MALDI-TOF and comparative genomic analysis of M-18 GAS strains associated with acute rheumatic fever outbreak in Northeast Italy, 2012-2013

Paolo Gaibani[1*], Erika Scaltriti[2], Claudio Foschi[1], Enrico Baggio[1], Maria Vittoria Tamburini[1], Roberta Creti[3], Maria Grazia Pascucci[4], Marco Fagioni[5], Simone Ambretti[1], Francesco Comandatore[6], Stefano Pongolini[2] and Maria Paola Landini[1]

[1] Operative Unit of Clinical Microbiology, St. Orsola[1] -Malpighi University Hospital, Bologna, Italy. [2] Sezione Diagnostica di Parma, Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna (IZSLER), Parma, Italy. [3] Dipartimento di Malattie Infettive, Parassitarie ed Immunomediate, Istituto Superiore di Sanità, Rome, Italy. [4] Public Health Authority Emilia Romagna, Italy [5] Bruker Daltonics s.r.l., Macerata, Italy, [6] Dipartimento di Scienze Veterinarie e Sanità Pubblica (DIVET), Università degli Studi di Milano, Milano, Italy

Key words: GAS, acute rheumatic fever (ARF), MALDI-TOF, comparative genomic analysis,

Running Title: MALDI-TOF and genomic analysis of M-18 GAS

*Corresponding Author:

Paolo Gaibani, PhD

Operative Unit of Microbiology, S.Orsola-Malpighi University Hospital, Regional Reference Centre for Microbiological Emergencies, 9 via G. Massarenti – 40138 Bologna, ITALY.

Telephone: +39 051 6364316 Fax: +39 051 6363076

e-mail: paolo.gaibani@unibo.it

ABSTRACT

Acute rheumatic fever (ARF) is a post-suppurative sequela caused by *Streptococcus pyogenes* infections affecting school-age children. Here, we describe the occurrence of an ARF outbreak that occurred in the Bologna province, Northeast Italy, between November 2012 and May 2013. Molecular analysis revealed that ARF-related group A *streptococcus* (GAS) strains belonged to the M-18 serotype, including subtypes *emm18.29* and *emm18.32*. All M-18 GAS strains shared the same antigenic profile, thus including SpeA, SpeB, SpeC, SpeL, SpeM and SmeZ.

MALDI-TOF analysis revealed that M-18 GAS strains grouped separately from other serotypes, thus suggesting a different *S.pyogenes* lineage. Single nucleotide polymorphisms and phylogenetic analysis based on whole genome sequencing showed that *emm18.29* and *emm18.32* GAS strains clustered in two distinct groups highlighting genetic variations between these subtypes. Comparative analysis revealed similar genome architecture between *emm18.29* and *18.32* strains that differed from non-invasive *emm18.0* strains. The major sources of differences between M-18 genomes were attributable to the prophage elements. Prophage regions contained several virulence factors that could have contributed to the pathogenic potential of *emm18.29* and *emm18.32* strains. Notably, phage ΦSPBO.1 carried erythrogenic toxin A gene (*speA1*) in six ARF-related M-18 GAS strains, but not in *emm18.0* strains. In addition, phage-encoded *hyaluronidase* gene (*hylP.2*) presented different variants among M-18 GAS strains by showing internal deletions located in the α-helical and TSβH regions.

In conclusion, our study yielded insights into the genome structure of M-18 GAS strains responsible for the ARF outbreak in Italy, thus expanding our knowledge of this serotype.

INTRODUCTION

Streptococcus pyogenes, group A streptococcus (GAS), is a gram-positive bacterium responsible for a wide spectrum of diseases ranging from moderate or mild infections to severe invasive diseases such as necrotizing fasciitis and toxic shock-like syndrome (TSLS). Several GAS infections can cause severe post-infectious sequelae, including acute post-streptococcal glomerulonephritis (APSGN), acute rheumatic fever (ARF), and rheumatic heart disease (RHD) (1).

ARF is a systemic disorder resulting from an autoimmune disease following a GAS infection that usually occurs in children between 5 and 15 years of age (2). During the last decades, the incidence of ARF cases has significantly declined in the United States and Western Europe, whereas it remains high in Eastern Europe, Asia and Australia (3). However, the resurgence of ARF in several geographical areas, including United States, is a matter of concern (4).

S.pyogenes possesses different virulence factors such as the M protein and Superantigens (SAgs) that contribute to the pathogenesis of GAS infection (1). On the basis of the high variability of the M protein among GAS strains, the 5'terminal sequence of the *emm* gene (*emm*-typing) is considered a reliable molecular marker commonly used for epidemiological studies (5).

Previous studies indicated that *emm*-types 1, 3, 5, 6, 18, 19, 24 and 29 have been isolated from ARF cases, suggesting a "rheumatogenic" role of certain serotypes (2). Epidemiological study of different ARF outbreaks in the US revealed a strict association with serotype M-18 GAS (6)

Recently, matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) has been introduced in microbiological laboratories for prompt highly accurate identification and classification of bacterial species (7). Several studies have now demonstrated the ability of MALDI-TOF MS to type and distinguish a wide range of bacterial species at subspecies or strain level (8,9).

On the other hand, whole genome sequencing is a consolidated procedure for epidemiological and evolutionary purposes (10-12). This technique is highly sensitive and can identify single nucleotide polymorphisms (SNPs) throughout the entire genome (13). In particular, the resolution of strains genetically indistinguishable by other molecular techniques (i.e. MLST, PFGE) made whole genome sequencing technology a powerful tool for the epidemiological investigations of related high clonal bacterial isolates (12,13).

Here, we report an epidemiological investigation based on both whole genome sequencing and MALDI-TOF MS on serotype M-18 GAS strains collected from primary school-age children in Bologna province during an ARF outbreak in early 2013.

MATERIAL AND METHODS

Epidemiological investigation

In February 2013, a notification of an ARF case following hospital admission was reported in an 11-year-old otherwise healthy white boy resident in Bologna province, Emilia-Romagna region. In this area, six ARF cases were diagnosed in the previous three months. After the last notification of ARF diagnosis, the Regional Health Agency instituted active epidemiological surveillance to monitor all potential ARF contact cases following WHO criteria (14). The active surveillance protocol required both clinical evaluation and culture screening of all classmates of ARF cases. At the same time, 14 GAS strains isolated from school-age children with pharyngotonsillitis in the Bologna metropolitan area were collected. Two months later, diagnosis of ARF was notified in a 4-year-old white boy resident in the same province. Subsequently, 14 GAS positive samples were collected from classmates of the second ARF case and from 34 symptomatic children resident in the same area.

The date of isolation, mucoid trait, *emm*-type, antimicrobial resistance, superantigen genes and epidemiological linkage for each strain are listed in Supplemental Table 1.

Bacterial isolation and identification

70 GAS strains were isolated from throat swabs collected at the bacteriology laboratory of St.Orsola-Malpighi Hospital, Bologna, except for the two thirty years old *emm18* GAS strains that were deposited in the bacterial collection bank of Istituto Superiore di Sanità (ISS). Bacteria were initially identified using standard methods and confirmed by MALDI-TOF 3.1 RTC (Bruker Daltonics, GmbH, Germany) following the manufacturer's instructions. Antimicrobial susceptibility to penicillin, ampicillin, tetracycline, chloramphenicol, erythromycin and clindamycin was tested by MicroScan semi-automated system (Siemens, Germany) and the results were interpreted according to the EUCAST criteria (15). GAS isolates were examined for the presence of a mucoid phenotype by culture visualization and categorized as mucoid or nonmucoid.

Emm-typing and SuperAntigen (SAg) genes

Genomic DNA from 70 GAS strains were extracted from pure cell bacterial culture using the manual DNeasy Blood&Tissue kit (Qiagen, Basel, Switzerland) according to the manufacturer's protocol. PCR amplification of the *emm* gene was performed as previously described (16). In order to assign the specific *emm* type and subtype, the first 240 nt of each sequence were compared to the *S.pyogenes emm*-database available at the CDC website (http://www.cdc.gov/streplab/index.html). Super Antigen (SAg) genes were analyzed by multiplex PCR assays, as previously described (17). The exotoxin genes *speB* and *speF* were used as PCR internal controls. The presence of the SAg genes (*speA, speC, speG, speH, speI, speJ, speK, speL, speM, ssa* and *smeZ*) was confirmed by single PCRs.

113

114 *Multilocus sequence typing*

115 To determine the genetic relationship between the eleven *S. pyogenes* isolates belonging to the
116 serotype M-18, multilocus sequence typing (MLST) based on seven housekeeping genes (*gki, gtr, ,*
117 *murI, mutS, recP, xpt, yiqL*) was performed (18). The allele numbers and the relative sequence type
118 (ST) were assigned using the *S.pyogenes* MLST database available at the webpage
119 (http://Spyogenes.mlst.net).

120

121 *MALDI-TOF MS sample preparation and analysis*

122 Sample preparation for MALDI-TOF MS was performed as previously described with minor
123 modifications (8). Briefly, colonies of fresh overnight culture derived from 49 GAS isolates were
124 resuspended at 1 McFarland and 1 ml of bacterial suspension was centrifuged at 5,000 g for 5 min.
125 Pellets were suspended with 300 µl distilled water and 900 µl absolute ethanol and pelletted again.
126 Then supernatants were discharged and cells were suspended in 20 µl formic acid (70%) and 20 µl
127 acetonitrile. Whole cell suspension was centrifuged at 12.000 g for 5 min and 1 µl of supernatants
128 was placed on a MALDI sample slide (Bruker-Daltonics, Bremen, Germany) and dried at RT. Then
129 sample was overlaid with 1 µl matrix solution (α-cyano-4-hydroxycinnamic acid in 50% acetonitrile
130 and 2.5% trifluoroacetic acid) and dried at RT. MALDI-TOF MS measurement was performed with
131 a Bruker MicroFlex MALDI-TOF MS (Bruker Daltonics, Germany) using FlexControl software
132 and a DH5-α *Escherichia coli* protein extract (Bruker- Daltonics) was deposited on the calibration
133 spot of the sample slide for external calibration. Spectra collected in the positive ion mode within a
134 mass range of 2,000 to 20,000 Da were analyzed using Bruker Biotyper automation control and the
135 Bruker Biotyper 3.1 software and library (database [DB] 5627 with 5,627 entries). The clustering
136 analysis of the GAS strains was performed by generation of the dendrogram based on the different

Journal of Clinical Microbiology

serotypes collected in this study. In detail, 49 strains representing 11 different serotypes included: M-1 [n=4], M-3 [n=2], M-5 [n=6], M-6 [n=9], M-9 [n=1], M-12 [n=1], M-18 [n=11], M-28 [n=9], M-44 [n=2], M-89 [n=3] and M-102 [n=1]. The main spectra (MSPs) of each strain were generated from ten technical replicates prior to manual visualization inspections by FlexAnalysis 3.4 software. The relationship between MSPs obtained from each strain was visualized in a score-oriented dendrogram using the average linkage algorithm implemented in the MALDI Biotyper 3.1 software (Bruker Daltonics, Germany).

Whole genome sequencing and comparative genomics

Whole genome sequencing was conducted on the two *S.pyogenes* isolated from the ARF case and 9 M-18 GAS strains included in this study (see Table S1 in the Supplemental materials). Libraries were prepared with the Nextera XT sample preparation kit (Illumina) and sequencing was performed on the Illumina MiSeq platform (Illumina, San Diego, USA) with a 2 X 250 paired-end run. All read sets were evaluated for sequence quality and read-pair length using FastQC (19) and then assembled with MIRA 4.0 using a de novo assembly mode (20).

For comparative studies, single nucleotide polymorphisms (SNPs) were identified with an in-house Perl pipeline based on the Mauve software (21). In this approach, the genome of *Streptococcus pyogenes*, group A Strain SP665Q, was used as reference, and the other eleven genome assemblies included in this study were aligned against it. All the alignments were then merged, and the coordinates of all nucleotide variations were detected on the basis of the annotated reference strain assembly (SP665Q). All the variations were then organized in a matrix to assess the presence/absence pattern in specific subsets of strains (e.g. *emm 18.29* and *emm18.32*). The core SNPs, defined as the non-degenerate SNPs present in all the twelve genomes and flanked by conserved positions,

Journal of Clinical Microbiology

were extracted and finally subjected to Bayesian phylogenetic analysis with MRBAYES (22). The Bayesian analysis was run on the GTR substitution model for 2000000 generations with a chain sampled every 1000th generation. The final parameter values and trees are summarized after discarding 25% of the posterior sample. The Bayesian tree is displayed and edited using FigTree v1.4.0 available at the website: http://tree.bio.ed.ac.uk/software/figtree.

Informative SNPs (i.e., present in at least two strains) were extracted from core SNPs using an in-house script. Additionally, among the strains with *emm-type* 18.29, all genes presenting at least one core SNP were selected and compared with the virulence factors of pathogenic bacteria (VFPB) database (23), using a blast search with a 10^{-5} value cutoff.

Prophages were detected and analyzed using the free web tool PHAST (PHAge Search Tool) (24). The comparative sequence circular maps of whole genomes and concatenated prophages of each strain of *S. pyogenes* M18GAS were generated using BRIG (25).

Accessions. The sequences of the 11 M-18 GAS genomes were deposited at EMBL-EBI under the accession numbers: **M18GASBO1065** (CDGV01000001-CDGV01000182), **M18GASBO665** (CDGO01000001-CDGO01000182), **M18GASBO9** (CDGY01000001-CDGY01000207), **M18GASBO8** (CDGW01000001-CDGW01000183), **M18GASBO7** (CDGX01000001-CDGX01000199), **M18GASBO6** (CDGM01000001-CDGM01000081), **M18GASBO5** (CDGQ01000001-CDGQ01000096) **M18GASBO4** (CDGN01000001-CDGN01000158), **M18GASBO3** (CDGS01000001-CDGS01000294), **M18GASBO2** (CDHA01000001-CDHA01000222), **M18GASBO1** (CDHB01000001-CDHB01000235) (Study project: **PRJEB7108**).

RESULTS

184 *Characterization of GAS isolates*

185 Eight ARF cases were recorded in the Bologna province between November 2012 and May 2013.
186 All patients were aged between 4 and 12 years. In the last years the number of ARF cases in this
187 province has ranged from one to four per year. At time of diagnosis, seven out of eight ARF cases
188 were culture negative, only one GAS strain being isolated from a patient resident in Bologna
189 province.

190 During the surveillance period, 70 GAS isolates were collected (see Table S1 in the supplemental
191 material). In detail, two consecutive isolates were collected from the first patient with ARF (i.e.
192 SPBO1 and SPBO2), six were isolated from classmates of the first case and 14 were isolated from
193 symptomatic children resident in the Bologna area at time of ARF diagnosis. Two months later, a
194 second collection comprised 14 GAS isolates collected from classmates of the second ARF case and
195 34 GAS obtained from symptomatic school-age children resident in the same area. However, no
196 GAS was isolated from second case of ARF.

197 Analysis of the *emm* sequence revealed that the GAS isolates obtained from patient with ARF was
198 *emm18.29*, as two out of six of the GAS isolated from contacts (Figure 1A). At the same time, GAS
199 isolated from symptomatic children of the community were: *emm28.0* (3 cases), *emm18.29* (2
200 cases), *emm89.0, emm1.0, emm56.0, emm9.0, emm6.4, emm102.3, emm12.0, emm5.3,* and *emm3.8*
201 (one case each), as shown in Figure 1B.

202 Molecular investigation conducted among the 14 GAS collected from the classmates of the second
203 ARF case showed that two isolates (14.3%) were *emm18.32*, four (28.5%) were *emm28.0* and eight
204 (57.2%) were *emm6.4*, as shown Figure 1C. Among the 35 GAS isolates collected from
205 symptomatic school-age children in the community, eight different *emm*-types were identified:
206 *emm*6.4 (23 cases), *emm*28.0 (2 cases), *emm*44.0 (3 cases), *emm1.0* (2 cases), and *emm5.3*,
207 *emm18.32, emm89.0 emm5.1* and *emm3.1* (one case each, respectively), as shown in Figure 1D.

Journal of Clinical Microbiology

Our data indicate that *emm18.29* was the dominant serotype among the isolates collected from the class of the first case, whereas *emm18.32* spread in the class of the second case. To investigate the relationship between the specific *emm*-type and the SAg profile, GAS isolates were evaluated by molecular analysis for the different SAg genes (see Table S1 in the supplemental material). Molecular analysis revealed eight different antigenic profiles among 70 GAS isolates. The chromosomally encoded *speB, speG* and *smeZ* genes were present in all isolates. In addition, all *emm-18* strains, including *emm18.29* and *emm18.32* and the two *emm18.0* derived from the ISS bank collection (SP665Q and SP1065Q) presented a characteristic SAg profile showing *speA*, *speC*, *speL* and *speM* genes. Overall, the isolates with the same *emm*-type shared a common SAg profile, with the exception of the *emm3* isolates. MLST analysis showed that all *emm18* GAS isolates belonged to sequence type (ST) 42.

To determine the association of the mucoid phenotype with *emm*-type, all GAS isolates were analyzed by culture visualization. Among GAS isolates, the M-102, M-9 and M-18 GAS strains showed the highest incidence (100%, 100% and 87.5%, respectively) of mucoid traits, while all *emm1* showed a nonmucoid colony type (see Table S1 in the supplemental material). Antimicrobial susceptibility testing showed that all isolates were susceptible to penicillin, ampicillin, clindamycin, chloramphenicol, tetracycline and clindamycin, while only two GAS belonging to the *emm89.0* and *emm5.3* serotypes resulted resistant to erythromycin

MALDI-TOF and clustering analysis

MALDI-TOF MS analysis of different *S.pyogenes* strains showed that 37 of 50 (74.0%) isolates clustered in accordance with the serotype group, as shown in Figure 2. In detail, the main spectra generated by MALDI-TOF MS analysis demonstrated a high overall discriminatory power of the strains belonging to the serotypes M-18, M-28 and M-3. However, different clustering groups were

Journal of Clinical Microbiology

observed for M-1, M-5, M-6, M-89 and M-44 strains by showing a different protein mass spectral profiling among isolates belonging to the same serotype (Figure 2). Notably, the score-oriented dendrogram showed that all isolates belonging to serotype M-18 formed a separate clustering group clearly distinguishable from other serotypes. Additionally, a different cluster grouping within the M-18 serotype was observed between *emm-18.0* and *18.32-18.29* subtypes with a critical distance of 500 (see Figure 2), respectively corresponding to the *S.pyogenes* strains isolated 30 years ago and the isolates involved in the ARF outbreak in Bologna province during 2013. However, MALDI-TOF MS analysis was not able to distinguish among *emm18.29* and *emm18.32* subtypes.

Whole genome sequencing and phylogenetic analysis of M-18 GAS isolates

The draft genomes of *emm18.29*, *emm18.32* and *emm18.0* GAS isolates were assembled into average 185 contigs with a G+C content of 38.6 % for a total of 1,929545 base pairs (Figure 3). Genome annotations predicted a total of 1881 open reading frames (ORFs).

To investigate the relationship among M-18 GAS isolates, a whole genome analysis was performed on the basis of core SNPs identified with the Mauve-based approach (593 core SNPs). Phylogenetic analysis indicated that two main clusters were highlighted with high posterior probabilities: the first cluster grouped the *emm18.29* strains, while the second grouped the *emm18.32* strains. In addition, the two e*mm18.0* GAS isolates derived from the ISS bank collection (SP665Q and SP1065Q) showed a closely relationship to the GAS8232 strain. Phylogenetic analysis indicated that *emm18.0* strains and GAS8232 differed by 198 informative SNPs and were nearer to *emm18.32* rather than *emm18.29* strains (Figure 4). Comparison of SNPs between M-18 strains isolated during the ARF outbreak in Bologna province showed that the two consecutive GAS isolates from the ARF case (SPBO1 and SPBO2) differed by 14 SNPs (0 informative SNPs) and were closely related to strains collected from classmates (SPBO3 and SPBO4) and from the community (SPBO5 and SPBO6), as

shown in Figure 4. Comparison of *emm18.29* genomes disclosed 216 different SNPs and five informative SNPs in this group. In addition, all *emm18.32* GAS strains collected from children during the second ARF episode were closely related (53 different SNPs and 0 informative SNPs) (Figure 4).

Further comparison of all M-18 genomes identified 73 and 207 SNPs exclusive to *emm18.29* and *emm18.32* clusters, respectively (see Table S2 in the supplemental material). Analysis of all unique SNPs of *emm18.29* cluster showed that 36.9% (27) were synonymous, 58.9% (43) were non-synonymous and 4.2% (3) were in intergenic regions. At the same time, the SNPs of *emm18.32* cluster were 41.6% (86) synonymous, 50.2% (104) non-synonymous and 8.2% (17) within intergenic regions, respectively. Analysis of synonymous and non-synonymous substitutions in the coding regions of the core genome revealed that SNPs were associated with different GAS virulent factors (see Table S2 in the supplemental material) present in the VFPB database (23). Notably, 46 SNPs substitutions occurred in prophage elements, including *hyaluronidase* and *gp58-like* genes both in *emm18.29* and *emm18.32* clusters (see table S2 in the supplemental material). In detail, five SNPs in the *hyaluronidase (hylP)* genes among the two cluster subtypes were synonymous and six were non-synonymous, most of them located in the N-terminal region of *hylP* gene.

Comparison of phage elements in M-18 GAS strains

Analysis of M-18 GAS chromosomes revealed that six regions contained prophage elements (ΦSPBO.1, ΦSPBO.2, ΦSPBO3, ΦSPBO4, ΦSPBO.5 and ΦSPBO.6) ranging from 7.6 to 75.8 Kb (Table 2). Genome distribution of the prophages across the GAS chromosome revealed that all M-18 strains shared a common localization of these elements, as shown in Figure 3. Comparison with other GAS genomes revealed that five prophage (ΦSPBO.1, ΦSPBO2, ΦSPBO3, ΦSPBO4 and

Journal of Clinical Microbiology

ΦSPBO.5) elements showed similar chromosome locations with GAS8232 genome (11), suggesting conservative site integrations of these regions across M-18 GAS strains (Figure 3). Moreover, examination of prophages elements showed a similar genetic architecture with GAS8232 (M-18) and MGAS315 (M-3) strains, as shown in figure 5 (11,26).

Genomes of M-18 GAS strains contained prophage regions harboring several virulence factors, including exotoxin type A (SpeA), exotoxin type C (SpeC), exotoxin type L (SpeL), exotoxin type M (SpeM), mitogenic factor (DNase) and streptodornase (Sdn) (Table 2). Comparison of the prophage elements showed that ΦSPBO.2,ΦSPBO.3 and ΦSPBO.5 regions were shared among all M-18 GAS strains. These three phage regions contained different virulence factors such as genes encoding SpeC, mitogenic factor, SpeL/SpeM, hyaluronidase and streptodornase. Interestingly, closely association between phages encoded SpeC and SpeL/speM were observed in seven M-18 GAS strains (Table 2).

Moreover, ΦSPBO.1 region was present in ten out of eleven of the M-18 GAS strains, lacking in the SP665Q isolate (*emm18.0* subtype). Analysis of virulence factors showed that ΦSPBO.1 region has variant of the *speA* gene (*speA1*) in six out of nine *emm18.29* and *emm18.32* strains, while was absent in *emm18.0* strains (Table 2). In addition, phage ΦSPBO.4 was lacking in *emm18.32* strains, whereas it was present in all *emm18.0* and *emm18.29* strains. This region contained a gene encoding mitogenic factor. Interestingly, each phage region had a *hyaluronidase* gene (Table 2). Similar findings were observed in a previous study demonstrating that all phages in M-3 GAS strainscontained a *hyaluronidase* gene (26).

Comparison of phage-encoded hyaluronidase (hylP.2) gene among M-18 GAS strains

To investigate the genetic variability in the *hylP.2* gene, the gene-encoded HylP.2 of the M-18 strains was compared with previously described GAS isolates. Our analysis indicated that the *hylP.2* gene was located in phage ΦSPBO.1 in all *emm18.29*, *emm18.32* and *emm18.0* GAS strains

(Table 2). Analysis of *hylP.2* gene derived from the *emm18.29, emm18.32* and *emm18.0* strains showed a high rate of homology with the M18 (MGAS8232) and M5 (Manfredo) GAS strains, two strains isolated from ARF cases (11,27). Comparison of deduced amino acid sequences with M5-GAS Manfredo strain demonstrated that HylP.2 were C-terminal truncated in all M-18 GAS isolates (including *emm18.29*, *emm18.32*, *emm18.0* and GAS8232 strain) (Figure 6). Of note, three *emm18.29* isolates (SPBO1, SPBO3 and SPBO6) showed an internal deletion located between the N-terminal and the TSβH domains. At the same time, the deduced amino acid sequence of *hylP.2* gene in the SPBO8 isolate was truncated in the TSβH region (Figure 6). Comparison analysis conducted on the *hylP.2* gene revealed different clustering groups according to the corresponding subtypes (data not shown).

DISCUSSION

Since late 2012, an outbreak of acute rheumatic fever (ARF) was observed in the Bologna province, Northeast Italy, where the annual frequency of ARF in resident children has ranged from 1 to 4 cases per year. From November 2012 to May 2013, eight cases of ARF were recorded, showing a significant increase of ARF in this area . Molecular analysis conducted among GAS collected from both contacts and the community during the outbreak indicated that M-18 represented one of the most prevalent serotype within classes of two unrelated ARF episodes. Our findings showed that two distinct subtypes, i.e. *emm18.29* and *emm18.32*, were observed to spread separately across the two classes of ARF cases. Analysis of GAS isolated from children in the community showed that the majority of the isolates were subtype *emm6.4.* Phenotypic analysis of GAS showed that the M-18 GAS strains presented a high frequency of mucoid strains, thus confirming previous findings (5). Indeed, GAS mucoid isolates have been observed to correlate with invasive infections and the

Journal of Clinical Microbiology

pathogenesis of rheumatic fever (1). Previous studies clearly demonstrated that higher capsule production has been observed in M-18 strains responsible for multiple ARF outbreaks (11). Although our study evidenced that GAS M-18 was responsible of an ARF case and that this serotype has spread in the two classes, we cannot exclude the possibility that others strains (i.e. serotypes) have been responsible of others ARF cases.

The present study evaluated the ability of MALDI-TOF MS to correctly identify GAS and to distinguish among the different serotypes. MALDI-TOF analysis demonstrated an excellent discrimination of M-18 GAS strains by showing a separate cluster. Overall, cluster analysis by MALDI-TOF showed a good concordance (74%) with *emm*-typing methods. In detail, a 100% concordance was observed with serotypes M-28 (9 out 9) and M-3 (2 out 2), whereas 75% and 83.4% concordances were found with M-1 (3/4) and M-5 (5/6), respectively. The discrepancies observed within these serotypes were attributable to two isolates, SPBO13 (*emm5.3*) and SPBO17 (*emm1.0*), grouped separately from strains belonging to the same serotypes. These results are in accordance with a previous study demonstrating the potential of the MALDI-TOF Biotyper System for GAS clustering analysis, thus showing a high discriminatory power among different serotypes (9). However, a poor discriminatory classification was observed for M-89 serotype. Based on these findings and for its rapidity and low cost analysis, we suggest the MALDI-TOF technique could be used successfully for the identification of the M-18 serotype among GAS strains.

In recent years, whole genome sequencing (WGS) has been applied for epidemiological purposes by showing a more accurate resolution than classical genotypic methods (13). WGS has been extensively used to explore the genetic organization of bacterial genomes and to compare the rearrangements between closely related strains (10,12,28). The present study described the complete genomes of nine M-18 GAS strains isolated during an ARF outbreak in northeastern part of Italy

Journal of Clinical Microbiology

and compared them with two M-18 *S.pyogenes* collected from non-invasive infections in the same area 30 years ago. SNP phylogenetic analysis revealed that *emm18.29* and *emm18.32* subtypes segregated in two separate clusters, whereas *emm18.0* GAS strains did not cluster in a distinct group. Genome polymorphism analysis demonstrated that isolates from ARF cases, community and class contacts were closely related showing a low number of informative SNPs both in *emm*-18.29 and *emm*-18.32 strains.

Our analysis revealed that M-18 strains possessed several virulence genes, including *speA*, *speC*, *speL*, *speM*, *smeZ*, *mitogenic factor*, and *hyaluronidase*, most of them located in prophage elements. Integrated prophages represent one of the most divergent tracts among GAS genomes and the majority of genetic variations among M-18 GAS strains (29).

Our results indicated that prophage elements are located in the same genomic locations among different M-18 strains collected in this study and in the MGAS8232 reference strain. In addition, genomic organization revealed that three prophage elements (ΦSPBO.2, ΦSPBO.3 and ΦSPBO.5) were common among M-18 GAS strains. Alignment of the prophage sequences showed a similar architecture between subtypes *emm18.29*, *emm18.32* and *emm18.0*, thus revealing a similar genomic architecture of M-18 GAS strains. It has been established that hyper-virulent GAS strains acquire virulence factors via prophage integrations (29). Recently, Bao *et al.* reported that prophage integrations represent one of the multiple genetic factors related to the pathogenic role of the M-23ND GAS strain (30). Our findings showed that three prophages (ΦSPBO.1, ΦSPBO.3, ΦSPBO.5) present in the M-18 GAS strains were similar with phages of M-3 serotypes (Φ315.1, Φ315.2 and Φ315.4) that have been previously associated with the emergence of virulent M-3 subclones (26) by different sequential acquisition.

371 We showed that the streptococcal pyrogenic exotoxin A gene (*speA1*) was located in the prophage
372 ΦSPBO.1 region in six out of nine *emm18.29* and *emm18.32* strains, whereas this gene was absent
373 in *emm18.0* strains. Also, we observed that the *speC* was present in all M-18 GAS strains
374 possessing *speL*/*speM,* thus showing a strictly correlation between these pro-phage encoded SAgs
375 genes. Based on these findings, we hypothesize that a different combination of phage-encoded
376 virulence factors could be related to the virulence of *emm18.29* and *emm18.32* strains isolated
377 during focal ARF that differed from non-invasive *emm18.0* strains collected from the same area 30
378 years ago.
379 Analysis of the phage-encoded virulence factors demonstrated that phage-encoded hyaluronidase
380 showed a higher number of synonymous and non-synonymous substitutions than other genes within
381 the M-18 GAS genomes. Previous studies reported that *hylP* and *hylP.2* genes were present with
382 different alleles among different serotypes from both invasive and non-invasive GAS isolates
383 (31,32). However, M-18 GAS strains have been observed to possess a unique *hylP.2* gene structure
384 among different isolates (33). Our findings demonstrated that the *hylP.2* gene possesses an internal
385 deletion located between the N-terminal and TSβH regions in different M-18 GAS strains.
386 Therefore, the truncated gene structures observed in several M-18 GAS isolates could be related to
387 a different or non-functional activity of the HylP.2 protein. Previous study showed that inactivation
388 in hyaluronate lyase (HylA) restored full encapsulation in partially encapsulated M-4 GAS strains,
389 thus demonstrating the mutually exclusive interaction between the hyaluronan capsule and active
390 hyaluronidase (32). In addition, Schommer *et al.* demonstrated in a mice model that the difference
391 in capsule size was regulated by bacterial hyaluronidase, and that the high molecular mass of the
392 hyaluronan capsule influences GAS virulence by facilitating deep tissue infections (34). Based on
393 our findings, we hypothesize that inactivation of HylP.2 could determine a different encapsulation
394 (i.e capsule sizing) of M-18 GAS strains, thus resulting in a more virulent clone. Therefore, the

Journal of Clinical Microbiology

internal deletion in the *hylP.2* gene observed in different isolates could reflect a different virulence potential among the M-18 GAS strains.

In conclusion, we investigated the molecular and epidemiological linkage between GAS strains isolated during an ARF outbreak in Bologna province in early 2013. Our study explored the genome sequence of M-18 GAS strains, thus providing a better understanding of the genetic architecture of the M-18 serotype and expanding our knowledge of the genetic elements related to the GAS infections.

ACKNOWLEDGMENTS

This work was supported by funds from Emilia-Romagna region

REFERENCES

1. **Cunningham MW**. 2014. Rheumatic fever, autoimmunity, and molecular mimicry: the streptococcal connection. Int. Rev. Immunol. **33**:314-329.

2. **Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, Sriprakash KS, Sanderson-Smith ML, Nizet V**. 2014. Disease manifestations and pathogenic mechanisms of group a *Streptococcus*. Clin. Microbiol. Rev. **27**:264-301.

3. **Tibazarwa KB, Volmink JA, Mayosi BM.** 2008. Incidence of acute rheumatic fever in the world: a systematic review of population-based studies. Heart. **94**:1534-1540.

4. **Wolfe RR.** 2000. Incidence of acute rheumatic fever: a persistent dilemma. Pediatrics. **105**:1375.

5. **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. Lancet Infect. Dis. **9**:611-616.

6. **Smoot JC, Korgenski EK, Daly JA, Veasy LG, Musser JM.** 2002. Molecular analysis of group A *Streptococcus* type *emm18* isolates temporally associated with acute rheumatic fever outbreaks in Salt Lake City, Utah. J. Clin. Microbiol. **40**:1805-1810.

7. **Sauer S, Kliem M.** 2010. Mass spectrometry tools for the classification and identification of bacteria. Nat. Rev. Microbiol. **8**:74-82.

8. **Mencacci A, Monari C, Leli C, Merlini L, De Carolis E, Vella A, Cacioni M, Buzi S, Nardelli E, Bistoni F, Sanguinetti M, Vecchiarelli A.** 2013. Typing of nosocomial outbreaks of *Acinetobacter baumannii* by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J. Clin. Microbiol. **51**:603-606.

9. **Wang J, Zhou N, Xu B, Hao H, Kang L, Zheng Y, Jiang Y, Jiang H.** 2012. Identification and cluster analysis of *Streptococcus pyogenes* by MALDI-TOF mass spectrometry. PLoS One. **7**:e47152.

10. **Sassera D, Comandatore F, Gaibani P, D'Auria G, Mariconti M, Landini MP, Sambri V, Marone A.** 2014. Comparative genomics of closely related strains of *Klebsiella pneumoniae* reveals genes possibly involved in colistin resistance. Ann. Microbiol. **64**:887–890

11. **Smoot JC, Barbian KD, Van Gompel JJ, Smoot LM, Chaussee MS, Sylva GL,Sturdevant DE, Ricklefs SM, Porcella SF, Parkins LD, Beres SB, Campbell DS, Smith TM, Zhang Q, Kapur V, Daly JA, Veasy LG, Musser JM.** 2002. Genome

437 sequence and comparative microarray analysis of serotype M18 group A *Streptococcus*
438 strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. USA.
439 **99**:4668-4673.

440 12. **Gaiarsa S, Comandatore F, Gaibani P, Corbella M, Dalla Valle C, Epis S, Scaltriti E,**
441 **Carretto E, Farina C, Labonia M, Landini MP, Pongolini S, Sambri V, Bandi C,**
442 **Marone P, Sassera D.** 2014. Genomic epidemiology of Klebsiella pneumoniae: the Italian
443 scenario, and novel insights into the origin and global evolution of resistance to carbapenem
444 antibiotics. Antimicrob. Agents Chemother., pii: AAC.04224-14.

445 13. **Aziz RK, Nizet V.** 2010. Pathogen microevolution in high resolution. Sci. Transl. Med.
446 **2**:16ps4.

447 14. **Carapetis JR, Parr J, Cherian T.** Standardization of epidemiologic protocols for
448 surveillance of post-streptococcal sequelae: acute rheumatic fever, rheumatic heart disease
449 and acute post-streptococcal glomerulonephritis. January 2006.
450 [http://www.niaid.nih.gov/topics/strepThroat/Documents/groupasequelae.pdf]

451 15. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for
452 interpretation of MICs and zone diameters. Version 4.0, 2014. [http://www.eucast.org]

453 16. **Beall B, Facklam R, Thompson T.** 1996. Sequencing emm-specific PCR products for
454 routine and accurate typing of group A Streptococci. J. Clin. Microbiol. 34:953–958.

455 17. **Friães A, Pinto FR, Silva-Costa C, Ramirez M, Melo-Cristino J.** 2013. Superantigen
456 gene complement of *Streptococcus pyogenes* relationship with other typing methods and
457 short-term stability. Eur. J. Clin. Microbiol. Infect. Dis. **32**:115-125.

18. **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. J. Bacteriol. **186**:4285-4294

19. **Andrews S.** FastQC A Quality Control tool for High Throughput Sequence Data

20. **Chevreux B, Wetter T, Suhai S.** 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99, pp. 45-56.

21. **Darling AE, Mau B, and Perna NT.** 2010. ProgressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. PLoS One. **5**:e11147.

22. **Huelsenbeck JP, Ronquist F.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. **17:**754–755.

23. **Chen LH, Xiong ZH, Sun LL, Yang J and Jin Q.** 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res. **40:**D641-645.

24. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Res. **39**:347-352.

25. **Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA.** 2011. Blast Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. **12:**402.

26. **Beres SB, Sylva GL, Barbian KD, Lei B, Hoff JS, Mammarella ND, Liu MY, Smoot JC, Porcella SF, Parkins LD, Campbell DS, Smith TM, McCormick JK, Leung DY, Schlievert PM, Musser JM.** 2002. Genome sequence of a serotype M3 strain of group A

Streptococcus: phage-encoded toxins, the high-virulence phenotype, and clone emergence. Proc. Natl. Acad. Sci. USA. **99**:10078-10083.

27. **Holden MT, Scott A, Cherevach I, Chillingworth T, Churcher C, Cronin A, Dowd L, Feltwell T, Hamlin N, Holroyd S, Jagels K, Moule S, Mungall K, Quail MA, Price C, Rabbinowitsch E, Sharp S, Skelton J, Whitehead S, Barrell BG, Kehoe M, Parkhill J.** 2007. Complete genome of acute rheumatic fever-associated serotype M5 *Streptococcus pyogenes* strain manfredo. J. Bacteriol. **189**:1473-1477.

28. **Parkhill J, Wren BW.** 2011. Bacterial epidemiology and biology lessons from genome sequencing. Genome Biol. **12**:230.

29. **Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H.** 2003. Prophage genomics. Microbiol Mol Biol Rev. **67**:238-276. Erratum in: Microbiol. Mol. Biol. Rev. **67**:473.

30. **Bao Y, Liang Z, Booyjzsen C, Mayfield JA, Li Y, Lee SW, Ploplis VA, Song H, Castellino FJ.** 2014. Unique Genomic Arrangements in an Invasive Serotype M23 strain of *Streptococcus pyogenes* Identify Genes that Induce Hypervirulence. J. Bacteriol. pii: JB.02131-14.

31. **Marciel AM, Kapur V, Musser JM.** 1997. Molecular population genetic analysis of a *Streptococcus pyogenes* bacteriophage-encoded hyaluronidase gene: recombination contributes to allelic variation. Microb. Pathog. **22**:209-217.

32. **Mylvaganam H, Bjorvatn B, Hofstad T, Osland A.** 2001. Molecular characterization and allelic distribution of the phage-mediated hyaluronidase genes *hylP* and *hylP2* among group A streptococci from western Norway. Microb. Pathog. **30**:311.

33. **Henningham A, Yamaguchi M, Aziz RK, Kuipers K, Buffalo CZ, Dahesh S, ChoudhuryB, Van Vleet J, Yamaguchi Y, Seymour LM, Ben Zakour NL, He L, Smith HV, Grimwood K, Beatson SA, Ghosh P, Walker MJ, Nizet V, Cole JN.** 2014. Mutual Exclusivity of Hyaluronan and Hyaluronidase in Invasive Group A *Streptococcus*. J. Biol. Chem. pii: jbc.M114.602847.

34. **Schommer NN, Muto J, Nizet V, Gallo RL.** 2014. Hyaluronan Breakdown Contributes to Immune Defense against Group A *Streptococcus*. J. Biol. Chem. **289**:26914-26921.

519 **Figure 1.** Distribution of *emm*-types among *S.pyogenes* strains collected during an ARF outbreak
520 occurred in Bologna province. (A) GAS strains isolated from the ARF case and class contacts
521 during the first episode. (B) GAS isolated from class children during the second ARF episode. (C)
522 Distribution of *emm*-types among GAS collected from the community at the time of the first ARF
523 case. (D) Distribution of *emm*-types among GAS collected from school-age children of the
524 community at time of the ARF episode.

525 **Figure 2.** Score-oriented dendrogram based on the main spectra (MSP) of 50 GAS strains obtained
526 by Bruker MALDI-TOF MS and analyzed with Biotyper 3.1 software. Correlation with the 11
527 different *emm*-types (M-1, M-3, M-5, M-6, M-9, M-12, M-18, M-28, M-44, M-89 and M-102)
528 shown. Dotted lines define a similarity cutoff value of 500, 300 and 200 used for clustering groups
529 of serotypes M-18, M-1, M-3, M-5, M-28, and M-6.

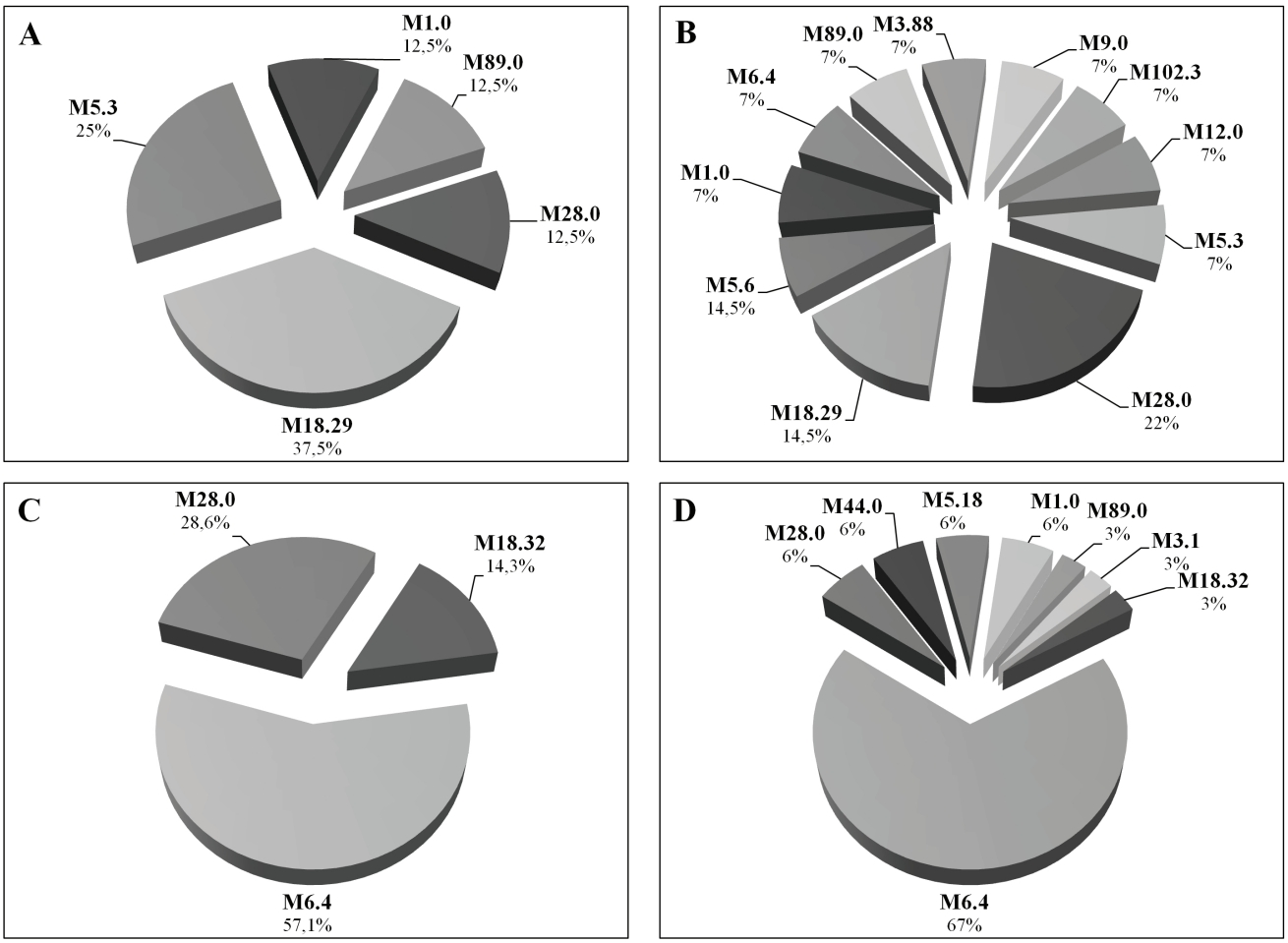
530 **Figure 3.** Comparison of the group A *Streptococcus* serotype M18 chromosomes. The circular
531 representation shows the genome comparison from centre to periphery respectively of the strains:
532 GAS8232, SPBO1, SPBO2, SPBO3, SPBO4, SPBO5, SPBO6, SPBO7, SPBO8, SPBO9, SP665Q
533 and SP1065Q (see legend for color association). The regions of differences within GAS genomes
534 are indicated with white gaps. The genomic localizations of the prophage elements (PEs) shared
535 with GASM8232 are indicated as black boxes outside the circular GAS chromosome maps.
536 Locations of the prophage regions (ΦSPBO.1, ΦSPBO.2, ΦSPBO.3, ΦSPBO.4, ΦSPBO.5,
537 ΦSPBO.6) of M-18 GAS isolates collected from Bologna province are indicated as red boxes.

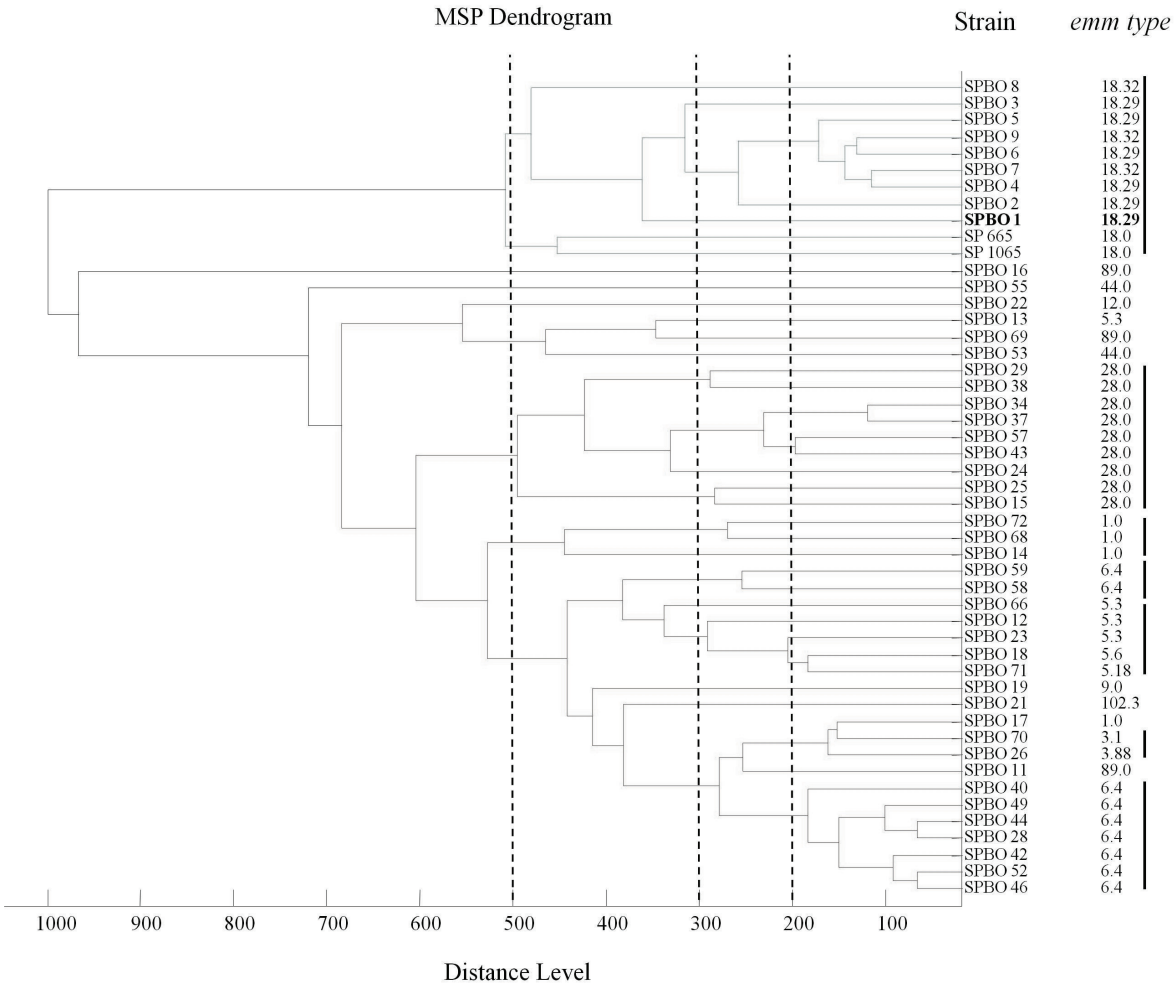
538 **Figure 4.** Bayesian tree-based on core SNPs identified with the Mauve-based approach. In each
539 node of the tree, posterior probabilities (>0.7) are indicated on the right of the node, while the N. of
540 different and informative SNPs located respectively up and down the branch are on the left of the
541 node.

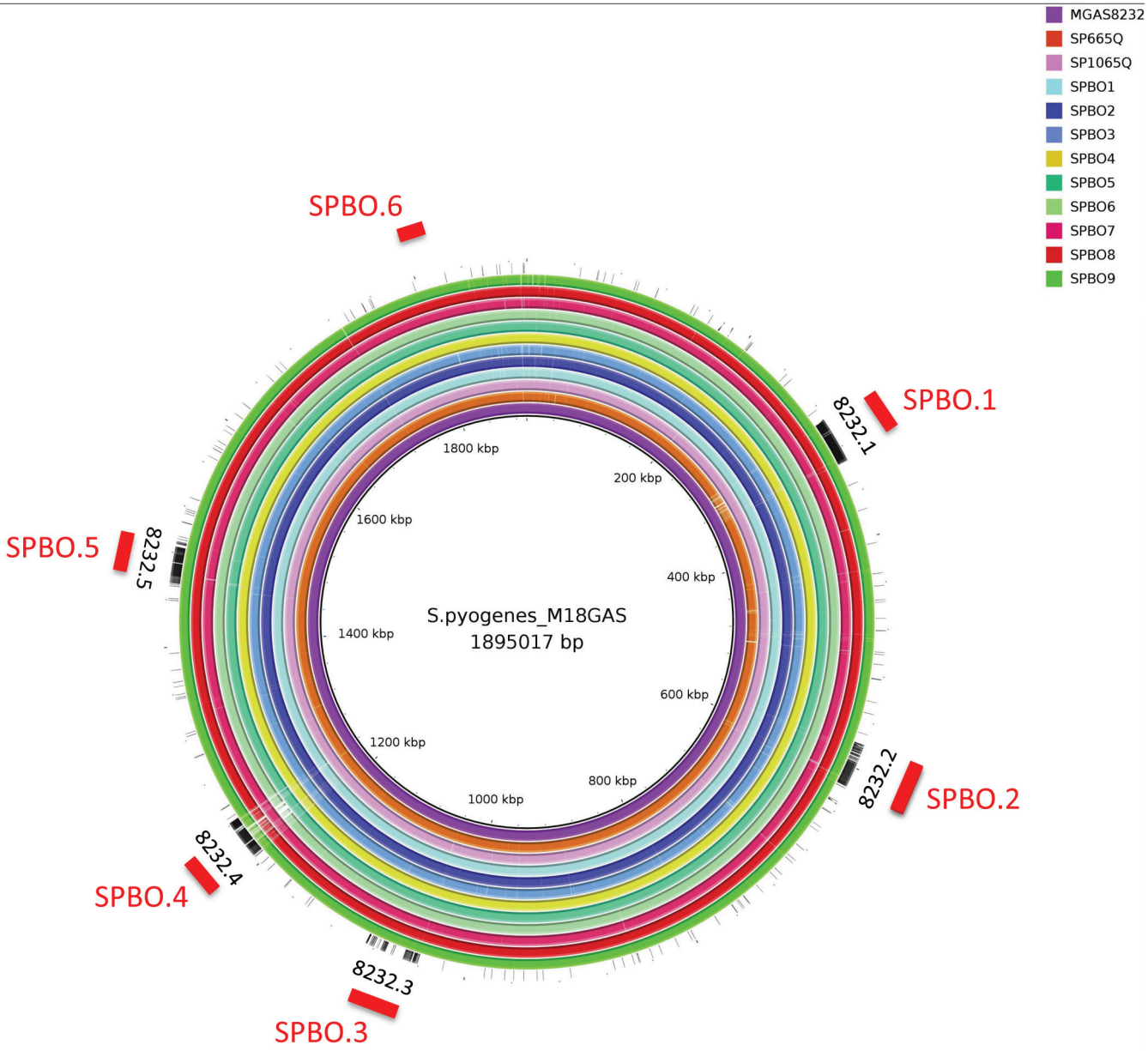
Journal of Clinical Microbiology

Figure 5. Circular representation of concatenated prophage elements (PEs) integrated in the genome of the M-18 GAS strains in comparison with M-3 GAS prophages . The outermost black circle represent the concatenated M-3 GAS prophages (Φ315.1, Φ315.2, Φ315.3, Φ315.4). The prophages of M-18 GAS strains collected in this study (SPBO1, SPBO2, SPBO3, SPBO4, SPBO5, SPBO6, SPBO7, SPBO8, SPBO9, SP1065Q and SP665Q) and reference strain (M18GAS8232) are indicate in red and blue, respectively. The areas of similarity and divergence are contrasted with white gapped areas indicating regions of highest variance.

Figure 6. Alignment of the phage-encoded *hyaluronidase* gene (*hylP.2*) derived from M-18 GAS (*emm18.29*, *emm18.32*, *emm18.0* subtypes). The *hylP.2* sequences from ARF-related M-18 (GAS8232) and M-5 (Manfredo) strains are shown. Dotted lines shown the common deleted regions between α-helical and TSβH domains between SPBO1, SPBO3 and SPBO6 strains. The deduced amino acid sequence of a 327 nt deletion in *emm18.29* strains is shown.



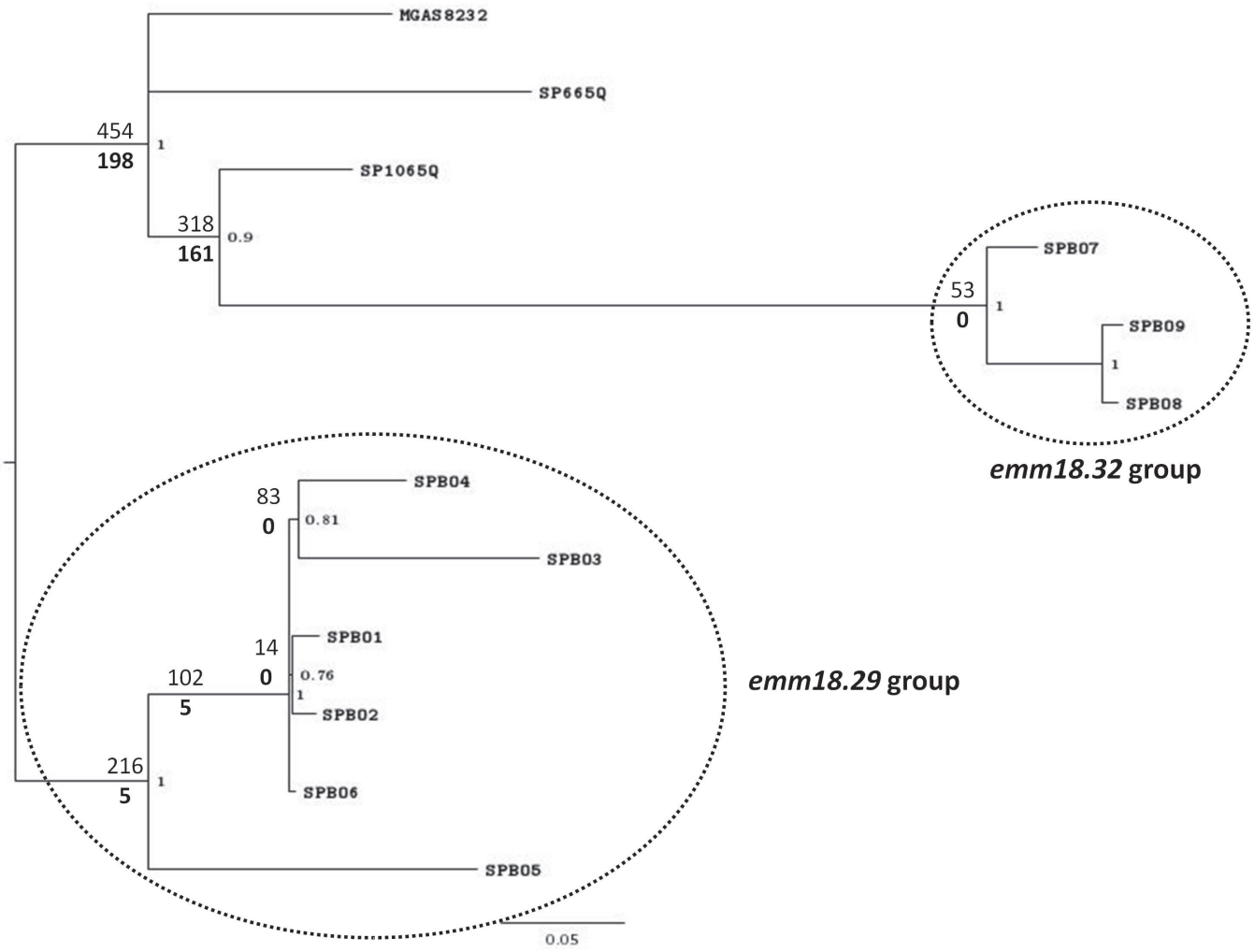


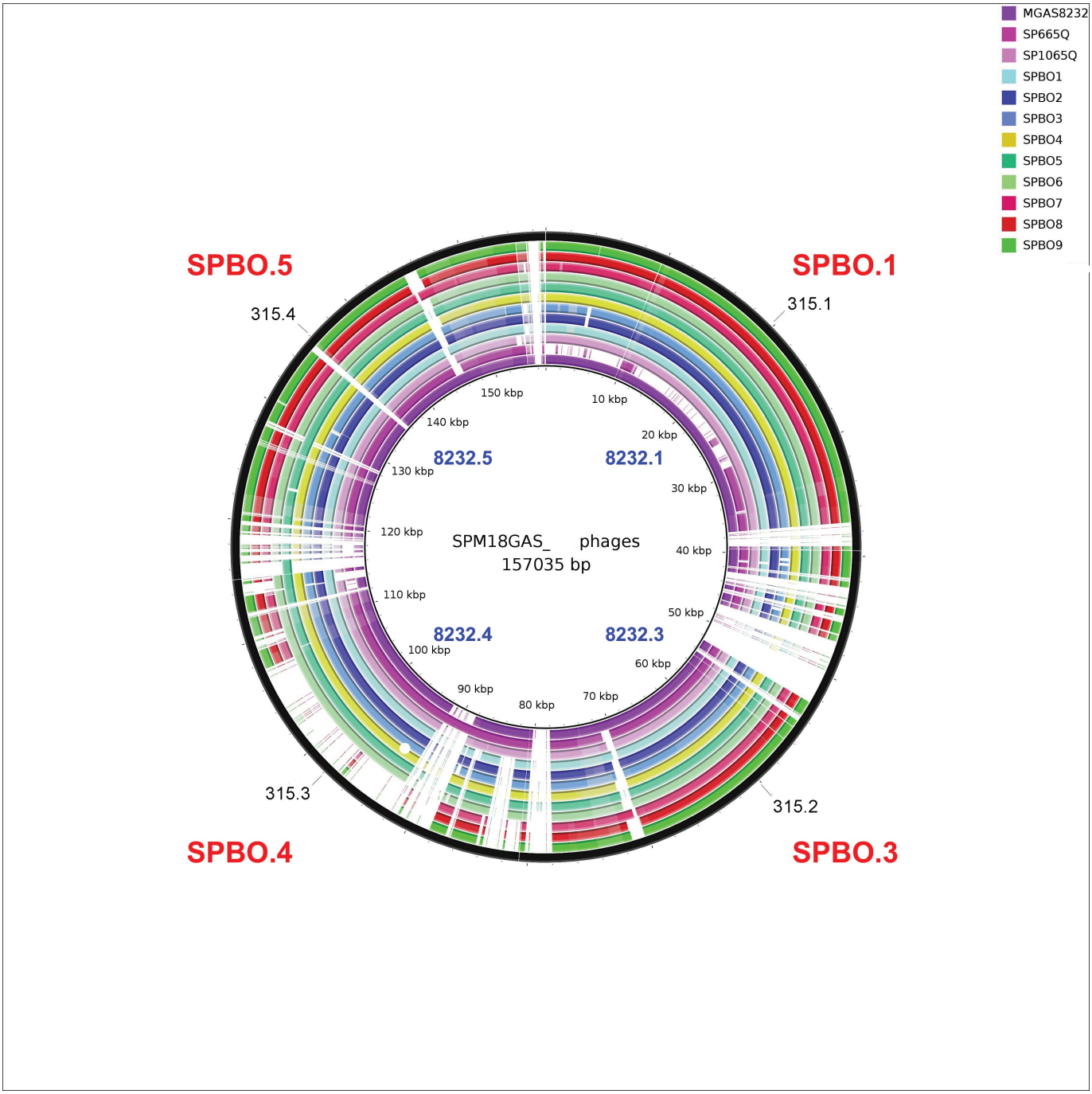


Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology





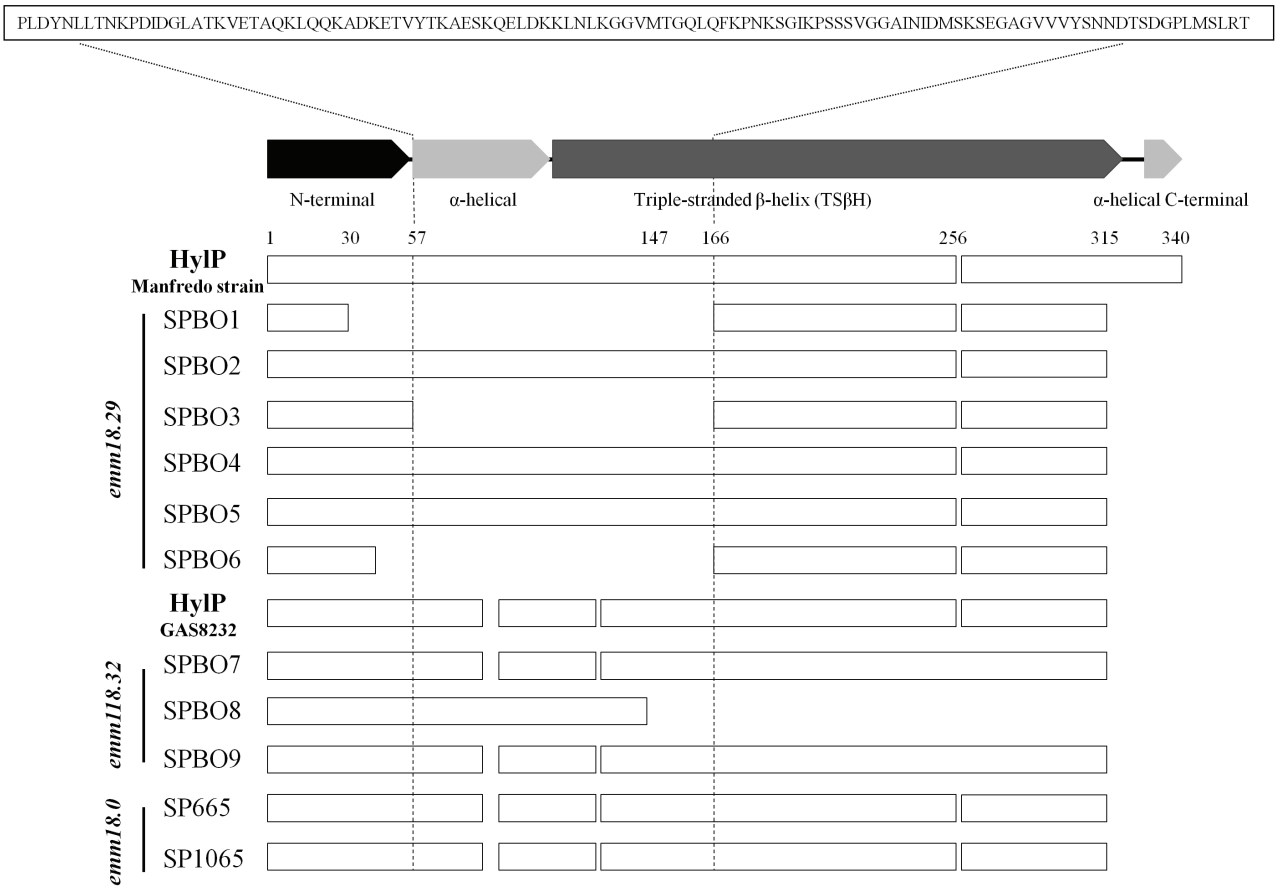


Table 1. Superantigens and *emm*-type of GAS strains collected during an ARF outbreak

emm subtypes	Superantigen genes (Number of isolates)											Antigenic Profile	Number of isolates
	SpeA	SpeC	SpeG	SpeH	SpeI	SpeJ	SpeK	SpeL	SpeM	ssa	smeZ		
6.4	-	-	32	-	-	-	32	-	-	-	32	**1**	32
28.0	-	-	9	-	-	9	-	-	-	-	9	**2**	9
18.29[b]	6	6	6	-	-	-	-	6	6	-	6	**3**	6
18.32	3	3	3					3	3		3	**3**	3
5.3	-	-	4	-	-	-	-	-	-	-	4	**4**	4
5.6			1								1	**4**	1
5.18			1								1	**4**	1
1.0	3	-	4	-	-	4	-	-	-	-	4	**5**	4
44.0	-	-	2	-	-	2	-	2	2	-	2	**6**	2
89.0	-	-	3	-	-	-	-	-	-	-	3	**4**	3
3.88	1	-	1	-	-	-	-	-	-	-	1	**7**	1
3.1			1				1				1	**1**	1
9.0	-		1	-	-	-	-	-	-	-	1	**4**	1
102.3	-		1	-	-	-	-	-	-	-	1	**4**	1
12.0	-		1	1	1	-	1	-	-	-	1	**8**	1
Total	13	9	70	1	1	15	34	11	11		70		70

[b] Two consecutive GAS strains were isolated from the first ARF case

Table 2. Pro-phage elements in M-18 GAS strains

	GAS strains											
	emm18.29						*emm18.32*			*emm18.0*		Virulence factors
Phage (Kb)	SPBO 1	SPBO 2	SPBO 3	SPBO 4	SPBO 5	SPBO 6	SPBO 7	SPBO 8	SPBO 9	SP665 Q	SP1065 Q	
ΦSPBO.1	56.4*	67.2	52.5	67.4*	64.9*	61.2*	68.7	75.8*	57.6*	-	66.1	*speA, hyaluronidase (hylP.2)*
ΦSPBO.2	32.5	38.5	36.9	39.5°	38.3°	37.2°	38	38.6°	37.3°	36.4°	39.6 °	*speC mitogenic factor, hyaluronidase*
ΦSPBO.3	47.8	68.5	59.4	57[#]	58.4[#]	59.4[#]	57.3	57.4[#]	59.1[#]	57.6[#]	67.1[#]	*speL, speM, hyaluronidase*
ΦSPBO.4	33	31.8	30.6	25.2	30.6	32.5	-	-	-	43.6	41.1	*hyaluronidase, mitogenic factor*
ΦSPBO.5	60	47.3	39.3	45.7	47.2	51	48.5	42.6	48	51.3	46.6	*hyaluronidase, streptodornase*
ΦSPBO.6	14.1	12.8	12.5	19.5	47.2	18.4	17	-	10.5	-	7.6	*hyaluronidase*

*Pro-phage containing variant of *speA* (*speA1*)

°Pro-phage containing *spec*

[#]Pro-phage containing *speL* and *speM*

Journal of Clinical Microbiology