Clinical Microbiology and Infection 26 (2020) 946.e5–946.e8

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Research note

Updated *emm*-typing protocol for *Streptococcus pyogenes*

H.R. Frost [1, 2], M.R. Davies [3], S. Velusamy [4], V. Delforge [1], A. Erhart [5], S. Darboe [5], A. Steer [2, 6], M.J. Walker [7], B. Beall [4], A. Botteaux [1], P.R. Smeesters [1, 2, 6, 8, *]

[1)] *Molecular Bacteriology Laboratory, Université Libre de Bruxelles, Brussels, Belgium*
[2)] *Tropical Diseases Research Group, Murdoch Children's Research Institute, Melbourne, Australia*
[3)] *Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia*
[4)] *National Centre for Immunization and Respiratory Diseases, Centres for Disease Control and Prevention, Atlanta, GA, USA*
[5)] *Medical Research Council The Gambia at the London School of Hygiene and Tropical Medicine, Fajara, Gambia*
[6)] *Department of Paediatrics, University of Melbourne, Melbourne, Australia*
[7)] *Australian Infectious Diseases Research Centre and School of Chemistry and Molecular Biosciences, University of Queensland, St Lucia, Australia*
[8)] *Academic Children's Hospital Queen Fabiola, Université Libre de Bruxelles, Brussels, Belgium*

ARTICLE INFO

Article history:
Received 10 January 2020
Received in revised form
12 February 2020
Accepted 19 February 2020
Available online 28 February 2020

Editor: G. Lina

Keywords:
emm-typing protocol
Group A Streptococcus
M protein
M-like protein
PCR
Streptococcus pyogene
Vaccine

ABSTRACT

Objectives: PCR-based typing of the *emm* gene *Streptococcus pyogenes* often results in the amplification of multiple bands. This has resulted in the misclassification of strains into types based on non-*emm* gene sequences. We aimed to improve the specificity of the *emm* typing PCR reaction using a primer called CDC3, the sequence for which has been previously used to identify *emm* genes *in silico*.
Methods: The proposed primer CDC3 was validated *in silico* from a global database of 1688 GAS genomes and *in vitro* with 32 isolates. PCR reactions were performed on genomic DNA from each isolate, using the published CDC1 forward primer with the CDC2 reverse primer or the new CDC3 reverse primer. The products were examined by gel electrophoresis, and representative PCR products were sequenced.
Results: In 1688 *S. pyogenes* genomes, the previous CDC2 reverse primer annealed *in silico* in 1671 *emm* genes and also in 2109 non *emm* genes in close proximity, whereas the new CDC3 primer annealed in 1669 *emm* genes only. The remaining 19 genes without a CDC3 binding site were chimeric *emm* genes. The PCR pair CDC1+CDC3 produced a single band at appropriate molecular weight in all 32 isolates tested, while the CDC1+CDC2 pair produced more than one band in 13 of 32 isolates (40%).
Conclusions: The new CDC3 primer is more specific for *emm* genes than the previous CDC2 primer and represents a simple solution to reduce the potential for mistyping *S. pyogenes* strains. **H.R. Frost, Clin Microbiol Infect 2020;26:946.e5–946.e8**

Introduction

Streptococcus pyogenes is a leading global bacterial pathogen which is categorized into *emm*-types by sequencing of the hypervariable 5′ end of the *emm* gene [1]. The CDC1 and CDC2 primers were first described by Whatmore et al. in 1994 [2] and the PCR-derived sequences associated with previously defined M serotypes by Beall et al. in 1996 [3]. This technology has greatly facilitated *S. pyogenes* typing and has been used in hundreds of epidemiological studies, identifying over 240 different *emm*-types [4,5]. The worldwide database of *emm*-type specific sequences is hosted and curated by the U.S. Centres for Disease Control and Prevention (CDC). Sequence data is generated either by PCR amplification and sequencing [3] using primers located in conserved regions of the signal peptide and the 3′ end of the *emm* gene [2], or derived from *de novo* assembly of whole genome sequencing (WGS) information [6].

Accurate *emm*-typing may not be achieved because of overlap in sequence between the *emm* genes and closely related *emm*-like genes called *mrp enn*, and *sph* [7]. The 3′ sequences of *emm*-like genes are sufficiently similar to *emm* genes such that the *emm* typing CDC 2 is able to anneal and facilitate amplification from these genes as well as *emm*. The non-specificity of primer CDC2 leads to observed double bands in gel electrophoresis and can

* Corresponding author. P. Smeesters, Molecular Bacteriology Laboratory, Route de Lennik 808, B-1070 Brussels, Belgium.
E-mail address: psmeeste@ulb.ac.be (P.R. Smeesters).

https://doi.org/10.1016/j.cmi.2020.02.026
1198-743X/

result in poor-quality sequencing or non-typeable strains. A recent large global WGS study enabled better definition of *emm* and *emm*-like genes [8,9]. Using this global database, we now propose an updated *emm*-typing protocol, using a new and more specific primer to replace the 3′-situated CDC2 for more effective and accurate typing of *S. pyogenes* strains globally.

Methods

Template preparation

The 32 isolates used to validate the new *emm*-typing protocol were isolated from five countries across four continents, and belonged to 30 different *emm*-types and 18 different *emm*-clusters [10,11] (Table 1). DNA was collected using a commercial DNA extraction kit (Qiagen), with an additional lysis step. Briefly, *S. pyogenes* colonies were inoculated in 5 mL Todd Hewitt Broth with 5% yeast extract (THY) and grown overnight at 37°C. Pellets were collected by centrifugation and resuspended in 0.2 mL enzymatic lysis buffer (Tris HCl, EDTA, Triton-X and lysozyme) and incubated at 37°C for 30 min. Lysates were mixed with 0.2 mL of the extraction kit lysis buffer plus 0.025 mL proteinase K, and incubated for a further 1 h at 56°C. After combining lysates with 0.2 mL absolute ethanol, the samples were added to spin columns and the extractions performed as per manufacturer's instructions.

PCR

PCR was performed in 25 μL total volume using GoTaq (Promega) polymerase and buffer, containing 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 10 nM forward and reverse primers, 100 ng DNA template and sterile, distilled water. The forward primer CDC1 (5′-TATTSGCTTAGAAAATTAA-3′) was used with reverse primers CDC2 (5′-GCAAGTTCTTCAGCTTGTTT-3′) [2] and the proposed new reverse primer, CDC3 (5′-TTCTTCAAGCTCTTTGTT-3′). A modified version of Primer 3 2.3.7 [12] was used to predict CDC3 primer binding sites and properties using default *in silico* PCR parameters. The following thermal cycling conditions were applied: denaturation at 94°C for 1 min, 30 cycles of 94°C for 15 s, annealing at 47°C for 30 s, extension at 72°C for 1 min 25 s, and final elongation at 72°C for 7 min. PCR products were visualized on 1% agarose gels stained with SafeRed (Carl Roth) dye with migration at 200V for 20 min.

Sequence analysis

The PCR products of 10 isolates, amplified with both the CDC1+CDC2 and CDC1+CDC3 primer pairs, were purified and sequenced using cycle sequencing on an Applied Biosystems 3730XL instrument by Eurofins genomics (https://www.eurofinsgenomics.eu/en/home/) with the emmseq2 primer (5′-TATTCGCTTAGAAAATTAAAAACAGG-3′). Sequences were analysed with the Streptococci Group A Subtyping Request form Blast 2.0 server hosted on the CDC website (https://www2a.cdc.gov/ncidod/biotech/strepblast.asp).

In silico analysis of PCR specificity

We used the Geneious Prime software package (Version 2020.0) to detect the primer sequences in a global WGS dataset of 1688

Table 1
Strains used for verification of the novel *emm* typing primer

Strain ID	Country	*emm*-type	*emm*-cluster	Number of bands on gel		PCR products sequenced
				CDC1+CDC2	CDC1+CDC3	
4235	Belgium	*M1	A-C3	1	1	Yes
4164	Belgium	M2	E4	1	1	
4152	Belgium	M3.1	A-C5	1	1	
SP115	Belgium	M6	M6	1	1	
4081	Belgium	M11	E6	1	1	
G147	Gambia	*M18.7	M18	3	1	
4048	Belgium	M12	A-C4	1	1	
I66	Brazil	M19.4	M19	1	1	
G28	Gambia	M25.1	E3	2	1	Yes
I108	Brazil	M28	E4	1	1	
NS13	Australia	M53	D4	1	1	
G236	Gambia	*M57	M57	2	1	
I75	Brazil	M58	E3	1	1	
G206	Gambia	*M65.4	E6	2	1	Yes
NS80	Australia	M70	D4	2	1	
G84	Gambia	M71.1	D2	2	1	Yes
4096	Belgium	M75	E6	1	1	
I29	Brazil	M82	E3	1	1	
G408	Gambia	M85.1	E6	1	1	Yes
SP18	Belgium	*M87	E3	2	1	Yes
4100	Belgium	M89	E4	1	1	
NS730	Australia	M90	E2	1	1	
G22	Gambia	*M95	M95	3	1	Yes
NS88.1	Australia	M98	D4	1	1	
NS192	Australia	M106	E2	1	1	
SP14	Belgium	*M108.1	D4	2	1	Yes
G391	Gambia	*M122	M122	2	1	Yes
G447	Gambia	*M164	M164	2	1	
I100	Brazil	M184	D5	1	1	
31126	Fiji	M222	M222	2	1	
G157	Gambia	*M225	E6	1	1	Yes
33141	Fiji	M238	A-C3	2	1	

* Agarose gels of PCR products are displayed in Fig. 1.

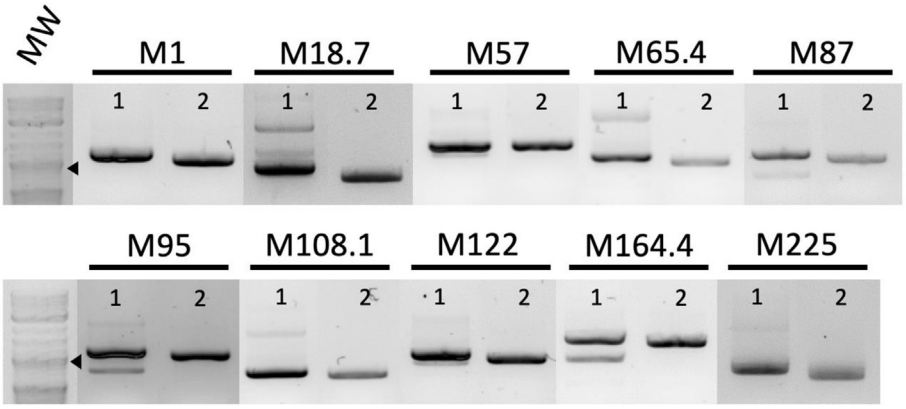

Fig. 1. Amplification of the *emm* typing sequence using CDC1+CDC2 primers (1) or CDC1+CDC3 primers (2) for 10 of the 32 strains. The 1 kb Plus DNA ladder (NEB) indicates sizes were variable but typically around 1 kb in length (indicated by black arrowheads).

genomes [8,9]. We allowed one mismatch in the primer annealing site except for in the 3′ terminal position, conditions typically permissive in PCR reactions [13].

Results

Overall, the combination of primers CDC1+CDC3 had the most consistent amplification of a single band at an appropriate molecular weight for amplification of the *emm* gene (Fig. 1). Of the 32 isolates, the CDC1+CDC3 primer combination produced only products with a single band, while CDC1+CDC2 produced 13 products with more than 1 band (Fig. 1, Table 1). CDC1+CDC3 was equally as effective at amplifying *emm* genes from strains containing no *emm*-like genes (e.g., M1, M3, M6, M12) as from strains containing the trio of *emm* and *emm*-like genes (e.g., M70, M65, M87). The identification of the correct *emm*-type by sequencing was identical between the two primer pairs assessed.

In silico analysis of the 1688 previously described *S. pyogenes* genomes [8,9] revealed that the new CDC3 primer is markedly more specific to *emm* genes than the CDC2 primer. The CDC2 primer sequence theoretically annealed in 1671 *emm* genes, but also in 658 *enn* genes, 1422 *mrp* genes and 29 *sph* genes. All of these genes are located immediately upstream or downstream of the *emm* gene and have a high degree of similarity in their signal peptide sequences. By contrast, the CDC3 primer sequence theoretically annealed in 1669 *emm* genes and no *emm*-like genes. The 19 and 21 *emm* genes that did not contain either the CDC2 or CDC3 primer sequence, respectively, were chimeric *emm* genes that contain the 3′ end of an *enn* gene [9].

Discussion

Our data support the proposal of an improved *emm*-typing protocol for *S. pyogenes*. The sequence of the proposed new CDC3 primer is based on a conserved and specific region within all *emm* genes that is not present in *emm*-like genes. The sensitivity of the CDC3 primer is equal to that of the previous CDC2 primer, however the specificity is greater, thereby improving sequencing quality and strain identification. The new *emm* typing protocol is available on the CDC website for broad dissemination.

Despite the increase in utilization of WGS for *emm* typing in some reference laboratories, PCR typing of *S. pyogenes* is routinely performed in both high- and low-income settings due to ease of use and low cost. Importantly, in low-income settings *S. pyogenes* is typically more genetically diverse [14,15], and the carriage of *emm*-like genes is more common [8,9]. Thus, we included isolates collected from five countries across four continents to validate this novel PCR assay. We found the CDC1+CDC3 primer combination the best for strains in all settings. Limitations of this study include the use of only 32 representative strains for validation and the need to determine the proportion of *S. pyogenes* strains that will remain non-typeable with the updated protocol.

Replacement of the CDC2 primer with the more *emm* specific CDC3 primer will improve accuracy of *S. pyogenes* typing. As *emm* typing is the primary epidemiologic marker for *S. pyogenes*, accurate documentation of the *emm* type is essential for public health diagnostics, outbreak tracking and predicting vaccine efficacy.

Transparency declaration

The authors do not declare any conflicts of interest. This work was supported by the Belgian Fonds National de la Recherche Scientifique research grants (PDR T.0255.16 and CDR J.0019.17). H.R.F. is supported by a Ph.D. fellowship granted by the Royal Academy of Medicine from Belgium and sponsored by GSK Vaccines. The funders had no role in study design and analysis.

Author contributions

Writing—original draft: H.R.F. and P.R.S. Writing—review & editing: M.R.D., S.V., V.D., A.E., S.D., A.S, M.J.W., B.B. and A.B. Conceptualization: P.R.S. Investigation: H.R.F., V.D. and A.B. Data curation and analysis: H.R.F., M.R.D., S.V., V.D., A.E., S.D., A.S, M.J.W., B.B., A.B. and P.R.S. Funding Acquisition and Supervision: A.B. and P.R.S.

References

[1] Botteaux A, Budnik I, Smeesters PR. Group A Streptococcus infections in children: from virulence to clinical management. Curr Opin Infect Dis 2018;31:224–30.
[2] Whatmore AM, Kehoe MA. Horizontal gene transfer in the evolution of group A streptococcal emm-like genes: gene mosaics and variation in Vir regulons. Mol Microbiol 1994;11:363–74.
[3] 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.
[4] Mcmillan DJ, Dreze PA, Vu T, Bessen DE, Guglielmini J, Steer AC, et al. Updated model of group A Streptococcus M proteins based on a comprehensive worldwide study. Clin Microbiol Infect 2013;19:E222–9.
[5] Bessen DE, Smeesters PR, Beall BW. Molecular epidemiology, ecology, and evolution of group A Streptococci. Microbiol Spectr 2018;6. https://doi.org/10.1128/microbiolspec.CPP3-0009-2018.
[6] Chochua S, Metcalf BJ, Li Z, Rivers J, Mathis S, Jackson D, et al. Population and whole genome sequence based characterization of invasive group A Streptococci recovered in the United States during 2015. MBio 2017;8:1–19.
[7] Frost HR, Sanderson-Smith M, Walker M, Botteaux A, Smeesters PR. Group A streptococcal M-like proteins: from pathogenesis to vaccine potential. FEMS Microbiol Rev 2018;42:193–204.

[8] Davies MR, Mcintyre L, Mutreja A, Lacey JA, Lees JA, Towers RJ, et al. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. Nat Genet 2019;51:1035–43.
[9] Frost HR, Davies MR, Delforge V, Lakhloufi D, Sanderson-Smith M, Srinivasan V, et al. Analysis of global collection of group A *Streptococcus* genomes reveals that the majority encode a trio of M and M-like proteins. mSphere 2020;5:e00806–19.
[10] 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.
[11] Frost HR, Laho D, Sanderson-Smith ML, Licciardi P, Donath S, Curtis N, et al. Immune cross-opsonization within emm clusters following group A Streptococcus skin infection: broadening the scope of type-specific immunity. Clin Infect Dis 2017;65:1523–31.
[12] Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res 2012;40:e115.
[13] Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, et al. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic Acids Res 1990;18:999–1005.
[14] Smeesters PR, Mardulyn P, Vergison A, Leplae R, Van Melderen L. Genetic diversity of Group A Streptococcus M protein: implications for typing and vaccine development. Vaccine 2008;26:5835–42.
[15] Smeesters PR, Dramaix M, Van Melderen L. The emm-type diversity does not always reflect the M protein genetic diversity—is there a case for designer vaccine against GAS? Vaccine 2010;28:883–5.