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
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 M-protein serotyping using rabbit sera was the original method of GAS typing in the 1950s, but was superseded by a polymerase chain reaction method (emm typing) and the associated functional classification of GAS (emm-cluster typing) [15–17].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): emm1, emm3, emm5, emm6, emm14, emm18, emm19, emm24, emm27, and emm29 [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 naïve 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.
1470 potentially relevant articles identified in Pubmed, Embase and Medline:
(filter « publication date » 01.01.1944 – 31.07.2018)
- 859 (((Streptococcus pyogenes) OR (Group A Streptococcus)) AND (rheumatic fever) AND
((M) OR (emn) OR (Streptococc M protein) OR (antibodies analysis)))
- 284 (outbreak) AND (rheumatic fever)
- 187 (resurgence) AND (rheumatic fever)
- 140 ((Streptococcus pyogenes) OR (Group A Streptococcus)) AND (rheumatogenic)

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 (emm11 [n = 28 cases], emm5 [n = 28], emm18 [n = 27], emm6 [n = 17], and emm3 [n = 16]) and 5 emm types not included in the classic rheumatogenic group (emm12 [n = 17], emm9 [n = 10], emm11 [n = 10], emm4 [n = 9], and emm74 [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 emm27 (Figure 2). Of note, emm27 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.
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
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.
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 No. T.0255.16/F).

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
28. Field N, Cohen T, Strue O, et al. Strengthening the reporting of molecular techniques, our understanding of rheumatic fever pathogenesis will be enhanced by the inclusion of strains originating from both LMICs and HICs in research efforts.

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

<table>
<thead>
<tr>
<th>Presence Among 73 ARF Associated emm Types</th>
<th>Presence Among 411 ARF Associated Isolates</th>
<th>Presence Among the 10 Classic Rheumatogenic emm Types</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classic rheumatogenic emm type</strong></td>
<td>9 (12.3%)</td>
<td>130 (31.6%)</td>
</tr>
<tr>
<td><strong>Nonclassic emm type</strong></td>
<td>64 (87.7%)</td>
<td>281 (68.4%)</td>
</tr>
<tr>
<td>PARF-motif</td>
<td>17 (41.1%)</td>
<td>1 (10%)</td>
</tr>
<tr>
<td>Rheumatogenic motifs</td>
<td>85 (20.1%)</td>
<td>5 (50%)</td>
</tr>
<tr>
<td>PARF or rheumatogenic motif</td>
<td>102 (24.8%)</td>
<td>6 (60%)</td>
</tr>
</tbody>
</table>

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


