

## Whole genome sequencing reveals extensive community-level transmission of group A *Streptococcus* in remote communities

A. C. BOWEN<sup>1,2,3</sup>, T. HARRIS<sup>1</sup>, D. C. HOLT<sup>1</sup>, P. M. GIFFARD<sup>1</sup>,  
J. R. CARAPETIS<sup>2,3</sup>, P. T. CAMPBELL<sup>4</sup>, J. McVERNON<sup>4</sup> AND S. Y. C. TONG<sup>1,5\*</sup>

<sup>1</sup> Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia

<sup>2</sup> Princess Margaret Hospital for Children, Perth, WA, Australia

<sup>3</sup> Telethon Kids Institute, University of Western Australia, Perth, WA, Australia

<sup>4</sup> University of Melbourne, Melbourne, VIC, Australia

<sup>5</sup> Royal Darwin Hospital, Darwin, NT, Australia

Received 19 September 2015; Final revision 11 December 2015; Accepted 16 December 2015;  
first published online 1 February 2016

### SUMMARY

Impetigo is common in remote Indigenous children of northern Australia, with the primary driver in this context being *Streptococcus pyogenes* [or group A *Streptococcus* (GAS)]. To reduce the high burden of impetigo, the transmission dynamics of GAS must be more clearly elucidated. We performed whole genome sequencing on 31 GAS isolates collected in a single community from children in 11 households with  $\geq 2$  GAS-infected children. We aimed to determine whether transmission was occurring principally within households or across the community. The 31 isolates were represented by nine multilocus sequence types and isolates within each sequence type differed from one another by only 0–3 single nucleotide polymorphisms. There was evidence of extensive transmission both within households and across the community. Our findings suggest that strategies to reduce the burden of impetigo in this setting will need to extend beyond individual households, and incorporate multi-faceted, community-wide approaches.

**Key words:** Impetigo, infectious disease epidemiology, *Streptococcus pyogenes*, transmission.

### INTRODUCTION

Impetigo is a common childhood infection [1] with an estimated 160 million prevalent cases globally [2]. Children living in Oceania have the highest documented prevalence [2], with the most severely affected group being Indigenous Australian children living in remote communities [3, 4], where the median childhood prevalence is 43% (95% CI 40.2–45.7) [2]. Impetigo is a non-benign disease that drives outbreaks of acute post-streptococcal glomerulonephritis [5] with

consequent chronic kidney disease [6, 7] and probably contributes to the highest reported rates of rheumatic heart disease in the world found in these communities [8, 9]. In a large randomized controlled trial, *Streptococcus pyogenes* [or group A *Streptococcus* (GAS)] was confirmed to be the primary driver of impetigo in an endemic context [10]. Therefore a clear understanding of the transmission dynamics of GAS is required to inform the design of interventions to reduce the prevalence of impetigo.

It has long been thought that housing conditions are a major contributor to the high prevalence of infectious and parasitic disease in children in remote Indigenous communities. For example, households are crowded in remote Indigenous communities with

\* Author for correspondence: A/Professor S. Y. C. Tong, Menzies School of Health Research, PO Box 41096, Casuarina, NT 0811, Australia.  
(Email: Steven.tong@menzies.edu.au)

a median of 3–7 persons per bedroom [10, 11] and a correlation exists between the level of crowding and prevalence of impetigo [11]. Interventions that improve housing quality have had only limited success in reducing the burdens of infectious disease [12], although interventions to reduce household crowding have not been tested. Part of the reason for the lack of impact of improved housing quality may be extensive transmission of infectious agents outside the household. Community interactions within Indigenous communities are complex, with considerable mobility of children between households, and much unstructured mixing opportunities in school and other community settings that may be more or less influential to infection risk than the home environment [13].

To better understand the relative contributions of household-level and community-level transmission, we obtained whole genome sequences (WGS) of 31 GAS isolates from household clusters of impetigo in a single community over a 3-day period. By assessing the relatedness of GAS strains associated with skin infections in multiple members of individual households, we sought to assess whether the transmission from other household members, or acquisition outside the household, was the likely source of infection, and thus whether individual households would likely be the most useful target for interventions to reduce acquisition of GAS and the subsequent burden of impetigo.

## METHODS

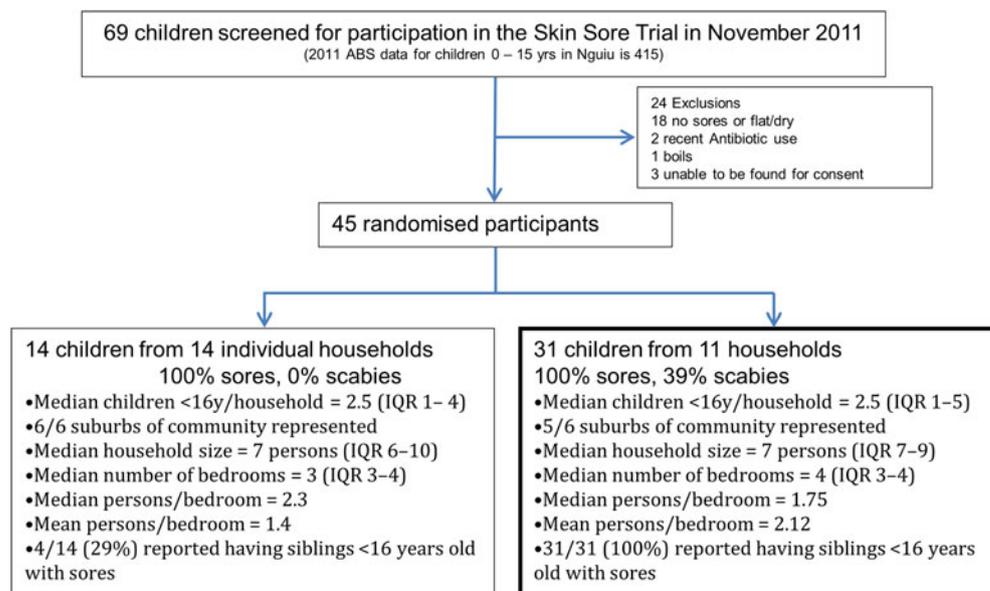
Isolates were collected from children with impetigo who were participants in a randomized controlled trial of oral trimethoprim-sulfamethoxazole *vs.* intramuscular benzathine benzylpenicillin G for the treatment of skin sores [10]. The trial recruited children aged 3 months to 13 years from seven remote communities in the Northern Territory (NT), Australia where the prevalence of impetigo is high year-round [14, 15]. Screening for eligibility in the trial occurred predominantly in schools. Following this, research nurses visited households to discuss the study with caregivers. Overall 65% of recruited children with impetigo were identified through school screening. During these household visits, siblings or relatives were also screened for participation (an additional 25% of recruitment). The remaining participants (10%) were referred directly from the clinic [10]. No attempt was made to screen children who did not attend school or were below school age, outside of the recruitment described above. Impetigo was graded as ‘mild’ if

there was one purulent or crusted sore and <5 sores in total, or ‘severe’ if there were  $\geq 2$  purulent or crusted sores or  $\geq 5$  sores in total. All children with impetigo in the trial had at least one microbiological swab collected [16]. To ascertain household crowding and number of other children with impetigo in the household a survey of the primary caregiver was conducted. Apart from children with impetigo, no swabs or clinical assessments of other household members were made.

For this WGS substudy, we concentrated on a single community of ~1500 population, at a single recruitment visit conducted between 8 and 11 November 2011. This community was chosen because it had the highest number of GAS isolates available from different sized household clusters. We included for sequencing isolates recovered from children residing in households where  $\geq 2$  children from the same household had culture-confirmed GAS impetigo. This sampling frame was selected to determine whether there was evidence of children acquiring GAS within the household or elsewhere. All isolates were collected prior to the commencement of antibiotic therapy.

Cotton swabs (Copan, Italy) were used to collect pus from skin sores for microbiological culture according to standard methods [17]. Where GAS was recovered, a single isolate from the agar plate was stored at  $-80^{\circ}\text{C}$ . DNA was extracted from the stored isolate using a QIAamp DNA Mini kit (Qiagen, Germany) according to the manufacturer’s instructions.

We obtained genome sequences with paired end libraries on the Illumina HiSeq (Illumina, USA) platform through Macrogen Inc. (South Korea). The reads have been deposited in NCBI Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>) with accession numbers SAMN03988118–SAMN03988148. Reads were mapped against the M1 GAS reference sequence of *S. pyogenes* isolates SF370 [18] (accession no. AE004092.2) using SPANDx [19] to define orthologous core genome single nucleotide polymorphisms (SNPs). To restrict the analysis to core genome SNPs, reads mapping to repetitive regions, regions with less than half the average depth, and regions with more than three times the average coverage of the entire genome were filtered out. We built a maximum likelihood tree using RaXML [20] with default settings and visualized with the Interactive Tree of Life [21]. The multi-locus sequence type (MLST) for each isolate was determined using SRST2 [22]. Short read data for each isolate was assembled using Velvet [23] with reference to SF370 (accession no. AE004092.2). Draft assemblies have been deposited at DDBJ/EMBL/GenBank



**Fig. 1.** Flow diagram of participants in the study.

under the accessions LRGF00000000–LRHJ00000000. The versions described in this paper are versions LRGF01000000–LRHJ01000000, and assembly metrics are provided in Supplementary Table S1. The best assembly for each sequence type (ST) (based on the smallest number of contigs and largest N50) was then used as a reference to map reads from the other strains of that ST to identify orthologous SNPs using SPANDx [19]. Reads from the assembled ‘reference’ strain were also mapped back to itself as a quality control procedure. SNPs were manually visualized in Artemis [24]. Two sets of short reads (SST2097-1, SST2090-1) did not assemble well (>300 contigs rather than 10–32 for other assemblies). However, the short reads from SST2097-1 and SST2090-1 mapped well against the reference assembly and alignments were manually checked. Epidemiological and genome sequence data were visualized using Circos [25].

### Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

### RESULTS

Of 69 children who were screened at the school for participation in the trial during the November 2011

recruitment, 45 (65%) with crusted or purulent impetigo were enrolled in the study. These 45 children had a median age of 7.4 years [interquartile range (IQR) 4.4–9.7 years] and 27 (60%) were female. Fourteen (31%) of these 45 children were the only member of their household with impetigo recruited in the trial and isolates recovered from these participants were not included for WGS. The remaining 31 children (69%) resided in 11 households, with a median of 3 (IQR 2–3) infected individuals in each household. Twenty-five (81%) of the 31 children in households where multiple infected children were identified had severe impetigo, compared to nine (64%) of the 14 children with impetigo from households without other cases ( $P = 0.4$ ). (See Fig. 1 for the study profile.)

Of the 31 isolates, there were nine STs that were clearly delineated following alignment against the reference GAS strain SF370 (Fig. 2). In order to provide improved resolution and finer mapping within each ST, the best *de novo* assembly within each ST was chosen to be the reference for that ST (median size of reference assemblies 1 801 299 bp, range 1 745 425–1 878 702 bp). Mapping of short reads from the reference isolate back to the reference assembly (i.e. itself) demonstrated no SNPs. Overall, the median depth of read coverage for all 31 sequences aligning against the reference assemblies was 133×. Within each ST, the number of SNPs for any one isolate compared to the reference assembly ranged from 0 to 3 (Table 1). Thus, isolates within each ST were essentially identical when considering orthologous SNPs at a



Table 1. Details of the 31 group A *Streptococcus* (*GAS*) strains

Isolate	Household	ST	<i>emm</i> type	SNP variant	SRA number
2090-1	A	10	70	No SNPs	SAMN03988125
2097-1	A	10	70	No SNPs	SAMN03988127
2096-1	A	332	166.1	CCGA	SAMN03988126
1023-1	B	332	166.1	CAGG	SAMN03988121
2245-1	B	332	166.1	TAGG‡	SAMN03988122
1024-1*	C	11	53		SAMN03988123
2232-1	C	176	58	No SNPs	SAMN03988124
1028-1	D	10	70	No SNPs	SAMN03988118
2237-1	D	10	70	No SNPs	SAMN03988119
2243-1	D	332	166.1	CCTG	SAMN03988120
2241-1*	E	205	230		SAMN03988130
2240-1	E	332	166.1	CCGA	SAMN03988129
2229-1	E	330†	164.3	No SNPs	SAMN03988128
2102-1*	F	304	108	CTA	SAMN03988132
2099-1	F	332	166.1	CCTG	SAMN03988131
2103-1*	F	332	166.1	CCGG	SAMN03988133
2095-1*	G	176	58	No SNPs	SAMN03988138
2091-1	G	304	108	TGG	SAMN03988134
2093-1	G	304	108	TGG	SAMN03988136
2094-1	G	304	108	CTA	SAMN03988137
2092-1	G	332	166.1	CCGG	SAMN03988135
1020-1	H	10	70	No SNPs	SAMN03988139
1021-1*	H	10	70	No SNPs	SAMN03988140
2098-1*	I	182	205	G	SAMN03988141
2242-1	I	182	205	A	SAMN03988142
1019-1	J	330†	164.3	No SNPs	SAMN03988143
2088-1*	J	330†	164.3	No SNPs	SAMN03988144
2233-1*	K	641	219	No SNPs	SAMN03988145
2235-1	K	641	44	No SNPs	SAMN03988146
2236-1	K	641	219	No SNPs	SAMN03988147
2238-1	K	641	219	No SNPs	SAMN03988148

SNP, Single nucleotide polymorphism; SRA, short read archive.

\* Isolate used as the 'reference' assembly for that sequence type (ST).

† Single locus variant of ST330.

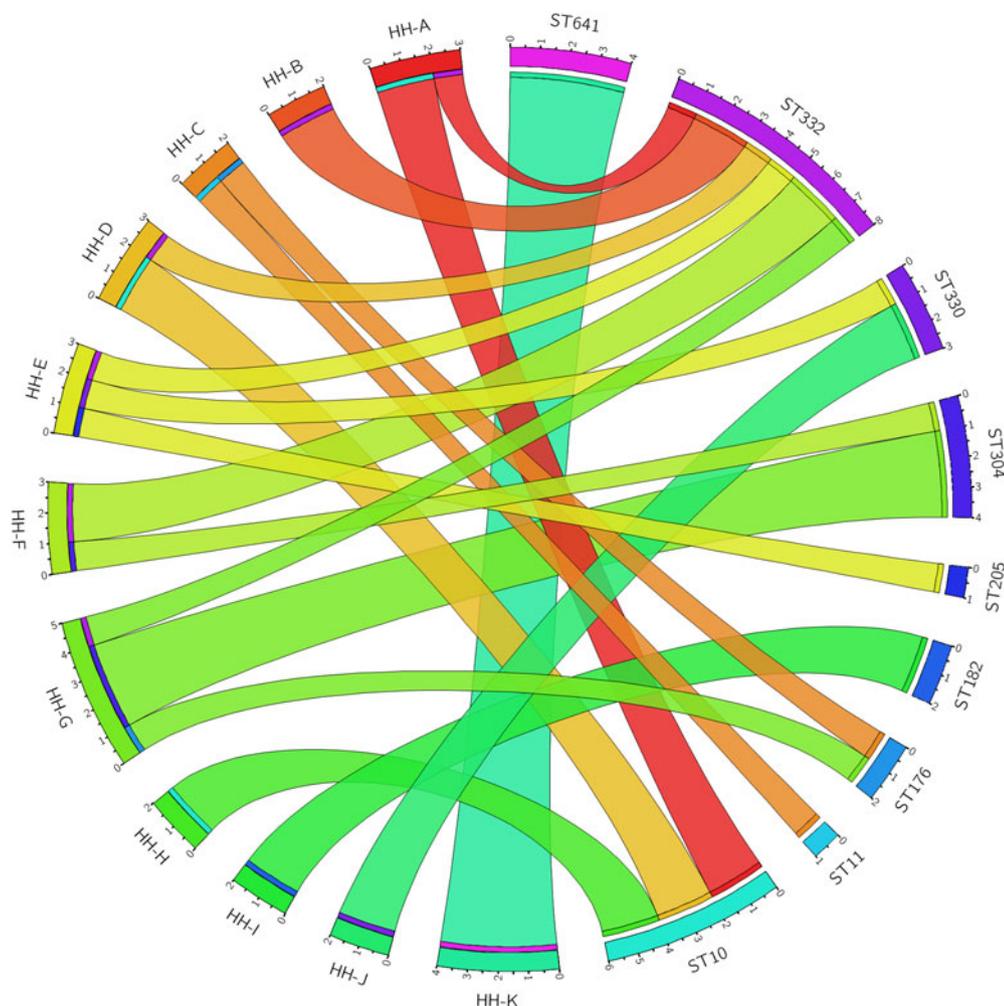
The SNP (single nucleotide variant) column refers to the SNP variants within each of the STs. (See Supplementary Table S1 for details of these SNPs.) Where there is only one representative of a ST, the cell is left blank.

This study was conducted in households with a heavy burden of disease. The households had a median of three children with impetigo at the time of screening and the impetigo was typically graded as severe. Previous studies in this regional endemic setting have demonstrated a considerable diversity of GAS when longitudinal sampling was performed in a small number of communities [26, 27], or with broader sampling for diversity across the entire region [28]. We show here that extensive diversity of STs exists even at a single time point (sampling over a 4-day period) and within a single community.

The previous studies in our region [26–28] were able to determine diversity at a MLST or *emm*-type level, but lacked the resolution now available with WGS

to understand the relationship of strains within these STs or *emm* types. We found that strains within each ST differed by at most three SNPs despite being in multiple households in some cases, representing clear evidence of recent transmission between several households (Fig. 3).

Most of the clones presented in this study are uncommon in the MLST database (<http://pubmlst.org/spyogenes/>). Specifically, *emm*166.1/ST332, *emm*219/ST641, *emm*164.3/ST330 (SLV) are all novel combinations not present in the MLST database; there are four *emm*70/ST10 strains – all from the NT; five *emm*53/ST11 strains – four from the NT, one from Brazil; two *emm*108/ST304 strains – both from the NT; one *emm*44/ST641 strain from Guyana; one *emm*205/



**Fig. 3.** Group A *Streptococcus* (GAS) sequence type (ST) distribution in households. Segments on the left of the circle represent 11 households (HH) and segments on the right represent nine GAS STs. Denoted within each segment are the number of participants with GAS recovered from an impetigo lesion within each household, and similarly the number of isolates within each ST.

ST182 strain from the United States; and one *emm230*/ST205 strain from Australia. *emm58*/ST176 appears to be a globally distributed clone with 36 strains in the MLST database from Australia (15), Europe (7), UK (3), the Americas (8) and Asia (3).

Although this report is the first to utilize WGS to examine transmission of GAS in impetigo lesions in an endemic region, the similarity of the clinical epidemiology and diversity of *emm* types from our region with other countries such as India [29, 30], Nepal [31], the Pacific Islands [32], Brazil [33] and Africa [34], suggests that our findings are likely to be applicable to other areas with a high burden of impetigo [2, 35, 36]. With regards to GAS population biology, despite the small size of this study and previous sampling in our region, we found novel combinations of *emm* types and STs. For example, *emm166*-1/ST332, *emm219*/ST641, and

*emm164*-3/ST330 (SLV) are all novel combinations not present in the MLST database. In addition, we found direct evidence of recombination at the *emm* locus in the four ST641 strains, with three being *emm219* and one *emm44*. Notably, ST182 was associated with *emm205* in this study, but ST182 has also been associated with *emm98* and *emm101* in strains from Australia. Taken together with the findings of multiple STs, each with little intra-lineage variation, it appears that in endemic, high-transmission settings such as in tropical northern Australia, not only is there frequent strain replacement but also present are the conditions that may facilitate recombination and horizontal gene transfer.

Limitations of this study include the use of a single time-point and a single community to assess for household clustering. However, even with this small sample size, it is clear that community-level

transmission of GAS is a key component of the dynamic epidemiology of GAS in an endemic setting. In addition, an inference of multiple importations of GAS into a household may be incorrect if individuals harboured multiple co-infecting strains, making attribution of a source other than the household spurious on the basis of a single isolate. Studies involving WGS of multiple *S. aureus* colonies from a single swab have demonstrated carriage of a cloud of variation [37, 38]; however, all colonies were of the same ST. Such studies have not yet been reported for GAS. Finally, our sampling in households was not complete and included only symptomatic children recruited into the study. Swabbing of all household members for either active infection or asymptomatic colonization over time would be required to understand underlying patterns of transmission in the household that may be influential, other than just observed disease.

Nonetheless, our findings suggest that interventions purely targeted at individual households (e.g. improved living conditions, reduced household crowding) may not be effective at reducing the prevalence of impetigo unless the majority of residents in a community benefit from the intervention and/or there are other community-wide intervention strategies. Therefore, it is not surprising that small-scale interventions targeting housing quality alone did not result in an appreciable reduction in infectious diseases [12].

## CONCLUSIONS

We report the first use of WGS to define the household transmission of GAS in a remote Indigenous community. It is likely that strategies aimed at reducing the burden of impetigo in remote Indigenous Australian children, and possibly in other high-burden settings, will need to include community-wide interventions.

## SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S095026881500326X>.

## ACKNOWLEDGEMENTS

We thank the participants, their families and study staff from Menzies School of Health Research for their contributions to this study.

An Australian National Health and Medical Research Council (NHMRC) project grant (545 346) and a Menzies School of Health Research Small

Grant supported this work. A.B. is an NHMRC Early Career Fellow (1 088 735) and S.T. (1 065 736) and J.McV. (1 061 321) are NHMRC Career Development Fellows.

## DECLARATION OF INTEREST

None.

## REFERENCES

1. **Koning S, et al.** Interventions for impetigo. *Cochrane Database of Systematic Reviews* 2012; **1**: CD003261.
2. **Bowen AC, et al.** The global epidemiology of impetigo: a systematic review of the population prevalence of impetigo and pyoderma. *PLoS ONE* 2015; **10**: e0136789.
3. **Carapetis JR, et al.** The global burden of group A streptococcal diseases. *Lancet Infectious Diseases* 2005; **5**: 685–694.
4. **Currie BJ, Carapetis JR.** Skin infections and infestations in Aboriginal communities in northern Australia. *Australasian Journal of Dermatology* 2000; **41**: 139–143.
5. **Marshall CS, et al.** Acute post-streptococcal glomerulonephritis in the Northern Territory of Australia: a review of 16 years data and comparison with the literature. *American Journal of Tropical Medicine and Hygiene* 2011; **85**: 703–710.
6. **White AV, Hoy WE, McCredie DA.** Childhood post-streptococcal glomerulonephritis as a risk factor for chronic renal disease in later life. *Medical Journal of Australia* 2001; **174**: 492–496.
7. **Hoy WE, et al.** Post-streptococcal glomerulonephritis is a strong risk factor for chronic kidney disease in later life. *Kidney International* 2012; **81**: 1026–1032.
8. **McDonald M, Currie BJ, Carapetis JR.** Acute rheumatic fever: a chink in the chain that links the heart to the throat? *Lancet Infectious Diseases* 2004; **4**: 240–245.
9. **Parnaby MG, Carapetis JR.** Rheumatic fever in indigenous Australian children. *Journal of Paediatrics and Child Health* 2010; **46**: 527–533.
10. **Bowen AC, et al.** Short-course oral co-trimoxazole versus intramuscular benzathine benzylpenicillin for impetigo in a highly endemic region: an open-label, randomised, controlled, non-inferiority trial. *Lancet* 2014; **384**: 2132–2140.
11. **McDonald MI, et al.** Low rates of streptococcal pharyngitis and high rates of pyoderma in Australian aboriginal communities where acute rheumatic fever is hyperendemic. *Clinical Infectious Diseases* 2006; **43**: 683–689.
12. **Baillie RS, Stevens M, McDonald EL.** The impact of housing improvement and socio-environmental factors on common childhood illnesses: a cohort study in Indigenous Australian communities. *Journal of Epidemiology and Community Health* 2012; **66**: 821–831.
13. **McDonald EL, Baillie RS, Morris PS.** Participatory systems approach to health improvement in Australian Aboriginal children. *Health Promotion International*.

- Published online: 4 February 2014. doi:10.1093/heapro/dau003.
14. **Carapetis JR, et al.** Success of a scabies control program in an Australian aboriginal community. *Pediatric Infectious Diseases Journal* 1997; **16**: 494–499.
  15. **Andrews RM, et al.** A regional initiative to reduce skin infections amongst aboriginal children living in remote communities of the Northern Territory, Australia. *PLoS Neglected Tropical Diseases* 2009; **3**: e554.
  16. **Bowen AC, et al.** The microbiology of impetigo in Indigenous children: associations between *Streptococcus pyogenes*, *Staphylococcus aureus*, scabies, and nasal carriage. *BMC Infectious Diseases* 2014; **14**: 727.
  17. **Bowen AC, et al.** Comparison of three methods for the recovery of skin pathogens from impetigo swabs collected in a remote community of Northern Territory, Australia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2013; **107**: 384–389.
  18. **Ferretti JJ, et al.** Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proceedings of the National Academy of Sciences USA* 2001; **98**: 4658–4663.
  19. **Sarovich DS, Price EP.** SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. *BMC Research Notes* 2014; **7**: 618.
  20. **Stamatakis A.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.
  21. **Letunic I, Bork P.** Interactive Tree Of Life v. 2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* 2011; **39** (Web Server issue): W475–478.
  22. **Inouye M, et al.** SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine* 2014; **6**: 90.
  23. **Zerbino DR, Birney E.** Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 2008; **18**: 821–829.
  24. **Carver T, et al.** Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 2012; **28**: 464–469.
  25. **Krzywinski M, et al.** Circos: an information aesthetic for comparative genomics. *Genome Research* 2009; **19**: 1639–1645.
  26. **Richardson LJ, et al.** Diversity of emm sequence types in group A beta-haemolytic *streptococci* in two remote Northern Territory Indigenous communities: implications for vaccine development. *Vaccine* 2010; **28**: 5301–5305.
  27. **McDonald MI, et al.** The dynamic nature of group A streptococcal epidemiology in tropical communities with high rates of rheumatic heart disease. *Epidemiology and Infection* 2008; **136**: 529–539.
  28. **Towers RJ, et al.** Extensive diversity of *Streptococcus pyogenes* in a remote human population reflects global-scale transmission rather than localised diversification. *PLoS ONE* 2013; **8**: e73851.
  29. **Dey N, et al.** High diversity of group A Streptococcal emm types in an Indian community: the need to tailor multivalent vaccines. *Clinical Infectious Diseases* 2005; **40**: 46–51.
  30. **Sagar V, et al.** Molecular heterogeneity among north Indian isolates of group A *Streptococcus*. *Letters in Applied Microbiology* 2004; **39**: 84–88.
  31. **Sakota V, et al.** Genetically diverse group A *streptococci* from children in far-western Nepal share high genetic relatedness with isolates from other countries. *Journal of Clinical Microbiology* 2006; **44**: 2160–2166.
  32. **Steer AC, et al.** emm and C-repeat region molecular typing of beta-hemolytic *Streptococci* in a tropical country: implications for vaccine development. *Journal of Clinical Microbiology* 2009; **47**: 2502–2509.
  33. **Smeesters PR, et al.** Differences between Belgian and Brazilian group A *Streptococcus* epidemiologic landscape. *PLoS ONE* 2006; **1**: e10.
  34. **Tewodros W, Kronvall G.** M protein gene (emm type) analysis of group A beta-hemolytic *streptococci* from Ethiopia reveals unique patterns. *Journal of Clinical Microbiology* 2005; **43**: 4369–4376.
  35. **Bessen DE, et al.** Molecular epidemiology and genomics of group A *Streptococcus*. *Infection and Genetic Evolution* 2015; **33**: 393–418.
  36. **Steer AC, et al.** Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. *Lancet Infectious Diseases* 2009; **9**: 611–616.
  37. **Tong SY, et al.** Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. *Genome Research* 2015; **25**: 111–118.
  38. **Harris SR, et al.** Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infectious Diseases* 2013; **13**: 130–136.