Clinical Infectious Diseases

MAJOR ARTICLE

The Limitations of the Rheumatogenic Concept for Group A Streptococcus: Systematic Review and Genetic Analysis

Gabrielle de Crombrugghe,[1,2] Noemie Baroux,[3] Anne Botteaux,[2] Nicole J. Moreland,[4] Deborah A. Williamson,[5] Andrew C. Steer,[3,6] and Pierre R. Smeesters[1,2,3,6]

[1]Academic Children Hospital Queen Fabiola, and [2]Molecular Bacteriology Laboratory, Université libre de Bruxelles, Brussels, Belgium; [3]Tropical Diseases Research Group, Murdoch Children's Research Institute, Melbourne, Australia; [4]Faculty of Medical and Health Sciences, The University of Auckland, New Zealand; [5]Microbiological Diagnostic Unit Public Health Laboratory, and [6]Centre for International Child Health, University of Melbourne, Australia

(See the Editorial Commentary by Norrby-Teglund and Siemens on pages 1461–2.)

Background. The concept that a minority of group A streptococcus (GAS) *emm* types are more "rheumatogenic" than others has been widely disseminated. We aimed to provide a comprehensive list of acute rheumatic fever–associated GAS isolates and assess the presence of associated rheumatogenic motifs.

Methods. Articles reporting GAS *emm*-type or *emm*-type–specific antibody responses associated with rheumatic fever were identified from 1 January 1944 to 31 July 2018. The revised Jones criteria were used to define rheumatic fever with a maximum period of 4 weeks between disease onset and microbiological characterization. A database of 175 representative M-protein sequences was used to analyze the protein diversity of rheumatic fever–associated strains in a phylogenetic tree and to identify the presence of 10 previously recognized rheumatogenic motifs.

Results. We included 411 cases of rheumatic fever, for which microbiological characterization identified 73 different *emm* types associated with the disease. The classic rheumatogenic *emm* types represented only 12.3% of the 73 emm types and were responsible for 31.6% of the 411 clinical cases. Rheumatic fever–associated *emm* types were disseminated throughout the phylogeny, suggesting they belong to various genetic backgrounds. Rheumatic fever–associated motifs were present in only 15.1% of the rheumatic fever–associated *emm* types and only 24.8% of clinical cases.

Conclusions. The concept of rheumatogenicity should be extended to include strains other than those classically described. Our results highlight significant knowledge gaps in the understanding of rheumatic fever pathogenesis and suggest that a GAS vaccine candidate should offer broad coverage against a variety of GAS genetic variants in order to protect against this serious sequela.

Keywords. Streptococcus pyogenes; rheumatic fever; pathogenesis; vaccine.

Acute rheumatic fever and rheumatic heart disease remain major public health issues, collectively affecting 33 million people and causing 319 400 deaths worldwide in 2015 [1]. Throat infection with *Streptococcus pyogenes* (group A streptococcus; GAS) causes an autoimmune response that leads to rheumatic fever and, secondarily, to rheumatic heart disease [2–4]. The available data from low- and middle-income countries (LMIC) support the hypothesis that GAS impetigo also plays a role in the pathogenesis of rheumatic heart disease [5]. In LMIC, rheumatic heart disease remains the principle cause of acquired heart disease in children, adolescents, and young adults. In 2018, the World Health Assembly called for renewed action to control rheumatic fever and rheumatic heart disease, including through acceleration of the development of a vaccine against GAS [6].

The autoimmune response that occurs during rheumatic fever appears to be mediated through molecular mimicry after a GAS throat infection in a predisposed individual [2–4]. In Oceanic populations, 2 recent genome-wide association studies identified potential host genetic susceptibilities [7, 8]. It has been generally accepted that only a limited number of GAS strains can trigger rheumatic fever [3, 9, 10]. These so-called "rheumatogenic" GAS strains are proposed to harbor key virulence determinants that induce an auto-immune reaction, while nonrheumatogenic strains do not [11]. The concept of rheumatogenicity has been widely disseminated in research, publications, and textbooks and has also had an impact on GAS vaccine development [3, 12].

Components of the M protein, a key GAS surface protein, and the group A carbohydrate antigen (N-acetyl-beta-D-glucosamine) are postulated to lead to molecular mimicry with human cardiac myosin and laminin on heart valves [2–4]. The antigenic structure of the group A carbohydrate does not vary across GAS strains and, therefore, is unlikely to explain the specificity of the rheumatogenic concept. More than 230 different M

Received 21 February 2019; editorial decision 12 April 2019; accepted 20 May 2019; published online July 23, 2019.

Correspondence: P. R. Smeesters, Hôpital Universitaire des Enfants Reine Fabiola, Avenue Jean Joseph Crocq 15, 1020 Brussels, Belgium (pierre.smeesters@huderf.be).

Clinical Infectious Diseases® 2020;70(7):1453–60

DOI: 10.1093/cid/ciz425

proteins have been described worldwide, potentially explaining variable capacity to induce rheumatic fever [13].

The epidemiologic surveillance of strains associated with GAS diseases, including rheumatic fever, is mostly based on M-protein analyses [14]. M-protein serotyping using rabbit sera was the original method of GAS typing in the 1950s, but was superseded by a polymerase chain reaction method (*emm* typing) and the associated functional classification of GAS (*emm*-cluster typing) [15–17]. Historically, a group of ten *emm* types have been epidemiologically associated with rheumatic fever (so-called classic rheumatogenic *emm* types): *emm*1, *emm*3, *emm*5, *emm*6, *emm*14, *emm*18, *emm*19, *emm*24, *emm*27, and *emm*29 [18].

There are 2 main theories that have been proposed for the auto-immune mechanism mediated by the M protein. First, M-protein epitopes may act to prime the immune system against heart proteins. GAS M protein contains antigenic epitopes similar to human cardiac myosin and laminin [19, 20]. Monoclonal antibodies against these antigens, derived from patients with rheumatic fever, cross-react with human myosin and valvular endothelium [21]. In a Lewis rat model of rheumatic fever, the passive transfer of M protein–specific T cells into a naive host initiated valvulitis, and immunization with recombinant M protein induced autoantibody formation and valvulitis [22]. This theory suggests that the rheumatogenicity of some strains correlates with the presence of rheumatogenic motifs on specific *emm* types [2–4]. In the literature, 9 rheumatogenic motifs have been widely proposed (Supplementary Methods; Supplementary Table S1) [20, 23, 24]. A recent genome-wide association study identified risk haplotypes concordant with several of these motifs [8].

Second, the complex of Type IV collagen bound to M protein may also trigger an autoimmune cascade. It has been proposed that the presence on the M protein of the collagen-binding motif peptide associated with rheumatic fever (PARF) leads to anticollagen antibodies that attack the subendothelial and perivascular connecting tissues, resulting in the inflammation of heart valve tissue [24, 25]. These antibodies have been detected in the sera of patients with rheumatic fever, but do not induce valvulitis in animal models [4].

We aimed to identify the GAS strains associated with rheumatic fever worldwide. We also aimed to analyze the genetic diversity of M protein of these strains, and the presence of rheumatogenic motifs. These data would have implications for future efforts to understand the pathophysiology of rheumatic fever and rheumatic heart disease, and for the formulation of GAS vaccines to ensure the coverage of strains known to be associated with rheumatic fever.

METHODS

We performed a systematic review of published clinical cases of rheumatic fever for which the associated *emm* type was available, and used this data set to do a genetic analysis of their associated M proteins, as well as an analysis of the presence of associated rheumatogenic motifs.

Data Sources

We searched all articles in the Pubmed, Medline, and Embase databases from 1 January 1944 (publication date of the original Jones criteria for diagnosis of rheumatic fever) [26] to 31 July 2018 for studies of rheumatic fever (Figure 1), following preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement [27]. We also reviewed the gray literature, including abstracts from the Lancefield Symposia on Streptococci and Streptococcal Infections, 1990 to 2017. Articles were analyzed by 2 independent authors, with review by a third author where there was disagreement. No language restrictions were used in the initial research. All included publications were assessed using a Strome ID [28]–derived matrix (Supplementary Table 2).

Case Definitions

For inclusion in our study, clinical cases had to satisfy the following criteria: (1) identification of a GAS strain in a patient with rheumatic fever who was diagnosed on the basis of the 2015 Jones criteria [29]; (2) a clearly identified period of less than 4 weeks between the microbiological characterization and the occurrence of rheumatic fever symptoms (as most cases of rheumatic fever occur 2 to 3 weeks after GAS pharyngeal infection); and (3) identification of the *emm* type, as described below. Clinical cases were excluded if there was a documented serological response to multiple *emm* types (see Supplementary Data for details).

Strains associated with rheumatic fever were categorized by country and development status, based on guidance from the Organization for Economic Cooperation and Development. For the purpose of this study, we modified the Organization for Economic Cooperation and Development's classification by assimilating indigenous populations (New Zealand Māori, New Caledonia Melanesians Kanak Community, and Hawaiian Polynesians: groups amongst which rheumatic fever remains endemic) into the LMIC grouping.

Allocation of *emm* Types Associated With Rheumatic Fever Cases

The *emm* type of each strain was identified by at least 1 of 2 techniques: (1) a bacterial culture, followed by *emm*-type identification by serotyping or *emm* typing; or 2) type-specific serology from the patient (by an enzyme-linked immunosorbent assay, Western blot, or indirect bactericidal assay). A significant serological response was determined by following the author's opinion in the articles. The *emm* sequence type information provided in some of the older studies was updated using the Centers for Disease Control and Prevention's *emm* database.

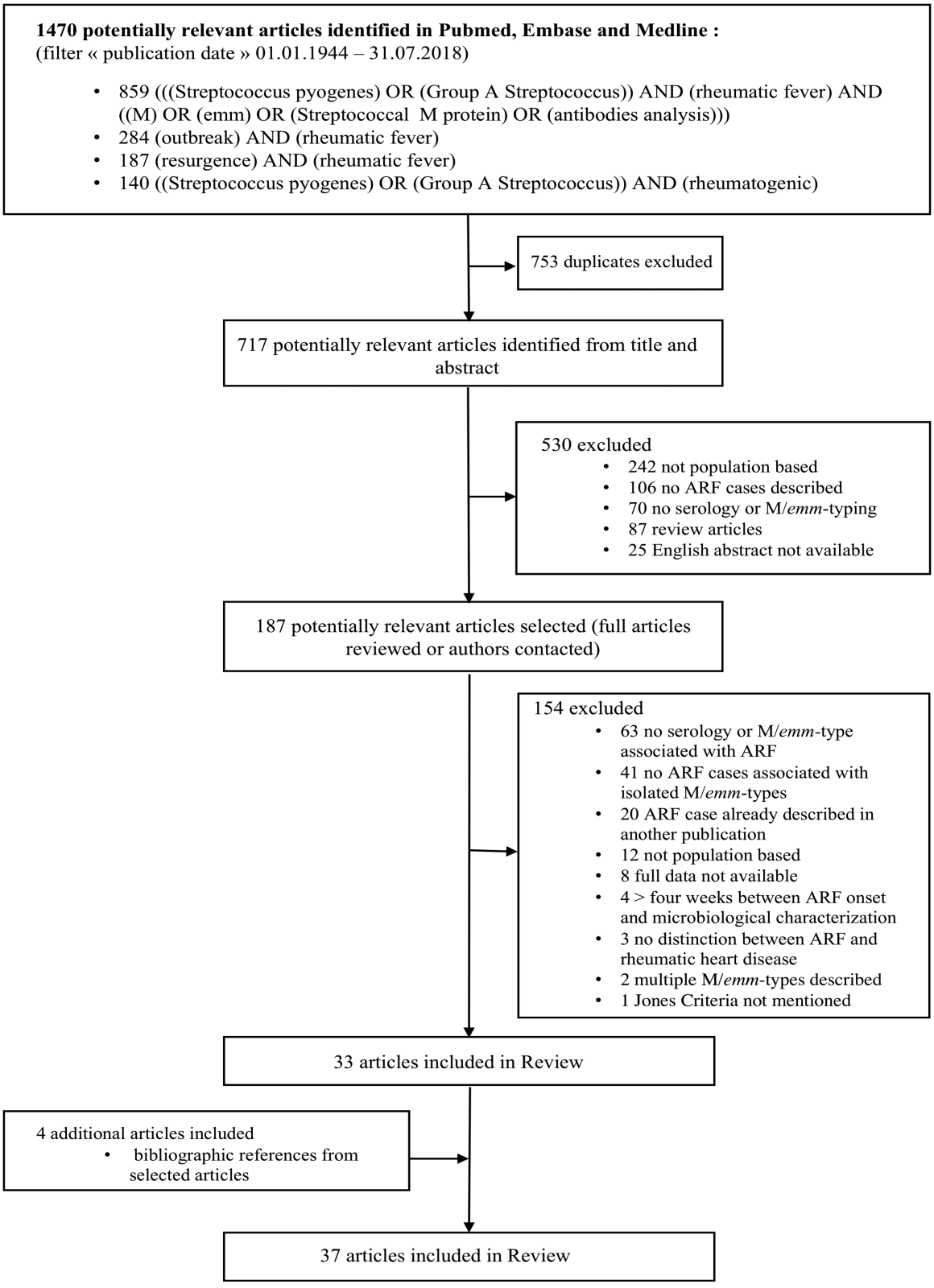

Figure 1. Summary of selection process and reasons for exclusion of studies. Abbreviation: ARF, acute rheumatic fever.

Distribution and Rheumatogenic Motifs of *emm* Types and *emm* Clusters

We used a database of M protein sequences, derived from a study of 175 representative *emm* types from 31 countries, to construct a phylogeny, as previously described [13, 17]. We analyzed the distribution of *emm* types against the 10 classic rheumatogenic motifs [18] and the 48 *emm* clusters [17]. The presence of the 9 M protein motifs linked to rheumatic fever and the PARF motif was searched for using the Geneious software (version R9; see Supplementary Methods and Supplementary Table S1 for details).

Patient and Public Involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for or in implementation of the study. No patients were asked to advise on interpretation of data or writing up results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Data Sharing Statement

We have added the complete data set of 411 acute rheumatic fever cases and associated GAS isolates in Supplementary Table S3.

Ethics Approval

Ethics approval was not needed for this study.

RESULTS

Our literature search identified 1470 potentially relevant articles and, after exclusions and bibliographic trailing, we included 37 articles. The data set included 411 clinical cases of rheumatic fever, originating from 14 countries, with 119 cases from 7 high-income countries (HICs; Belgium, Italy, Japan, New Zealand, Taiwan, United Kingdom, and United States) and 292 cases from 7 LMIC and indigenous populations (LMICs: Chile, India, Kuwait, France [New Caledonia Melanesians Kanak Community], Thailand, Trinity-Tobago, and Tunisia; indigenous populations: New Zealand Māori and Hawaiian Polynesians; Supplementary Table S3).

A total of 73 different *emm* types were identified, including 13 identified by both culture and serology, 60 by culture alone, and 0 by serology alone (Table 1). There were 67 *emm* types originating from LMICs, compared to 22 from HICs (Table 1).

The 73 *emm* types were disseminated across the whole phylogeny (Figure 2). They were observed in the 10 most frequent *emm* clusters globally (E4, AC3, AC4, E6, E3, E1, D4, AC5, E2, and M6), which are responsible for approximately 92% of worldwide GAS infections [17, 30, 31]. The 10 most frequent *emm* types associated with rheumatic fever in our review included 5 of the 10 classic rheumatogenic *emm* types (*emm*1 [$n = 28$ cases], *emm*5 [$n = 28$], *emm*18 [$n = 27$], *emm*6 [$n = 17$], and *emm*3 [$n = 16$]) and 5 *emm* types not included in the classic rheumatogenic group (*emm*12 [$n = 17$], *emm*9 [$n = 10$], *emm*11 [$n = 10$], *emm*4 [$n = 9$], and *emm*74 [$n = 9$]; Table 1; Figure 2). Overall, the 10 classic rheumatogenic *emm* types represented 9 of the 73 *emm* types associated with rheumatic fever and 31.6% of all 411 rheumatic fever cases, with a lower representation among LMIC cases (14.7%) than among HIC cases (73.1%; Table 2). The classic rheumatogenic strains were all situated in the same part of the phylogenetic tree (Clade Y), except for *emm*27 (Figure 2). Of note, *emm*27 was also the only classic rheumatogenic strain for which our literature review did not identify a single case of rheumatic fever.

The 9 rheumatogenic motifs were found among 18 *emm*-types, including 9 of the 73 *emm* types associated with rheumatic fever (85/411 isolates, 20.1%; Figure 2; Table 2; Supplementary Table S1). It follows that the remaining 9 *emm* types these motifs were found in, were not associated with rheumatic fever in this review and were concentrated in subclade Y1, where the majority of classic rheumatogenic strains were also located (Figure 2). However, they were absent from the other clades and, notably, from clade X (49% of the 73 *emm* type associated with rheumatic fever) and subclade Y2 (23% of the rheumatic fever *emm* types). The PARF motif was found among 5 *emm* types, including 2 of the 73 *emm* types associated with rheumatic fever (17/411 isolates, 4.1%; Figure 2; Table 2; Supplementary Table S1). The PARF motif was only found in subclades Y2 and in outlier *emm* types.

DISCUSSION

In this study, encompassing 73 years of global data, the distribution of strains associated with rheumatic fever is very broad, and the so-called classic rheumatogenic strains are responsible for a limited proportion of the rheumatic fever cases worldwide, concentrated in HICs. These data concur with contemporary reports from Hawaii and New Zealand [32, 33] and call into question the exactitude of the rheumatogenic concept.

New epidemiologic research into GAS disease over the past 2 decades has observed that GAS strains circulating in LMICs have far greater diversity than those in HICs [30, 31]. In HICs, the 25 most frequent *emm* types accounted for 90.3% of all GAS isolates, while in Africa, for example, the 25 most frequent *emm* types accounted for only 62.5% of all isolates [30]. The results we present here indicate that the diversity of strains associated with rheumatic fever is also greater in LMICs than in HICs. In addition to the classic rheumatogenic *emm* types, this study identified more than 60 *emm* types associated with rheumatic fever in LMICs. If rheumatic fever has nearly disappeared from HICs since the mid-twentieth century, the disease remains frequent in many LMICs, where the greatest burden of rheumatic heart disease occurs [1]. Further research is therefore warranted to better understand the molecular epidemiology of GAS disease in these settings.

Table 1. Acute Rheumatic Fever–associated *emm* Types Worldwide, 1944–2018

emm Type	*emm* Cluster	Number of Isolates			Presence in HIC	Presence in LMIC
		Culture	Serology	Total		
M1[a]	A-C3	28	0	28	Yes	Yes
M2	E4	8	0	8	Yes	Yes
M3[a]	A-C5	15	1	16	Yes	Yes
M4	E1	7	2	9	Yes	Yes
M5[a]	M5	14	14	28	Yes	Yes
M6[a]	M6	15	2	17	Yes	Yes
M8	E4	1	0	1	/	Yes
M9	E3	7	3	10	Yes	Yes
M11	E6	10	0	10	/	Yes
M12	A-C4	17	0	17	Yes	Yes
M14[a]	M14	1	0	1	/	Yes
M15	E3	1	0	1	/	Yes
M17	M17	2	0	2	Yes	/
M18[a]	M18	20	7	27	Yes	Yes
M19[a]	M19	6	0	6	Yes	Yes
M22	E4	3	3	6	Yes	Yes
M24[a]	M24	1	0	1	Yes	/
M25	E3	1	5	6	/	Yes
M28	E4	1	0	1	/	Yes
M29[a]	M29	5	1	6	Yes	Yes
M30	A-C2	3	0	3	Yes	/
M33	D4	7	0	7	/	Yes
M36	D1	2	0	2	/	Yes
M39	A-C4	1	0	1	/	Yes
M41	D4	5	0	5	Yes	Yes
M43	D4	2	0	2	/	Yes
M44	E3	2	0	2	/	Yes
M46	A-C1	1	0	1	Yes	/
M49	E3	2	0	2	/	Yes
M50	E2	1	0	1	Yes	/
M53	D4	8	0	8	/	Yes
M54	D1	2	0	2	/	Yes
M55	M55	1	0	1	/	Yes
M57	M57	2	0	2	/	Yes
M58	E3	4	1	5	Yes	Yes
M59	E6	3	1	4	/	Yes
M60	E1	1	0	1	/	Yes
M63	E6	2	0	2	/	Yes
M65	E6	5	0	5	/	Yes
M66	E2	1	0	1	/	Yes
M67	E6	1	0	1	/	Yes
M68	E2	2	0	2	/	Yes
M71	D2	4	0	4	/	Yes
M73	E4	1	1	2	/	Yes
M74	M74	9	0	9	/	Yes
M75	E6	2	1	3	Yes	Yes
M76	E2	2	0	2	/	Yes
M77	E4	2	0	2	Yes	Yes
M78	E1	1	0	1	Yes	/
M80	D4	1	0	1	/	Yes
M81	E6	3	0	3	/	Yes
M85	E6	1	0	1	/	Yes
M87	E3	1	0	1	/	Yes
M89	E4	4	0	4	/	Yes
M90	E2	2	0	2	/	Yes
M91	D4	1	0	1	/	Yes
M92	E2	3	0	3	/	Yes
M93	D4	2	0	2	/	Yes
M95	M95	1	0	1	/	Yes
M98	D4	1	0	1	/	Yes
M103	E3	2	0	2	/	Yes
M104	E2	1	0	1	/	Yes
M108	D4	3	0	3	/	Yes
M113	E3	1	0	1	/	Yes
M116	D4	1	0	1	/	Yes
M118	E3	1	0	1	/	Yes
M122	M122	1	0	1	/	Yes
M123	D3	1	0	1	/	Yes
M153	NT	1	0	1	/	Yes
M197	A-C2	1	0	1	/	Yes
M218	M218	1	0	1	/	Yes
M232	E4	3	0	3	/	Yes
M238	A-C3	1	0	1	/	Yes
NT	…	64	25	89	Yes	Yes
Total	…	344	67	411	…	…

Abbreviations: /, no; HIC, high-income countries; LMIC, low- and middle-income countries; NT, nontypable.

[a] Identifies "classical rheumatogenic" *emm* types.

A causal relationship between a GAS isolate and a rheumatic fever case remains challenging, because of the 2- to 4-week delay between a GAS infection and disease onset. A positive culture demonstrates live bacteria but does not differentiate between carriage and infection. Positive serology is suggestive of an active infection; however, baseline preinfecting antibody titres were absent in most of the included studies, making interpretation more difficult. The overall congruence of the bacterial culture results with the human serological responses supports a temporal association between an episode of rheumatic fever and infection with the strain isolated. Nearly one-fifth of the 73 *emm* types were identified by both culture and serology (Figure 2) and, in the phylogenetic tree, the overall distribution of the *emm* types identified by both microbiological techniques was similar.

It has been proposed that some motifs within the M protein are able to trigger rheumatic fever [2–4]. However, our review found that only 15.1% of the *emm* types and 24.8% of the isolates associated with rheumatic fever possessed at least 1 of these proposed rheumatogenic motifs (either 1 of the 9 M-protein motifs or the PARF motif). None of the rheumatogenic motifs were found in any of the 36 *emm* types associated with rheumatic fever located in clade X, nor the 10 *emm* types in *emm* cluster D4. These clade X and *emm* cluster D4–related *emm* types are epidemiologically relevant in many parts of the world, including HICs; are frequently associated with skin infections; and their rheumatogenicity is, so far, unexplained [17, 30, 31, 33]. This

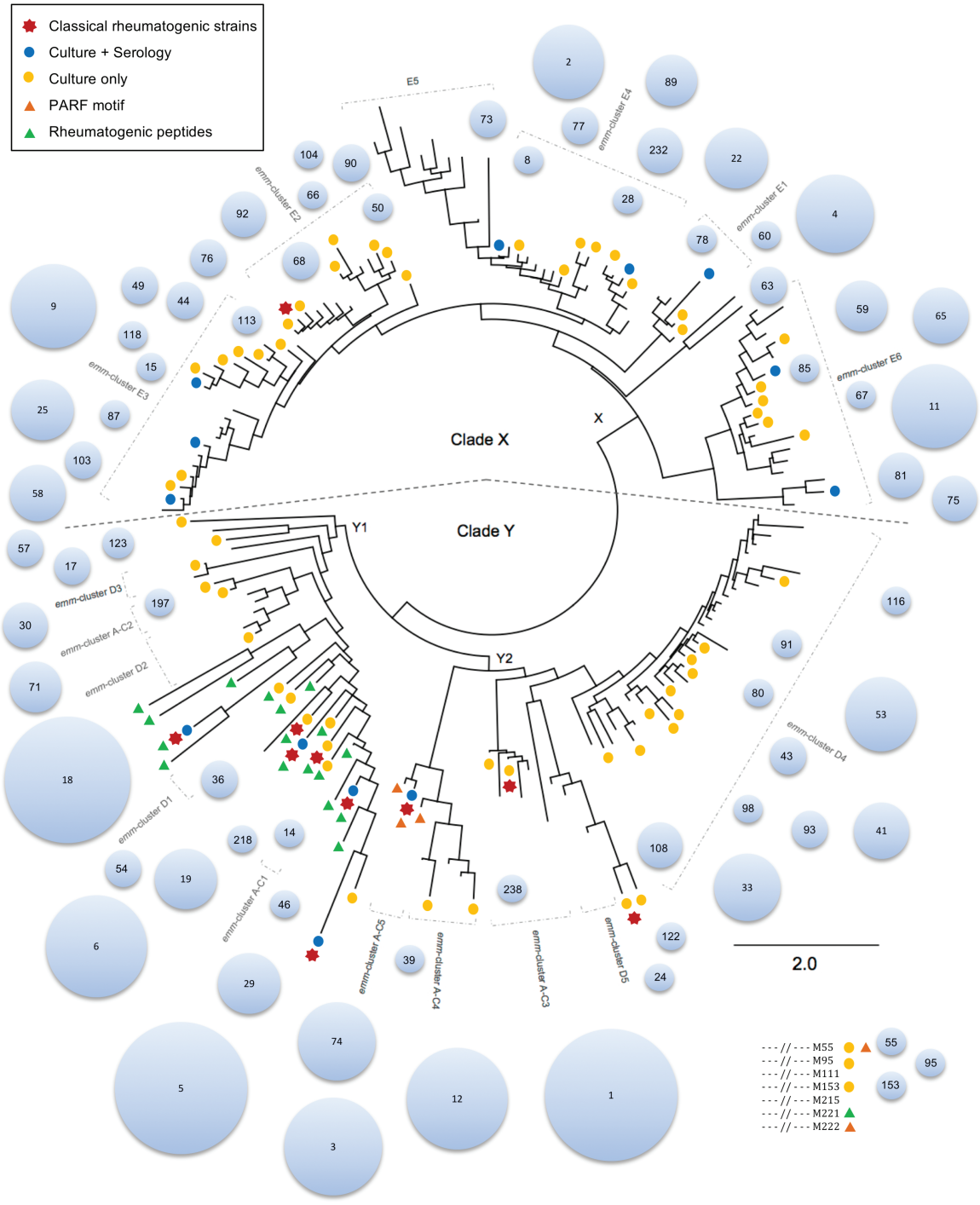

Figure 2. PhyML phylogenetic tree, adapted from a previous study [17]. The tree is drawn to scale, with branch lengths in the same units (number of amino acid substitutions per site) as those of the evolutionary distances used for the phylogenetic tree. The tree is supported by an aLRT (>80% for the *emm* clusters, as previously described; values are not displayed on the figure for clarity reasons) [17]. The tree is divided into 2 clades: clade X, composed of 6 *emm* clusters; and clade Y, divided into 2 subclades (Y1 and Y2) and composed of 10 *emm* clusters. There are 7 outlier *emm* types, illustrated by dashed lines. The classical rheumatogenic strains are marked by red stars. The blue and yellow dots correspond to isolates associated with rheumatic fever identified in our literature review. The strains containing the PARF motif are identified with orange triangles, and those containing 1 of 9 rheumatogenic motifs are identified with green triangles. The size of the gray circles are proportional to the frequency of each *emm* type. Abbreviations: aLRT, approximate likelihood-ratio test; PARF, peptide associated with rheumatic fever.

suggests that other motifs and/or antigens, such as the group A carbohydrate antigen, or others yet to be discovered, may also play a role in rheumatic fever pathogenesis; it highlights the need for microbiological studies using streptococcal strains originating from a broader variety of genetic and epidemiological backgrounds.

Table 2. Frequency of *emm* Types and Motifs Associated With Rheumatic Fever

	Presence Among 73 ARF Associated *emm* Types	Presence Among 411 ARF Associated Isolates	Presence Among the 10 Classic Rheumatogenic *emm* Types
Classic rheumatogenic *emm* type	9 (12.3%)	130 (31.6%)	/
Nonclassic *emm* type	64 (87.7%)	281 (68.4%)	/
PARF-motif	2 (2.7%)	17 (4.1%)	1 (10%)
Rheumatogenic motifs	9 (12.3%)	85 (20.1%)	5 (50%)
PARF or rheumatogenic motif	11 (15.1%)	102 (24.8%)	6 (60%)

Abbreviations: /, 0; ARF, acute rheumatic fever; PARF, peptide associated with rheumatic fever.

Our study found a far higher number of strains associated with rheumatic fever than previously reported and generally accepted. A vaccine candidate that protects only against the classic rheumatogenic *emm* types would have limited impact on the rheumatic fever incidence in endemic countries. The strains associated with rheumatic fever belonged to various genetic backgrounds, and many were associated with rheumatic fever in several studies and by different identification methods. Our finding calls into question the widely accepted concept that only a limited number of strains are capable of inducing rheumatic fever. With the availability of new, large-scale biological techniques, our understanding of rheumatic fever pathogenesis will be enhanced by the inclusion of strains originating from both LMICs and HICs in research efforts.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Author contributions. P. R. S. and A. C. S. designed the study. G. d. C., N. B., A. B., P. R. S., and A. C. S. acquired the data. G. d. C., N. B., A. B., N. J. M., D. A. W., A. C. S., and P. R. S. analyzed and interpreted the data. G. d. C. and P. R. S. wrote the first draft, and N. B., A. B., N. J. M., D. A. W., and A. C. S. revised it. All authors approved the submitted version. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Disclaimer. The funder had no role in the study design and analysis.

Financial support. This work was supported by the Belgian Fonds National de la Recherche Scientifique (grant number FNRS PDR N° T.0255.16F).

Potential conflicts of interest. The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Watkins DA, Johnson CO, Colquhoun SM, et al. Global, regional, and national burden of rheumatic heart disease, 1990–2015. N Engl J Med **2017**; 377:713–22.
2. Carapetis JR, Beaton A, Cunningham MW, et al. Acute rheumatic fever and rheumatic heart disease. Nat Rev Dis Primers **2016**; 2:15084.
3. Hurst JR, Kasper KJ, Sule AN, McCormick JK. Streptococcal pharyngitis and rheumatic heart disease: the superantigen hypothesis revisited. Infect Genet Evol **2018**; 61:160–75.
4. Karthikeyan G, Guilherme L. Acute rheumatic fever. Lancet **2018**; 392:161–74.
5. Parks T, Smeesters PR, Steer AC. Streptococcal skin infection and rheumatic heart disease. Curr Opin Infect Dis **2012**; 25:145–53.
6. Available at: http://www.who.int/news-room/detail/25-05-2018-seventy-first-world-health-assembly-update-25-may.
7. Parks T, Mirabel MM, Kado J, et al; Pacific Islands Rheumatic Heart Disease Genetics Network. Association between a common immunoglobulin heavy chain allele and rheumatic heart disease risk in Oceania. Nat Commun **2017**; 8:14946.
8. Gray LA, D'Antoine HA, Tong SYC, et al. Genome-wide analysis of genetic risk factors for rheumatic heart disease in aboriginal Australians provides support for pathogenic molecular mimicry. J Infect Dis **2017**; 216:1460–70.
9. Stollerman GH. Nephritogenic and rheumatogenic group A streptococci. J Infect Dis **1969**; 120:258–63.
10. Shulman ST, Stollerman G, Beall B, Dale JB, Tanz RR. Temporal changes in streptococcal M protein types and the near-disappearance of acute rheumatic fever in the United States. Clin Infect Dis **2006**; 42:441–7.
11. Bisno AL, Pearce IA, Stollerman GH. Streptococcal infections that fail to cause recurrences of rheumatic fever. J Infect Dis **1977**; 136:278–85.
12. Steer AC, Carapetis JR, Dale JB, et al. Status of research and development of vaccines for *Streptococcus pyogenes*. Vaccine **2016**; 34:2953–8.
13. McMillan DJ, Drèze PA, Vu T, et al. Updated model of group A streptococcus M proteins based on a comprehensive worldwide study. Clin Microbiol Infect **2013**; 19:E222–9.
14. Smeesters PR, McMillan DJ, Sriprakash KS. The streptococcal M protein: a highly versatile molecule. Trends Microbiol **2010**; 18:275–82.
15. Lancefield RC. A serological differentiation of human and other groups of hemolytic streptococci. J Exp Med **1933**; 57:571–95.
16. Facklam R, Beall B, Efstratiou A, et al. *emm* typing and validation of provisional M types for group A streptococci. Emerg Infect Dis **1999**; 5:247–53.
17. Sanderson-Smith M, De Oliveira DM, Guglielmini J, et al; M Protein Study Group. 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.
18. Bisno AL. The concept of rheumatogenic and non-rheumatogenic group A streptococci. In: Read SE, Zabriskie JB, eds. Streptococcal diseases and the immune response. New York, New York: Academic Press, **1980**:789–803.
19. Cunningham MW. Streptococcus and rheumatic fever. Curr Opin Rheumatol **2012**; 24:408–16.
20. Guilherme L, Faé KC, Oshiro SE, Tanaka AC, Pomerantzeff PM, Kalil J. T cell response in rheumatic fever: crossreactivity between streptococcal M protein peptides and heart tissue proteins. Curr Protein Pept Sci **2007**; 8:39–44.
21. Galvin JE, Hemric ME, Ward K, Cunningham MW. Cytotoxic mAb from rheumatic carditis recognizes heart valves and laminin. J Clin Invest **2000**; 106:217–24.
22. Kirvan CA, Galvin JE, Hilt S, Kosanke S, Cunningham MW. Identification of streptococcal M-protein cardiopathogenic epitopes in experimental autoimmune valvulitis. J Cardiovasc Transl Res **2014**; 7:172–81.
23. Cunningham MW, Antone SM, Smart M, Liu R, Kosanke S. Molecular analysis of human cardiac myosin-cross-reactive B- and T-cell epitopes of the group A streptococcal M5 protein. Infect Immun **1997**; 65:3913–23.
24. Reissmann S, Gillen CM, Fulde M, et al. Region specific and worldwide distribution of collagen-binding M proteins with PARF motifs among human pathogenic streptococcal isolates. PLOS One **2012**; 7:e30122.
25. Tandon R, Sharma M, Chandrashekhar Y, Kotb M, Yacoub MH, Narula J. Revisiting the pathogenesis of rheumatic fever and carditis. Nat Rev Cardiol **2013**; 10:171–7.
26. Jones TD. Diagnosis of rheumatic fever. JAMA **1944**; 126:481–4.
27. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ **2009**; 339:b2535.
28. Field N, Cohen T, Struelens MJ, et al. Strengthening the reporting of molecular epidemiology for infectious diseases (STROME-ID): an extension of the STROBE statement. Lancet Infect Dis **2014**; 14:341–52.

29. Gewitz MH, Baltimore RS, Tani LY, et al; American Heart Association Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease of the Council on Cardiovascular Disease in the Young. Revision of the Jones criteria for the diagnosis of acute rheumatic fever in the era of Doppler echocardiography: a scientific statement from the American Heart Association. Circulation **2015**; 131:1806–18.
30. 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.
31. Smeesters PR, McMillan DJ, Sriprakash KS, Georgousakis MM. Differences among group A streptococcus epidemiological landscapes: consequences for M protein-based vaccines? Expert Rev Vaccines **2009**; 8:1705–20.
32. Erdem G, Mizumoto C, Esaki D, et al. Group A streptococcal isolates temporally associated with acute rheumatic fever in Hawaii: differences from the continental United States. Clin Infect Dis **2007**; 45:e20–4.
33. Williamson DA, Smeesters PR, Steer AC, et al. M-protein analysis of streptococcus pyogenes isolates associated with acute rheumatic fever in New Zealand. J Clin Microbiol **2015**; 53:3618–20.