Original Article

Genetic Diversity and Allelic Variation in South Indian Isolates of Group A Streptococci Causing Invasive Disease

Reena Rajkumari, John Melbin Jose[1], Kootallur Narayanan Brahmadathan[2]

Division of Integrative Biology, School of Bio-Sciences and Technology, VIT University, [1]Department of Microbiology, Christian Medical College, Vellore, [2]Microbiological Laboratory, Coimbatore, Tamil Nadu, India

Abstract

Background: Reported literature on invasive group A streptococcal isolates in India is very scanty. This study was undertaken to determine the molecular heterogeneity of such isolates as seen in a tertiary care center. **Materials and Methods:** Thirty two blood culture isolates and 18 from other sterile body fluids were characterized by emm gene sequencing and multilocus sequence typing. **Results:** Forty two emm types were identified including 25 from 32 blood isolates and 17 from 18 other body fluid isolates. Types 110, 74, 63, 85, 102, 105, 124 and st854.1 were common to both groups and accounted for 40% of the isolates. Two types namely, stKNB6 and stKNB9 were newly identified types. MLST identified forty eight sequence types (MLST - ST) of which 31 were from 32 blood isolates and 17 from 18 body fluid isolates; thirty three of them were hitherto unrecognized at the time of identification. Two blood isolates of emm 85 had the same MLST - ST 484 while three blood isolates of emm 110 had three different STs namely, ST 493, 494 and 497. Two types, ST 493 and ST497 had single locus variation while ST 497 had a double locus variation. **Conclusions:** Our study shows that subtle allelic variations in the house keeping genes results in the development of new strains in a given emm type and contribute significantly to the existing high diversity of strains circulating in the community.

Keywords: Allelic variation, emm types, invasive Group A streptococcus, multilocus sequence typing – sequence type, molecular epidemiology

Introduction

Group A streptococcus (GAS), also known as *Streptococcus pyogenes,* is a significant global pathogen capable of causing uncomplicated primary infection, debilitating post-streptococcal sequelae and fatal invasive disease.[1] While non-suppurative complications are major problems in the developing world, invasive GAS (iGAS) disease is of concern in the Western countries.[2,3] Although timely antibiotic prophylaxis can prevent later complications, non-compliance to treatment is an issue in their prevention in many countries of the world.[4] Therefore, development of an appropriate vaccine is an alternate choice for the control and prevention of both non-suppurative sequelae as well as iGAS disease.[5]

Typing of GAS helps in the identification of diversity or clonality of GAS strains which will help in the design of an appropriate GAS vaccine. For many years, this was done by conventional M typing based on antigenicity of the hypervariable region of M protein molecule. This was cumbersome because of difficulty in preparing M antisera and was later replaced by sequencing of *emm* gene through amplification of the hypervariable region of this gene.[6] Using this method, >200 emm types have been identified globally till date.[7] Since M protein induces type-specific protective immunity, vaccines based on this molecule have attracted a lot of attention. Identification of their distribution will help to determine if such M-based vaccine will be effective in different populations.[5]

Multilocus sequence typing (MLST) is a method to study genetic variation of bacteria by studying allelic variations in housekeeping genes.[8] In case of GAS, this is done through amplification of internal fragments of the seven housekeeping genes that code for enzymes necessary for the viability of the cell.[9] Following amplification, the sequences are aligned,

Access this article online

Quick Response Code:

Website: www.ijmm.org

DOI: 10.4103/ijmm.IJMM_17_298

Address for correspondence: Dr. Kootallur Narayanan Brahmadathan, Microbiological Laboratory, 12A Cowley Brown Road, R. S. Puram, Coimbatore - 641 002, Tamil Nadu, India.
E-mail: brahmadathan1979@gmail.com

How to cite this article: Rajkumari R, Jose JM, Brahmadathan KN. Genetic diversity and allelic variation in south Indian isolates of Group A streptococci causing invasive disease. Indian J Med Microbiol 2017;35:575-9.

analysed, edited and trimmed exactly to the respective allele of each primer. Single locus query or multiple locus options available at http://spyogenes.mlst.net determines an allele number for each sequence trimmed and pasted for all seven loci of a test strain into the corresponding boxes. The sequence of each fragment is compared with all previously identified sequences at that locus available in the gene bank. If the sequence corresponds to a known allele, a number is given to that sequence. If the sequence is a new allele, it will be compared with the most similar allele for that locus to check for nucleotide differences and if new, forward and reverse sequences are submitted to the database at d.godoy@imperial.ac.uk. If new sequence is found, a new number is given and added to the database. The combination of seven allele number forms the allelic profile of the strain. Each profile is known as a sequence type (ST). The MLST database is available at http://www.mlst.net.

Materials and Methods

Fifty iGAS isolates including 32 blood culture isolates and 18 from other sterile body fluids were selected for the study. The latter included isolates from peritoneal fluid (n = 6), cerebrospinal fluid (n = 5), pleural fluid (n = 4) and one each from fluid from necrotising fasciitis, bile fluid and synovial fluid. Invasive isolates were defined as those recovered in pure culture from blood and other sites which are otherwise sterile.

Isolates were subjected to emm typing and MLST by amplification techniques. The emm typing was carried out by the Centers for Disease Control (CDC) protocol and as standardised in our laboratory.[6] Briefly, multiple colonies were immersed in 300 µl of normal saline, centrifuged and suspended in 50 µl of TE containing 300 IU of mutanolysin and 30ug/ml of hyaluronidase. The Final lysate was prepared after incubation at 37°C for 30 min and heating at 100°C for 10 min. The extracted DNA was amplified using 20 µl master mix (TAQPCR CORE KIT, Qiagen, Hilden, Germany) that contained 0.4 µl each of forward (TATT(C/G)GCTTAGAAAATTAA) and reverse (GCAAGTTCTTCAGCTTGTTT) primers (Sigma-Aldrich, Bangalore). After pre-and post-sequencing clean-up, DNA was subjected to sequencing with Big Dye Terminator Kit in an ABI prism 310 automated sequencer (Applied Biosystems, Warrington, UK). The emm gene sequence was searched for homology by BLAST search analysis (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi) through CDC website (http://www.cdc.gov/ncidod/biotech/strep/strepblast.htm). Strains showing >95% sequence homology with the reference strain in the CDC Gene Bank database were selected and designated particular parental emm type. For subtype assignment, database of trimmed 180 base entries corresponded to the first 50 residues of the mature M protein and the adjacent 10 C terminal residues of the signal sequence. If a perfect 180/180 match was obtained to an entry from the type-specific BLAST option, the subtype was reported to be correctly identified. If a perfect match to bases 31–180 is combined with 3 or fewer mismatches to bases 1–30 was found, this also indicated identification of the specific subtype. If there was any mutation in the sequence corresponding to the first 50 residues of the mature protein, it was considered as a new subtype.

MLST was carried out using primers for seven housekeeping genes [Table 1].[10] They are internal fragments of the following enzymes whose primers are as follows (Sigma-Aldrich, Bangalore).

DNA extraction was done using a commercial kit (Ql Aamp DNA BLOOD mini kit, Qiagen, Hilden, Germany) as described above. Fifty µl of master mix was prepared from the extracted DNA was used for amplification using each of the above seven sets of primers using the commercial kit (TAQ PCR CORE KIT, Qiagen, Hilden, Germany) as described above. The PCR parameters were 95°C for 5 min, followed by 95°C for 1 min, 55°C for 1 min, 72°C for 1 min and 72°C for 5 min for 28 cycles. The amplified products were prepurified and checked for purity by gel electrophoresis. The sequence of each fragment was obtained using the respective primers used in the initial amplification. As stated earlier, the ST was identified at http://spyogenes.mlst.net, d.godoy@imperial.ac.uk. and http://www.mlst.net. Strains with new alleles and ST were submitted to the database and given new numbers and STs. This study was approved by the Institutional Research Committee of the Christian Medical College, Vellore, Tamil Nadu.

Results

Forty-two emm types were identified including twenty-five from 32 blood culture isolates and 17 from 18 other sterile body fluid isolates [Table 2]. Eight emm types, namely, 110,

Table 1: Primer sequences for seven housekeeping genes

Enzyme (gene)	Primer sequence (up)	Primer sequence (down)
gki	GGCATTGGAATGGGATCACC	TCTCCTGCTGCTGACAC
gtr	GAGGTTGTGGTGATTATTGG	GCAAAGCCCATTTCATGAGTC
murI	TGCGACTCAAAATGTTAAAATGATTG	GATGATAATTCACCGTTAATGTCAAAATAG
mutS	GAAGAGTCATCTAGTTTAGAATACGAT	AGAGAGTTGTCACTTGCGCGTTTGATTGCT
recP	GCAAATTCTGGACACCCAGG	CTTTCACAAGGATATGTTGCC
xpt	TTACTTGAAGAACGCATCTTA	ATGAGGTCACTTCAATGCCC
Yqil	TGCAACAGTATGGACTGACCAGAGAACAAGATGC	CAAGGTCTCGTGAAACCGCTAAAGCCTGAG

gki: Glucose kinase, *gtr*: Glutamine transporter protein, *murI*: Glutamate racemase, *mutS*: DNA mismatch repair protein, *recP*: Transketolase, *xpt*: Xanthine phosphoribosyl transferase, *Yqil*: Acetyl-coA transferase

Table 2: Distribution of emm and multilocus sequence typing sequence types among invasive group A streptococcus isolates

emm type	Source	MLST-ST	emm type	Source	MLST-ST
8	Blood	470*	110 (4)	Blood	522*
				Blood	493*
				Blood	494*
				Bile fluid	515*
18	Blood	523*	112	Blood	497*
43.3	Blood	475*	118	Blood	511*
44	Blood	476*	124 (2)	Blood	544*
				CSF	544*
49.4	Blood	478*	st1731	Blood	504*
53	Blood	363	st6735 (2)	Blood	506*
				Blood	507*
55	Blood	100	st854.1 (3)	Blood	519*
				Peritoneal fluid	224
				Peritoneal fluid	224
63 (2)	Blood	338	stKNB6**	Blood	517*
	Peritoneal fluid	338			
74 (3)	Blood	480*	60	CSF	193
	Blood	520*			
	CSF	480*			
81.2	Blood	521*	66	CSF	249
85 (3)	Blood	484*	119	CSF	237
	Blood	484*			
	Pleural fluid	514*			
86 (2)	Blood	518*	25	Peritoneal fluid	350
	Blood	503*			
93	Blood	502*	28	Peritoneal fluid	473*
100 (2)	Blood	486*	stKNB9**	Peritoneal fluid	114
	Blood	487*			
102 (2)	Blood	489*	15	Pleural fluid	513*
	Pleural fluid	489*			
104	Blood	353	56	Pleural fluid	512*
105 (2)	Blood	490*	58.8	Synovial fluid	516*
	Fluid from necrotising fasciitis	151			

Numbers in parentheses indicate number of strains. *Newly recognised MLST-STs, **Newly recognised emm types. MLST: Multilocus sequence typing, ST: Sequence type

74, 63, 85, 102, 105, 124 and st854.1 were common to both groups and accounted for 42% of the isolates. Thus, in all, 34 emm types were identified from 50 iGAS isolates indicating extreme diversity among them. Two types, namely, stKNB6 and stKNB9 were hitherto unrecognised types at the time of their identification.

Forty-eight ST were identified among 50 isolates that included 31 from 32 blood isolates and 17 from 18 sterile site isolates. MLST-STs 544, 338 and 489 were common to both groups; thus 45 ST were identified among 50 isolates that represented 34 emm types. Multiple ST was identified among eight emm types. Thus, four ST were identified in emm110, three ST in emm 74 and two among emm 85, 86, 100, 105, st854.1 and st6735 each. Thirty-seven of 48 ST were hitherto unrecognised at the time of identification.

Strain variations among eight emm types were studied with respect to their alleles in the housekeeping genes [Table 3]. Double locus variations (DLV) were seen in emm 85 and 110 with respect to glutamine transporter protein *(gtr)* and Acetyl-coA transferase and emm 86 with respect to *gtr* and DNA mismatch repair protein. Other five showed single locus variations (SLV) which were all respect to *gtr*. It is interesting that in all eight strains, variation included that of *gtr* gene.

Discussion

Our study shows extensive diversity among iGAS isolates and reflects the circulation of a wide variety of GAS strains that are capable of causing invasive disease in this population. Identification of 25 emm types among 32 blood isolates and 17 types among 18 isolates from other sterile sites shows complete lack of clonality among them. Such heterogeneity is a unique feature of GAS isolates from developing countries.[7] In a study on GAS invasive disease in Fiji islands, Steer *et al*.[7] reported 38 different emm types among 55 iGAS isolates seen in 64 cases during a 2-year period. This is

Table 3: Allelic variations among different emm types

emm type	Source	ST	*gki*	*Gtr*	*Muri*	*muts*	*recp*	*Xpt*	*yqil*
74	Blood	480	43	2	2	7	1	3	80
	Blood	520	43	32	2	7	1	3	1
	CSF	480	43	2	2	7	1	3	80
85	Blood	484	42	2	8	22	33	3	4
	Blood	484	42	2	8	22	33	3	4
	Pl.fluid	514	42	37	8	22	33	3	63
86	Blood	518	4	12	36	11	4	23	1
	Blood	503	4	2	36	5	4	23	1
100	Blood	486	2	12	8	3	2	32	1
	Blood	487	2	3	8	3	2	32	1
105	Blood	490	2	31	2	17	4	5	1
	Necrotising fasciitis fluid	151	2	28	2	17	4	5	1
110	Blood	493	25	37	8	6	30	25	4
	Blood	494	25	4	8	6	30	25	4
	Blood	497	25	3	8	6	30	25	4
	Bile fluid	515	25	2	8	6	30	25	39
st6735	Blood	506	105	6	2	6	1	49	4
	Blood	507	105	3	2	6	1	49	4
st854.1	Blood	519	4	2	54	6	34	8	44
	Peritoneal fluid (2)	224	4	3	54	6	34	8	44

Highlighted numbers indicate single and double locus variations. *gki*: Glucose kinase, *gtr*: Glutamine transporter protein, *murI*: Glutamate racemase, *mutS*: DNA mismatch repair protein, *recP*: Transketolase, *xpt*: Xanthine phosphoribosyl transferase, *Yqil*: Acetyl-coA transferase

in contrast to reports from Western countries where iGAS isolates are usually restricted to a few types.[10] The only other Indian study involving invasive disease had shown similar diversity with significant variations in their emm type distribution between south and north Indian isolates.[3] Significant differences in the emm type distribution was also highlighted between developing and developed countries in a study published in 2009 implying that multivalent M type-based vaccine will not be equally effective in different regions of the world.[7,11] The 30-valent Dale vaccine based on emm types that account for 98% of all cases of pharyngitis in the US and Canada, 90% of invasive disease in the US and 78% of invasive disease in Europe evoke bactericidal antibodies against all 30 vaccine serotypes.[5] A study reported in 2009 observed that there was a higher diversity of strains in lower to middle-income settings as compared to high-income settings and that the theoretical coverage of a multivalent vaccine would be favourable in developed countries (>72%) than regions with more serious GAS disease; for example Africa 39% and Pacific 24%.[11] Of 34 emm types identified in our study, 7 (20.6%) were represented in the 30-valent Dale vaccine and nine (27.3%) were represented in the 33 non-vaccine-types.[12] Thus in total, 16 (47.06%) emm types in our study would be represented in both groups leaving >50% of the types uncovered by the current 30-valent vaccine. Recently, a classification based on 48 emm-clusters containing closely related M proteins that share binding and structural properties has been proposed.[13] It also proposed that the cross-protection observed in the Dale vaccine occurs within these 48 emm-clusters. If that is so, there is hope for a broadly effective vaccine that will cover larger number of M types and therefore more effective in many more regions globally.

MLST is a very useful nucleotide sequence-based method to study genetic relationships between organisms of a bacterial species.[14] Since it is associated with sequences, it gives unambiguous results and is easily portable between laboratories.[15] Housekeeping genetic sequences are used for analysis because they are present in every organism and their products serve vital functions in the cell. Further, mutations within them are largely believed to be selectively neutral. It would also help to identify the nature and magnitude of development of new strains in a bacterial population. MLST have been used for epidemiological typing of a variety of pathogenic microbes.[16] Considering that one need to use seven forward and an equal number of reverse primers, MLST can be an expensive technique in resource crunch situations, probably showing the very few studies reported on MLST for any bacterium. To the best of our knowledge, this is the first Indian report on MLST profile of GAS strains. Forty-five STs were identified among 34 emm types showing that MLST is a good indicator of strain variation among GAS strains. Eight of the 34 emm types showed >1 ST's with emm 110 showing the maximum variation in the MLST profile [Table 2]. Three of the four emm 110 isolates were blood culture isolates, all with different MLST profiles showing that they were three different strains. Of the nineteen ST identified among eight emm types, 6 had DLV while the remaining 13 had only SLV. Thus, strain variations among emm types are predominantly due to subtle variations with respect to one allele. This may be due to minor but varying host environmental conditions

that the bacterium encounters *in vivo*. One incidental finding is that all SLV and DLV involved the gene *gtr* which codes for *gtr* that is essential for transport of this amino acid across cell membrane. This finding may be significant because glutamine is involved in nitrogen/ammonium transport across membrane and is essential for the viability of any bacterial cell.[17]

Conclusion

Our study highlights the extensive diversity of South Indian iGAS isolates and the molecular mechanism behind it. The MLST profiles and the allelic variations show how strain variations occur among GAS population. In the context of relatively low rates of recombination among GAS strains, the high rate of SLVs is probably due to point mutations arising out of subtle but hostile host environment. Development of large number of closely related strains resulting in high diversity has direct implications while designing candidate vaccines.

Acknowledgement

This study formed a part of the PhD thesis of RR and was conducted at the Department of Clinical Microbiology, Christian Medical College and Hospital, Vellore, Tamil Nadu, during 2004–2007. We wish to thank the staff of the department for their full support and cooperation for the work. Financial support from the ASSIST Program of the European Union during 2004–2007 and Senior Research Fellowship for RR from ICMR are gratefully acknowledged.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. Infect Immun 2001;69:2416-27.
2. Shah B, Sharma M, Kumar R, Brahmadathan KN, Abraham VJ, Tandon R, *et al.* Rheumatic heart disease: Progress and challenges in India. Indian J Pediatr 2013;80 Suppl 1:S77-86.
3. Haggar A, Nerlich A, Kumar R, Abraham VJ, Brahmadathan KN, Ray P, *et al.* Clinical and microbiologic characteristics of invasive *Streptococcus pyogenes* infections in North and South India. J Clin Microbiol 2012;50:1626-31.
4. Rolston DD, Brahmadathan KN, Koshi G, Cherian G. Patient compliance with prophylactic Benzathine penicillin for rheumatic fever. Med J Aust 1981;2:160-1.
5. Dale JB, Batzloff MR, Cleary PP, Courtney HS, Good MF, Grandi G, *et al*. Current approaches to Group A streptococcal vaccine development. In: Ferretti JJ, Stevens DL, Fischetti VA, editors. *Streptococcus pyogenes* Basic Biology to Clinical Manifestations. Oklahoma City (OK), USA: University of Oklahoma Health Sciences Center; 2016. p. 938-84.
6. Beall B, Facklam R, Thompson T. Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci. J Clin Microbiol 1996;34:953-8.
7. Steer AC, Carapetis JR, Dale JB, Fraser JD, Good MF, Guilherme L, *et al.* Status of research and development of vaccines for *Streptococcus pyogenes*. Vaccine 2016;34:2953-8.
8. Spratt BG. Multilocus sequence typing: Molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Curr Opin Microbiol 1999;2:312-6.
9. McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B, *et al.* Multilocus sequence typing of *Streptococcus pyogenes* representing most known emm types and distinctions among subpopulation genetic structures. J Bacteriol 2004;186:4285-94.
10. Luca-Harari B, Darenberg J, Neal S, Siljander T, Strakova L, Tanna A, *et al.* Clinical and microbiological characteristics of severe *Streptococcus pyogenes* disease in Europe. J Clin Microbiol 2009;47:1155-65.
11. 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. Lancet Infect Dis 2009;9:611-6.
12. Dale JB, Fischetti VA, Carapetis JR, Steer AC, Sow S, Kumar R, *et al.* Group A streptococcal vaccines: Paving a path for accelerated development. Vaccine 2013;31 Suppl 2:B216-22.
13. Sanderson-Smith M, De Oliveira DM, Guglielmini J, McMillan DJ, Vu T, Holien JK, *et al.* A systematic and functional classification of *Streptococcus pyogenes* that serves as a new tool for molecular typing and vaccine development. J Infect Dis 2014;210:1325-38.
14. Maiden MC. Multilocus sequence typing of bacteria. Annu Rev Microbiol 2006;60:561-88.
15. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, *et al.* Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 1998;95:3140-5.
16. Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, *et al.* MLST revisited: The gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 2013;11:728-36.
17. Kleiner D. Bacterial ammonium transport. FEMS Microbiol Rev 1985;32:87-100.