Longitudinal analysis of group A *Streptococcus emm* types and *emm* clusters in a high prevalence setting reveals past infection does not prevent future infection

40 word summary: We examined longitudinal patterns of *emm* types in Group A *Streptococcus* samples collected from Fijian schoolchildren. In a setting where impetigo is the dominant mode of transmission, we found no evidence that infection history modifies future infection.

Patricia Therese Campbell[1,2], Steven Y. C. Tong[3,4], Nicholas Geard[1,5], Mark R. Davies[6], Kate A. Worthing[6], Jake A. Lacey[7], Pierre R. Smeesters[2,8,9,10], Michael R. Batzloff[11], Joseph Kado[12,13,14,15], Adam W.J. Jenney[10,13], Jodie McVernon[1,2,16], Andrew C. Steer[2,10,17]

[1]Victorian Infectious Diseases Reference Laboratory Epidemiology Unit, The Peter Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne, Melbourne, Australia, 3000

[2]Murdoch Children's Research Institute, The Royal Children's Hospital, Melbourne, Australia, 3000

[3]Victorian Infectious Disease Service, The Royal Melbourne Hospital, and Doherty Department University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia, 3000

[4]Menzies School of Health Research, Charles Darwin University, Darwin, Australia, 0810

[5]School of Computing and Information Systems, Melbourne School of Engineering, The University of Melbourne, Australia, 3000

[6]Department of Microbiology and Immunology, The University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia, 3000

[7]Doherty Department at the Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia, 3000

[8]Molecular Bacteriology Laboratory, Université Libre de Bruxelles, Brussels, Belgium

[9]Department of Pediatrics, Academic Children Hospital Queen Fabiola, Université Libre de Bruxelles, Brussels, Belgium

[10]Centre for International Child Health, University of Melbourne, Melbourne, Australia, 3000

[11]Institute for Glycomics, Gold Coast campus, Griffith University, QLD 4222, Australia

[12]Department of Paediatrics, Colonial War Memorial Hospital, Suva, Fiji

[13]College of Medicine, Nursing and Health Sciences, Fiji National University, Suva, Fiji

[14]Fiji Rheumatic Heart Disease Control Program, Suva, Fiji

[15]Telethon Kids Institute, University of Western Australia, Perth, Australia, 6009

[16]Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Australia, 3053

[17]Department of Paediatrics, University of Melbourne, Royal Children's Hospital Melbourne, Parkville, Australia, 3000

FOOTNOTE PAGE

1) The authors declare that they do not have a commercial or other association that might pose a conflict of interest.

2) This study was funded by the National Institute of Allergy and Infectious Diseases (grant U01AI60579). SYCT is supported by an Australian National Health and Medical Research Council (NHMRC) Career Development Fellowship (GNT1145033). JM is supported by NHMRC Principal Research Fellowship (GNT1117140). The work was supported by NHMRC project grants GNT1098319 and GNT1130455.

3) Author for correspondence: Dr Patricia T. Campbell, The Peter Doherty Institute for Infection and Immunity, 792 Elizabeth St, Melbourne, Australia, 3000, email: patricia.campbell@unimelb.edu.au, telephone +61 3 9035 6050

ABSTRACT

Group A *Streptococcus* (GAS) is a pathogen of global importance, but despite the ubiquity of GAS infections, there remains incomplete understanding of the relationship between infection, colonization and immunity. The M protein, encoded by the *emm* gene, is a major virulence factor and vaccine candidate and forms the basis of a number of classification systems. Longitudinal patterns of *emm* types collected from 457 Fijian schoolchildren over a ten-month period were analyzed. No evidence of tissue tropism was observed, and there was no apparent selective pressure or constraint of *emm* types. Patterns of *emm* type acquisition suggest limited, if any, modification of future infection based on infection history. Where impetigo is the dominant mode of transmission, circulating *emm* types either may not be constrained by ecological niches or population immunity to the M protein, or may require several infections over a longer period of time to induce such immunity.

Keywords:

Streptococcus pyogenes; emm-cluster; immunity; skin infection

Background

Group A *Streptococcus* (GAS) is a pathogen of global importance, responsible for more than 700 million superficial infections and at least 500,000 deaths per year [1, 2]. Almost all diseases caused by GAS are most common in developing regions, from superficial conditions such as pyoderma (including impetigo) and pharyngitis, to severe sequelae including invasive disease, rheumatic heart disease (RHD) and post-streptococcal glomerulonephritis [1]. Despite the ubiquity of GAS, there remains incomplete understanding of its immunobiology and the relationship between infection, colonization and immunity, hindering efforts aimed at sustainable control including the development of safe and effective vaccines.

The GAS M protein is a major virulence factor that elicits antibody production and enables the bacteria to inhibit phagocytosis in the absence of antibodies, making it a prime vaccine candidate [3]. This M protein forms the basis of a number of classification systems for GAS: *emm* typing, *emm* patterns and *emm* clusters, where *emm* refers to the *emm* gene encoding this surface M protein [4]. Over 240 different *emm* types have been identified based on the variable N-terminus part of the protein, contributing to the complexity of epidemiologic studies [5]. Based on the structure of *emm* and *emm*-like genes in the GAS genome, *emm* types may be further grouped into *emm* patterns, referred to as A-C, D and E [6].

Recently, *emm* types have been grouped into 48 *emm* clusters based on closely related sequences, shared structural characteristics and similar binding capacities [7, 8]. It has been hypothesized that cross-protective immunity may occur between *emm* types that exist within the same *emm* cluster. Preliminary and limited laboratory studies have shown that *in vitro* cross-protection does occur within certain *emm* clusters in Fijian children [8]. The existence of cross-protective immunity within *emm* clusters could substantially aid vaccine development against this multi-strain pathogen. However, it remains to be seen whether these findings translate to population level protection.

The relationship between *emm* types and disease burden is important when selecting priority strains for prevention. It has long been believed that different GAS strains preferentially cause either impetigo or pharyngitis [9]. Based on a number of population-based surveys, GAS strains with the *emm* pattern A-C display a tropism for the throat, while D have a tropism for the skin and E are found in both tissue sites [6, 10]. However, the mechanism responsible for different disease manifestations remains to be identified [11]. Distribution of *emm* types and clinical manifestation differs between settings, with fewer *emm* types circulating in low prevalence settings (typically dominated by pharyngitis) compared to high prevalence settings (typically dominated by impetigo), and some *emm* types common in developed countries rarely found in in developing countries [12-14].

In this study, we investigated longitudinal data on *emm* types, patterns and clusters at individual, school and regional levels to evaluate *emm* immunobiology within a streptococcal disease endemic setting. We examined *emm* types isolated from children over time for evidence of immune protection at either the *emm* type or *emm* cluster level. First, we hypothesized that if immunity to prevalent *emm* types did develop, we would observe the subsequent disappearance of these *emm* types at individual and population levels as opportunities for further transmission diminished, followed by the appearance of new *emm* types to which the population had no prior immunity. Second, we also hypothesized that if immunity to *emm* types did develop, we would observe different *emm* types circulating in children infected at a single timepoint compared to those infected at multiple timepoints, as immunity would limit the acquisition of *emm* types that could cause a subsequent infection. Third, we framed similar hypotheses in terms of *emm* clusters.

Methods

Ethical approval

This study was approved by the Fiji National Research Ethics Review Committee, the Fiji National Health Research Committee, and the University of Melbourne Human Research Ethics Committee.

Written informed consent was required from participants or a parent or guardian prior to collection of information.

Setting and participants

This was a prospective longitudinal cohort study conducted in three schools in the Central Division of Fiji from February 2006 to November 2006. Two of the schools were rural, with all children of iTaukei (Indigenous Fijian) ethnicity, while the third school was located in Suva, with most children of Indo-Fijian ethnicity. Enrolment rates for the three schools were: 96.4% (rural school 1); 80.6% (rural school 2); and 53.2% (urban school) [15]. For skin screening, each school was visited every two months over a ten-month period, a total of six visits per school. Children aged 5–15 years were screened for skin sores and a swab taken from crusted or purulent sores [15]. At the first skin screening visit, children without sore throat symptoms had their throats swabbed for evidence of asymptomatic colonization [16]. Over the ten-month period, each school was visited twice per week and throat swabs collected from children reporting sore throat symptoms within the preceding seven days [16].

Laboratory methods

Skin and throat swabs were collected, transported and tested for the presence of beta-hemolytic colonies using standard methods [15, 16]. *emm*-typing of the dominant beta-hemolytic colony was undertaken according to US Centers for Disease Control and Prevention (CDC) standard methods (https://www.cdc.gov/streplab/groupa-strep/emm-background.html) [15, 16].

Data preparation and analysis

Characterization and analysis of emm types

Data were restricted to Lancefield Group A isolates. *emm* types were assigned to *emm* clusters as per Sanderson-Smith et al [7] and analyzed by time period, school, disease type and participant. Each skin screening visit took place over a period of approximately two weeks, followed by a window of around six weeks before the next round of skin screening visits commenced. The study period was

therefore divided into six 'screening time periods': five of approximately two months' duration, each running from the date of the first skin screening associated with that visit to the day before the first skin screening associated with the next visit; and a sixth of approximately two weeks' duration, covering the final skin screening visit. Throat isolates were assigned to the screening time period within which they were collected.

Diversity and prevalence

At the child level, the number of distinct *emm* types and *emm* clusters was tabulated against the number of positive swabs. Simpson's index of diversity, the probability that two randomly selected *emm* types are different, was calculated for each screening time period by school [17].

The total number of isolates of each *emm* cluster and each *emm* type were calculated. Prevalence of each *emm* type per 1000 children was calculated at each time point by dividing the number of isolates by the number of children participating in that visit, stratified to the same level. *emm* clusters were ordered by the highest number of samples overall, and within each *emm* cluster, *emm* types were similarly ordered.

Individual level exposure responses

Children with more than one isolate of the same *emm* type during the study were identified and their screening, pharyngitis and *emm* type history reported, noting that the absence of an isolate for a screening visit may represent either no sores, sores present but no swab taken (i.e. because not crusted, or purulent, as specified in the study protocol), or a swab was taken but did not grow GAS. For these children, we also report the *emm* clusters corresponding to their *emm* types in the Supplementary Material.

The ordering of *emm* types and *emm* clusters from highest to lowest number of isolates was compared between two groups of children: those with a single positive isolate and those with multiple positive isolates.

Data were prepared in Stata (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.) and MATLAB 2017b (The Mathworks). Analyses were conducted in R version 3.4.4.

Results

Demographics

A total of 457 children were enrolled in the study, with a minimum of 400 children seen at any of the six skin screening visits, and 80% of children seen at all of them. The number of children screened per visit per school ranged from 73–80 (rural school 1), 160–175 (rural school 2) and 161–202 (urban school). All 255 children enrolled from rural school 1 and rural school 2 identified as iTaukei (Indigenous Fijian), with the population at the urban school consisting of 99 (49%) Indo-Fijian, 67 (33%) iTaukei and 36 (18%) other ethnicities. Sex was evenly distributed, with 229 of the 457 children female. Median age was 9.9 years (interquartile range (IQR) 7.9–12.0 years).

Clinical data

There were 451 GAS-positive isolates collected from 245 children during the study. Of these, 379 (84%) were from impetigo samples, 45 (10%) from pharyngitis samples and 27 (6%) from asymptomatic throat colonization samples, which were collected only at the first visit. Median prevalence of GAS infection per 1000 children per screening time period (excluding colonization) was highest in rural school 2 (227.5, IQR 199.2–247.4), followed by rural school 1 (167.6, IQR 135.1–237.5) and the urban school, with the lowest observed prevalence (97.0, IQR 78.4–135.5).

Of the 245 children with GAS-positive isolates, most (195, 80%) returned either one or two GAS-positive swabs (Figure 1A). The highest number of GAS-positive swabs (impetigo, pharyngitis and asymptomatic throat colonization combined) for any child was six, returned by two children (< 1%).

Diversity and prevalence

The five most frequently observed *emm* types were *emm*70, *emm*33, *emm*25, *emm*93.3, and *emm*11, accounting for around 30% of positive isolates. This ranking held both for children infected only once and for children infected more than once. The five most frequently observed clusters were D4, E3, E6, E4, and E2, for children infected only once and for children infected more than once, accounting for 77% of positive isolates in both groups.

Of the 128 children returning more than one GAS-positive swab, 100 had a different *emm* type and 83 had a different *emm* cluster isolated from each of their swabs. More children with two positive swabs had two different *emm* clusters (65/78 children, 83%) than a single cluster. Most children with three or more positive swabs had at least one repeated *emm* cluster (36/50 children, 72%) (Figure 1B).

A wide variety of *emm* types circulated during all six screening time periods (range of 19 to 37 *emm* types per screening time period). There was no evidence of prevalent *emm* types disappearing, followed by the appearance of new dominant *emm* types (Figure 2). Rather, the most prevalent *emm* types were consistently present in a given setting (e.g. *emm*70, *emm*33) and less prevalent *emm* types were only detected during one or two screening time periods. There was no evidence of competitive exclusion of *emm* types at the population level, with each *emm* cluster having several isolates circulating concurrently in the same setting. Urban and rural schools appeared to have different patterns of circulation, with *emm*70 highly prevalent in both rural schools throughout the entire study, but completely absent from the urban school. In contrast to the rural schools, there were no dominant *emm* types at the urban school, and no *emm* types consistently isolated across all screening time periods.

Despite differences in overall prevalence of GAS-positive swabs for each screening time period, high diversity was observed across both rural and urban settings (Figure 2), with Simpson's index of diversity ranging from 0.77 to 0.99 (Supplementary Figure 1). No clear association was observed between prevalence and diversity, and we did not perform a formal test of association given the

small number of data points and observed differences per setting. The prevalence of infection by cluster related directly to the number of *emm* types categorized within each cluster. For example the most prevalent cluster in our study was D4, which is the largest cluster with 32 *emm* types. The next most prevalent cluster was E3, which is the second largest cluster with 19 *emm* types (Supplementary Figure 2).

The vast majority of *emm* types (56/62 detected) were recovered from impetigo samples (Figure 3). Even for those *emm* types consistently present throughout the study period, isolation from pharyngitis was rare and generally not repeated over more than two consecutive screening time periods (with the exception of *emm*44, *emm*92 *emm*101 and *emm*238.3). Throat colonization isolates collected at the first skin screening visit did not greatly increase the *emm* type diversity, adding only two *emm* types (*emm*137, *emm*14.4). While cluster D4, typically associated with skin infection, dominated the impetigo isolates, it was also present in pharyngitis isolates during each time screening time period, along with cluster E3 (Supplementary Figure 3). Throat colonization isolates collected at the first skin screening visit were mostly from the clusters with the highest prevalence of skin isolates at that time, reflected in *emm* pattern D comprising 63% of the colonization isolates (Supplementary Table 1). Conversely, *emm* pattern A-C, typically associated with throat infection, was isolated from around 16% of pharyngitis isolates.

Individual level exposure responses

For the 28 children with repeated isolation of the same *emm* type, a variety of longitudinal patterns were observed (Figure 4). Four children had the same *emm* type isolated from pharyngitis and impetigo samples at different times (children 9, 11, 13 and 16), and three children had the same *emm* type isolated from throat colonization and impetigo samples on the same day (children 22, 24 and 27). For skin infections, re-acquisition of the same *emm* type following a documented skin screening without GAS infection was observed in six children (children 1, 10, 17, 18, 19 and 28). Two different *emm* types were isolated from impetigo samples taken on the same day in four children

(children 2, 4, 7 and 19). For these same 28 children (Supplementary Figures 4 and 5), we observed different *emm* types from the same cluster isolated at the same time (child 4), and at different times (children 1, 2, 4, 7, 17 and 26). In addition, three of the 28 children had *emm* types from different clusters isolated concurrently (children 2, 7 and 19).

Discussion

Our analysis of longitudinal GAS *emm* types, patterns and clusters in Fijian schoolchildren provides insights into the links between prevalence, diversity and tissue tropism. We found no evidence of tissue tropism in this setting, with the *emm* types isolated from pharyngitis and throat colonization samples reflecting those isolated from impetigo samples, the dominant mode of infection in this cohort. There was no evidence of displacement of one *emm* type by another over time, with common *emm* types generally present throughout the study. There was no apparent selective pressure or constraint of *emm* types within clusters, with multiple *emm* types from each cluster circulating concurrently. Despite differences in prevalence by site and by time, high levels of diversity were observed in both urban and rural settings. The patterns of acquisition in the small number of children having the same *emm* type isolated more than once suggest limited, if any, modification of future infection based on infection history in these children. It may be that the small number of children experiencing repeated infection with the same *emm* type is indicative that most children developed immunity to the *emm* type that caused their infection. We cannot be certain whether the limited number of repeat infections we observed is due to immune protection or a lack of re-exposure to the same *emm* type. Our observation that the same *emm* types and clusters circulated in children infected at single and multiple time points raises questions about the immunity induced by the M protein in this tropical setting and its capacity to limit the collection of *emm* types or clusters that could cause subsequent infections. It is important to note that our conclusions regarding individual and population immunity to GAS are based on infection patterns rather than measurement of M type-specific or other antibodies to GAS. However, measurement and

understanding of the GAS immune response is not straightforward, with inconsistent development of antibodies following acquisition and a lack of evidence for a protective effect [18-20].

A particular strength of our study is the novel longitudinal analysis of *emm* types and clusters, with limited loss to follow up. More than 80% of children in our cohort were seen at all six skin screening visits, with swabs taken from all crusted and purulent skin sores, providing rich data on the bacterial population. As with all studies, there are limitations. With a likely resolution time of less than one month (Adrian Marcato, The Peter Doherty Institute for Infection and Immunity, personal communication), it is possible that we might have missed detecting some impetigo infections that arose and resolved within our two-month screening time period. Additional infections could add to the level of observed strain diversity, making our estimates a lower bound of diversity. Further, while children may have had as many as two swabs collected from different impetigo lesions, only the dominant colony from each swab was *emm* typed. We may have missed detecting multiple *emm* types within a single lesion, again meaning our results provide an underestimate of overall diversity. The sporadic detection of many *emm* types in our age-limited population sample suggests that the relevant mixing pool is wider than schools, and infection may be readily acquired from elsewhere in the community, including households. However, as only symptomatic infections were swabbed after the first skin screening visit, we were unable to investigate any relationship between throat colonization and skin infection and it is possible that these *emm* types may have been present asymptomatically in our cohort throughout the study. It is possible that GAS recovered from throat swabs during pharyngitis episodes was also asymptomatic colonization. For the single time point with throat colonization data, all but two of the *emm* types found in asymptomatic children were also present in impetigo and/or pharyngitis isolates. However, without longitudinal data on asymptomatic colonization in this cohort we cannot estimate the overall contribution of colonization to strain diversity.

The large number of concurrently circulating *emm* types (between 19 and 37 *emm* types per screening time period) is consistent with observations from other settings with high prevalence of impetigo [13, 21-23]. Studies in a mouse model of skin infection suggest re-infection with the same *emm* type within a short period is required to stimulate type-specific immunity following skin infection [24]. If this requirement holds true for human infections, it would be anticipated that this phenomenon would positively select for diversity in high prevalence settings, as a large number of circulating *emm* types makes re-exposure to the same *emm* type less likely. The observation that a very small number of children (28/457, 6.1%) had repeated isolation of the same *emm* type within the study period suggests that there would be limited opportunity for the development of type-specific immunity, should re-exposure be necessary. In an earlier study in a low prevalence population, where colonization was the dominant source of GAS isolates and pharyngitis the dominant disease manifestation, children were equally likely to acquire the same *emm* type as a different *emm* type in a single year [25]. This observation further supports the notion that low prevalence/low diversity settings afford greater opportunity for the development of natural immunity, given the greater likelihood of re-exposure to the same strain.

We did not observe the classic epidemiologic picture of A-C pattern strains dominating pharyngitis isolates and pattern D representing a very small percentage of pharyngitis isolates [6]. Rather, we observed patterns D and E dominant for both impetigo and pharyngitis (and pattern D dominant for colonization), consistent with studies in other high prevalence settings [12-14, 26, 27]. Causal mechanisms for this tropism remain uncertain [6, 28], but appear to be overwhelmed by transmission pressure in high prevalence settings.

An *in vitro* study of a subset of the children in our analysis found that skin infection may elicit a functional immune response in some children with M type-specific and cross-reactive immune responses following skin infection [8]. However, results varied depending on the *emm* cluster analyzed and the evidence was not strong for the D4 cluster, the most prevalent in our population

[8]. Our observations of re-acquisition of the same *emm* type or cluster during the period of our study suggest that the duration of protective immunity following infection may be short or even absent in some children. One could hypothesize that other, non-M protein, GAS antigens may be needed to protect against GAS infection due to D4 cluster in particular, and due to skin associated *emm* types in a tropical setting in general. A GAS vaccine would definitively need to provide a substantially longer duration of protection than the one observed here for some children to have an impact on prevalence.

This longitudinal study suggests that in settings where impetigo is the dominant mode of transmission, circulating *emm* types either may not be constrained by ecological niches or population immunity to the M protein, or may require several infections over a longer period of time to induce such immunity. With limited evidence of M immunity to GAS apparent in our data, further work is needed to understand how settings have transitioned from high prevalence to low prevalence (as has happened in many developed countries), as this may provide clues for future control.

References

1. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infectious Diseases **2005**; 5:685-94.

2. Sanyahumbi AS, Colquhoun S, Wyber R, Carapetis JR. Global disease burden of Group A Streptococcus. In: Ferretti JJ, Stevens DL, Fischetti VA, eds. *Streptococcus pyogenes* Basic biology to clinical manifestations. Oklahoma City: The University of Oklahoma Health Sciences Center, **2017**:661-704.

3. Fischetti VA. M protein and other surface proteins on Streptococci. In: Ferretti JJ, Stevens DL, Fischetti VA, eds. *Streptococcus pyogenes* Basic biology to clinical manifestations. Oklahoma City: The University of Oklahoma Health Sciences Center, **2017**:27-54.

4. Bessen DE. Molecular basis of serotyping and the underlying genetic organization of Streptococcus pyogenes. In: Ferretti JJ, Stevens D, Fischetti VA, eds. *Streptococcus pyogenes* Basic biology to clinical manifestations. Oklahoma City: The University of Oklahoma Health Sciences Center, **2017**:97-108.

5. Bessen DE, Smeesters PR, Beall BW. Molecular epidemiology, ecology, and evolution of Group A Streptococci. Microbiology Spectrum **2018**; 6:CPP3-0009-2018.

6. Bessen DE. Tissue tropisms in group A *Streptococcus*: What virulence factors distinguish pharyngitis from impetigo strains? Current Opinion in Infectious Diseases **2016**; 29:295-303.

7. Sanderson-Smith M, De Oliveira DMP, Guglielmini J, et al. A systematic and functional classification of *Streptococcus pyogenes* that serves as a new tool for molecular typing and vaccine development. The Journal of Infectious Diseases **2014**; 210:1325-38.

8. Frost HR, Laho D, Sanderson-Smith ML, et al. Immune cross-opsonization within emm clusters following group A *Streptococcus* skin infection: broadening the scope of type-specific immunity. Clinical Infectious Diseases **2017**; 65:1523-31.

9. Wannamaker LW. Differences between streptococcal infections of the throat and of the skin. The New England Journal of Medicine **1970**; 282:78-85.

10. Bessen DE, Lizano S. Tissue tropisms in group A streptococcal infections. Future Microbiology **2010**; 5:623-38.

11. Smeesters PR, McMillan DJ, Sriprakash KS, Georgousakis MM. Differences among group A streptococcus epidemiological landscapes: consequences for M protein-based vaccines? Expert Review of Vaccines **2009**; 8:1705-20.

12. 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**; 43:4369-76.

13. Smeesters PR, Vergison A, Campos D, de Aguiar E, Miendje Deyi VY, Van Melderen L. Differences between Belgian and Brazilian group A Streptococcus epidemiologic landscape. PLoS ONE **2006**; 1:e10.

14. Sakota V, Fry AM, Lietman TM, Facklam RR, Li Z, Beall B. 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**; 44:2160-6.

15. Steer AC, Jenney AWJ, Kado J, et al. High burden of impetigo and scabies in a tropical country. PLoS Neglected Tropical Diseases **2009**; 3:e467.

16. Steer AC, Jenney AWJ, Kado J, et al. Prospective surveillance of streptococcal sore throat in a tropical country. The Pediatric Infectious Disease Journal **2009**; 28:477-82.

17. Magurran AE. Measuring biological diversity. Oxford: Blackwood Science Ltd, **2004**.

18. Lancefield RC. Persistence of type-specific antibodies in man following infection with group A Streptococci. Journal of Experimental Medicine **1959**; 110:271-92.

19. Quinn RW, Van der Zwaag R, Lowry PN. Acquisition of Group A streptococcal M protein antibodies. Pediatric Infectious Disease **1985**; 4:374-8.

20. Hysmith ND, Kaplan EL, Cleary PP, Johnson DR, Penfound TA, Dale JB. Prospective longitudinal analysis of immune responses in pediatric subjects after pharyngeal acquisition of Group A Streptococci. Journal of the Pediatric Infectious Diseases Society **2017**; 6:187-96.

21. Richardson LJ, Towers RJ, Cheng AC, 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**; 28:5301-5.

22. Speers DJ, Levy A, Gichamo A, Eastwood A, Leung MJ. M protein gene (emm type) analysis of group A *Streptococcus* isolates recovered during an acute glomerulonephritis outbreak in northern Western Australia. Pathology **2017**; 49:765-9.

23. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. The Lancet Infectious Diseases **2009**; 9:611-6.

24. Pandey M, Ozberk V, Calcutt A, et al. Streptococcal immunity is constrained by lack of immunological memory following a single episode of pyoderma. PLoS Pathogens **2016**; 12:e1006122.

25. Martin JM, Green M, Barbadora KA, Wald ER. Group A Streptococci among school-aged children: clinical characteristics and the carrier state. Pediatrics **2004**; 114:1212-9.

26. McDonald MI, Towers RJ, Fagan P, Carapetis JR, Currie BJ. Molecular typing of *Streptococcus pyogenes* from remote Aboriginal communities where rheumatic fever is common and pyoderma is the predominant streptococcal infection. Epidemiology and Infection **2007**; 135:1398-405.

27. Williamson DA, Smeesters PR, Steer AC, et al. Comparative M-protein analysis of *Streptococcus pyogenes* from pharyngitis and skin infections in New Zealand: Implications for vaccine development. BMC Infectious Diseases **2016**; 16:561.

28. Loh JMS, Tsai JYC, Proft T. The ability of Group A streptococcus to adhere to immortalized human skin versus throat cell lines does not reflect their predicted tissue tropism. Clinical Microbiology and Infection **2017**; 23:677.e1-.e3.

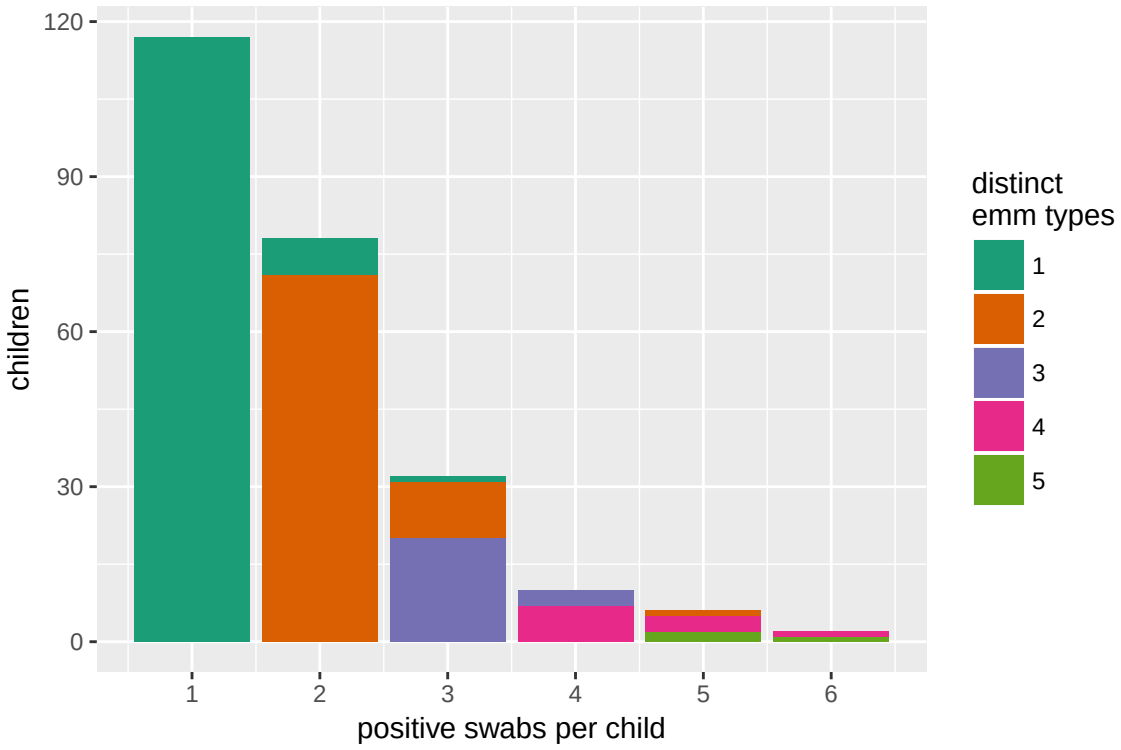
Figure legend

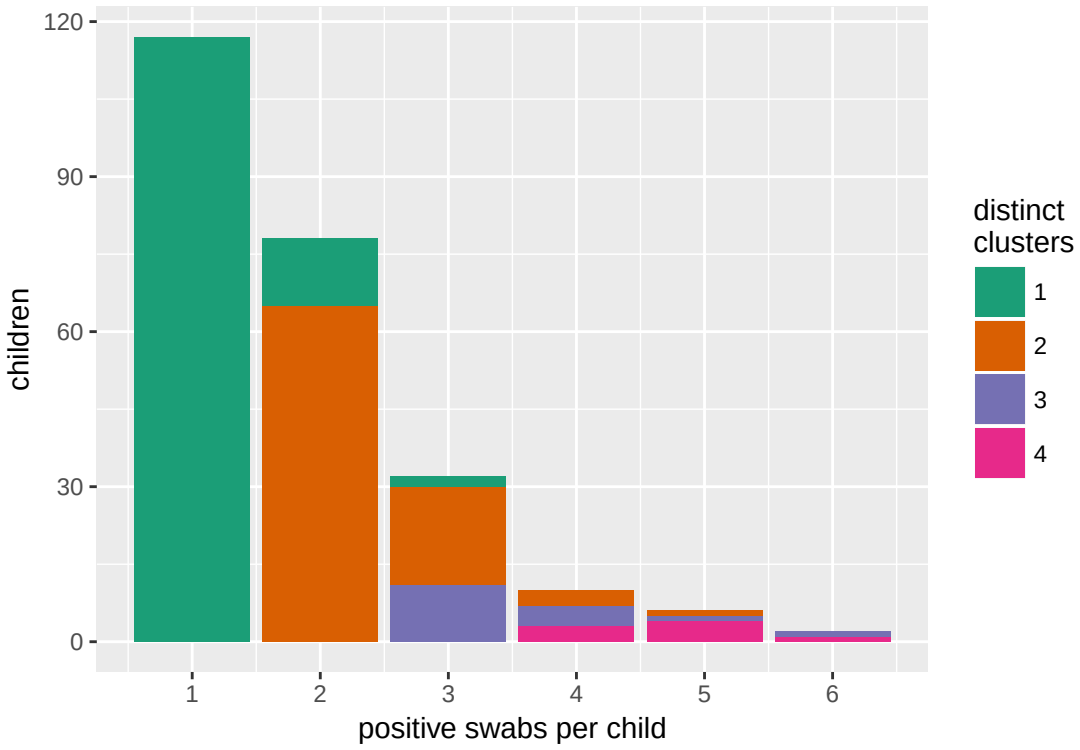
Figure 1: Number of GAS-positive swabs per child, colored by: A) the number of distinct *emm* types present in these swabs; and B) the number of distinct *emm* clusters present in these swabs.

Figure 2: Prevalence of *emm* types per 1000 children for each screening time period, grouped by cluster and stratified by school. *emm* clusters (labelled on y-axis) are ordered by overall number of isolates, and within each *emm* cluster, *emm* types are ordered by overall number of isolates. Labels for each school include the median population, number of positive swabs and overall prevalence, together with their ranges.

Figure 3: Prevalence of *emm* types per 1000 children for each time screening time period, grouped by cluster and stratified by specimen type (impetigo, pharyngitis, colonization). *emm* clusters (labelled on y-axis) are ordered by overall number of isolates, and within each *emm* cluster, *emm* types are ordered by overall number of isolates.

Figure 4: Longitudinal information for the subset of 28 children with repeated isolation of the same *emm* type. Solid lines connect the same *emm* type isolated on more than one occasion, with dotted lines indicating the same, but with a different *emm* type isolated between the two samples. Colors represent unique *emm* types, with the symbols representing the type of sample (S = skin; T = throat (with symptoms); C = throat (without symptoms)). Each 'X' represents a skin screening visit attended by the child during which no GAS was isolated, either because there were no sores present, there were sores present but no swab was taken, or a swab was taken but did not grow GAS.





	rural 2	rural 1	urban
population	167(160–175)	74(73–80)	181(161–202)
positive swabs	38(26–56)	14(7–20)	18(7–53)
prevalence	228(158–320)	180(93–274)	97(43–262)

D4 n=146: emm70, emm33, emm93.3, emm56, emm101, emm52.1, emm230, emm86.2, emm98.1, emm53, emm224, emm116.1, emm93.4

E3 n=78: emm25, emm82.1, emm103, emm44, emm49, emm25.1, emm58, emm219

E6 n=55: emm11, emm69.1, emm177, emm85, emm42, emm75.1, emm11.1, emm73, emm89, emm232.1

E4 n=42: emm77, emm112.2, emm89.11, emm109.1

E2 n=26: emm76.4, emm92, emm104

D1 n=19: emm207.1, emm54.1

D2 n=13: emm71, emm100

A-C4 n=13: emm39.1, emm39.4, emm218.1, emm74, emm4.5, emm60.1, emm165, emm105, emm123, emm238.3, emm19.4, emm57, emm222, emm14.4, emm122.2, emm205, emm137, emm97.1, emm55, emm197

1 2 3 4 5 6 | 1 2 3 4 5 6 | 1 2 3 4 5 6

time period

isolates per 1000 children

120
90
60
30

impetigo
pharyngitis
colonization
D4 n=146
E3 n=78
E6 n=55
E4 n=42
E2 n=26
D1 n=19
D2 n=13
A-C4 n=13
emm70
emm33
emm93.3
emm56
emm101
emm52.1
emm230
emm86.2
emm98.1
emm53
emm224
emm116.1
emm93.4
emm25
emm82.1
emm103
emm44
emm49
emm25.1
emm58
emm219
emm11
emm69.1
emm177
emm85
emm42
emm75.1
emm11.1
emm73
emm89
emm232.1
emm77
emm112.2
emm89.11
emm109.1
emm76.4
emm92
emm104
emm207.1
emm54.1
emm71
emm100
emm39.1
emm39.4
emm218.1
emm74
emm4.5
emm60.1
emm165
emm105
emm123
emm238.3
emm19.4
emm57
emm222
emm14.4
emm122.2
emm205
emm137
emm97.1
emm55
emm197
1 2 3 4 5 6 1 2 3 4 5 6 1
time period
isolates per 1000 children
30
20
10