Analysis of virulence factors and *emm* typing of *Streptococcus pyogenes* clinical isolates.

Authors: Gustavo E. Sambrano[1], Gustavo P. Riboldi[2], Keli C. Reiter[1], Thiago G. S. Paim[1], Neidimar C. C. Tolfo[1], Renata O. Soares[1], Pedro A. d'Azevedo[1].

[1]Laboratório de Cocos Gram-positivos, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

[2] Center for Medicinal Chemistry of Open Innovation, UNICAMP, Campinas, São Paulo, Brazil.

Corresponding author:

Gustavo Enck Sambrano;

245, Sarmento Leite Street, Porto Alegre, RS, Zip code: 90050-170, Brazil.

E-mail address: gustavo.sambrano@gmail.com.

Abstract

Background**:** *Streptococcus pyogenes*, a Group A streptococci (GAS), is an important human pathogen that causes a wide range of infections.

Methods: Twenty five clinical isolates of *S. pyogenes* were submitted to an *emm* typing and to a Real-time PCR analysis for 23 important virulence factors.

Results: Fourteen *emm* types were found and the *emm1* type was the most prevalent. The majority of the isolates were classified as *emm* pattern E, followed by A-C3. No pattern D was found. Among the virulence factors, the most prevalent were SpeG, Slo, C5a-peptidase and SPNA. Phage encoded virulence genes were also found among the strains, such as *mf-2, SpeJ* and *SpeL*.

Discussion: The *emm1* type was the most prevalent while the 13 others M types were distributed along the strains. No tissue tropism was found on the isolates. The virulence factors analysis demonstrated that chromosomally and phage-encoded genes were found, which confers a potential for high virulent micro-organisms.

Introduction

Streptococcus pyogenes, a Group A streptococci (GAS), is an important human pathogen that causes a wide range of infections, from mild to severe invasive diseases. GAS are responsible for more than 500,000 deaths annually, most of them occurring in low and middle-income countries (Steer et al., 2016).

Considered one of the most important virulence factors, the M protein promotes host interactions and adherence to human epithelial cells; specially helping bacteria to escape from host immune response by inhibition of phagocytosis (McMillan et al., 2013). The discrimination of GAS into more than 250 *emm* types is possible due to the sequencing of the hypervariable region of the *emm* gene encoding the M protein (Friães et al., 2012). Although this is considered an effective method to distinguish these micro-organisms, it still comprises a high number of *emm* types. Moreover, it has contributed to the development of at least two more classifications systems. The first one is called *emm* pattern and serves as a genotypic marker to classify GAS strains according to their tissue tropism; A-C throat specialists, D skin specialists and E generalists. This method is based on the organization of *emm* and *emm*-like genes located in the *mga* locus (McGregor et al., 2004). The second is called *emm* cluster and is based on the entire sequence of the M protein. A phylogenetic tree was built using sequences of 175 different *emm* types, and they were classified by their sequence similarity. This classification was able to divide the strains into two major clades, X and Y. The clade Y was divided in two main subclades (Y1 and Y2), and these subclades were further subdivided into 48 *emm*-clusters along with clade X (Sanderson-Smith et al., 2014).

Many virulence factors contribute to the pathogenesis of *S. pyogenes* diseases, each one with a specific function and with a determined degree of importance (Bisno, Brito & Collins, 2003; Yang et al., 2013). Superantigens (SAgs) are toxins capable of activating a large set of human T cells, resulting

in a massive production of proinflammatory cytokines (Reglinski & Sriskandan, 2014). The ability to hemolyze red blood cells (β-hemolysis) is due to the presence of two distinct toxins streptolysin S (SLS) and the unrelated large cholesterol-dependent, oxygen-sensitive streptolysin O (SLO) (Molloy et al., 2011). Among this great variety of virulence factors described in GAS strains, some are directly linked to phage carriage. The importance of phages for genetic variability, especially by transferring virulence genes, has been extensively studied and is well consolidated (Banks, Beres & Musser, 2002).

Over the exposure, the aim of this study was to describe the genetic profile of clinical isolates of *Streptococcus pyogenes*.

Materials and Methods

Bacterial strains were collected from two hospitals in the South of Brazil during the 2000-2013 period. This amount ended up to have most of the samples collected in the last two years. The collection method used was convenience sampling depending on the hospital's availability. This study was approved by the Ethics Committee of institutions involved. The clinical isolates were identified as *Streptococcus pyogenes* by the API 20 Strep system (bioMérieux SA, Marcy L'Etoile, France) and by 16S rRNA conventional PCR (Lintges et al., 2007).

Fresh colonies, 24h growth on trypticase soy agar supplemented with 5% sheep blood were used to DNA extraction by the QIAamp DNA Mini Kit (QIAGEN) as described by the manufacturers. The DNA product was quantified in the NanoSpec (Nanometrics). Then, a PCR reaction for the *emm* gene was performed in triplicate for each isolate, according to protocols available at the Centers for Disease and Control (CDC) website (www.cdc.gov/streplab/). PCR products were purified with Wizard® Genomic DNA Purification Kit (Promega) and measured to guarantee the correct DNA

concentration. The amount of DNA ranged from 20-40 ng/µl. The complete *emm* gene DNA was sequenced in triplicate, forward and reverse, using primers described by CDC Strep Lab in a concentration of 4 pmol. Three sequencing results from each isolate were analysed using the software Geneious 9.1.8. The consensus sequence obtained was manually adjusted to ensure a minimum phred score of 20 for each base. The nucleotide sequences were submitted to comparison with the CDC *emm* database in order to obtain the correct *emm* type and *emm* cluster for each isolate. The genes listed in Table 1 were used to analyse virulence factors, using a SYBR Green Master Mix and the platform ABI7500 (Applied Biosystems). All reactions were performed in duplicate, with a positive and a negative control for each primer. A positive amplification demonstrated similar melting curves in both reactions compared to its respectively positive control (Tm ± 1°C). In addition, an analysis of the quantification plot was made to ensure that the melting peak represented a true amplification.

The data analysis was mostly descriptive and no statistical analysis was performed due to the limited number of isolates.

Results

The conventional PCR for the *emm* gene showed a positive reaction for all 25 isolates. BLASTn results were acceptable regarding query cover (range from 91-100%), E-value (zero), and identity (range from 93-100%).

Fourteen different *emm* types were identified among the 25 clinical isolates. The most frequent was *emm1* 7/26 (28%), followed by *emm12, emm27, emm57, emm60* and *emm68* 2/25 each, (8%). Other *emm* types were found in only 1/25 (4%) strains, namely *emm8, emm22, emm44, emm58, emm59, emm73, emm90* and *emm92*. The *emm* pattern analysis demonstrated that 15/25 (60%) of the isolates were classified as *emm* pattern type E (*emm* cluster, E2, E3, E4 and E6) while 10/25 (40%)

were pattern A-C (*emm* cluster, A-C3, A-C4 and M57). No isolate was classified as *emm* pattern D. The sites of isolation were as follows: skin lesion (8/25), blood (7/25), oropharynx (6/25) and the last four clinical samples were obtained from pleural fluid, sputum, and abscess(2).

Regarding the 23 virulence factors, twelve belonged to the superantigens class, and were distributed as follows: *speA1-3* (40%), speA-5 (52%), *speC* (44%), *speJ* (76%), *speK* (24%), *speL* (88%), *speH* (56%), *speI* (64%), *speM* (68%), *smeZ* (80%) and *ssa* (28%). The only SAg present in all clinical specimens was *speG*, and the only absent was the *speA4* allele. Among the non-SAgs virulence genes, three mithogenic factors were verified, and the most frequent was *mf2* (92%), followed by *mf3* (68%) and *dnaseB* (*SpeF*) (56%). The prevalence of the two-serum opacity factor portions were: *sofC* (68%) and *sofN* (48%). The last six virulence factors analyzed were reported individually. Three of them were found in all isolates: streptolysin O (*slo*), C5a-peptidase (*scpa*), and a nuclease A (*spna*). The Dnase (*sda1*) was present in 88% of the isolates, and the NAD-glycohydrolase (*nga*) in 96%. The streptococcal inhibitor of the complement (*sic*) was present in 80% of the isolates.

The superantigens were similarly distributed among the *emm* types and the sites of infection, showing the diversity of the isolates included in this study (Fig 1).

Discussion

Streptococcus pyogenes is estimated to be responsible for over 600 million new cases of pharyngitis each year. In addition to triggering autoimmune sequelae, such as acute rheumatic fever, nasopharyngeal infection with *S. pyogenes* represents the principal reservoir for invasive diseases such as necrotizing fasciitis, pneumonia and toxic shock syndrome (Osowicki et al., 2018).

Numerous typing schemes have been used to characterise and measure the genetic diversity among isolates of *S. pyogenes*. One of the most used is *emm* typing, in which the hypervariable portion of the *emm* gene encoding M protein is sequenced (Steer et al., 2009). Recently, a study from São

Paulo, Brazil, demonstrated that the most predominant *emm* types were *emm1*, *emm6*, *emm12*, *emm22*, *emm77* and *emm87* (De Amicis et al., 2014). Although our sample size was not adequate for an epidemiological study, we were able to verify that emm1 was predominant. However, the other 13 emm types on this work did not stand out and were found to be well distributed among the strains.

Sixty percent of the isolates were classified as *emm* pattern E and 40% were classified as pattern A-C. This classification has been shown to correlate significantly with tissue tropism; patterns A-C are the throat specialists, D is the skin specialists and E is generalists. However, only a small proportion of specific *emm* pattern strains have been extensively studied (Chiang-Ni et al., 2016). Here, even with a variety of infections sites, no pattern D was found, and no relatedness was found among the site of infection and the *emm* pattern or even the *emm* cluster.

Streptococcus pyogenes is recognized by its arsenal of virulence factors, which targets and impairs the immune system. Many of these are scattered throughout the genome, and are not strictly part of pathogenicity island. For the most part, these virulence genes are conserved amongst strains, apart from the ones carried by phages (Ibrahim et al., 2016). In our study, four virulence genes were present in all 25 isolates, a superantigen (*speG*), a streptolysin O (*slo*), C5a-peptidase (*scpa*), and a nuclease A (*spna*). The first three are usually present within the GAS chromosomes while the last one is prophage encoded. The mithogenic factor (*mf2*) and the superantigens (*speL* and *speJ*) are all phage encoded, and were also present in a high number of isolates. In addition, the streptococcal inhibitor of complement (*sic*), the NAD-glycohydrolase (*nga)*, which is directly linked to SLO production and the superantigen SMEZ, are all chromosomally encoded and were also present in a high number of isolates (Banks, Beres & Musser, 2002; Ibrahim et al., 2016). These results can be considered evidence that horizontal gene transfer occurs among our isolates and that most of them have a high number of virulence factors, which demonstrates a potential to cause severe diseases.

An important review has brought to light many studies about SAgs and its association with invasive and non-invasive diseases and asymptomatic carriage. However, it is still a controversial topic, because associations are shown in some studies but not in others (Commons et al., 2014). In our study, SAgs were distributed along the diverse clinical isolates without showing a pattern, which could led us to believe in associations between them and *emm* types or site of infection.

This investigation was performed with a number of strains that is considered low for an epidemiological approach. However, we believe that the results are in accordance with our aims, and also could be used as preliminary data for other studies.

Conclusions

In summary, the *emm* types found here are too diverse to define a tendency. The *emm* pattern and *emm* cluster analysis demonstrated that the majority of the isolates were classified as generalists considering the site of infection, and no isolate was classified as skin specialists. The analysis of the virulence factors demonstrated that chromosomally and phage-encoded genes were found, which confers a potential for high virulent micro-organism. No relatedness could be found among virulence factors and *emm* typing or site of infection, reinforcing the controversial character of this subject.

References

De Amicis KM, Freschi de Barros S, Alencar RE, Postól E, Martins C de O, Arcuri HA, Goulart C, Kalil J, Guilherme L. 2014. Analysis of the coverage capacity of the StreptInCor candidate vaccine against Streptococcus pyogenes. *Vaccine* 32:4104–4110. DOI: 10.1016/j.vaccine.2013.08.043.

Banks DJ, Beres SB, Musser JM. 2002. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. *Trends in Microbiology*. DOI: 10.1016/S0966-842X(02)02461-7.

Bisno AL, Brito MO, Collins CM. 2003. Molecular basis of group A streptococcal virulence. *Lancet Infectious Diseases* 3:191–200. DOI: 10.1016/S1473-3099(03)00576-0.

Chiang-Ni C, Zheng P-X, Wang S-Y, Tsai P-J, Chuang W-J, Lin Y-S, Liu C-C, Wu J-J. 2016. Epidemiology Analysis of Streptococcus pyogenes in a Hospital in Southern Taiwan by Use of the Updated emm Cluster Typing System. *Journal of clinical microbiology* 54:157–62. DOI: 10.1128/JCM.02089-15.

Commons RJ, Smeesters PR, Proft T, Fraser JD, Robins-Browne R, Curtis N. 2014. Streptococcal superantigens: categorization and clinical associations. *Trends in Molecular Medicine* 20:48–62. DOI: 10.1016/J.MOLMED.2013.10.004.

Friães A, Pinto FR, Silva-Costa C, Ramirez M, Melo-Cristino J, Portuguese Group for the Study of Streptococcal Infections. 2012. Group A streptococci clones associated with invasive infections and pharyngitis in Portugal present differences in emm types, superantigen gene content and antimicrobial resistance. *BMC microbiology* 12:280. DOI: 10.1186/1471-2180-12-280.

Ibrahim J, Eisen JA, Jospin G, Coil DA, Khazen G, Tokajian S. 2016. Genome Analysis of Streptococcus pyogenes Associated with Pharyngitis and Skin Infections. *PLOS ONE* 11:e0168177. DOI: 10.1371/journal.pone.0168177.

Lintges M, Arlt S, Uciechowski P, Plümäkers B, Reinert RR, Al-Lahham A, Lütticken R, Rink L. 2007. A new closed-tube multiplex real-time PCR to detect eleven superantigens of Streptococcus pyogenes identifies a strain without superantigen activity. *International journal of medical microbiology : IJMM* 297:471–8. DOI: 10.1016/j.ijmm.2007.03.015.

McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B, Bessen DE. 2004. Multilocus sequence typing of Streptococcus pyogenes representing most known emm types and distinctions among subpopulation genetic structures. *Journal of bacteriology* 186:4285–94. DOI: 10.1128/JB.186.13.4285-4294.2004.

McMillan DJ, Drèze P-A, Vu T, Bessen DE, Guglielmini J, Steer AC, Carapetis JR, Van Melderen L, Sriprakash KS, Smeesters PR, Group the MPS. 2013. Updated model of group A Streptococcus M proteins based on a comprehensive worldwide study. *Clinical microbiology and infection : the*

official publication of the European Society of Clinical Microbiology and Infectious Diseases 19:E222-9. DOI: 10.1111/1469-0691.12134.

Molloy EM, Cotter PD, Hill C, Mitchell DA, Ross RP. 2011. Streptolysin S-like virulence factors: The continuing sagA. *Nature Reviews Microbiology* 9:670–681. DOI: 10.1038/nrmicro2624.

Osowicki J, Vekemans J, Kaslow DC, Friede MH, Kim JH, Steer AC. 2018. WHO/IVI global stakeholder consultation on group A Streptococcus vaccine development: Report from a meeting held on 12–13 December 2016. *Vaccine*. DOI: 10.1016/J.VACCINE.2018.02.068.

Reglinski M, Sriskandan S. 2014. The contribution of group A streptococcal virulence determinants to the pathogenesis of sepsis. *Virulence* 5:127–136. DOI: 10.4161/viru.26400.

Sanderson-Smith M, De Oliveira DMP, Guglielmini J, Mcmillan DJ, Vu T, Holien JK, Henningham A, Steer AC, Beall BW, Walker MJ, Parker MW. 2014. 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*:1325–38. DOI: 10.1093/infdis/jiu260.

Steer AC, Carapetis JR, Dale JB, Fraser JD, Good MF, Guilherme L, Moreland NJ, Mulholland EK, Schodel F, Smeesters PR. 2016. Status of research and development of vaccines for Streptococcus pyogenes. *Vaccine* 34:2953–2958. DOI: 10.1016/j.vaccine.2016.03.073.

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. *The Lancet Infectious Diseases* 9:611–616. DOI: 10.1016/S1473-3099(09)70178-1.

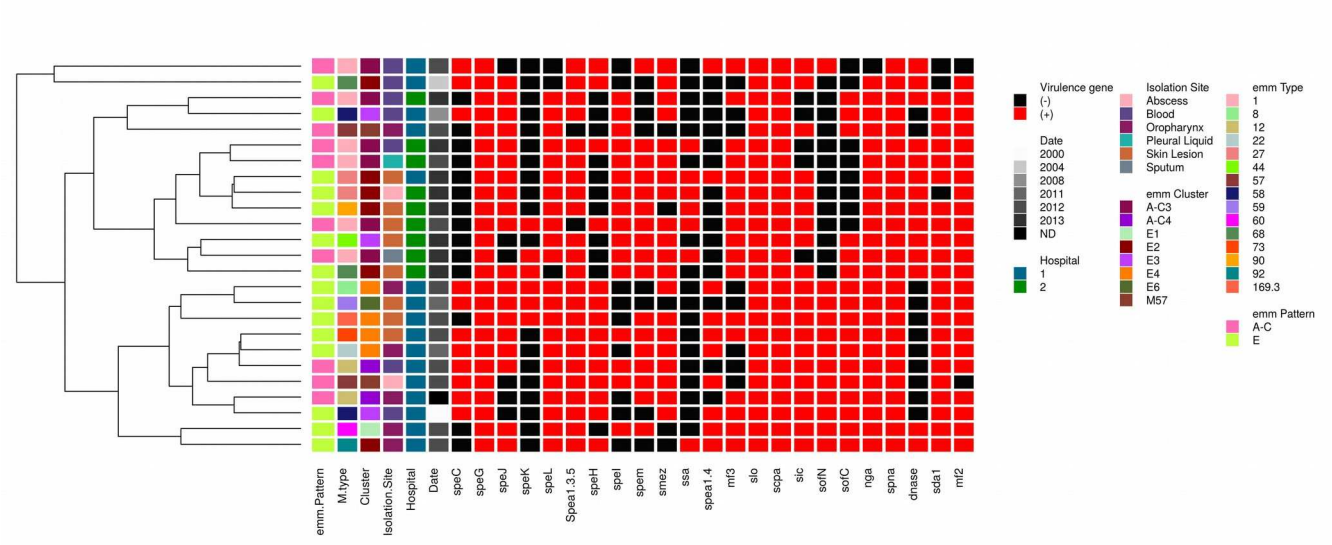
Yang P, Peng X, Zhang D, Wu S, Liu Y, Cui S, Lu G, Duan W, Shi W, Liu S, Li J, Wang Q. 2013. Group a Streptococcus strains circulating during scarlet fever epidemic, Beijing, China, 2011. *Emerging Infectious Diseases* 19:909–915. DOI: 10.3201/eid1906.121020.

245 Table 1. Primers used in this study.

Target	Primer Sequences	Tm (^{0}C)	References
16S rRNA	F: GTGAGTAACGCGTAGGTAACCTACCTCATAG	–	Lintges et. al, 2007
	R: CCCAGGCGGAGTGCTTAATG		
spec	F: GGTAAATTTTTCAACGACACACACATTAAA	74.27	Lintges et. al, 2007
	R: TGTTGAGATTCTCCCGAAATAAATAGAT		
speg	F: GCTATGGAAGTCAATTAGCTTATGCAGAT	74.87	Lintges et. al, 2007
	R: TTATGCGAACAGCCTCAGAGG		
spej	F: CAATTAAATTACGCATACGAAATCATACCAGTA	73.51	Lintges et. al, 2007
	R: ACGAGTAAATATGTACGGAAGACCAAAAATA		
spek	F: TATCGCTTGCTCTATACACTACTGAGAGT	73.27	Lintges et. al, 2007
	R: CCAAACTGTAGTATTTTCATCCGTATTAAA		
spel	F: GGACGCAAGTTATTATGGATGCTCA	73.95	Lintges et. al, 2007
	R: TTAAATAAGTCAGCACCTTCCTCTTTCTC		
spea1-3,5	F: GGTATTTGCTCAACAAGACCCCGAT	77.29	Lintges et. al, 2007
	R: TGTGTTTGAGTCAAGCGTTTCATTATCT		
speh	F: TCTATCTGCACAAGAGGTTTGTGAATGTCCA	78.71	Lintges et. al, 2007
	R: GCATGCTATTAAAGTCTCCATTGCCAAAA		
spei	F: AAGGAAAAATAAATGAAGGTCCGCCAT	75.10	Lintges et. al, 2007
	R: TCGCTTAAAGTAATACCTCCATATGAATTCTTT		
spem	F: GCTTTAAGGAGGAGGAGGTTGATATTTATGCTCTA	62.13	Lintges et. al, 2007
	R: CAAAGTGACTTACTTTACTCATATCAATCGTTTC		
smez	F: CAATAATTTCTCGTCCTGTGTTTGGAT	78.19	Lintges et. al, 2007
	R: GATAAGGCGTCATTCCACCATAG		
ssa	F: AATTATTATCGATTAGTGTTTTTGCAAGTA	62.67	Lintges et. al, 2007
	R: AGCCTGTCTCGTACGGAGAATTATTGAACTC		
spea1-4	F: CAAGAAGTATTTGCTCAACAAGACCCCA	72.28	Lintges et. al, 2007
	R: TTAGATGGTCCATTAGTATATAGTTGCTTGTTATC		
sof_N	F: AGCCTGACACACTTGGTTGGGT	82.89	This study.
	R: GCGGCGCTCAAAATGGTGTGGT		
sof_C	TCGGCGCCTTCGTCAATTTGCT	78.37	This study.
	ACCCCAAGCACAGACAGCTCCA		
nga	F: AGCCTTGTCAGAGGTCACTTTTGG	83.21	This study.
	R: GGCAACCAGGAACGTTGAGCGA		
spna	F: ACAGCAGGTGTTTTGTTTGGACCT	78.04	This study.
	R: ACTGCTTTCTCTTGCATTTACAGCCT		
dnase	F: AGTCCAGCCTGCGGGGTTTTGA	63.15	This study.
	R: ACTGCACTGGCACGACAAACACA		
sda1	F: ACGGAAATGTTCGTAACGGCTACC	78.91	This study.
	R: GTGCATCCGTGTCAGGCTCGTT		
mf2	F: GCACCTGTGTTTAGCCAGGCTGT	81.24	This study.
	R: TGACCCTGTGGGATGGCACAAC		
mf3	F: ACGAAGGCGACTGAGACACCAG	82.60	This study.
	R: TGGCCACGGTCCATTAACCAAGT		
slo	F: TGTCAGCAATGAAGCTCCGCCA	78.23	This study.
	F: GCGGTTTTTCTGGCACACGAGG		
sic	R: TGAGTGGCCTCAGTGGAATGGCT	83.11	This study.
	F: ACGGGTCTTGCCCCCAAAGCTA		
scpa	R: GCTGTCGATCAAGAGCACGGCA	84.20	This study.

246 Primers and it is respective melting temperatures used in the Real-time reactions.

247 Figure 1.



249 Schematic representation of the virulence factors distribution among all 25 clinical isolates of *S. pyogenes*. The tree was
250 built based on the presence/absence of each virulence gene using similarity via the Jaccard method. Also, a representation
251 of the main parameters used to describe the genetic profile of the micro-organisms included in this study.