Invasive *Streptococcus dysgalactiae* subspecies *equisimilis* compared with *Streptococcus pyogenes* in Australia, 2011–23, and the emergence of a multi-continent *stG62647* lineage: a retrospective clinical and genomic epidemiology study

Ouli Xie, Leo Featherstone, An N T Nguyen, Andrew J Hayes, Miranda E Pitt, Stephanie Spring, Alice Liu, Gerry Tonkin-Hill, Ravindra Dotel, Neela Joshi Rai, Alexander Rofe, Sebastian Duchêne, Deborah C Holt, Louise M Judd, Lachlan J M Coin, Vicki L Krause, Matthew V N O'Sullivan, Robert W Baird, Katherine Bond, Benjamin P Howden, Tony M Korman, Bart J Currie, Mark R Davies, Steven Y C Tong**

Summary

Background *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE) is closely related to *Streptococcus pyogenes*, with overlapping disease manifestations. We compared the clinical and genomic epidemiology of invasive SDSE with invasive *S pyogenes* across different settings in Australia and phylogenetically contextualised the SDSE sequences within a global cohort of genomes.

Methods In this retrospective clinical and genomic epidemiology study, cases of invasive SDSE isolated from normally sterile sites were identified and whole-genome sequenced across five hospital networks in temperate southeast Australia (Melbourne and Sydney) and the tropical Top End of the Northern Territory. SDSE disease incidence, case demographics, clinical outcomes, and longitudinal lineage dynamics were compared between southeast Australia and the Top End and to co-collected invasive *S pyogenes* cases in each region. SDSE genomes and lineages were also contextualised within 1166 global SDSE sequences. Genomic transmission clusters (not necessarily direct transmission) were inferred between isolates from different individuals by single-linkage clustering at a single nucleotide polymorphism threshold of less than or equal to seven for SDSE and less than or equal to five for *S pyogenes* based on previous transmission analyses.

Findings Between Jan 1, 2011, and Feb 28, 2023, there were 693 invasive SDSE cases and 995 invasive *S pyogenes* cases. Invasive SDSE occurred almost exclusively in adults. The overall invasive SDSE incidence in southeast Australia was similar to invasive *S pyogenes* (incidence rate ratio [IRR] 1·15, 95% CI 0·91–1·46; p=0·26) and increased over the study period (IRR 1·06 per year, 95% CI 1·05–1·08; p<0·0001) from 1·30 cases per 10 000 admissions in 2011 to 3·72 cases per 10 000 admissions in the first 2 months of 2023 (95% CI 2·13–6·07). In southeast Australia, where stringent COVID-19 non-pharmaceutical interventions (NPIs) were implemented between 2020 and 2021, the SDSE incidence plateaued during 2020–21 but did not significantly decline (IRR 1·09 compared with 2017–19, 95% CI 0·88–1·35; p=0·47). By contrast, *S pyogenes* incidence substantially declined in 2020–21 in southeast Australia (IRR 0·35 compared to 2017–19, 95% CI 0·22–0·52; p=0·017). In the Top End, SDSE incidence was lower than *S pyogenes* (IRR 0·24, 95% CI 0·19–0·31; p<0·0001). However, crude incidence remained higher than southeast Australia (crude IRR 1·24, 95% CI 1·07–1·42; p=0·0037) and disproportionately affected First Nations Australians in the Top End compared with non-First Nations individuals (IRR 3·36, 95% CI 2·33–4·85; p<0·0001). Comparing 2020–21 with 2017–19, there was no decline in SDSE (IRR 1·27, 95% CI 0·73–2·24; p=0·45) or *S pyogenes* (IRR 0·97, 95% CI 0·80–1·18; p=0·81) incidence in the Top End, which did not implement prolonged stringent COVID-19 NPIs. Analysing the available genomes of invasive cases and in lineages for which more than or equal to five invasive cases occurred, only 24 (6%) of 384 SDSE cases were assigned to genomic transmission clusters, compared with 271 (52%) of 524 *S pyogenes* cases. An *stG62647* lineage encompassed 113 (26%) of 436 sequenced SDSE genomes. Analysis of available SDSE sequences from Australia, western Europe, and North America inferred concurrent international expansion of the *stG*62647 lineage in all three regions between 1990 and 2005.

Interpretation We identified a substantial burden of invasive SDSE, dominated by the emergent *stG62647* lineage. The contrasting epidemiology between species in the different Australian regions, during COVID-19 NPIs, and genomic infection patterns indicates transmission dynamic, pathogen population, and host–pathogen interaction differences between SDSE and *S pyogenes* and indicates implications for disease control measures.

Funding Australian National Health and Medical Research Council.

Lancet Microbe 2025; 6: 101182

Published Online October 7, 2025
https://doi.org/10.1016/j.lanmic.2025.101182

*Contributed equally

Department of Infectious Diseases, University of Melbourne, The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia (O Xie PhD, A Liu FRACP, K Bond FRACP, Prof S Y C Tong PhD); **Monash Infectious Diseases, Monash Health, Melbourne, VIC, Australia** (O Xie, S Spring FRACP, T M Korman FRACP); **Global and Tropical Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia** (O Xie, D C Holt PhD, Prof B J Currie PhD); **Department of Microbiology and Immunology, University of Melbourne, The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia** (L Featherstone PhD, A N T Nguyen PhD, A J Hayes MPhil, M E Pitt PhD, G Tonkin-Hill PhD, S Duchêne PhD, L M Judd PhD, Prof L J M Coin PhD, K Bond, M R Davies PhD); **Division of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, ACT, Australia** (L Featherstone); **Australian Institute for Microbiology and Infection, University of Technology Sydney, Sydney, NSW, Australia** (M E Pitt); **Infectious Diseases Department, Royal Darwin Hospital, Darwin, NT, Australia** (S Spring, Prof B J Currie); **Victorian Infectious Diseases Service, The Royal Melbourne Hospital, The Peter Doherty Institute for Infection and**

Immunity, Melbourne, VIC, Australia (A Liu, A Rofe MBBS, K Bond, Prof S Y C Tong); Peter MacCallum Cancer Centre, Melbourne, VIC, Australia (G Tonkin-Hill); Department of Biostatistics, Faculty of Medicine, University of Oslo, Oslo, Norway (G Tonkin-Hill); Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, VIC, Australia (G Tonkin-Hill); Department of Infectious Diseases, Blacktown Hospital, Sydney, NSW, Australia (R Dotel PhD); Department of Infectious Diseases, Westmead Hospital, Sydney, NSW, Australia (N Joshi Rai MPH, M V N O'Sullivan PhD); Department of Microbiology, The Royal Melbourne Hospital, Melbourne, VIC, Australia (A Rofe, K Bond); Department of Computational Biology, Institut Pasteur, Paris, France (S Duchêne); Faculty of Health, Charles Darwin University, Darwin, NT, Australia (D C Holt); Centre for Pathogen Genomics, The University of Melbourne, Melbourne, VIC, Australia (L M Judd, Prof B P Howden PhD); Northern Territory Centre for Disease Control, Northern Territory Department of Health, Darwin, NT, Australia (V L Krause FAFPHM); Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia (M V N O'Sullivan); NSW Health Pathology, Westmead Hospital, Sydney, NSW, Australia (M V N O'Sullivan); Territory Pathology, Northern Territory Department of Health, Royal Darwin Hospital, Darwin, NT, Australia (R W Baird FRACP); Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, University of Melbourne, The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia (Prof B P Howden); Department of Infectious Diseases & Immunology, Austin Health, Melbourne, VIC, Australia (Prof B P Howden); Centre for Inflammatory Diseases, Monash University, Melbourne, VIC, Australia (T M Korman)

Correspondence to: Prof Steven Y C Tong, Department of Infectious

Introduction

Streptococcus dysgalactiae subspecies *equisimilis* (SDSE) is increasingly recognised as a cause of serious human disease and is closely related to the more well-studied pathogen, *Streptococcus pyogenes* (group A streptococcus).[1] SDSE causes a similar spectrum of disease to *S pyogenes*.[1] Both pathogens share ecological niches in the human throat and skin, although rates of gastrointestinal or genital carriage can differ, and are thought to transmit by the respiratory route or direct contact.[1–3] Examining the clinical and genomic epidemiology of SDSE compared with *S pyogenes* might provide insights into evolutionary patterns, biology, and burden of disease, and inform disease control measures.

In high-income regions, the incidence of invasive SDSE disease in adults is similar to, and in some countries exceeds, invasive *S pyogenes* incidence.[4–6] However, there is a paucity of data on the epidemiology of SDSE in socioeconomically disadvantaged regions. In tropical northern Australia, remote First Nations communities have a disproportionate burden of *S pyogenes* disease, driven by underlying inequities in social determinants of health.[7] It is unclear whether there is a similar burden of SDSE disease in this region.

Multiple high-income countries have recently reported the emergence of an SDSE *stG62647 emm*-type population, which anecdotally has been associated with more severe disease than other *emm*-type lineages.[6,8,9] However, the same *emm*-type can be found across distant lineages in SDSE, and few whole-genome resolution studies have tracked global SDSE evolution and epidemiology.[10]

We aimed to compare the clinical and genomic epidemiology of invasive SDSE between 2011 and 2023 in Australia across high and low streptococcal disease burden

Research in context

Evidence before this study

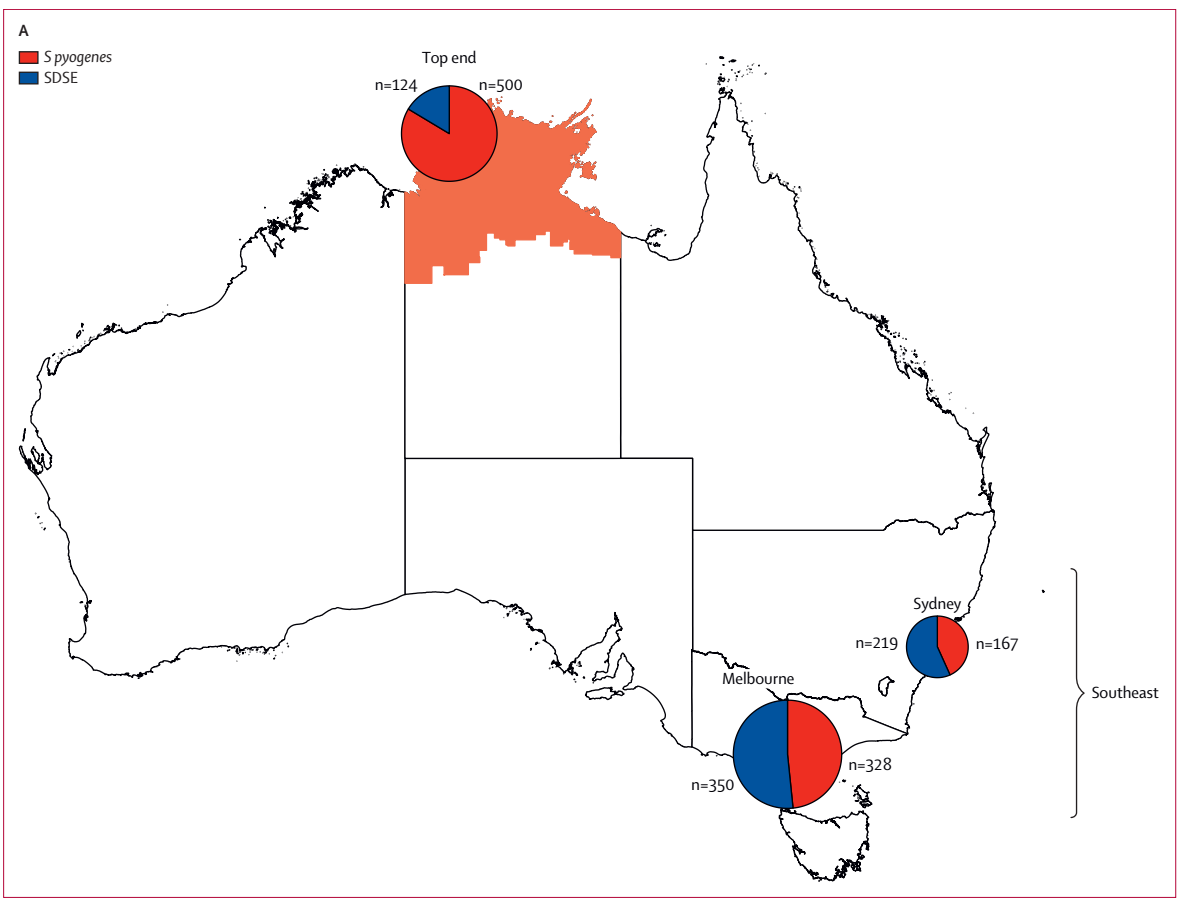
Streptococcus dysgalactiae subspecies *equisimilis* (SDSE) is closely related to *Streptococcus pyogenes* (group A streptococcus) and shares many disease manifestations and virulence mechanisms. Invasive SDSE infection mostly affects adults and often occurs in those with comorbidities, such as immunocompromise, medical implants, obesity, and diabetes. Examination of SDSE clinical, epidemiological, and genomic dynamics in comparison to *S pyogenes* could shed light on cross-species biology. We searched PubMed from database inception to June 14, 2024, using the search terms "*Streptococcus*" AND "*dysgalactiae*" AND "*equisimilis*" AND "genome". We repeated the search using ("group G" OR "group C) in place of "*equisimilis*". Limited to studies with at least ten genomes, we found ten studies examining SDSE epidemiology in humans at whole-genome resolution. Most studies described only *emm*-type, which discriminates lineages poorly for SDSE, examined restricted longitudinal or geographical diversity, or were limited to specific disease manifestations, antimicrobial resistant strains, or individual lineages. No studies described the combined clinical and genomic epidemiology of invasive SDSE in socioeconomically disadvantaged regions, which have a high burden of streptococcal disease.

Added value of this study

By undertaking a cross-species clinical and genomic epidemiological comparison, we show a substantial but contrasting burden of invasive SDSE incidence between disparate regions in Australia—the sparsely populated tropical Top End of the Northern Territory of Australia and more densely populated urban southeast Australia. The incidence of invasive SDSE in adults was similar to *S pyogenes* in southeast Australia. In the Top End, where *S pyogenes* is hyperendemic, SDSE incidence was four-times lower than *S pyogenes*, but crude incidence remained higher than southeast Australia and disproportionately affected First Nations people. Genomically, we show that few invasive SDSE cases were inferred to be part of genomic transmission clusters compared with *S pyogenes*. In individuals with multiple episodes of invasive SDSE infection, subsequent episodes were frequently caused by near-identical strains compared with the index episode. Geographically, more SDSE and *S pyogenes* cases in the Top End were part of transmission clusters compared with southeast Australia. When combined with epidemiological observations, these findings show transmission and host–pathogen differences leading to invasive disease across species and disease settings. At a pathogen population level, SDSE disease in Australia was dominated by a lineage predominantly carrying the *stG62647 emm*-type (*stG62647* lineage), which we inferred to have expanded largely concurrently across Australia, western Europe, and North America between 1990 and 2005.

Implications of all the available evidence

There is a considerable burden of SDSE disease, including in an Australian setting hyperendemic for *S pyogenes* disease, which highlights the need for expanded SDSE disease surveillance and ongoing measures to address health and socioeconomic inequities associated with disproportionate streptococcal disease in disadvantaged settings. The observed SDSE incidence and genomic infection patterns build upon existing epidemiological reports and indicate transmission differences compared with *S pyogenes*, and suggest that disease control measures, such as close contact antimicrobial prophylaxis to prevent secondary invasive cases, might have different effects across these species. Further studies are needed to understand SDSE carriage and immunity in susceptible populations and molecular drivers of emergent SDSE lineages to address the growing burden of disease.



(Figure 1 continues on next page)

settings, and concomitantly collected invasive *S pyogenes* cases in each region.[11] A detailed analysis of the invasive *S pyogenes* data is described separately.[11] We also aimed to place SDSE epidemiology within the context of a global SDSE population genomic framework, with an emphasis on the emergence of a lineage carrying the *stG62647 emm*-type.

Methods

Study design and isolates

In this retrospective clinical and genomic epidemiology study, invasive SDSE cases were identified between Jan 1, 2011, and Feb 28, 2023, at five tertiary or quaternary health networks in geographically diverse regions across Australia (figure 1A). One network covered the tropical Top End of the Northern Territory, a largely remote region of approximately 475 000 km^2, and two networks each in Melbourne and Sydney represented high-resourced urban regions in temperate southeast Australia. Three networks admitted adults and children, and one network in each of Melbourne and Sydney admitted only adults. Invasive *S pyogenes* cases at the same networks were identified and data were collected using identical case report forms.[11] Between 2020 and 2021, Melbourne and Sydney implemented stringent non-pharmaceutical interventions (NPIs), including prolonged stay-at-home orders, in response to the COVID-19 pandemic. By contrast, the Top End of the Northern Territory did not implement prolonged stay-at-home orders, but border access and travel to remote communities were restricted. Further details of the hospital networks are described in appendix 1 (pp 1–2).

See Online for appendix 1

Invasive SDSE was defined as culture of presumed *Streptococcus dysgalactiae* from a normally sterile site (eg, blood, synovial fluid, or cerebrospinal fluid) and identified by matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (Bruker Daltonics; Bremen, Germany) or phenotypic testing (appendix 1 p 2). Whole-genome sequencing confirmation of subspecies was done for isolates retrieved from –80°C storage (appendix 1 p 3). Only cases that were culture-positive and from the bloodstream were included in Melbourne and Sydney, whereas all sterile sites were included from the Top End to mirror the previous *S pyogenes* study design.[11] Isolation of SDSE or *S pyogenes* from the same individual within 14 days was considered the same episode. Episodes after 14 days were considered new subsequent episodes (encompassing relapse or reinfection).

This study was approved by the Royal Melbourne Hospital Human Research Ethics Committee (HREC/80105/MH-2021) and the Human Research Ethics Committee of the Northern Territory Department of Health and Menzies

Diseases, University of Melbourne, The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC 3000, Australia
steven.tong@unimelb.edu.au

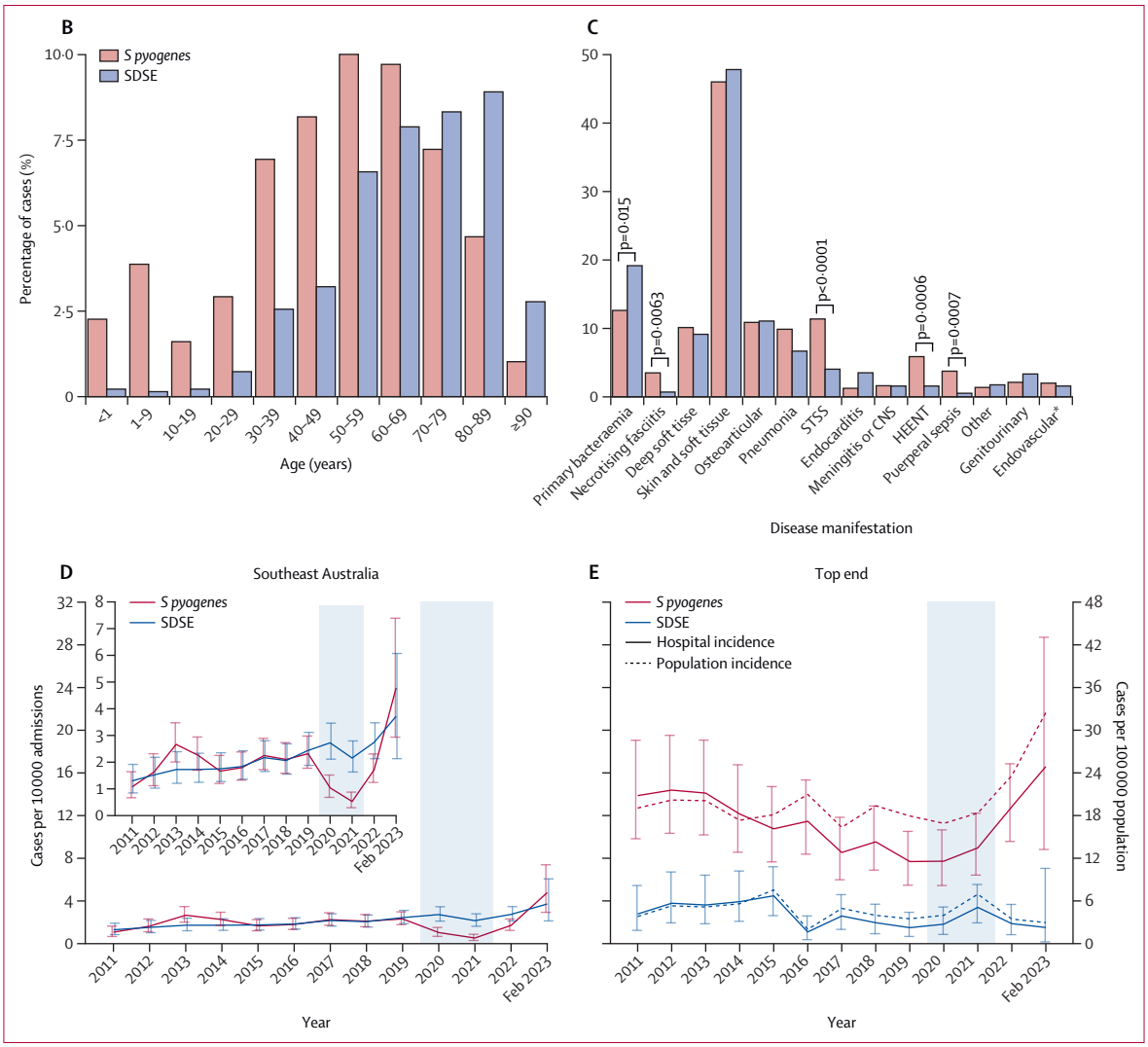


Figure 1: **Geographical sampling, incidence, and clinical characteristics of invasive SDSE compared with *Streptococcus pyogenes* disease**
(A) Geographical distribution of all cases from the tropical Top End of the Northern Territory and four hospital networks across temperate Melbourne and Sydney in the southeast of Australia from January, 2011, to February, 2023. The geographical area covering the Top End is highlighted in orange. (B) Age distribution of cases of bloodstream invasive SDSE and *S pyogenes* disease in the Top End and southeast Australia. (C) Disease manifestations associated with bloodstream invasive SDSE and *S pyogenes* infection in the Top End and southeast Australia. Each case could have more than one disease manifestation. For individuals with more than one episode of infection, only the first case was included. Bonferroni-corrected p values are indicated above for comparisons with p<0·05. Yearly incidence of invasive SDSE compared with *S pyogenes* in southeast Australia (D; inset graph is rescaled incidence plot) and the Top End (E). Only bloodstream cases are included. Incidence reported as cases per 10 000 acute hospital admissions per year (hospital incidence). Incidence in the Top End also reported as cases per 100 000 population per year (population incidence). Population incidence was not calculated for sites in southeast Australia, as two hospitals served as statewide referral centres and an accurate catchment population could not be estimated. Error bars represent 95% CIs for hospital incidence. As data were only collected until February, 2023, incidence in 2023 represents only the first 2 months of the year. The blue shaded period from 2020 to 2021 approximates the interval of the most stringent non-pharmaceutical interventions during the COVID-19 pandemic. HEENT=head, eye, ear, nose, and throat. SDSE=*Streptococcus dysgalactiae* subspecies *equisimilis*. STSS=streptococcal toxic shock syndrome. *Endovascular does not include infective endocarditis.

School of Health Research (2021–4181). Individual participant consent was waived, as all data were available from routine clinical care.

Clinical data and outcomes

Clinical data were collected from medical records including demographics, self-reported First Nations identification, and clinician-determined disease manifestation(s) (appendix 1 p 3). Outcomes included intensive care unit (ICU) admission, length of stay, in-hospital mortality, and 30-day mortality. Streptococcal toxic shock syndrome was classified per the 2010 US National Notifiable Diseases Surveillance System definition but with isolation of SDSE.[12] Puerperal sepsis was defined as infection within 28 days postpartum.

Genomic sequencing and analysis

Bacterial isolates were whole-genome sequenced using the NextSeq 550 platform (Illumina; San Diego, CA, USA) to

generate 150-base pair paired-end reads. Further microbiological methods are described in appendix 1 (p 2).

A standardised framework was used to quality control sequences, assemble genomes, and infer *emm*-type, multilocus sequence type, and Lancefield group carbohydrate (appendix 1 pp 3–4).[10] A maximum likelihood phylogeny was inferred from a single nucleotide polymorphism (SNP) alignment (Snippy version 4.6.0) against reference GGS_124 (NC_012891.1) using IQ-tree (version 2.2.2.7; appendix 1 p 4).[13]

Global genomic sequence clusters were assigned by PopPUNK (version 2.6.0) using a previously described model.[10,14] Genomic sequence clusters were highly concordant with the phylogeny and are henceforth referred to as lineages. Sublineages were assigned using PopPIPE (version 1.0.0; appendix 1 pp 4–5). Genomic analyses included all cases with available genomes (including non-bloodstream cases from the Top End) except for temporal analyses, which excluded isolates from Sydney as these were only available from August, 2020.

Reference-independent SNP distances between isolates were calculated using SKA (version 1.0). Genomic transmission clusters (not necessarily direct transmission) were inferred between isolates from different individuals by single-linkage clustering at an SNP threshold of less than or equal to seven for SDSE and less than or equal to five for *S pyogenes* based on previous transmission analyses.[2,15] Only genomes from well sampled lineages with at least five representatives were included. For clustering analyses, cases or isolates from the same sublineage were merged within individuals. A sensitivity analysis was done with clustering within each region separately to account for geographical variation.

Complete genomes were generated for ten SDSE isolates representative of major lineages in this study using Oxford Nanopore long-read sequencing with R10.4.1 flow cells (Oxford, UK; appendix 1 p 5). Assembly was done using Trycycler (version 0.5.4; appendix 1 p 5).[16]

Global genomic analysis

A global database of SDSE genomes (n=1166 from 19 countries) was generated from all publicly available sequences from the National Center for Biotechnology Information Sequence Read Archive as of Jan 15, 2024, and building on a previously published SDSE global genomic diversity dataset (appendix 3).[2] Genomes isolated from animal hosts were excluded, and all sequences were quality controlled using the same criteria as newly sequenced genomes from this study (appendix 1 pp 3–4). A global phylogeny was inferred using the same methods as above.[13]

To infer the population trajectory of the *stG62647* lineage, genomes that formed a monophyletic group with ST20-*stG62647* (appendix 2) sequences were included.[8] The evolutionary timescale of the global *stG62647* lineage was inferred using BEAST (version 1.10.4; appendix 1 pp 5–6). The ancestral trajectory in the effective population size (a measure of genetic diversity, with increasing values indicating a growing number of infections) was estimated using a Bayesian coalescent skyline model (appendix 1 p 6).[17]

Carriage and gene truncations in the *sil* two-component regulatory locus were investigated by mapping against reference MMC404 (CP160427) and BLAST-based search (appendix 1 pp 6–7). Mobile genetic element and accessory gene gain or loss was determined using Panaroo (version 1.3.3) and a previously described mobile genetic element pipeline (appendix 1 pp 7–8).[10,18,19]

For more on **Snippy** see https://github.com/tseemann/snippy

Statistical analysis

Only bloodstream cases were included for incidence analyses to allow equal comparison across regions. Clinical and demographic analyses were further limited to bloodstream cases with complete clinical data and only the first episode for individuals with multiple infections. All comparisons were made between SDSE and *S pyogenes* unless otherwise stated.

For more on **PopPIPE** see https://github.com/bacpop/PopPIPE

We used the Wilcoxon rank-sum test to compare continuous demographic variables and Fisher's exact test for categorical demographic variables and disease manifestations at an α of 0·05. Bonferroni correction was applied to disease manifestations for multiple testing. Odds ratios (ORs) and 95% CIs of ICU admissions, in-hospital mortality, and 30-day mortality were estimated using logistic regression. Length of stay was compared by quasi-Poisson regression for overdispersion (rate ratio, 95% CI). Outcome comparisons were adjusted (adjusted OR, 95% CI), with age as a continuous variable and sex, self-identified First Nations status, region, and year as covariates. The association between the *stG62647* lineage and severe disease (composite of ICU admission and in-hospital mortality) was assessed by logistic regression compared with all other lineages (OR, 95% CI). A supplementary analysis comparing SDSE disease in the Top End to southeast Australia was done using the same methods.

Disease incidence was calculated from bloodstream cases per 10 000 acute hospital admissions per year. Population incidence was also calculated for the Top End, which was served by a single hospital network. The population denominator was extracted from Australian Bureau of Statistics (ABS) mid-year statistics and from the nearest ABS census year for the First Nations population.[20] Population incidence was not calculated for southeast Australia, as an accurate catchment population could not be calculated. Temporal trends were estimated by quasi-Poisson regression using the natural log of yearly admissions as an offset, except for incidence affecting First Nations people and sensitivity analyses in the Top End where population was used (incidence rate ratio [IRR], 95% CI). Crude IRR (95% CI) comparing the Top End and southeast Australia, and across species within each region, was estimated using quasi-Poisson regression with year as a covariate.

See **Online** for appendix 3

See **Online** for appendix 2

The Morisita index as implemented in vegan (version 2.6) was used to quantify geographical lineage composition dissimilarity, where zero indicates identical populations

	SDSE (n=569)	*Streptococcus pyogenes* (n=800)	Unadjusted p value	Unadjusted OR or RR (95% CI)	Adjusted p value	Adjusted OR or RR (95% CI)
Age, years	69 (55–81)	53 (36–67)	<0·0001	..	..	..
Age (excluding adult only sites)	70 (57–81); 365	52 (33–66); 631	<0·0001	..	..	..
Sex			0·0006	..	..	..
Female	220 (39%)	384 (48%)	..	..	..	..
Male	349 (61%)	416 (52%)	..	..	..	..
First Nations*						
Top End	49/89 (55%)	250/383 (65%)	..	..	..	..
Southeast	6/480 (1%)	5/417 (1%)	..	..	..	..
Length of stay, days	9 (5–20)	8 (5–17)	0·46	0·95 (0·82–1·09)	0·61	1·04 (0·89–1·22)
In-hospital mortality	59 (10%)	80 (10%)	0·67	0·96 (0·67–1·37)	<0·0001	2·06 (1·38–3·09)
30-day mortality	63 (11%)	79 (10%)	0·47	0·88 (0·62–1·25)	0·0044	1·78 (1·20–2·65)
Intensive care admission	94 (17%)	190 (24%)	0·0012	1·57 (1·20–2·08)	<0·0001	1·95 (1·44–2·66)

Data are median (IQR), median (IQR); n, n (%), or n/N (%). Unadjusted OR and adjusted OR for in-hospital mortality, 30-day mortality, and intensive care admission and RR for length of stay were calculated with SDSE as a reference. OR=odds ratio. RR=risk ratio. SDSE=*Streptococcus dysgalactiae* subspecies *equisimilis*. *p value not calculated due to different underlying populations.

Table: **Demographics and outcomes of individuals with invasive SDSE infection compared with invasive *Streptococcus pyogenes* infection**

and one indicates completely different populations.[21] Statistical analyses were done in R (version 4.3.1).

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

Between Jan 1, 2011, and Feb 28, 2023, there were 569 invasive SDSE cases from 545 individuals in southeast Australia and 124 invasive SDSE cases from 117 individuals in the Top End of the Northern Territory (figure 1A; appendix 1 pp 9–10; appendix 2A). In the Top End, 12 (10%) of 124 invasive SDSE cases were isolated without concurrent positive blood cultures. Over the same period, there were 495 invasive *S pyogenes* cases from 491 individuals in southeast Australia and 500 cases from 444 individuals in the Top End (appendix 2C). Invasive *S pyogenes* cases are described in detail separately.[11]

Complete clinical data were available for 569 (87%) of 651 individuals with bloodstream SDSE infection. 563 (99%) of 569 individuals were older than 18 years. By contrast, there was a bimodal age distribution for *S pyogenes* (figure 1B). The crude 30-day mortality of SDSE was 11% (63 of 569 individuals) and was similar between both species (table). However, after adjustment for covariates, *S pyogenes* cases were associated with significantly greater 30-day mortality, in-hospital mortality, and ICU admission (appendix 1 p 22). The spectrum of manifestations overlapped between pathogens, but frequencies varied (figure 1C; appendix 1 p 23). SDSE comparisons between southeast Australia and the Top End are presented in appendix 1 (pp 24–26). Individuals with invasive SDSE were significantly older in southeast Australia (median age 69 years, IQR 57–82) than in the Top End (61 years, 48–76; p=0·0026), but length of stay, 30-day mortality, in-hospital mortality, and ICU admission did not significantly differ by region, even after adjusting for demographics.

The relative incidence of bloodstream invasive SDSE compared with *S pyogenes* varied between regions. The overall incidence of invasive SDSE in southeast Australia was similar to invasive *S pyogenes* (IRR 1·15, 95% CI 0·91–1·46; p=0·26; figure 1D, appendix 1 pp 11, 27) and increased (IRR 1·06 per year, 95% CI 1·05–1·08; p<0·0001) during the study period from 1·30 cases per 10 000 admissions in 2011 to 3·72 cases per 10 000 admissions in the first 2 months of 2023 (95% CI 2·13–6·07). Although the increasing trend in SDSE incidence plateaued in 2020–21 during COVID-19 NPIs, there was no significant decline in 2020–21 (IRR 1·09 compared with 2017–19, 95% CI 0·88–1·35; p=0·47) compared with a substantial reduction in *S pyogenes* incidence in southeast Australia over the same period (IRR 0·35 for 2020–21 compared with 2017–19, 95% CI 0·22–0·52; p=0·017). By contrast, in the Top End where *S pyogenes* is hyperendemic, the incidence of invasive SDSE was lower than *S pyogenes* (crude IRR 0·24, 95% CI 0·19–0·31; p<0·0001; figure 1E). However, the invasive SDSE crude incidence in the Top End remained greater than southeast Australia (IRR 1·24, 95% CI 1·07–1·42; p=0·0037) and disproportionately affected First Nations individuals (IRR 3·36 compared with non-First Nations individuals in the Top End, 95% CI 2·33–4·85; p<0·0001; appendix 1 p 12). Comparing 2020–21 with 2017–19, there was no decline in the incidence of SDSE (IRR 1·27, 95% CI 0·73–2·24; p=0·45) or *S pyogenes* (IRR 0·97, 95% CI 0·80–1·18; p=0·81) in the Top End. In the Top End, where population-based incidence was available, findings were similar when comparing 2020–21 with 2017–19 using population-based incidence (SDSE IRR 1·32, 95% CI 0·81–2·15; p=0·35; *S pyogenes* 0·99, 0·86–1·13; p=0·85).

Examining individuals with multiple SDSE infections, most were from southeast Australia (23 [82%] of

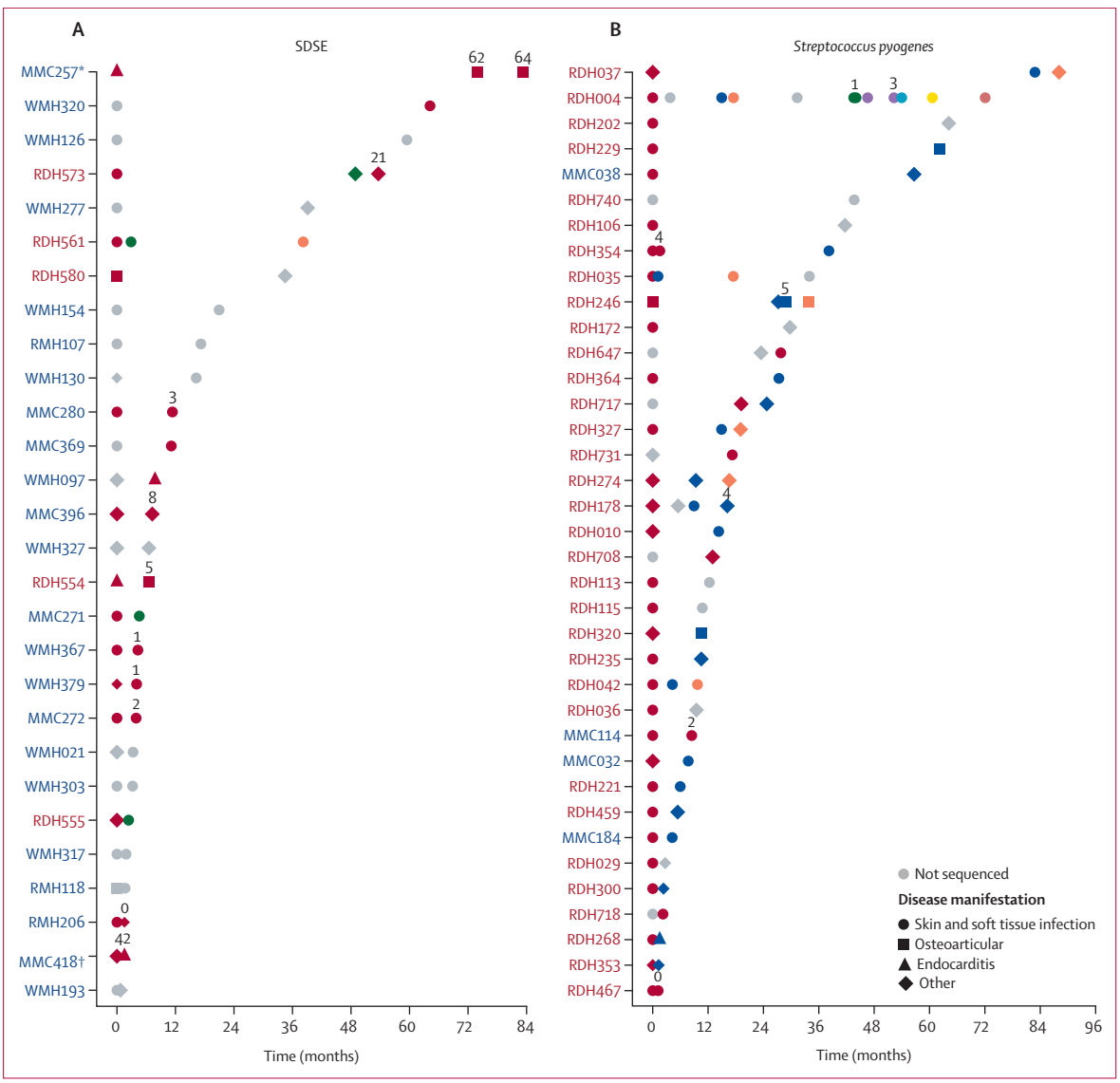


Figure 2: **Timeline of individuals with multiple episodes of invasive infection**
Individuals with multiple episodes of invasive SDSE infection (A) and with multiple episodes of *Streptococcus pyogenes* infection (B). Each row represents an individual with multiple episodes of infection, with labels in red representing cases from the Top End of the Northern Territory and labels in blue representing cases from the southeast of Australia. Strains defined by whole-genome sequencing are coloured arbitrarily within each individual, and the same strain (within the same sublineage) shares the same colour. Episodes without an isolate available for sequencing are coloured grey. The pairwise number of SNPs inferred by SKA between a case and the first isolate from the same sublineage is labelled above episodes. SNPs between cases from different sublineages are not shown (different colours). Select disease manifestations are indicated by the shape of points. SDSE=*Streptococcus dysgalactiae* subspecies *equisimilis*. SNP=single nucleotide polymorphism. *For MMC257, both subsequent infections have 50 SNPs compared with the index episode within a 14-kilobase region, indicative of recombination contributing most of the SNP distance. †For MMC418, 21 SNPs compared with the index episode occur over two separate regions (approximately 2 kilobases each), indicative of recombination.

28 individuals), whereas those with multiple *S pyogenes* infections were mostly from the Top End (33 [89%] of 37 individuals). Subsequent SDSE infections occurred up to 84 months after the index episode (median 7 months, IQR 3–19; figure 2A; appendix 1 p 13). Of 31 subsequent SDSE isolations in 28 individuals, ten (77%) of 13 individuals with sequences available were infected by an isolate from the same sublineage during a subsequent episode, differing by a median of four SNPs (IQR 2–26). By contrast, 60 subsequent isolations of *S pyogenes* occurred in 37 individuals (median 7 months, IQR 4–15). For 25 individuals with subsequent *S pyogenes* isolates sequenced, only six (24%) had isolates of the same sublineage (figure 2B).

From all available genomes, 395 (91%) of 436 SDSE genomes and 530 (83%) of 642 *S pyogenes* genomes were assigned to lineages with five or more representatives and were included in genomic transmission cluster analyses. Of these, 11 (3%) of 395 SDSE genomes and six (1%) of

530 *S pyogenes* genomes from subsequent infections were merged with index cases to form a total of 384 SDSE and 524 *S pyogenes* cases for transmission clustering. Inference of genomic transmission clusters found 24 (6%) of 384 SDSE cases were assigned to ten clusters compared with 271 (52%) of 524 *S pyogenes* cases in 72 clusters (figure 3A, B). No SDSE and only five *S pyogenes* clusters were shared across the Top End and southeast. Within clusters, the median time between cases was 130 days for SDSE (IQR 12–214) and 99 days (29–221) for *S pyogenes* (appendix 1 p 13). Sensitivity analyses by region were consistent, although a higher proportion of SDSE and *S pyogenes* cases in the Top End were assigned to transmission clusters compared with southeast Australia (figure 3C–F). In the Top End sensitivity analysis, 12 (15%) of 78 SDSE cases and 180 (66%) of 271 *S pyogenes* cases were assigned to genomic transmission clusters. In the southeast Australia sensitivity analysis, 12 (4%) of 306 SDSE cases and 85 (34%) of 253 *S pyogenes* cases were assigned to genomic transmission clusters.

Examining the sequenced SDSE pathogen population (436 [63%] of 693 bloodstream and non-bloodstream cases), there were 35 lineages encompassing 96 sublineages, 33 *emm*-types, 64 *emm*-subtypes, and 88 sequence types (figure 4A). Cases with available genome sequences were similar to those without sequences except with regard to deep soft tissue infection (40 [16%] 244 of missing cases *vs* 14 [3%] of 409 cases with available sequences; appendix 1 pp 28–29). Sublineages were generally well separated by SNP distance (appendix 1 pp 15–16). *emm*-type poorly aligned with lineages, with 17 *emm*-types found in more than one lineage (appendix 1 p 17). In contrast to *S pyogenes*, where lineage composition in the Top End was distinct from those in southeast Australia (Morisita distance 0·80), SDSE lineages overlapped between the regions (Morisita distance 0·33; figure 4A).[11]

A sublineage in this study, corresponding to the previously described *stG62647* lineage, was dominant and comprised 113 (26%) of 436 sequenced SDSE genomes, of which 97 (86%) were index bloodstream cases with complete clinical data (figure 4B).[8] Three *stG62647* lineage genomes were *emm*-type *stG6* due to recombination. The *stG62647* lineage was not associated with severe clinical outcomes (30 cases with severe outcomes and 67 cases with non-severe outcomes) when compared with all other lineages (69 severe outcomes and 187 non-severe outcomes; OR 1·21, 95% CI 0·72–2·01; p=0·46).

Compared with *S pyogenes*, for which there was serial lineage replacement over time in the Top End, the SDSE composition showed long-term maintenance of lineages, except for the emergence of an *stGLP1* sublineage in place of the *stG62647* lineage and a lineage associated with the *stGM220* *emm*-type since 2019 in the Top End (appendix 1 pp 18–19).[11] Prevalence of antimicrobial resistance genes and virulence factors in SDSE genomes are presented in appendix 2. When placed in a global context, SDSE genomes from this study were representative of most major global lineages (appendix 1 p 20).

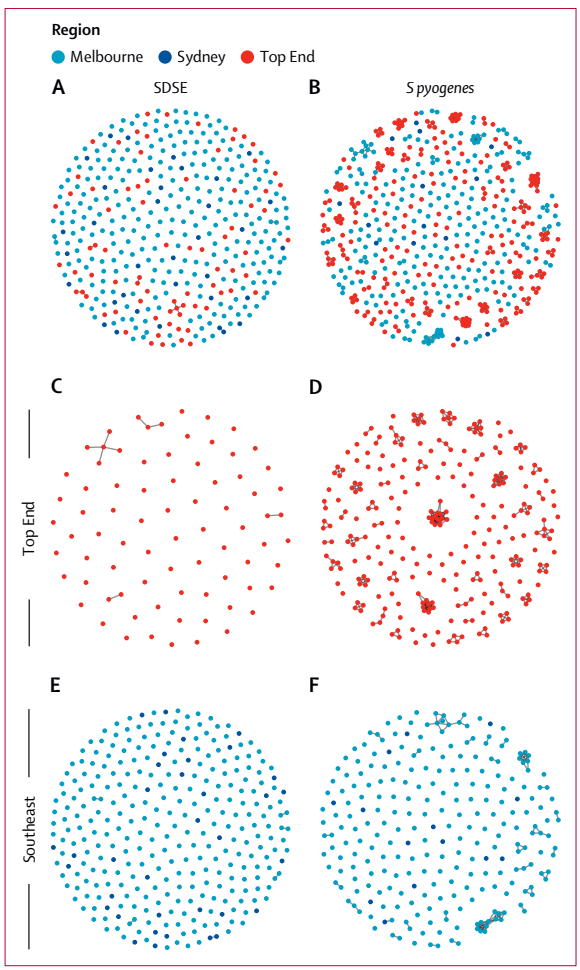


***Figure 3*: Genomic transmission clusters of cases from January, 2011, to February, 2023**

Invasive SDSE (A) and invasive *Streptococcus pyogenes* (B). All sequenced invasive cases, regardless of anatomical site of isolation, were included. SDSE (C) and *S pyogenes* (D) sensitivity analysis clustering cases from only the Top End. SDSE (E) and *S pyogenes* (F) sensitivity analysis clustering cases from only southeast Australia. Genomic transmission clusters are inferred as recent community transmission networks and not necessarily direct transmission links. Only genomes from lineages with at least five isolates were included. Isolates from the same sublineage in individuals with multiple episodes of infection are merged into a single case. Each node in the graph represents a case coloured by region, with Melbourne and Sydney in southeast Australia distinguished separately to provide greater resolution. Each link denotes a pairwise single nucleotide polymorphism distance less than the species-specific clustering threshold. SDSE=*Streptococcus dysgalactiae* subspecies *equisimilis*.

Examining 236 *stG62647* lineage genomes from seven countries and three continents collected between 2003 and 2022, we found evidence of local expansions with clustering of genomes by continent and international transmissions, with intermixing of regions at the tips of the tree (figure 5A). No representatives were found in genomes from east Asia (n=159), South America (n=3), or Africa (n=3). The

stG62647 lineage was estimated to have emerged around 1956 (95% highest posterior density [HPD] 1942 to 1967). The mean molecular clock rate was $5{\cdot}80\times10^{-7}$ nucleotide substitutions per site per year (95% HPD $4{\cdot}91\times10^{-7}$ to $-6{\cdot}67\times10^{-7}$). There was a rapid increase in effective population size between 1990 and 2005, a finding also recovered within each geographical region (figure 5B, C). There were no non-mobile genetic element-associated accessory genes or mobile genetic element gain or loss events unique to the *stG62647* lineage. All genomes within the lineage had a previously described disrupted negative regulator *silB* gene, with insertion of an IS1548 element after nucleotide position 243 indicative of ancestral acquisition.[8] *silB* was not disrupted in the nearest non-*stG62647* lineage genome (MMC234, ~500 recombination-masked SNPs distant; appendix 1 p 21).

Discussion

In this study, we identify an under-recognised burden of invasive SDSE disease in Australia, particularly in remote tropical northern Australia and First Nations populations, without a substantial decline during COVID-19 NPIs; we show that genomically, few SDSE cases were assigned to genomic transmission clusters compared with *S pyogenes* and subsequent SDSE infections in individuals with multiple infections were frequently caused by near-identical isolates; and we portray the multi-continent emergence of a dominant SDSE *stG62647* lineage.

The incidence of invasive SDSE was similar to *S pyogenes* in urban southeast Australia, except during COVID-19 NPIs in 2020–21. By contrast, SDSE incidence was four-times lower than *S pyogenes* in the largely remote Top End of the Northern Territory, yet crude SDSE incidence in the Top End remained higher than in southeast Australia. Like *S pyogenes*, invasive SDSE disease disproportionately affected First Nations Australians in the Top End, where remote First Nations communities have a disproportionate burden of poor health driven by social determinants such as household overcrowding, poor sanitation and water supply, and restricted access to health care.[7] These disparities are consistent with a previous data-linkage study from the state of Western Australia.[5]

Cross-species differences in incidence during COVID-19 NPIs in southeast Australia, genomic transmission clusters, and tendency for near-identical SDSE strains to cause multiple episodes of infection in individuals suggest there might be SDSE and *S pyogenes* transmission and host–pathogen differences leading to invasive disease. In southeast Australia, like international reports, COVID-19

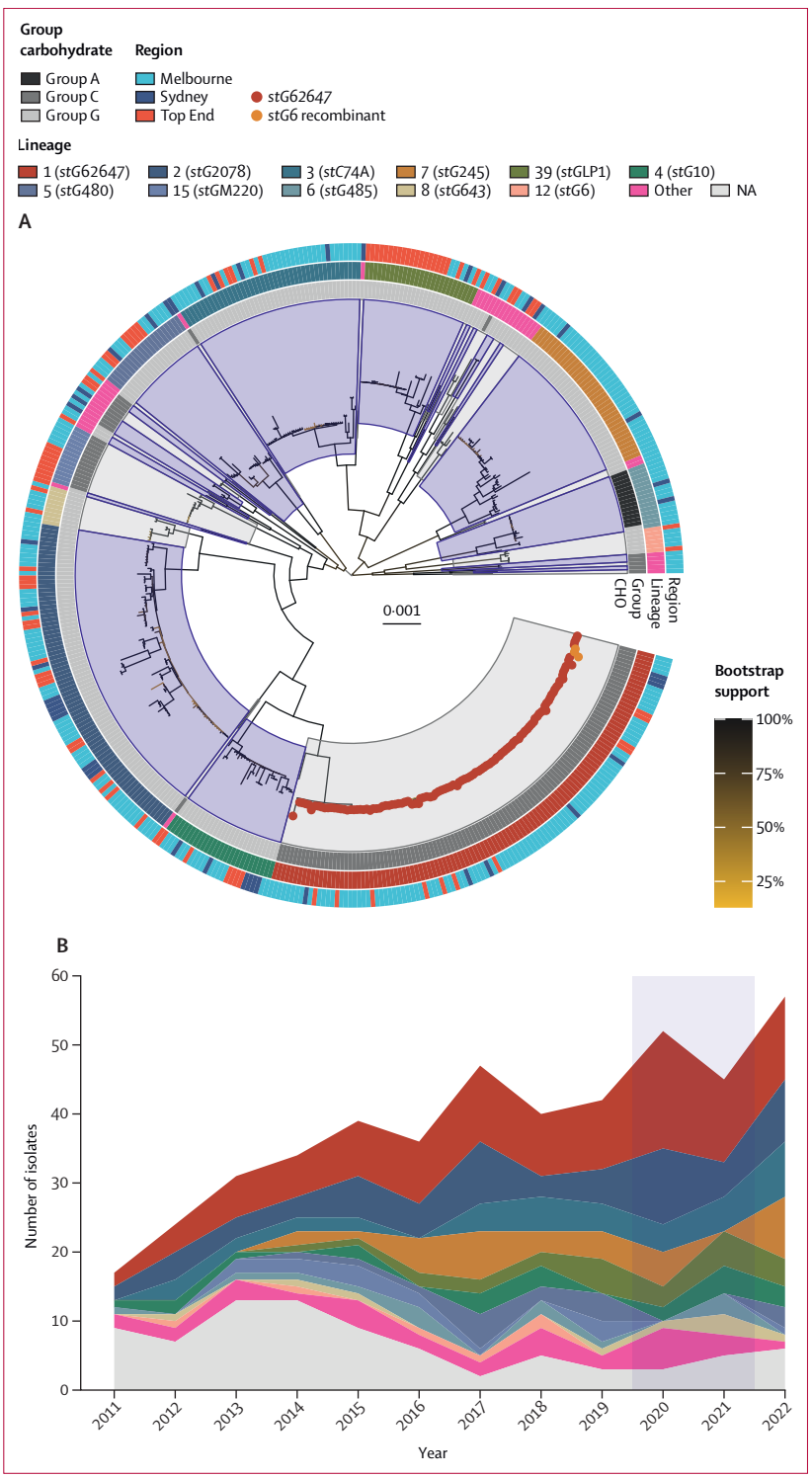


Figure 4: **Genomic population structure and temporal dynamics of invasive SDSE**
(A) Maximum likelihood phylogenetic tree of 436 SDSE genomes from Melbourne, Sydney, and the Top End of the Northern Territory inferred under a GTR+F+G4 model. Distinct lineages are highlighted by alternating blue and grey shades from internal nodes. From inner to outer rings, CHO predicted from the carbohydrate synthesis locus, highly represented lineages, and geographical region of isolation are labelled. A dominant sublineage carrying the *stG62647 emm*-type is labelled at the tips of the tree, including three *stG6* recombinant genomes. Lineages are labelled by the most frequent *emm*-type present within the lineage in brackets. Branches of the phylogeny are coloured by bootstrap support. The scale bar represents substitutions per site. (B) Longitudinal population dynamics until the end of 2022 from hospital networks in Melbourne and the Top End combined and coloured by lineage. Isolates from the first 2 months of 2023 are not included to preserve scale. Isolates not available for sequencing are shaded grey. The blue shading in 2020–21 approximates the interval of most stringent non-pharmaceutical interventions during the COVID-19 pandemic. CHO=Lancefield group carbohydrate. NA=not available. SDSE=*Streptococcus dysgalactiae* subspecies *equisimilis*.

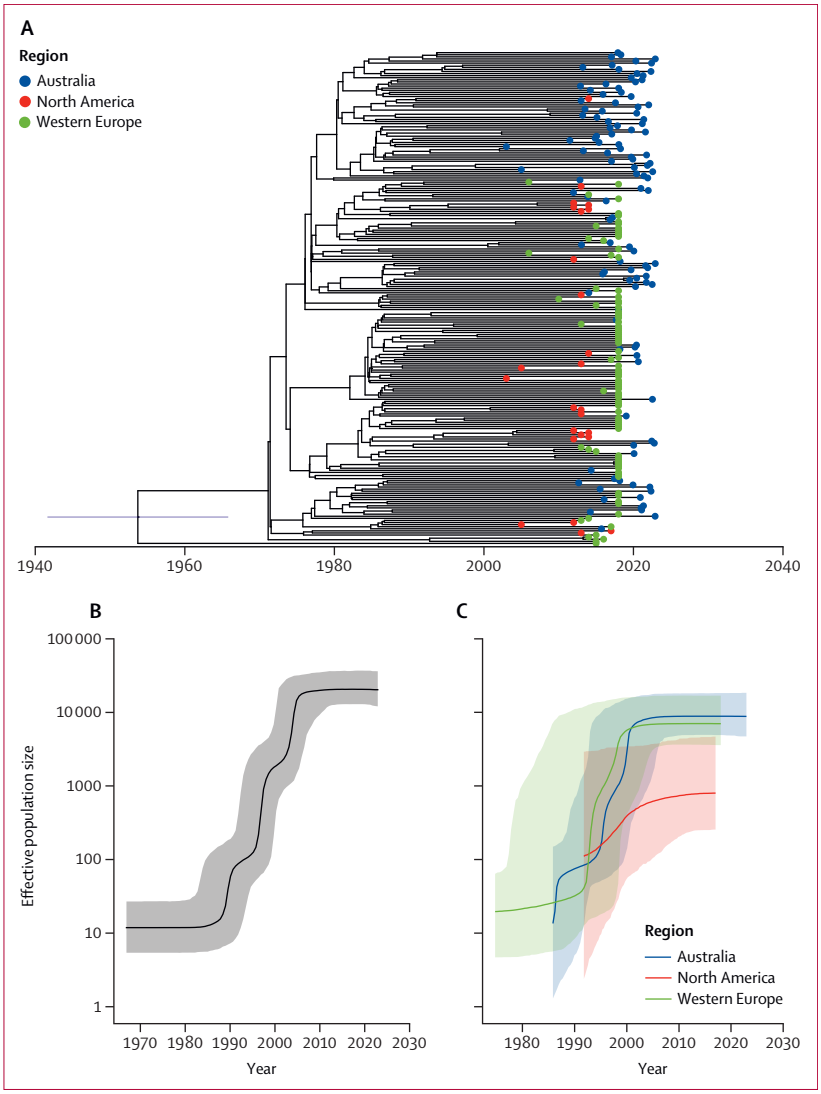


***Figure 5*: Time-scaled phylogeny of global *stG62647*-lineage genomes and estimated effective population size**
(A) Maximum clade credibility time-scaled phylogeny of 236 globally disseminated *stG62647* lineage genomes with mean inferred root date 1956. (B) Median inferred effective population size of global *stG62647* lineage genomes with shading representing 95% highest posterior density. The effective population size is a measure of diversity, with increasing values indicating a growing pathogen population. (C) Median inferred effective population size of *stG62647*-lineage genomes isolated from Australia (n=115), western Europe (n=96), and North America (n=25). Shading represents 95% highest posterior density in each region. The alignment was subset and effective population size inferred independently using genomes from each region for the region-specific analysis and thus represents the pathogen population captured from strains circulating in each region. The effective population size is estimated from inferred coalescent events to the time of most recent sample, resulting in different time ranges between the regions. Comparisons between regions should be interpreted with caution due to different sampling strategies and numbers of isolates.

NPIs interrupted *S pyogenes* transmission with historically low incidence in 2020–21, followed by a surge after lifting of NPIs.[11,22] Although the increasing trend in invasive SDSE incidence in southeast Australia plateaued during 2020–21, there was no substantial decline, which could be due to cross-species differences in carriage duration (whereby prolonged carriage would be less affected by transmission interruption) or differential cross-species impacts of NPIs on transmission. A similar incidence trend has been noted in western Europe.[4,6] There was no change in invasive SDSE or *S pyogenes* incidence in 2020–21 in the Top End, where prolonged stringent NPIs such as stay-at-home orders were not implemented.

Genomic transmission clustering showed that very few SDSE cases were closely related compared with *S pyogenes*. Geographically, more SDSE and *S pyogenes* cases were clustered in the Top End compared with southeast Australia, suggesting an interaction with host demography and disease burden. Based on published transmission models and given similar molecular evolutionary rates for SDSE and *S pyogenes* found in this study and previous studies, pairwise SNP distance and clustering would be expected to correlate with time between sampled genomes, number of intermediate hosts, and within-host evolution from carriage.[23–25] Therefore, these observations could be consistent with more unsampled intermediate hosts without invasive disease for SDSE or prolonged carriage with a lower transmission rate. Clinically, these findings suggest that close contact antimicrobial prophylaxis, which is used for invasive *S pyogenes* to prevent secondary cases and implies close or direct transmission between cases, is less likely to be beneficial for invasive SDSE. Accordingly, reports of secondary invasive SDSE disease in close contacts are rare.[26]

Examining individuals with multiple episodes of invasive disease, SDSE cases were frequently caused by near-identical isolates, indicating relapse associated with prolonged carriage or reinfection with or reacquisition of the same strain. This finding is concordant with published case series.[27,28] Data on the duration of carriage for SDSE are scarce, but carriage might be more than 12 months in some individuals.[2] Invasive SDSE disease occurs almost exclusively in older individuals and those with comorbidities who might have poorer immune responses to infection.[1,29] Data from a small uncontrolled study suggest individuals with comorbidities might not generate strain-specific opsonising immune responses after SDSE bacteraemia, which could contribute to prolonged carriage or recurrent disease with the same strain.[30] Taken together, our combined epidemiological and genomic observations suggest that invasive SDSE infection might arise from long-term carriage rather than after acute transmission events associated with clusters of invasive disease.

Invasive SDSE disease in this study was dominated by a previously described *stG62647* lineage, which we inferred to have expanded across multiple continents between 1990 and 2005, likely facilitated by multi-continent international dissemination.[8] Molecular events, such as disruption of the negative regulator gene *silB*, have been hypothesised to convey a selective advantage.[8] However, the lack of *stG62647* lineage genomes before the inferred expansion prohibited identification of evolutionary signatures associated with emergence and enhanced fitness of contemporary strains. In contrast to a previous report of 19 *stG62647* lineage cases and data from a mouse necrotising myositis model, we did not find an association between *stG62647* lineage infection and severe clinical outcomes.[8,9]

This study has limitations. We were unable to age-adjust incidence, as population-based incidence could not be estimated for sites in southeast Australia. Given the younger age profile of the Top End and First Nations populations and the higher incidence of SDSE in older people, our comparisons likely underestimate the IRR in the Top End compared with southeast Australia and in First Nations individuals. Additionally, population-based surveillance would improve point estimates of incidence, particularly during COVID-19 NPIs, when there were fluctuations in total acute hospitalisations. However, these fluctuations in overall acute hospitalisations during COVID-19 in our study were small and unlikely to have masked incidence trends. Larger case numbers with inclusion of additional clinical data, such as comorbidities, would improve the precision and power of cross-species and population comparisons and allow statistical testing for incidence trend changes with the introduction of COVID-19 NPIs. SDSE whole-genome sequencing is constrained in many regions and biased towards invasive disease. As such, global lineage diversity, including *stG62647* lineage dynamics, could not be inferred for the continents of Africa, Asia, and South America.

In conclusion, through comparison of co-collected SDSE and *S pyogenes* datasets in different epidemiological settings, we highlight the elevated burden and disparities of invasive SDSE disease in Australia and the multi-continent expansion of the *stG62647* lineage. Primordial disease prevention measures remain crucial to addressing the disproportionate burden of streptococcal disease in susceptible populations. Complementarily, increased global genomic surveillance and targeted carriage and immunological profiling of susceptible populations are required to provide a more comprehensive understanding of SDSE disease and confirm disease and transmission patterns raised by this study. These findings could inform disease control measures to combat streptococcal disease, including consideration of possible disparate effects of these measures across pathogens.

Contributors

OX worked on study design, microbial sequencing, data analysis, data interpretation, and initial manuscript preparation. MRD and SYCT contributed to conception of the project and data interpretation. LF did the phylodynamic analyses. LF and SD contributed to phylodynamic data interpretation. ANTN, MEP, and LJMC contributed to long-read sequencing and generation of new reference genomes. AJH and GT-H contributed to design of data analysis pipelines and data interpretation. SS, AL, RD, NJR, and AR contributed to clinical data collection, retrieval of bacterial specimens, and data curation. LMJ contributed to data curation and short-read sequencing. DCH, VLK, MVNO, RWB, KB, TMK, and BJC contributed to project design and data interpretation. All authors contributed to manuscript preparation and review. OX, MRD, and SYCT have verified the underlying data of the study. All authors had full access to all the data in the study and accept responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

Accession numbers for SDSE reads are listed in appendix 2 and are available in sequence read archive Bioprojects PRJNA1133199 and PRJNA857543. Sequences for complete genomes are available on National Center for Biotechnology Information GenBank with accession numbers listed in appendix 2. Accession numbers for *S pyogenes* reads and genomes are available from Bioprojects PRJNA1087113 and PRJNA857543 as reported previously.[11] An updated SDSE PopPUNK database (version 2) including all 436 genomes from this study and 1166 publicly available genomes is available at https://www.bacpop.org/poppunk. Global genomic sequence clusters 8 and 23 are merged in the updated database but were reported separately in this study. Full de-identified clinical data are available from the corresponding author on reasonable request and subject to ethical approval.

Acknowledgments

We thank Despina Kotsanas, Kathryn Cisera, Marcia Scida, Celeste Woerle, Vanessa Rigas, Mitchell Brown, and the clinical and research microbiology teams at Monash Health, Royal Melbourne Hospital, Menzies School of Health Research, Territory Pathology, and Westmead Hospital for their assistance in retrieving bacterial isolates. We also thank Taylor Harshegyi and Doherty Applied Microbial Genomics for sample processing and short-read Illumina sequencing. OX was supported by the Australian National Health and Medical Research Council postgraduate scholarship (GNT2013831) and Avant Foundation Doctors in Training Research Scholarship (2021/000017).

References

1 Xie O, Davies MR, Tong SYC. *Streptococcus dysgalactiae* subsp. *equisimilis* infection and its intersection with *Streptococcus pyogenes*. *Clin Microbiol Rev* 2024; **37:** e0017523.

2 Xie O, Zachreson C, Tonkin-Hill G, et al. Overlapping *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subspecies *equisimilis* household transmission and mobile genetic element exchange. *Nat Commun* 2024; **15:** 3477.

3 Trell K, Rignér S, Wierzbicka M, Nilson B, Rasmussen M. Colonization of β-hemolytic streptococci in patients with erysipelas-a prospective study. *Eur J Clin Microbiol Infect Dis* 2019; **38:** 1901–06.

4 Oppegaard O, Glambek M, Skutlaberg DH, et al. *Streptococcus dysgalactiae* bloodstream infections, Norway, 1999–2021. *Emerg Infect Dis* 2023; **29:** 260–67.

5 Wright CM, Moorin R, Pearson G, et al. Invasive infections caused by Lancefield groups C/G and A *Streptococcus*, Western Australia, Australia, 2000–2018. *Emerg Infect Dis* 2022; **28:** 2190–97.

6 Itzek A, Weißbach V, Meintrup D, et al. Epidemiological and clinical features of *Streptococcus dysgalactiae* ssp. *equisimilis stG*62647 and other *emm* types in Germany. *Pathogens* 2023; **12:** 589.

7 May PJ, Bowen AC, Carapetis JR. The inequitable burden of group A streptococcal diseases in Indigenous Australians. *Med J Aust* 2016; **205:** 201–03.

8 Oppegaard O, Mylvaganam H, Skrede S, Lindemann PC, Kittang BR. Emergence of a *Streptococcus dysgalactiae* subspecies *equisimilis stG*62647-lineage associated with severe clinical manifestations. *Sci Rep* 2017; **7:** 7589.

9 Beres SB, Olsen RJ, Long SW, et al. Analysis of the genomics and mouse virulence of an emergent clone of *Streptococcus dysgalactiae* subspecies *equisimilis*. *Microbiol Spectr* 2023; **11:** e0455022.

10 Xie O, Morris JM, Hayes AJ, et al. Inter-species gene flow drives ongoing evolution of *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis*. *Nat Commun* 2024; **15:** 2286.

11 Xie O, Chisholm RH, Featherstone L, et al. Temporal and geographical lineage dynamics of invasive *Streptococcus pyogenes* in Australia from 2011 to 2023: a retrospective, multicentre, clinical and genomic epidemiology study. *Lancet Microbe* 2025; **6:** 101053.

12 US Centers for Disease Control and Prevention. Streptococcal toxic shock syndrome (STSS) (*Streptococcus pyogenes*) 2010 case definition. 2010. https://ndc.services.cdc.gov/case-definitions/streptococcal-toxic-shock-syndrome-2010/ (accessed Aug 6, 2022).

13 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.

14 Lees JA, Harris SR, Tonkin-Hill G, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 2019; **29:** 304–16.

15 Lacey JA, Marcato AJ, Chisholm RH, et al. Evaluating the role of asymptomatic throat carriage of *Streptococcus pyogenes* in impetigo transmission in remote Aboriginal communities in Northern Territory, Australia: a retrospective genomic analysis. *Lancet Microbe* 2023; **4:** e524–33.

16 Wick RR, Judd LM, Cerdeira LT, et al. Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* 2021; **22:** 266.

17 Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 2005; **22:** 1185–92.

18 Tonkin-Hill G, MacAlasdair N, Ruis C, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020; **21:** 180.

19 Jespersen MG, Hayes A, Davies MR. Corekaburra: pan-genome post-processing using core gene synteny. *J Open Source Softw* 2022; **7:** 4910.

20 Australian Bureau of Statistics. Regional population by age and sex. 2022. https://www.abs.gov.au/statistics/people/population/regional-population-age-and-sex/2022 (accessed Dec 11, 2023).

21 Morisita M. Measuring of interspecific association and similarity between assemblages. *Mem Fac Sci Kyushu Univ Ser E Biol* 1959; **3:** 65–80.

22 Guy R, Henderson KL, Coelho J, et al. Increase in invasive group A streptococcal infection notifications, England, 2022. *Euro Surveill* 2023; **28:** 2200942.

23 Stimson J, Gardy J, Mathema B, et al. Beyond the SNP threshold: identifying outbreak clusters using inferred transmissions. *Mol Biol Evol* 2019; **36:** 587–603.

24 Duchêne S, Holt KE, Weill FX, et al. Genome-scale rates of evolutionary change in bacteria. *Microb Genom* 2016; **2:** e000094.

25 Didelot X, Fraser C, Gardy J, Colijn C. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. *Mol Biol Evol* 2017; **34:** 997–1007.

26 Sunagawa K, Shirafuji T, Sun G, et al. Intra-familial transmission of *Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE): a first case report and review of the literature. *J Infect Chemother* 2022; **28:** 819–22.

27 Trell K, Sendi P, Rasmussen M. Recurrent bacteremia with *Streptococcus dysgalactiae*: a case-control study. *Diagn Microbiol Infect Dis* 2016; **85:** 121–24.

28 Senneby E, Hallström B, Rasmussen M. Genetic relatedness of *Streptococcus dysgalactiae* isolates causing recurrent bacteraemia. *J Med Microbiol* 2021; **70**. https://doi.org/10.1099/jmm.0.001330.

29 Nevanlinna V, Huttunen R, Aittoniemi J, Luukkaala T, Rantala S. Major risk factors for *Streptococcus dysgalactiae* subsp. *equisimilis* bacteremia: a population-based study. *BMC Infect Dis* 2023; **23:** 43.

30 Bläckberg A, de Neergaard T, Frick IM, et al. Lack of opsonic antibody responses to invasive infections with *Streptococcus dysgalactiae*. *Front Microbiol* 2021; **12:** 635591.