distribution of samples for each major BAPS level 2 group, defined as ≥20 samples and posterior support 0·98 for MRCA, shown for each group by the dark grey bubble. Grey bars on these bubbles indicate the HPD intervals. (B–F) Associations of the major transmission groups with epidemiological risk factors and genotypic macrolide susceptibility. Box plots indicate median and IQR, with the whiskers representing the highest and lowest values within 1·5 × IQR of the upper and lower quartiles, and the dots representing outlier values. (G) The median pairwise SNP distances represent the overall diversity within each cluster. BAPS=Bayesian analysis of population structure. HPD=highest posterior density. MRCA=most recent common ancestor. MSMO=men who have sex with men only. MSMW=men who have sex with men and women. MSW=men who have sex with women only. SNP=single nucleotide polymorphism. WSM=women who have sex with men.

Strait Islander people living largely in remote Australia, with outbreaks since 2015 also increasing in women and heterosexual men in Australian cities such as Melbourne.[29] Here, we observed that rather than one clonal outbreak, the syphilis epidemic in Australia comprises distinct sublineages of TPA, with co-circulation of globally prevalent strains in both remote and urban settings.

Consistent with other findings globally,[3,4,30,31] most (74%) TPA sequences in Australia belonged to the SS14 lineage, with limited genetic diversity between Australian SS14 genomes and those circulating globally. Moreover, the emergence and expansion of the major SS14 BP1 lineage in Australia in the late 20th century is in keeping with the emergence of the major SS14 sublineage in other continents.[3] Similarly, the major group within the Nichols lineage (BP3) emerged in the 1980s with a stable population circulating in Australia. Most (87%) genomes in the BP3 lineage were from Australia, with the remaining genomes from Europe and North America. Although it is possible the Nichols BP3 lineage might be more common in Australia, it is also possible this finding represents deeper sampling in our study. However, regardless of sampling, our findings are consistent with recent hypotheses of a marked population bottleneck and expansion of the contemporary population structure of TPA in the late 20th century.[3] In addition to syphilis, there has also been an increase in Australia of other bacterial sexually transmitted infections (STIs) over the past two decades.[32] Importantly, a characteristic feature of recent gonorrhoea and shigellosis epidemics in Australia has been clonal expansion of azithromycin-resistant lineages of *Neisseria gonorrhoeae* and *Shigella* spp, respectively.[10,11] In this current study, 87% of TPA strains were genotypically macrolide-resistant, further highlighting macrolide-resistance as an important feature of several bacterial STIs in the 21st century. It is plausible that azithromycin resistance in TPA (as in other STIs) has been partly driven by use of azithromycin (whether for STI treatment or for other infectious diseases), with a resultant selection pressure for the emergence of azithromycin resistance.

Horizontal gene transfer in TPA has not been reported as a mechanism for antimicrobial resistance gene acquisition; as such, ongoing assessment of mutational changes in putative resistance elements is important. Hence, in addition to genotypic assessment of macrolide resistance, we also looked at amino acid sequence variation in penicillin-binding proteins (Pbp1, Pbp2, and MrcA) and the putative β-lactamase Tp47. Sequence variation in penicillin-binding proteins has

been proposed as a potential mechanism of reduced susceptibility to β-Lactam antibiotics in *Streptococcus*,[33] with modifications in *pbp2* in *N gonorrhoeae* associated with reduced susceptibility to extended spectrum cephalosporins.[34] In this study, we identified three mutations in *mrcA* (Gly708Ser, Ala506Val, and Pro44Leu) and one mutation in *pbp1* (Pro564Leu) that might be associated with functional changes. However, in the absence of reported treatment failures associated with penicillin therapy, the clinical significance of these findings is unclear and additional research would be required to characterise the phenotype of potentially relevant non-synonymous mutations.

One important recent feature of the syphilis epidemic in Australia has been an increase in syphilis notifications in heterosexual populations.[29] Previous work in our setting has shown bridging of *N gonorrhoeae* strains across different sexual networks, with major lineages of *N gonorrhoeae* circulating in both MSM and heterosexual populations.[10] In this study, although most contemporary subgroups of TPA in Australia were predominantly MSM-associated, women and heterosexual men were identified in all major subgroups, with one subgroup (BP1·5) being mainly associated with heterosexuals. Of note, four of the major six subgroups included people living with and without HIV infection, including one subgroup (BP1·3) that emerged in the late 2010s. In this subgroup, approximately 50% of individuals were living with HIV. Although we did not have information on HIV pre-exposure prophylaxis (PrEP) use, which became available in Australia in 2017,[35] it is possible that the recent emergence of this subgroup in our setting reflects increased disassortative sexual mixing, in the context of minimal risk of HIV transmission due to PrEP and antiretroviral treatment.[36] Our data also suggest that the current syphilis epidemic in northern Australia might be driven by multiple lineages of TPA, rather than one homogenous outbreak. Although we only included 25 genomes from the Northern Territory, these are the first whole genomes of TPA sampled from this remote region, and were distributed across three global lineages with other genomes from Australia and overseas. Based on our data, it is possible that the syphilis epidemic in northern Australia is driven by multiple incursions of TPA lineages with subsequent dissemination; however, additional sampling from the Northern Territory and elsewhere in Australia is required to confirm this hypothesis.

Important strengths of our study include the contemporary sampling frame (from October, 2005, to December, 2020); integration of genomic and epidemiological data, and use of a stringent quality control process, meaning that only high-quality genomic sequences were included in our analyses. Although we only included genomes from two parts of Australia, previous studies have indicated high interconnectivity of sexual networks in urban Australia, meaning our findings are likely to have broader national applicability.[37] An important caveat of our work (and indeed other published studies using whole-genome sequencing for syphilis) is the low evolutionary rate of TPA, meaning there is relatively limited genetic diversity between genomes across space and time. Accordingly, whole-genome sequencing data should not be used in real-time to infer direction of transmission in sexual networks, but rather to identify epidemiological associations with distinct TPA subgroups, as in this current study. Moreover, the laboratory processes for TPA whole-genome sequencing are comparatively resource-intense compared with whole-genome sequencing of culturable pathogens such as *N gonorrhoeae*.

The multi-locus sequence typing scheme described in 2018 for TPA includes four targets (TP0136, TP0548, TP0705, and 23S rRNA);[38] however, a single gene target may simplify routine typing in many settings. In this work, we show strong concordance of single-locus typing of the TP0548 allele with genomic subgroups derived from whole-genome sequencing data. In some settings, TP0548 typing could be initially used on samples from individuals with a clinical presentation compatible with syphilis to provide a rapid broad overview of subtypes circulating within a population, pending detailed whole-genome sequencing analysis. However, regardless of typing methodology, clinical samples are still required for amplification and sequencing of TPA. Research in 2021 in our setting has shown high detection of TPA from mucocutaneous surfaces (including the mouth and anus) using PCR in individuals with primary or secondary syphilis, even in the absence of visible lesions.[39] Enhancing TPA PCR testing in patients with primary and secondary syphilis could therefore increase the samples available for genomic analysis and provide additional insights into TPA circulation.

In conclusion, we provide the first genomic insights into the current syphilis epidemic in Australia and demonstrate that this epidemic is driven by multiple, cocirculating strains that are linked to globally disseminated lineages. Although the priority for syphilis reduction efforts in Australia should be urgently improving public health resources (eg, improved access to testing and treatment) for priority populations, our findings highlight the need for a focus on controlling overlapping sub-epidemics of syphilis, rather than a one-size-fits-all public health approach.

Contributors

DAW, GT, SP, and MLT designed the study. FA, BG, EM, and MG provided the clinical samples. MAB and NRT shared RNA enrichment bait designs and advised sequencing methods. BZ, YWL, GT, MLT, and SP performed DNA extraction and library preparation. CKF, EPFC, MYC, AL, and NH provided sample metadata. GT, MLT, SH, and CMS performed bioinformatic analysis. MLT, SH, and DJI prepared the figures. DJI, SD, and DAW provided input into data analysis. All authors had full access to all the data in the study and had final responsibility for the decision to

submit for publication. MLT, GT, DAW, and SP verified the underlying data. DAW, MLT, and DJI prepared the manuscript, which was revised by all authors. DAW, BPH, CSB, JSH, and JP provided guidance and commentary on clinical aspects of this work.

Declaration of interests

We declare no competing interests.

Data sharing

The accessions for all isolates and associated metadata used in this study can be found in appendix 2. Short-read data for all Australian isolates sequenced in this study are available from the National Center for Biotechnology Information Sequence Read Archive (BioProject PRJNA754263). Data supporting the findings of this study are available within the text and in appendix 2. Source data are provided in BioProject (PRJNA754263).

Acknowledgments

This study was funded by the National Health and Medical Research Council Australia (NHMRC; GNT1147735). SD was funded by the Australian Research Council (DE190100805). NRT and MAB were funded by the Wellcome Trust (grant 206194). CKF is supported by an NHMRC Investigator Grant (APP1172900). DJI is supported by an NHMRC Investigator Grant (APP1195210). DAW is supported by an NHMRC Investigator Grant (APP1174555). Part of this research was undertaken using the Linkage Infrastructure, Equipment and Facilities high performance computer General-Purpose Graphics Processing Unit Facility hosted at the University of Melbourne. This facility was established with the assistance of LIEF Grant LE170100200. We gratefully acknowledge Shellee Williams from the Northern Territory Centre for Disease Control for assistance with identifying syphilis cases, and Kevin Freeman from Territory Pathology for assistance identifying pathology specimens for inclusion in the study.

References

1 Rowley J, Vander Hoorn S, Korenromp E, et al. Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. *Bull World Health Organ* 2019; **97:** 548–562P.

2 Hook EW 3rd. Syphilis. *Lancet* 2017; **389:** 1550–57.

3 Beale MA, Marks M, Cole MJ, et al. Global phylogeny of *Treponema pallidum* lineages reveals recent expansion and spread of contemporary syphilis. *Nat Microbiol* 2021; **6:** 1549–60.

4 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.

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

6 Department of Health. National Notifiable Disease Surveillance System. http://www9.health.gov.au/cda/source/cda-index.cfm (accessed July 1, 2021).

7 Read P, Fairley CK, Chow EP. Increasing trends of syphilis among men who have sex with men in high income countries. *Sex Health* 2015; **12:** 155–63.

8 Spiteri G, Unemo M, Mårdh O, Amato-Gauci AJ. The resurgence of syphilis in high-income countries in the 2000s: a focus on Europe. *Epidemiol Infect* 2019; **147:** e143.

9 Aung ET, Chen MY, Fairley CK, et al. Spatial and temporal epidemiology of infectious syphilis in Victoria, Australia, 2015–2018. *Sex Transm Dis* 2021; **48:** e178–82.

10 Williamson DA, Chow EPF, Gorrie CL, et al. Bridging of *Neisseria gonorrhoeae* lineages across sexual networks in the HIV pre-exposure prophylaxis era. *Nat Commun* 2019; **10:** 3988.

11 Ingle DJ, Easton M, Valcanis M, et al. Co-circulation of multidrug-resistant *Shigella* among men who have sex with men in Australia. *Clin Infect Dis* 2019; **69:** 1535–44.

12 Nishiki S, Lee K, Kanai M, Nakayama SI, Ohnishi M. Phylogenetic and genetic characterization of *Treponema pallidum* strains from syphilis patients in Japan by whole-genome sequence analysis from global perspectives. *Sci Rep* 2021; **11:** 3154.

13 Chen W, Šmajs D, Hu Y, et al. Analysis of *Treponema pallidum* strains from China using improved methods for whole-genome sequencing from primary syphilis chancres. *J Infect Dis* 2021; **223:** 848–53.

14 Sun J, Meng Z, Wu K, et al. Tracing the origin of *Treponema pallidum* in China using next-generation sequencing. *Oncotarget* 2016; **7:** 42904–18.

15 Pinto M, Borges V, Antelo M, et al. Genome-scale analysis of the non-cultivable *Treponema pallidum* reveals extensive within-patient genetic variation. *Nat Microbiol* 2016; **2:** 16190.

16 Grillová L, Oppelt J, Mikalová L, et al. Directly sequenced genomes of contemporary strains of syphilis reveal recombination-driven diversity in genes encoding predicted surface-exposed antigens. *Front Microbiol* 1691; **2019:** 10.

17 Parliament of Victoria. Public Health and Wellbeing Act. 2008. https://www.legislation.vic.gov.au/in-force/acts/public-health-and-wellbeing-act-2008/053 (accessed July 1, 2021).

18 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.

19 Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018; **34:** 3094–100.

20 Pětrošová H, Zobaníková M, Čejková D, et al. Whole genome sequence of *Treponema pallidum* ssp. *pallidum*, strain Mexico A, suggests recombination between yaws and syphilis strains. *PLoS Negl Trop Dis* 2012; **6:** e1832.

21 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017; **14:** 587–89.

22 Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020; **37:** 1530–34.

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

24 Duchene S, Lemey P, Stadler T, et al. Bayesian evaluation of temporal signal in measurably evolving populations. *Mol Biol Evol* 2020; **37:** 3363–79.

25 Marra C, Sahi S, Tantalo L, et al. Enhanced molecular typing of *Treponema pallidum*: geographical distribution of strain types and association with neurosyphilis. *J Infect Dis* 2010; **202:** 1380–88.

26 Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research* 2012; **40:** W452–57.

27 Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 2015; **31:** 2745–47.

28 Adzhubei I, Jordan DM, Sunyaev SR. predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; **76:** unit7.20.

29 Kirby Institute. HIV, viral hepatitis and sexually transmissible infections in Australia: annual surveillance report 2018. Sydney: Kirby Institute, UNSW Sydney, 2018.

30 Venter JME, Müller EE, Mahlangu MP, Kularatne RS, Diekema DJ. *Treponema pallidum* macrolide resistance and molecular epidemiology in southern Africa, 2008 to 2018. *J Clin Microbiol* 2021; **59:** e0238520.

31 Gallo Vaulet L, Grillová L, Mikalová L, et al. Molecular typing of *Treponema pallidum* isolates from Buenos Aires, Argentina: frequent Nichols-like isolates and low levels of macrolide resistance. *PLoS One* 2017; **12:** e0172905.

32 Jasek E, Chow EP, Ong JJ, et al. Sexually transmitted infections in Melbourne, Australia from 1918 to 2016: nearly a century of data. *Commun Dis Intell Q Rep* 2017; **41:** e212–22.

33 Vannice KS, Ricaldi J, Nanduri S, et al. *Streptococcus pyogenes* pbp2x mutation confers reduced susceptibility to β-lactam antibiotics. *Clin Infect Dis* 2020; **71:** 201–04.

34 Unemo M, Nicholas RA. Emergence of multidrug-resistant, extensively drug-resistant and untreatable gonorrhea. *Future Microbiol* 2012; **7:** 1401–22.

35 Traeger MW, Cornelisse VJ, Asselin J, et al. Association of HIV preexposure prophylaxis with incidence of sexually transmitted infections among individuals at high risk of HIV infection. *JAMA* 2019; **321:** 1380–90.

36 Chow EPF, Phillips TR, Bradshaw CS, Chen MY, Fairley CK. Brief report: sexual mixing patterns by HIV status and PrEP use among men who have sex with men partnerships in Melbourne, 2011–2018. *J Acquir Immune Defic Syndr* 2020; **83:** 99–102.

37 Whiley DM, Trembizki E, Buckley C, et al. Molecular antimicrobial resistance surveillance for *Neisseria gonorrhoeae*, Northern Territory, Australia. *Emerg Infect Dis* 2017; **23:** 1478–85.

38 Grillová L, Bawa T, Mikalová L, et al. Molecular characterization of *Treponema pallidum* subsp. *pallidum* in Switzerland and France with a new multilocus sequence typing scheme. *PLoS One* 2018; **13:** e0200773.

39 Towns JM, Leslie DE, Denham I, et al. *Treponema pallidum* detection in lesion and non-lesion sites in men who have sex with men with early syphilis: a prospective, cross-sectional study. *Lancet Infect Dis* 2021; **21:** 1324–31.