

1 **Title:**

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

4

5 **Short Title:**

6 **Transmission of Group A *Streptococcus***

7

8 **Authors:**

9 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.  
10 Campbell<sup>4</sup>, J. McVernon<sup>4</sup>, S.Y.C. Tong<sup>1,5</sup>

11

12 <sup>1</sup>Menzies School of Health Research, Charles Darwin University, Darwin, Northern  
13 Territory, AUSTRALIA

14 <sup>2</sup>Princess Margaret Hospital for Children, Perth, Western Australia, AUSTRALIA

15 <sup>3</sup>Telethon Kids Institute, University of Western Australia, Perth, Western Australia,  
16 AUSTRALIA

17 <sup>4</sup>University of Melbourne, Melbourne, Victoria, AUSTRALIA

18 <sup>5</sup>Royal Darwin Hospital, Darwin, Northern Territory, AUSTRALIA

19

20 **Corresponding Author and request for reprints**

21 A/Prof Steven Tong

22 Menzies School of Health Research

23 PO Box 41096 Casuarina NT AUSTRALIA

24 Steven.tong@menzies.edu.au

25

26 **Summary:**

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

42

43

44 **Introduction**

45 Impetigo is a common childhood infection [1] with an estimated 160 million  
46 prevalent cases globally [2]. Children living in Oceania have the highest documented  
47 prevalence [2,3], with the most severely affected group being Indigenous Australian  
48 children living in remote communities [4,5], where the median childhood prevalence

49 is 43% (95% CI 40.2 – 45.7%)[2]. Impetigo is a non-benign disease that drives  
50 outbreaks of acute post-streptococcal glomerulonephritis [6] with consequent chronic  
51 kidney disease [7,8] and probably contributes to the highest reported rates of  
52 rheumatic heart disease in the world found in these communities [9,10]. In a large  
53 randomized controlled trial, *Streptococcus pyogenes* was confirmed to be the  
54 primary driver of impetigo in an endemic context [11]. Therefore a clear  
55 understanding of the transmission dynamics of *S. pyogenes* is required to inform the  
56 design of interventions to reduce the prevalence of impetigo.

57

58 It has long been thought that housing conditions are a major contributor to the high  
59 prevalence of infectious and parasitic disease in children in remote Indigenous  
60 communities. Households are crowded in remote Indigenous communities with a  
61 median of 3–7 persons per bedroom [11,12] and a correlation exists between the  
62 level of crowding and prevalence of impetigo [12]. However, interventions that  
63 improve housing quality have had only limited success in reducing the burdens of  
64 infectious disease [13]. Part of the reason for that may be extensive transmission of  
65 infectious agents outside the household. Community interactions within Indigenous  
66 communities are complex, with considerable mobility of children between  
67 households, and much unstructured mixing opportunities in school and other  
68 community settings that may be more or less influential to infection risk than the  
69 home environment [14].

70

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

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

78

## 79 **Methods**

80 Isolates were collected from children with impetigo who were participants in a  
81 randomized controlled trial of oral trimethoprim-sulphamethoxazole versus  
82 intramuscular benzathine benzylpenicillin G for the treatment of skin sores [11]. The  
83 trial recruited children aged three months to 13 years from seven remote  
84 communities in the Northern Territory, Australia. Screening for eligibility in the trial  
85 occurred predominantly in schools. Following this, research nurses visited  
86 households to discuss the study with caregivers. Overall 65% of recruited children  
87 with impetigo were identified through school screening. During these household  
88 visits, siblings or relatives were also screened for participation (an additional 25% of  
89 recruitment). The remaining participants were referred directly from the clinic (10%)  
90 [11]. No attempt was made to screen children who did not attend school or were  
91 below school-age, outside of the recruitment described above. All children with  
92 impetigo in the trial had at least one microbiological swab collected [15]. To  
93 ascertain household crowding and number of other children with impetigo in the  
94 household, a qualitative survey of the primary caregiver was conducted. No swabs or  
95 clinical assessments of these household members were made.

96

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

99 chosen because it had the highest number of *S. pyogenes* isolates available from  
100 different sized household clusters. We included for sequencing isolates recovered  
101 from children residing in households where  $\geq 2$  children from the same household  
102 had culture confirmed *S. pyogenes* impetigo. All isolates were collected prior to the  
103 commencement of antibiotic therapy.

104

105 Cotton swabs (Copan, Italy) were used to collect pus from skin sores for  
106 microbiological culture according to standard methods [16]. Where *S. pyogenes* was  
107 recovered, a single isolate from the agar plate was stored at  $-80^{\circ}\text{C}$ . DNA was  
108 extracted from the stored isolate using a QIAamp DNA Mini Kit (Qiagen, Germany)  
109 according to the manufacturer's instructions.

110

111 We obtained genome sequences with paired end libraries on the Illumina HiSeq  
112 (Illumina, USA) platform through Macrogen Inc (South Korea). The reads have been  
113 deposited in NCBI Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>)  
114 with accession numbers SAMN03988118–SAMN03988148. Reads were mapped  
115 against an M1 GAS reference sequence (accession number AE004092.2) using  
116 SPANDx [17] to define orthologous single nucleotide polymorphisms (SNPs) and a  
117 maximum likelihood tree was built using RaXML [18] with default settings and  
118 visualized with the Interactive Tree of Life [19]; and the multilocus sequence type  
119 (MLST) for each isolate was determined using SRST2 [20]. Short read data for each  
120 isolate was assembled using Bowtie2 [21] with reference to an M1 GAS reference  
121 sequence (accession number AE004092.2). The best assembly for each ST (based on  
122 the smallest number of contigs and largest N50) was then used as a reference to map  
123 reads from the other strains of that ST to identify orthologous SNPs using SPANDx

124 [17]. Reads from the assembled ‘reference’ strain were also mapped back to itself as  
125 a quality control procedure. SNPs were manually visualized in Artemis [22].  
126 Epidemiological and genome sequence data were visualized using Circos [23].

127

## 128 **Results**

129 Of 69 children who were screened at the school for participation in the trial during  
130 the November 2011 recruitment trip, 45 (65%) with crusted or purulent impetigo  
131 were enrolled in the study. These 45 children had a median age of 7.4 years  
132 (interquartile range, IQR 4.4 – 9.7 years) and 27 (60%) were female. Fourteen of  
133 these 45 (31%) children were the only member of their household with impetigo  
134 recruited in the trial and isolates recovered from these participants were not included  
135 for whole genome sequencing. The remaining 31 children (69%) resided in 11  
136 households, with a median of 3 (IQR 2–3) infected individuals in each household.  
137 Twenty-five (81%) of the 31 children in households where multiple infected children  
138 were identified had severe impetigo, compared to 9 (64%) of the remaining 14  
139 children. See Figure 1 for the study profile.

140

141 Of the 31 isolates, there were nine STs that were clearly delineated following  
142 alignment against the reference GAS M1 strain (Figure 2). De novo assemblies of  
143 each isolate were obtained. The best assembly within each ST was chosen to be the  
144 ‘reference’ assembly for that ST and these assemblies had a median size of  
145 1,801,299 bp (range 1,745,425–1,878,702). The median depth of read coverage  
146 aligning against the ‘reference assemblies’ was 133x. Mapping of short reads from  
147 the ‘reference’ isolate back to the ‘reference’ assembly (i.e., itself) demonstrated no  
148 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

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

177

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

186

187 Limitations of this study include the use of a single time point and a single  
188 community to assess for household clustering. However, even with this small sample  
189 size, it is clear that community level (between household) transmission of *S.*  
190 *pyogenes* is a key component of the dynamic epidemiology of *S. pyogenes* in an  
191 endemic setting. In addition, an inference of multiple importations into a household  
192 may be incorrect if individuals harboured multiple co-infecting strains, making  
193 attribution of a source other than the household spurious on the basis of a single  
194 isolate. Studies involving WGS of multiple *S. aureus* colonies from a single swab  
195 have demonstrated carriage of a cloud of variation [25,26], however all colonies  
196 were of the same sequence type. Such studies have not yet been reported for *S.*  
197 *pyogenes*. Finally, our sampling in households was not complete and included only  
198 symptomatic children recruited into the study. Swabbing of all household members



199 for either active infection or asymptomatic colonization over time would be required  
200 to understand underlying patterns of transmission in the household that may be  
201 influential, other than just observed disease.

202

203 Nonetheless, our findings suggest that interventions purely targeted at a household  
204 level (e.g., improved housing conditions, reduced household crowding) may not be  
205 effective at reducing the prevalence of impetigo if only a minority of residents in a  
206 community benefit from the intervention and/or there is an absence of other  
207 community-wide intervention strategies. Therefore, it is not surprising that small  
208 scale interventions targeting household crowding alone did not result in an  
209 appreciable reduction in infectious diseases [13].

210

## 211 **Conclusions**

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

216

## 217 **Acknowledgements**

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

220

221 **Financial Support**

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

226

227 **Declaration of Interest**

228 None.

229

230 **Ethical Standards**

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

234

235 **References**

- 236 1. Koning S, van der Sande R, Verhagen AP, van Suijlekom-Smit LW, Morris AD,  
237 et al. (2012) Interventions for impetigo. *Cochrane Database Syst Rev* 1:  
238 CD003261.
- 239 2. Bowen AC, Mahe A, Hay RJ, Andrews RM, Steer AC, et al. (2015) The Global  
240 Epidemiology of Impetigo: A Systematic Review of the Population  
241 Prevalence of Impetigo and Pyoderma. *PLoS One* 10: e0136789.
- 242 3. Carapetis JS, A. Mulholland, K. (2005) The current evidence for the burden of  
243 group A streptococcal diseases. *Discussion papers on Child Health*. Geneva:  
244 WHO. pp. 52.
- 245 4. Carapetis JR, Steer AC, Mulholland EK, Weber M (2005) The global burden of  
246 group A streptococcal diseases. *Lancet Infect Dis* 5: 685-694.
- 247 5. Currie BJ, Carapetis JR (2000) Skin infections and infestations in Aboriginal  
248 communities in northern Australia. *Australas J Dermatol* 41: 139-143; quiz  
249 144-135.
- 250 6. Marshall CS, Cheng AC, Markey PG, Towers RJ, Richardson LJ, et al. (2011)  
251 Acute post-streptococcal glomerulonephritis in the Northern Territory of  
252 Australia: a review of 16 years data and comparison with the literature. *Am J*  
253 *Trop Med Hyg* 85: 703-710.

- 254 7. White AV, Hoy WE, McCredie DA (2001) Childhood post-streptococcal  
255 glomerulonephritis as a risk factor for chronic renal disease in later life. *Med*  
256 *J Aust* 174: 492-496.
- 257 8. Hoy WE, White AV, Dowling A, Sharma SK, Bloomfield H, et al. (2012) Post-  
258 streptococcal glomerulonephritis is a strong risk factor for chronic kidney  
259 disease in later life. *Kidney Int* 81: 1026-1032.
- 260 9. McDonald M, Currie BJ, Carapetis JR (2004) Acute rheumatic fever: a chink in  
261 the chain that links the heart to the throat? *Lancet Infect Dis* 4: 240-245.
- 262 10. Parnaby MG, Carapetis JR (2010) Rheumatic fever in indigenous Australian  
263 children. *J Paediatr Child Health* 46: 527-533.
- 264 11. Bowen AC, Tong SY, Andrews RM, O'Meara IM, McDonald MI, et al. (2014)  
265 Short-course oral co-trimoxazole versus intramuscular benzathine  
266 benzylpenicillin for impetigo in a highly endemic region: an open-label,  
267 randomised, controlled, non-inferiority trial. *Lancet* 384: 2132-2140.
- 268 12. McDonald MI, Towers RJ, Andrews RM, Bengler N, Currie BJ, et al. (2006) Low  
269 rates of streptococcal pharyngitis and high rates of pyoderma in Australian  
270 aboriginal communities where acute rheumatic fever is hyperendemic. *Clin*  
271 *Infect Dis* 43: 683-689.
- 272 13. Bailie RS, Stevens M, McDonald EL (2012) The impact of housing improvement  
273 and socio-environmental factors on common childhood illnesses: a cohort  
274 study in Indigenous Australian communities. *J Epidemiol Community Health*  
275 66: 821-831.
- 276 14. McDonald EL, Bailie RS, Morris PS (2014) Participatory systems approach to  
277 health improvement in Australian Aboriginal children. *Health Promot Int.*
- 278 15. Bowen AC, Tong SY, Chatfield MD, Carapetis JR (2014) The microbiology of  
279 impetigo in Indigenous children: associations between *Streptococcus*  
280 *pyogenes*, *Staphylococcus aureus*, scabies, and nasal carriage. *BMC Infect*  
281 *Dis* 14: 727.
- 282 16. Bowen AC, Tong SY, Chatfield MD, Andrews RM, Carapetis JR (2013)  
283 Comparison of three methods for the recovery of skin pathogens from  
284 impetigo swabs collected in a remote community of Northern Territory,  
285 Australia. *Trans R Soc Trop Med Hyg* 107: 384-389.
- 286 17. Sarovich DS, Price EP (2014) SPANDx: a genomics pipeline for comparative  
287 analysis of large haploid whole genome re-sequencing datasets. *BMC Res*  
288 *Notes* 7: 618.
- 289 18. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and  
290 post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.
- 291 19. Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and  
292 display of phylogenetic trees made easy. *Nucleic Acids Res* 39: W475-478.
- 293 20. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, et al. (2014) SRST2:  
294 Rapid genomic surveillance for public health and hospital microbiology labs.  
295 *Genome Med* 6: 90.
- 296 21. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat*  
297 *Methods* 9: 357-359.
- 298 22. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA (2012) Artemis: an  
299 integrated platform for visualization and analysis of high-throughput  
300 sequence-based experimental data. *Bioinformatics* 28: 464-469.
- 301 23. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: an  
302 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 methicillin-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.

<b>Isolate</b>	<b>Household</b>	<b>ST</b>	<b><i>emm</i> Type</b>	<b>SNP variant</b>	<b>SRA number</b>
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 <sup>#</sup>	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 <sup>#</sup>	E	205	230		SAMN03988130
2240-1	E	332	166.1	GGCA	SAMN03988129
2229-1	E	330*	164.3	No SNPs	SAMN03988128
2102-1 <sup>#</sup>	F	304	108	CAT	SAMN03988132
2099-1	F	332	166.1	GGAG	SAMN03988131
2103-1 <sup>#</sup>	F	332	166.1	GGCG	SAMN03988133
2095-1 <sup>#</sup>	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 <sup>#</sup>	H	10	70	No SNPs	SAMN03988140

2098-1 <sup>#</sup>	I	182	205	G	SAMN03988141
2242-1	I	182	205	A	SAMN03988142
1019-1	J	330*	164.3	No SNPs	SAMN03988143
2088-1 <sup>#</sup>	J	330*	164.3	No SNPs	SAMN03988144
2233-1 <sup>#</sup>	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

316

317 **Note:**

318 # isolate used as the 'reference' assembly for that sequence type (ST).

319 \* single locus variant of ST330

320 The SNP (single nucleotide variant) column refers to the SNP variants within each of  
321 the STs. Where there is only one representative of an ST, the cell is left blank.

322 SRA (short read archive).

323

324 **Figure 1:** Flow diagram of participants in the study.

325

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

329

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