Author/s:
Bowen, AC;Harris, T;Holt, DC;Giffard, PM;Carapetis, JR;Campbell, PT;McVernon, J;Tong, SYC

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

Date:
2016-07-01

Citation:
Bowen, A. C., Harris, T., Holt, D. C., Giffard, P. M., Carapetis, J. R., Campbell, P. T., McVernon, J. & Tong, S. Y. C. (2016). Whole genome sequencing reveals extensive community-level transmission of group A Streptococcus in remote communities. Epidemiology and Infection, 144 (9), pp.1991-1998. https://doi.org/10.1017/S095026881500326X.

Title:

Whole genome sequencing reveals extensive household-to-household transmission of Group A *Streptococcus* in remote communities

Short Title:

Transmission of Group A *Streptococcus*

Authors:

A.C. Bowen[1,2,3], T. Harris[1], D.C. Holt[1], P.M. Giffard[1], J.R. Carapetis[2,3], P.T. Campbell[4], J. McVernon[4], S.Y.C. Tong[1,5]

[1]Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, AUSTRALIA

[2]Princess Margaret Hospital for Children, Perth, Western Australia, AUSTRALIA

[3]Telethon Kids Institute, University of Western Australia, Perth, Western Australia, AUSTRALIA

[4]University of Melbourne, Melbourne, Victoria, AUSTRALIA

[5]Royal Darwin Hospital, Darwin, Northern Territory, AUSTRALIA

Corresponding Author and request for reprints

A/Prof Steven Tong

Menzies School of Health Research

PO Box 41096 Casuarina NT AUSTRALIA

Steven.tong@menzies.edu.au

Summary:

Impetigo is common in remote Indigenous children of Northern Australia, with the primary driver in this context being *Streptococcus pyogenes*. To reduce the high burden of impetigo, the transmission dynamics of *S. pyogenes* must be more clearly elucidated. We performed whole genome sequencing (WGS) on 31 *S. pyogenes* isolates collected from children from households with >1 *S. pyogenes* infected child from the same community. We aimed to determine whether transmission was more or less likely at the household level. Children from households with more than 1 child suffering *S. pyogenes* impetigo were more likely to have severe impetigo than children with no affected siblings. The 31 isolates consisted of nine STs with evidence of both within and between household transmissions. This is the first time WGS has been used to define the household transmission of *S. pyogenes* in a remote Indigenous community. Given the evidence both of within and between household transmission, strategies to reduce the burden of impetigo in this setting will need to be multi-faceted and community-wide, with targeting including and beyond household overcrowding.

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 [2,3], with the most severely affected group being Indigenous Australian children living in remote communities [4,5], 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 [6] with consequent chronic kidney disease [7,8] and probably contributes to the highest reported rates of rheumatic heart disease in the world found in these communities [9,10]. In a large randomized controlled trial, *Streptococcus pyogenes* was confirmed to be the primary driver of impetigo in an endemic context [11]. Therefore a clear understanding of the transmission dynamics of *S. pyogenes* 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. Households are crowded in remote Indigenous communities with a median of 3–7 persons per bedroom [11,12] and a correlation exists between the level of crowding and prevalence of impetigo [12]. However, interventions that improve housing quality have had only limited success in reducing the burdens of infectious disease [13]. Part of the reason for that 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 [14].

To better understand the relative contributions of household level and community level transmission, we obtained whole genome sequences (WGS) of 31 *S. pyogenes* isolates from household clusters of impetigo in a single community over a three-day

period. By assessing the relatedness of *S. pyogenes* strains associated with skin infections in multiple members of individual households, we sought to assess whether the household was likely to be the most useful target for interventions to reduce acquisition and subsequent burden of impetigo.

Methods

Isolates were collected from children with impetigo who were participants in a randomized controlled trial of oral trimethoprim-sulphamethoxazole versus intramuscular benzathine benzylpenicillin G for the treatment of skin sores [11]. The trial recruited children aged three months to 13 years from seven remote communities in the Northern Territory, Australia. 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 were referred directly from the clinic (10%) [11]. No attempt was made to screen children who did not attend school or were below school-age, outside of the recruitment described above. All children with impetigo in the trial had at least one microbiological swab collected [15]. To ascertain household crowding and number of other children with impetigo in the household, a qualitative survey of the primary caregiver was conducted. No swabs or clinical assessments of these household members were made.

For this WGS sub-study, we concentrated on a single community at a single recruitment visit conducted between November 8 and 11, 2011. This community was

chosen because it had the highest number of *S. pyogenes* isolates available from different sized household clusters. We included for sequencing isolates recovered from children residing in households where ≥2 children from the same household had culture confirmed *S. pyogenes* impetigo. 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 [16]. Where *S. pyogenes* was recovered, a single isolate from the agar plate was stored at -80ºC. 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 an M1 GAS reference sequence (accession number AE004092.2) using SPANDx [17] to define orthologous single nucleotide polymorphisms (SNPs) and a maximum likelihood tree was built using RaXML [18] with default settings and visualized with the Interactive Tree of Life [19]; and the multilocus sequence type (MLST) for each isolate was determined using SRST2 [20]. Short read data for each isolate was assembled using Bowtie2 [21] with reference to an M1 GAS reference sequence (accession number AE004092.2). The best assembly for each 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

[17]. Reads from the assembled ‘reference’ strain were also mapped back to itself as a quality control procedure. SNPs were manually visualized in Artemis [22]. Epidemiological and genome sequence data were visualized using Circos [23].

Results

Of 69 children who were screened at the school for participation in the trial during the November 2011 recruitment trip, 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 of these 45 (31%) children were the only member of their household with impetigo recruited in the trial and isolates recovered from these participants were not included for whole genome sequencing. 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 9 (64%) of the remaining 14 children. See Figure 1 for the study profile.

Of the 31 isolates, there were nine STs that were clearly delineated following alignment against the reference GAS M1 strain (Figure 2). De novo assemblies of each isolate were obtained. The best assembly within each ST was chosen to be the ‘reference’ assembly for that ST and these assemblies had a median size of 1,801,299 bp (range 1,745,425–1,878,702). The median depth of read coverage aligning against the ‘reference assemblies’ was 133x. Mapping of short reads from the ‘reference’ isolate back to the ‘reference’ assembly (i.e., itself) demonstrated no SNPs. Within each ST, the number of SNPs for any one isolate compared to the

149 'reference' assembly ranged from 0 to 3 (Table 1). Thus, isolates within each ST
150 were essentially identical when considering orthologous SNPs at a whole genome
151 level.
152
153 There was evidence of both within and between household transmissions (Figure 3).
154 Five of the 11 households had a single circulating ST, consistent with transmission
155 within the household. There were only two instances where a household was found
156 to have an ST that was not found in any other household (i.e., the household was
157 uniquely identified with an ST and vice versa – household I with ST182, and
158 household K with ST641). There was also abundant evidence of transmission of STs
159 between households. The six ST10 isolates were identical (i.e., 0 SNPs were
160 identified) and spread across three different households. The two ST176 isolates
161 were identical and from different households. The three ST330 isolates were
162 identical and from two different households. Of the four ST304 isolates, two variants
163 (differing by 3 SNPs from each other) were identified. Three of the ST304 isolates
164 were from a single household, where both variants were present. Within the eight
165 ST332 isolates, four SNP positions resulted in five variants, with a maximum of two
166 SNPs between any two variants. No household had more than one isolate of any
167 ST332 variant.
168

169 ## Discussion

170 Our initial exploration of the relatedness of GAS isolates between concurrently
171 infected family members reveals marked diversity of strains within the identified
172 household units. This finding suggests the independent introduction of multiple
173 strains acquired from outside of the household, rather than dominance of a single

strain resulting from close contact transmission within the home. It is noteworthy that the majority of index study participants were recruited from the community school, a setting likely to be important for cross-household transmission.

Children from households containing more than one affected child were more likely to have severe impetigo than those with only one child with impetigo. In addition, these households had a median of three children with impetigo at a time. Despite this household burden, the use of WGS did not demonstrate the expected household relatedness of isolates. Further studies to disentangle the transmission dynamics are needed. One possibility is using global positioning system (GPS) technology to track children's movements over a short time period to observe the social dynamics which may drive *S. pyogenes* transmission [24].

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 (between household) transmission of *S. pyogenes* is a key component of the dynamic epidemiology of *S. pyogenes* in an endemic setting. In addition, an inference of multiple importations 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 [25,26], however all colonies were of the same sequence type. Such studies have not yet been reported for *S. pyogenes*. 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 a household level (e.g., improved housing conditions, reduced household crowding) may not be effective at reducing the prevalence of impetigo if only a minority of residents in a community benefit from the intervention and/or there is an absence of other community-wide intervention strategies. Therefore, it is not surprising that small scale interventions targeting household crowding alone did not result in an appreciable reduction in infectious diseases [13].

Conclusions

We report the first use of WGS to define the household transmission of *S. pyogenes* in a remote Indigenous community. It is likely that strategies aimed at reducing the burden of impetigo in remote Indigenous Australian children will need to include community-wide interventions.

Acknowledgements

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

Financial Support

An Australian National Health and Medical Research Council (NHMRC) project grant (545346) and a Menzies School of Health Research Small Grant supported this work. AB is an NHMRC Early Career Fellow (1088735) and ST and JMcV are NHMRC Career Development Fellows (1065736).

Declaration of Interest

None.

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.

References

1. Koning S, van der Sande R, Verhagen AP, van Suijlekom-Smit LW, Morris AD, et al. (2012) Interventions for impetigo. Cochrane Database Syst Rev 1: CD003261.
2. Bowen AC, Mahe A, Hay RJ, Andrews RM, Steer AC, et al. (2015) The Global Epidemiology of Impetigo: A Systematic Review of the Population Prevalence of Impetigo and Pyoderma. PLoS One 10: e0136789.
3. Carapetis JS, A. Mulholland, K. (2005) The current evidence for the burden of group A streptococcal diseases. Discussion papers on Child Health. Geneva: WHO. pp. 52.
4. Carapetis JR, Steer AC, Mulholland EK, Weber M (2005) The global burden of group A streptococcal diseases. Lancet Infect Dis 5: 685-694.
5. Currie BJ, Carapetis JR (2000) Skin infections and infestations in Aboriginal communities in northern Australia. Australas J Dermatol 41: 139-143; quiz 144-135.
6. Marshall CS, Cheng AC, Markey PG, Towers RJ, Richardson LJ, et al. (2011) Acute post-streptococcal glomerulonephritis in the Northern Territory of Australia: a review of 16 years data and comparison with the literature. Am J Trop Med Hyg 85: 703-710.

7. White AV, Hoy WE, McCredie DA (2001) Childhood post-streptococcal glomerulonephritis as a risk factor for chronic renal disease in later life. Med J Aust 174: 492-496.
8. Hoy WE, White AV, Dowling A, Sharma SK, Bloomfield H, et al. (2012) Post-streptococcal glomerulonephritis is a strong risk factor for chronic kidney disease in later life. Kidney Int 81: 1026-1032.
9. McDonald M, Currie BJ, Carapetis JR (2004) Acute rheumatic fever: a chink in the chain that links the heart to the throat? Lancet Infect Dis 4: 240-245.
10. Parnaby MG, Carapetis JR (2010) Rheumatic fever in indigenous Australian children. J Paediatr Child Health 46: 527-533.
11. Bowen AC, Tong SY, Andrews RM, O'Meara IM, McDonald MI, et al. (2014) 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 384: 2132-2140.
12. McDonald MI, Towers RJ, Andrews RM, Benger N, Currie BJ, et al. (2006) Low rates of streptococcal pharyngitis and high rates of pyoderma in Australian aboriginal communities where acute rheumatic fever is hyperendemic. Clin Infect Dis 43: 683-689.
13. Bailie RS, Stevens M, McDonald EL (2012) The impact of housing improvement and socio-environmental factors on common childhood illnesses: a cohort study in Indigenous Australian communities. J Epidemiol Community Health 66: 821-831.
14. McDonald EL, Bailie RS, Morris PS (2014) Participatory systems approach to health improvement in Australian Aboriginal children. Health Promot Int.
15. Bowen AC, Tong SY, Chatfield MD, Carapetis JR (2014) The microbiology of impetigo in Indigenous children: associations between Streptococcus pyogenes, Staphylococcus aureus, scabies, and nasal carriage. BMC Infect Dis 14: 727.
16. Bowen AC, Tong SY, Chatfield MD, Andrews RM, Carapetis JR (2013) Comparison of three methods for the recovery of skin pathogens from impetigo swabs collected in a remote community of Northern Territory, Australia. Trans R Soc Trop Med Hyg 107: 384-389.
17. Sarovich DS, Price EP (2014) SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. BMC Res Notes 7: 618.
18. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.
19. Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39: W475-478.
20. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, et al. (2014) SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med 6: 90.
21. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359.
22. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28: 464-469.
23. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19: 1639-1645.

303 24. Vazquez-Prokopec GM, Bisanzio D, Stoddard ST, Paz-Soldan V, Morrison AC,
304 et al. (2013) Using GPS technology to quantify human mobility, dynamic
305 contacts and infectious disease dynamics in a resource-poor urban
306 environment. PLoS One 8: e58802.
307 25. Tong SY, Holden MT, Nickerson EK, Cooper BS, Koser CU, et al. (2015)
308 Genome sequencing defines phylogeny and spread of methicillin-resistant
309 Staphylococcus aureus in a high transmission setting. Genome Res 25: 111-
310 118.
311 26. Harris SR, Cartwright EJ, Torok ME, Holden MT, Brown NM, et al. (2013)
312 Whole-genome sequencing for analysis of an outbreak of meticillin-resistant
313 Staphylococcus aureus: a descriptive study. Lancet Infect Dis 13: 130-136.
314

315 **Table 1:** Details of the 31 Group A *Streptococcus* (GAS) strains.

Isolate	Household	ST	*emm* 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	GGCA	SAMN03988126
1023-1	B	332	166.1	GTCG	SAMN03988121
2245-1	B	332	166.1	ATCG	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	GGAG	SAMN03988120
2241-1#	E	205	230		SAMN03988130
2240-1	E	332	166.1	GGCA	SAMN03988129
2229-1	E	330*	164.3	No SNPs	SAMN03988128
2102-1#	F	304	108	CAT	SAMN03988132
2099-1	F	332	166.1	GGAG	SAMN03988131
2103-1#	F	332	166.1	GGCG	SAMN03988133
2095-1#	G	176	58	No SNPs	SAMN03988138
2091-1	G	304	108	TCC	SAMN03988134
2093-1	G	304	108	TCC	SAMN03988136
2094-1	G	304	108	CAT	SAMN03988137
2092-1	G	332	166.1	GGCG	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

Note:

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. Where there is only one representative of an ST, the cell is left blank.

SRA (short read archive).

Figure 1: Flow diagram of participants in the study.

Figure 2: Maximum likelihood tree of 31 Group A *Streptococcus* (GAS) isolates aligned against a reference M1 isolate (accession number AE004092.2). The multilocus sequence types are indicated by the outer circle.

Figure 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.