Variability in the Distribution of Genes Encoding Virulence Factors and Putative Extracellular Proteins of *Streptococcus pyogenes* in India, a Region with High Streptococcal Disease Burden, and Implication for Development of a Regional Multisubunit Vaccine

Vivek Sagar, René Bergmann, Andreas Nerlich, David J. McMillan, D. Patric Nitsche Schmitz, and Gursharan S. Chhatwal

Department of Medical Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany, and Queensland Institute for Medical Research, Brisbane, Australia

*Streptococcus pyogenes* causes a wide variety of human diseases and is a significant cause of morbidity and mortality. Attempts to develop a vaccine were hampered by the genetic diversity of *S. pyogenes* across different regions of the world. This study sought to identify streptococcal antigens suitable for a region-specific vaccine in India. We used a two-step approach, first performing epidemiological analysis to identify the conserved antigens among Indian isolates. The second step consisted of validating the identified antigens by serological analysis. The 201 streptococcal clinical isolates from India used in this study represented 69 different *emm* types, with *emm*12 being the most prevalent. Virulence profiling of the North and South Indian *S. pyogenes* isolates with a custom-designed streptococcal virulence microarray identified seven conserved putative vaccine candidates. Collagen-like surface protein (SCI), putative secreted 5'-nucleotidase (PSNT), and C5a peptidase were found in 100% of the isolates, while R28, a putative surface antigen (PSA), and a hypothetical protein (HYP) were found in 90% of the isolates. A fibronectin binding protein, SfbI, was present in only 78% of the isolates. In order to validate the identified potential vaccine candidates, 185 serum samples obtained from patients with different clinical manifestations were tested for antibodies. Irrespective of clinical manifestations, serum samples showed high antibody titers to all proteins except for SCI and R28. Thus, the data indicate that PSNT, C5a peptidase, PSA, HYP, and SfbI are promising candidates for a region-specific streptococcal vaccine for the different parts of India.

*Streptococcus pyogenes* (group A streptococcus [GAS]) is exclusively a human pathogen and the etiological agent of a wide variety of diseases that vary in clinical severity, while also being a significant cause of morbidity and mortality (7). These diseases include pharyngitis, impetigo, scarlet fever, poststreptococcal glomerulonephritis, invasive diseases, rheumatic fever (RF), and rheumatic heart disease (RHD) (9). While rheumatic fever and rheumatic heart disease are the greatest cause of mortality in developing nations, deaths in developed nations are mainly attributable to invasive diseases (7, 26).

The differences in the prevalence and molecular epidemiology of GAS isolates reflect the differences in the importance of RF/RHD and invasive diseases in these populations. GAS carriage and infection are prevalent in many developing nations, with a large number of different *emm* types circulating at one time and no one type being dominant. In contrast, a limited number of specific *emm* types are predominant in developed nations and are often associated with specific clinical manifestations (2, 11, 31–33, 37, 41).

Despite the majority of GAS-associated deaths occurring in developing nations, the majority of comparative genetic and genomic studies have focused on isolates from developed nations (3, 5, 13, 20, 38). These studies report a high degree of genetic diversity between isolates with divergent *emm* types, as well as diversity within an *emm* type. In regions with high streptococcal disease burden, where GAS isolates are more likely to come into direct contact, the probability of lateral gene transfer (LGT) is increased. Greater inter- and intra-*emm*-type genetic diversity is likely to exist. Such diversity and potential changes in population structure are important considerations when designing vaccine candidates that should provide coverage against the entire GAS population.

In this study, we used a two-step approach in order to identify promising candidates for a region-specific vaccine. In the first step, we conducted epidemiological analysis in order to identify antigens conserved in different parts of India; in the second step, we purified these antigens and performed serological analysis using convalescent-phase serum samples. We used virulence gene profiling to assess genetic diversity and population structure of GAS in India, a country where the streptococcal disease burden is high (1, 7, 11, 31, 32). We found a high degree of genetic diversity between isolates of different *emm* types, but relatively conserved genotypes within an *emm* type. We assessed the distribution of seven genes conserved in Indian isolates encoding current vaccine targets in this population and report on the serological responses against each of these antigens.
TABLE 1 Oligonucleotides used in this study

| Primera | Sequence (5′–3″)b | Restriction site | Annealing temp (°C) |
|---------|------------------|------------------|--------------------|
| ScF     | GCGAATTCCAGGAGTCTTCTTCTACGACTATGA | EcoRI | 65 |
| ScR     | GCGCTGACAGCTCTGTTGGTGTGGCTA | SalI | |
| PSNTF   | GCCTAATTTTGTGATTTACAGGATAATAGATTT | EcoRI | 65 |
| PSNTR   | GCGCTGACAGCTGAGCTGATGGCACTCCCTCCACAT | SalI | |
| HYPF    | GCAATTTCCGATCGATGCGATGCGATGC | SalI | 65 |
| HYPR    | GCGCTGACCTGACTCATGGGGCCCTAA | EcoRI | |
| PSAF    | GCCTAATTTTGTGATTTACAGGATAATAGATTT | SalI | 65 |
| PSAR    | GCGCTGACCTTATTTGCGGAGTTGCTGGTACCT | EcoRI | |
| R28F    | GCAATTTCCGATCGATGCGATGCGATGC | SalI | |
| R28R    | GCGCTGACCTTATTTGCGGAGTTGCTGGTACCT | SalI | |
| ScPF    | CCGTGTGACTATTATCGCCGTCTGCTGCTTGGCG | BamHI | 55 |
| ScPR    | CCGTGTGACTATTATCGCCGTCTGCTGCTTGGCG | SalI | |

a F, forward; R, reverse.
b Restriction sites are indicated in boldface.

MATERIALS AND METHODS

Bacterial strains and human sera. The bacterial isolates and sera used in the present study were collected as part of the ASSIST program, funded by the European Commission, with three Indian and three European partners (http://www.helmholtz-hzi.de/en/research/research_projects/view/project/projekt/assist/). The samples were collected from 2007 to 2010 in two defined areas in Chandigarh (northern India) and Vellore (southern India), which are about 3,000 km apart, have high streptococcal disease burdens, and different climatic conditions. Besides hospital patients, 3,000 and 2,400 school children were screened in Chandigarh and Vellore, respectively. Sixty-five isolates from Chandigarh and 136 from Vellore were included here. The isolates were also classified on the basis of respiratory illness (http://www.helmholtz-hzi.de/en/research/research_projects/view/project/projekt/assist/) and clinical exposure in the survey areas due to high disease burden, we included here. The isolates were also classified on the basis of recovery from the throats of asymptomatic carriers (n = 44) or patients presenting with pharyngitis (n = 20). Another 32 isolates were collected from the skin in these surveys. Thirty-four isolates were collected from patients presenting with invasive disease at clinics in Chandigarh and Vellore. The isolation site of 71 isolates was unknown. Serum samples were collected from individuals in northern (n = 110) and southern (n = 75) India. Eighty-two of these samples were obtained from patients with RHD, 24 were collected from patients with RF, and 9 were obtained from patients with invasive disease. Another 15 were obtained from patients presenting with symptoms of pharyngitis, 30 were asymptomatic patients positive for GAS, and 25 sourced from healthy people from the survey areas with no current signs of streptococcal infection. Because of the possibility of subclinical exposure in the survey areas due to high disease burden, we included 10 serum samples collected from healthy people from Germany as controls. The serum samples and streptococcal isolates were from the same patients in case of invasive diseases. For all other clinical manifestations, the isolates and serum samples were from the defined survey areas, but not from the same patients.

DNA extraction and emm typing. Genomic DNA was isolated using zirconium beads in combination with a DNeasy kit (Qiagen, Hilden, Germany). The emm type of individual strains was determined by amplification and sequencing the 5′-end region of the emm gene described by Beall et al. (4) and comparison with the emm gene database (http://www.cdc.gov/ncidod/biotech/strep/strepblast.htm).

Microarray, hybridization, and data analysis. The microarray used here has been described in detail in a previous study (19). Sample preparation, array hybridization, and data processing were performed as described previously (27). Briefly, isolated genomic DNA was digested with AluI and labeled with biotin/streptavidin-Cy5, and fluorescence signals were quantified after hybridization using ImaGene software (BioDiscovery). A two-component mixture model was fitted to the background-corrected and log2-transformed data by a maximum-likelihood method. A discriminant function was used to represent the propensity of a gene for being present or absent in a particular isolate. Discriminant values were stored in a signal probability matrix and used to construct dendrograms using Bayesian agglomerative hierarchical clustering (34). All routines for statistical calculations were implemented in the R statistics package (www.r-project.org).

Cloning and expression of cell surface proteins. Of 219 genes on the array, 7 were selected that showed 75% or more frequency in 201 strains and encode for cell wall proteins. These proteins, which included collagen-like surface protein (SCI), putative secreted 5′-nucleotidase (PSNT), C5a peptidase (C5a), R28 (R28), putative surface antigen (PSA), and a hypothetical transposase (HYP), were amplified (Table 1) and cloned into the pGEX-6P-1 vector (GE Healthcare). The cloning of streptococcal fibronectin binding protein 1 (SfbI) was described by Talay et al. (42). The expression of the recombinant glutathione S-transferase (GST) fusion proteins in Escherichia coli BL21(DE3) was induced by the addition of IPTG (isopropyl-β-D-thiogalactopyranoside) to 0.5 mM. Cultures were harvested and lysed using a French press (SLM Instruments, Inc.). Cellular debris was removed by centrifugation, and the presence of recombinant proteins in supernatants was confirmed by SDS-PAGE.

Antibody responses to putative vaccine candidates. Enzyme-linked immunosorbent assay (ELISA) was performed using glutathione-coated microtiter plates (Thermo Fisher Scientific) with E. coli lysate (0.25 µg/ml) containing GST-tagged recombinant proteins according to the method of Sehr et al. (35). Phosphate-buffered saline (PBS) and E. coli lysate containing GST alone served as controls. A peptide representing the IgG binding motif from FOG, a streptococcal protein with IgG-Fc fragment binding capacity (23), was used as the normalization control. Horseradish peroxidase-conjugated goat antihuman IgG antibody and ABTS [2,2′-azinobis(3-ethylbenzthiazolesulfonic acid)] were used for color development that was measured at 405 nm. Absorbance values were subsequently normalized against values obtained from wells containing PBS, GST, and the FOG IgG binding peptide. Differences in absorbance between groups were examined using one-way analysis of variance (ANOVA). Because of the high number of serum samples and eight proteins, the experiment was designed in such a way that the maximum number of serum samples could be covered by one experiment in order to avoid experiment to experiment variations. Therefore, we chose to monitor the immunogenicity of proteins in all serum samples at the dilution of 1:200 and expressed the result as the optical density.

RESULTS

emm sequence type distribution of GAS isolates from northern and southern India. The 201 isolates collected represent 69 different emm sequence types (see Table S1 in the supplemental material) and 83 different emm sequence subtypes (data not shown).
The 20 most prevalent emm types (Fig. 1) were emm12 (6.5%), emm11 (5.5%), emm49 (4.5%), emm28, and emm80, st1389 (3.5% each), emm3 (3.0%), and emm4, emm44, emm75, and emm112 (2.5% each). The types emm2, emm22, emm69, emm74, emm77, emm93, emm110, emm104, and emm108 accounted for 2% of all isolates. The 20 prevalent emm types represented 57.7% of all GAS isolates in both regions.

Of these 20 prevalent emm types, 10 (emm4, emm11, emm12, emm44, emm49, emm74, emm80, emm110, emm112, and st1389) were found both in northern and southern India. emm3, emm28, emm69, emm77, emm93, emm104, and emm108 were only found in southern India, whereas emm2, emm22, and emm75 were specific for northern India. With 55 different emm sequence types, the southern Indian isolates showed a greater heterogeneity compared to 32 different emm sequence types found within the northern Indian isolates. This is reflected in the Simpson’s index of diversity (D) for emm types in southern India (D = 0.981; 95% confidence interval, 0.976 to 0.986) compared to northern India (D = 0.957; 95% confidence interval, 0.936 to 0.977).

Overview of gene distribution in Indian isolates. Of the 219 streptococcal virulence factors and extracellular surface proteins (VF/ECP) represented on the array, 91 (41.5%) were found in all isolates (see Table S2 in the supplemental material), and 150 (68.5%) were found in more than 80% of the isolates. In order to identify region-specific virulence genes, we statistically analyzed the distribution of VF/ECP genes with respect to geographic region, 102 genes (46.6%) and 100 genes (45.7%) were present in all northern and southern Indian isolates, respectively. In order to identify region-specific virulence genes, we statistically analyzed the distribution of VF/ECP genes in both regions. We identified one gene in the northern Indian isolates and 15 genes in the southern Indian isolates with a statistically significant difference in distribution (Fisher exact test, P < 0.05, Table 2), reflecting the higher emm type heterogeneity in southern India. However, none of the 16 genes was found exclusively in one of the regions.

Genetic relationships among isolates. emm typing is still largely accepted as sufficient for defining related strains in GAS research. However, epidemiological and molecular observations also demonstrate that LGT involving the emm gene occurs, suggesting that isolates of the same emm type that are temporally or geographically displaced may be genetically diverse and are more closely related to isolates of a different emm type (6, 12, 22, 25). To investigate the relationship between all isolates, we used a Bayesian agglomerative hierarchical cluster algorithm (BHC) to construct a dendrogram of all isolates that were represented by ≥3 isolates per emm type (n = 154 isolates), based on the presence or absence of all genes represented on the array. Using this approach, we found isolates of the same emm type to predominantly cluster together (Fig. 2). However, there were instances where this did not occur, suggesting that LGT of the emm gene may have occurred in these instances. We next analyzed the association of the clusters with regard to disease type. As shown in Fig. 2, there was only one distinct subcluster solely composed of invasive isolates consisting of one emm4, one emm28, and three emm49 strains. Using BHC clustering, we did not find any significant clustering of virulence factors with invasive and noninvasive isolates. However, when we analyzed the association of virulence factors with invasiveness using the Fisher exact test (P < 0.05), we found 13 genes positively associated with invasiveness (see Table S3 in the supplemental material). Six genes were negatively associated with invasiveness (odds ratio < 1).

Serological response to vaccine antigens. We examined the distribution of the shortlisted seven GAS vaccine candidates in the Indian population and found them to be highly conserved across the population (Table 3). Collagen-like surface protein (SCI), putative secreted 5'-nucleotidase (PSNT), and C5a peptidase were found in 100% of isolates. R28, the putative surface antigen (PSA), and the hypothetical protein (HYP) were found in >90% of the isolates. Streptococcal fibronectin binding protein I (SfbI) was found in 78.5% of isolates.

---

**FIG 1** Distribution of the 20 most common GAS emm types. Black bars represent emm types found in northern India, and open bars represent isolates found in southern India. The 20 prevalent emm types only represented 58% of all GAS isolated in both regions.
Serum antibodies indicate that epitopes of a particular protein are visible to the host immune system, and these proteins are likely to be recognized by antibodies generated against corresponding vaccine candidates. Therefore, we next investigated serological responses to these seven identified GAS vaccine candidates. We expressed these proteins as fusions with GST and used them in ELISAs with human sera (Fig. 3).

Control healthy sera had very low immune response against all of the proteins. The majority of serum samples from India had elevated responses to PSNT, C5a peptidase, and PSA, demonstrating the high incidence of GAS infection in these populations. PSA-specific IgG responses were statistically more frequent in the pharyngitis group than in healthy and RHD groups (P < 0.05). Lower antigen specific antibody responses were observed with HYP and SfbI. For these two proteins, the highest responses were observed with HYP and SfbI. For these two proteins, the highest responses were observed with sera collected from pharyngitis patients and patients with GAS throat colonization at the time of sera collection. For HYP, IgG responses in the pharyngitis group were statistically significantly higher than in the healthy groups from nonsurvey and survey areas and in the RF group (P < 0.05). The SfbI response for the carrier group and pharyngitis group was statistically significantly higher than control and RHD groups (P < 0.05) (see Table S4 in the supplemental material). The majority of sera did not respond or responded weakly to R28 and SCI. No differences in responses were observed in sera collected from northern and southern India. For 4 of 7 of the proteins used here the response in pharyngitis sera was significantly higher than nonsurvey area healthy sera (see Table S4 in the supplemental material), indicating the possibility that these proteins are most likely expressed in the early phase of infection.

**DISCUSSION**

This study is the most comprehensive assessment of the genetic repertoire of GAS carried out in a country with a high streptococcal disease burden. The 20 most common emm types accounted for 57.7% of the isolates collected, reinforcing the high degree of diversity in the population based on emm typing and is consistent with results of other studies of GAS population structure carried out in developing nations (41). Within the Indian population, 91 genes were conserved in 100% of isolates. Using the same array, we previously reported that 129 and 125 genes were 100% conserved in GAS isolates from the Netherlands and the United States (18, 19), respectively. The majority (80 genes) of conserved genes in the isolates from Europe and India were similar, indicating that they are part of the GAS core genome and encode proteins that are critical for virulence or confer biological functions essential to the fitness of the organism and are therefore promising vaccine candidates.

The M protein remains a favored GAS vaccine candidate. The predominant bactericidal antibody response raised after GAS infection targets the amino-terminal of the M protein, the same region that is used for emm typing. However, antibodies raised against this region have traditionally been thought to be type specific, i.e., antibodies raised against the amino terminus of one M protein will not recognize the amino terminus of other M proteins. Candidate vaccines that target this region therefore must include amino termini from multiple M proteins. Of the 69 emm sequences reported in the present study, 14 are represented in the 26-valent amino-terminal GAS vaccine candidate (14) and represent 31.8% of isolates recovered. The highest proportion of isolates included in the vaccine was found in southern Indian GAS isolates (17.4%) and included 10 vaccine-related emm types. Of the northern Indian isolates, 14.4% were covered by the vaccine and included six vaccine-related emm types. Recently, a similar 30-valent vaccine has also been reported (10). Nineteen of the emm types present in the 30mer vaccine were present in isolates in the present study. That same study also reported that cross-reactivity was observed with another 24 M proteins not present in the vaccine. Fourteen of these cross-reactive emm types were also present here. In theory, the 30-valent vaccine would induce antibodies that are effective against 58% of the isolates recovered here. The emm types present in the 26-valent vaccine and 30-valent vaccine were chosen based on their importance in North American and European contexts. A tailored vaccine, containing

**TABLE 2 Genes differentially conserved in northern and southern Indian GAS populations**

| Gene  | Gene name                                           | Conservation (%) |
|-------|-----------------------------------------------------|------------------|
|       |                                                     | Northern India   | Southern India  |
| Spy0116 | Hypothetical protein                               | 56.9             | 91.9            |
| Spy0317 | Conserved hypothetical protein                     | 55.4             | 92.6            |
| Spy1006 | Putatively lysin-phage associated                   | 96.9             | 85.3            |
| Spy2009 | Hypothetical protein (transposase)                  | 83.1             | 94.9            |
| SpyM3_0130 | Streptolysin O                                     | 63.1             | 94.1            |
| SpyM3_0304 | Conserved hypothetical protein                      | 78.5             | 94.9            |
| SpyM3_0343 | Hypothetical protein                              | 93.8             | 100.0           |
| SpyM3_0653 | Putative ABC transporter substrate-binding protein | 87.7             | 98.5            |
| SpyM3_0815 | Putative hemolysin III                             | 70.8             | 90.4            |
| SpyM3_0823 | Hypothetical protein                              | 83.1             | 97.1            |
| SpyM3_0833 | Putative citrate lyase beta subunit                | 86.2             | 97.1            |
| SpyM3_0862 | Putative DNA/pantothenate metabolism flavoprotein | 40.0             | 91.9            |
| SpyM3_0999 | Hypothetical protein                              | 93.8             | 100.0           |
| SpyM3_1390 | Putative penicillin-binding protein 1A            | 92.3             | 99.3            |
| SpyM3_1718 | Surface lipoprotein DppA                        | 67.7             | 91.2            |
| SpyM3_1762 | Hypothetical protein                             | 93.8             | 99.3            |

a Fisher exact test, P < 0.05.
amino termini from M types common in India or other regions where streptococcal disease is prevalent, may increase vaccine coverage by this approach. However, given the large numbers of \textit{emm} types circulating and the differences in M types present in northern and southern India, such a vaccine may prove difficult to design.

Both traditional and reverse vaccinology approaches have been used to identify alternatives to the M protein (8). Rodriguez-Ortega et al. (29) used a proteomic approach for identifying streptococcal surface-exposed proteins for their use as vaccine candidates. Those researchers, however, used only three \textit{emm} types that were not prevalent in India. Reverse vaccinology, a genome-based approach to vaccine development, has also been used to identify streptococcal vaccine candidates (17, 36). That approach, however, requires the whole genome sequence of the isolates and, to date, no prevalent \textit{emm}-type strain in India has been sequenced.

Proteins that confer essential virulence and biological functions and are therefore encoded in the core genome are attractive vaccine targets. In addition to functionality through bactericidal or neutralizing activity, antibodies raised against critical proteins may contribute to prevention of infection by inhibiting protein function. Multisubunit vaccines containing several proteins have added appeal, since the abolition of multiple biological activities...
FIG 3  Serological responses to streptococcal antigens in sera. Samples were collected from healthy individuals, asymptomatic carriers, and patients presenting with pharyngitis, RF, RHD, and invasive diseases. Control sera (non-endemic healthy) were collected from patients living in Germany. Statistical differences between the mean absorbance for each group were determined using one-way ANOVA and are presented in Table S4 in the supplemental material.
may attenuate the virulence of the organism further. The targeting of several proteins also potentially has an additive effect with respect to opsonic or neutralizing antibodies. A multisubunit approach to vaccine development can also reduce the probability of vaccine escape that may occur after the loss of epitopes targeted by the vaccine that may occur through mutation or lateral gene transfer (24).

The seven proteins that we examined serologically have all been proposed as potential vaccine candidates and were used here because they represent both well-characterized and relatively new targets. All of these proteins are also predicted to be surface associated. Our array data support this selection because six of these proteins were conserved in all of the invasive isolates tested here. R28 is a cell surface virulence protein that has many repetitive sequences and has homology with a protein found in group B streptococcus (39, 40). SCI is a collagen-like protein that is known to have a collagen-like sequence (16). HYP has homology to FbaA proteins (43), suggesting that it has fibronectin binding ability. PSNT is a putative surface nucleotidase protein likely to be involved in nutrient acquisition (44). The corresponding PSNT of Haemophilus influenzae has been reported to have protective efficacy in a rat infection model (44). PSA is a homologue to Spy0843 of S. pyogenes MGAS5005, which contributed to a protective host immune response in a mouse infection model (28).

The identified antigens were validated for their vaccine potential by serological analysis. For invasive disease, we had the matching isolates and serum samples, because these were acute manifestations that required hospitalization. For other manifestations, which were not acute, it was not possible to have the matching isolate serum samples, so we had to rely on the samples from defined survey areas. The immunogenicities of the identified proteins were tested using ELISA in order to determine their suitability as vaccine candidates. For simplicity, antibodies were determined using a single dilution and results were expressed as optical density values. For further studies on these antigens, however, the endpoint titers or the use of a reference standard serum will be required (21). Our data demonstrate that antibodies to five of the seven GAS proteins are present in sera in individuals living in a region where streptococcal disease is common. In general, there was no difference between antibody levels between individuals, irrespective of an individual’s disease status. However, two proteins, HYP and SfbI, had higher antibody responses in sera from pharyngitis and asymptomatic throat carriers compared to sera from RF/RHD and asymptomatic noncarriers. Fibronectin binding proteins have been shown to be involved in adherence to epithelial cells (30, 42) and may therefore be important in colonization and early infection. The responses to PSA, which binds to the surfaces of epithelial cells (15), were also significantly higher in sera from pharyngitis patients. Given the increased antibody response to these three antigens (HYP, SfbI, and PSA) in carriers and patients with pharyngitis, they may be excellent candidates for inclusion in a vaccine designed to prevent initial colonization. In the present study, a lower immune response was generally detected against all of the proteins with serum samples from invasive diseases. This could be due to the fact that the invasive disease is an acute manifestation, where the time is too short for the generation of sufficient antibodies to give a high response.

In conclusion, PSNT, C5a peptidase, PSA, HYP, and SfbI are promising candidates for a region-specific streptococcal vaccine for the Indian continent. Among these five candidates, PSA and PSNT have not been well studied and can be further explored in future studies. The two-step approach used here to identify vaccine candidates with regional specificity seems very promising. However, more isolates and more defined serum samples should be analyzed in order to more fully evaluate this technique.

ACKNOWLEDGMENTS

We thank N. Janze for excellent technical assistance. We are grateful to the ASSIST epidemiology team members, especially A. Kumar, K. N. Brahmadathan, V. Abraham, and Y. Sharma, for providing streptococcal isolates and serum samples. We thank A. P. Osley for carefully reading the manuscript.

This study was supported by the European Community’s Sixth Framework Programme ASSIST under contract 032390.

REFERENCES

1. Agarwal AK, Yunus M, Ahmad J, Khan A. 1995. Rheumatic heart disease in India. J. R. Soc. Health 115:303–309.
2. Asrat D, et al. 2006. High diversity of group A streptococcal emm types among healthy schoolchildren in Ethiopia. Clin. Infect. Dis. 42:1362–1367.
3. Banks DJ, et al. 2004. Progress toward characterization of the group A streptococcus metagenome: complete genome sequence of a macrolide-resistant serotype M1 strain. J. Infect. Dis. 190:727–738.
4. Beall B, Facklam R, Thompson T. 1996. Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci. J. Clin. Microbiol. 34:953–958.
5. Beres SB, et al. 2006. Molecular genetic anatomy of inter- and intrasertype variation in the human bacterial pathogen group A streptococcus. Proc. Natl. Acad. Sci. U. S. A. 103:7059–7064.
6. Bessen DE, Hollingshead SK. 1995. Horizontal transfer and mosaic-like emm gene structures in group A streptococci. Dev. Biol. Stand. 85:169–173.
7. Carapetis JR, Steer AC, Mulholland KE, Weber M. 2005. The global burden of group A streptococcal diseases. Lancet Infect. Dis. 5:685–694.
8. Cole JN, Henningham A, Gillen CM, Ramachandran V, Walker MJ. 2008. Human pathogenic streptococcal proteomics and vaccine development. Proteomics Clin. Appl. 2:387–410.
9. Cunningham MW. 2000. Pathogenesis of group A streptococcal infections. Clin. Microbiol. Rev. 13:470–511.
10. Dale JB, Penfold TA, Chiang EY, Walton WJ. 2011. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococcus. Vaccine 29:8175–8178.
11. Dey N, et al. 2005. High diversity of group A Streptococcus emm types in an Indian community: the need to tailor multivalent vaccines. Clin. Infect. Dis. 40:46–51.
12. Eshaghi M, Ali AM, Jamal F, Yusoff K. 2002. Existence of two emm-like “mrp” and “emm” genes in the mga regulon of the Streptococcus pyogenes strain ST4347. J. Biochem. Mol. Biol. Biophys. 6:23–28.
13. Ferretti JJ, et al. 2001. Complete genome sequence of an M1 strain of Streptococcus pyogenes. Proc. Natl. Acad. Sci. U. S. A. 98:4658–4663.
14. Hu MC, et al. 2002. Immunogenicity of a 26-valent group A streptococcal vaccine. Infect. Immun. 70:2171–2177.
15. Loimaranta V, et al. 2009. Leucine-rich repeats of bacterial surface proteins serve as common pattern recognition motifs of human scavenger receptor gp340. J. Biol. Chem. 284:18614–18623.
16. Lukomska S, et al. 2000. Identification and characterization of the scl gene encoding a group A streptococcus extracellular protein virulence factor with similarity to human collagen. Infect. Immun. 68:6542–6553.
17. Maione D, et al. 2005. Identification of a universal group B streptococcus vaccine by multiple genome screen. Science 309:148–150.
18. McMillan DJ, et al. 2006. Genes for the majority of group A streptococcal virulence factors and extracellular surface proteins do not confer an increased propensity to cause invasive disease. Clin. Infect. Dis. 43:884–891.
19. McMillan DJ, et al. 2007. Variations in the distribution of genes encoding virulence and extracellular proteins in group A streptococci are largely restricted to 11 genomic loci. Microbes Infect. 9:259–270.
20. McShan WM, et al. 2008. Genome sequence of a nephritogenic and highly transformable M49 strain of Streptococcus pyogenes. J. Bacteriol. 190:7773–7785.
21. Miura K, et al. 2008. Development and characterization of a standardized ELISA including a reference serum on each plate to detect antibodies induced by experimental malaria vaccines. Vaccine 26:193–200.

22. Musser JM, et al. 1995. Genetic diversity and relationships among Strep-tococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease. Infect. Immun. 63:994–1003.

23. Nitsche-Schmitz DP, et al. 2007. Group G streptococcal IgG binding molecules FOG and protein G have different impacts on opsonization by C1q. J. Biol. Chem. 282:17530–17536.

24. Pai R, et al. 2005. Postvaccine genetic structure of Streptococcus pneumoniae serotype 19A from children in the United States. J. Infect. Dis. 192:1988–1995.

25. Panchaud A, et al. 2009. M-protein and other intrinsic virulence factors of Streptococcus pyogenes are encoded on an ancient pathogenicity island. BMC Genomics 10:198. doi:10.1186/1471-2164-10-198.

26. Pfoh E, Wessels MR, Goldmann D, Lee GM. 2008. Differences between Belgian and Brazilian group A streptococcus strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. U. S. A. 99:4668–4673.

27. Sette A, Rappuoli R. 2010. Reverse vaccinology: developing vaccines in the era of genomics. Immunity 33:530–541.

28. Smeesters PR, et al. 2006. Differences between Belgian and Brazilian group A streptococcus epidemiologic landscape. PLoS One 1:e10. doi: 10.1371/journal.pone.0000010.

29. Smoot JC, et al. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A streptococcus strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. U. S. A. 99:4668–4673.

30. Stalhammar-Carlemalm M, Areschoug T, Larsson C, Lindahl G. 1999. The R28 protein of Streptococcus pyogenes is related to several group B streptococcal surface proteins, confers protective immunity and promotes binding to human epithelial cells. Mol. Microbiol. 33:208–219.

31. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. 2009. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. Lancet Infect. Dis. 9:611–616.

32. Talay SR, et al. 2000. Co-operative binding of human fibronectin to Sfb1 protein triggers streptococcal invasion into respiratory epithelial cells. Cell Microbiol. 2:521–535.

33. Zagursky RJ, et al. 2000. Identification of a Haemophilus influenzae 5’-nucleotidase protein: cloning of the nucl gene and immunogenicity and characterization of the NucA protein. Infect. Immun. 68:2525–2534.