Genomic analysis of a *Streptococcus pyogenes* strain causing endocarditis in a child

M. Beye[1], K. El Karkouri[1], N. Labas[1], D. Raoult[1,2] and P.-E. Fournier[1]

1) *URMITE, UM63, CNRS7278, IRD198, Inserm1095, Institut Hospitalo-Universitaire Méditerranée-infection, Aix-Marseille Université, Faculté de Médecine, Marseille, France and* 2) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdul Aziz University, Jeddah, Saudi Arabia*

Abstract

We sequenced the genome of *Streptococcus pyogenes* strain G773 that caused an infective endocarditis in a 4-year-old boy suffering from acute endocarditis. The 1.9-Mb genome exhibited a specific combination of virulence factors including a complete integrative and conjugative element, sp2905, previously described as incomplete in *S. pyogenes*, and five bacteriocin-coding genes. However, strain G773 lacked a CRISPR-Cas system.

Keywords: Endocarditis, genome comparison, real-time genome sequencing, *Streptococcus pyogenes*, virulence
Original Submission: 20 November 2016; **Revised Submission:** 3 January 2017; **Accepted:** 11 January 2017
Article published online: 28 January 2017

Corresponding author: P.-E. Fournier, URMITE, UM63, CNRS7278, IRD198, Inserm1095, Institut Hospitalo-Universitaire Méditerranée-infection, Aix-Marseille Université, Faculté de Médecine, 27 Bd Jean Moulin, 13385 Marseille cedex 5, France
E-mail: pierre-edouard.fournier@univ-amu.fr

Introduction

Streptococcus pyogenes, also referred to as Group A streptococcus, is a major human pathogen [1]. This bacterium colonizes the throat or skin and is responsible for a number of purulent infections [1,2], including pharyngitis, impetigo, necrotizing fasciitis and streptococcal toxic shock syndrome [2]. In addition, Group A streptococcus may induce autoimmune diseases such as acute post-streptococcal glomerulonephritis, rheumatic fever and rheumatic heart disease [2]. In addition, rare cases of *S. pyogenes* endocarditis have been reported in the literature [3]. Many virulence determinants have been identified using genomic and molecular analyses of *S. pyogenes* strains, many of which have a redundancy in the processes of adhesion and colonization, innate immune resistance, and ability to facilitate degradation of tissue barriers and spread in the human body [2].

In the present study, we report a case of acute *S. pyogenes* endocarditis in a child. To determine whether this strain had specific virulence factors, we sequenced its genome and compared it with those of other *S. pyogenes* strains available in public databases.

Case report

On 10 October 2014, a 4.5-year-old boy was admitted to the emergency room in the Timone University Hospital, Marseille, France, for drowsiness and severe dehydration that complicated a gastroenteritis episode. In the past week, he had developed abdominal pain, vomiting and profuse watery diarrhoea. On abdominal ultrasonography, an ileitis was diagnosed. A brain magnetic resonance imaging was normal. He was admitted to the paediatric ward where a jugular vein catheter was implanted for rehydration. On 23 October 2014, he developed a fever of 40 °C. Blood tests revealed a leucocyte count of 42 × 10^9/L (81% polymorphonuclear cells) and a C-reactive protein level of 214 mg/L. A thrombosis of the jugular catheter required its withdrawal. Subsequently, a trans-thoracic echocardiogram revealed a mitral valve vegetation. Three blood cultures taken at 30-minute intervals were positive for *S. pyogenes*. Following an intravenous treatment with amoxicillin, 200 mg/kg/day, and gentamicin, 3 mg/kg/day, for 2 weeks, and then oral amoxicillin, 200 mg/kg/day, for 4 weeks, the patient recovered fully.

http://dx.doi.org/10.1016/j.nmni.2017.01.007

Materials and methods

Genome sequencing

Genomic DNA (gDNA) from *S. pyogenes* strain G773, isolated from the patient's blood, was sequenced using the Paired-End strategy on the MiSeq sequencer (Illumina Inc, San Diego, CA, USA) with 16 other genomic projects using the Nextera XT DNA sample prep kit (Illumina). The Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA) was used to quantify the gDNA at 7.34 mg/L. To prepare the Paired-End library, the gDNA was diluted to obtain 1 ng of each genome as input. The "tagmentation" step fragmented and tagged the DNA. Then limited cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter Inc, Fullerton, CA, USA), the libraries were normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled for sequencing on the MiSeq sequencer. The pooled single-strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and Paired-End sequencing with dual index reads were performed in a single 39-hour run in 2 × 250-bp.

Total information of 7.8 Gb was obtained from a 871 K/mm^2 cluster density with a cluster passing quality control filters of 80.5% (18 857 000 clusters). Within this run, the index representation for the *S. pyogenes* genome was determined to be 6.37%. The 966 853 Paired-End reads were trimmed and filtered according to the read qualities.

Phylogenetic analyses

To identify the most closely related strains to *S. pyogenes* strain G773, phylogenetic trees based on sequences from the M protein (*emm* gene) [4–6] were performed using the maximum likelihood (ML) method with LG (+F) models and the neighbour-joining (NJ) method with JTT model in the MEGA version 6 software. Bootstrap replicates were set to 100 and 1000 for the ML and NJ trees, respectively. This was carried out after a BLASTP search for homologous proteins and multiple sequence alignment using the Muscle algorithm [7]. ML and NJ analyses were first performed using M protein sequences from strain G733 and other *S. pyogenes* strains with closely related protein sequences in GenBank, as well as 32 representative sequences of the M protein from each of the 16 emm-clusters and the X and Y clades defined in the study from Sanderson-Smith *et al.* [6] (see Supplementary material, Figs S1 and S2). We also performed ML and NJ phylogenetic analyses focusing on M protein sequences from *S. pyogenes* strains classified in the X clade (Fig. 1).

Genomic analysis

The genomic reads obtained from sequencing were assembled using the A5 assembler [8]. Then, a step of finishing was done using the Mauve software and CLC bioserver [9]. After assembly and finishing, the genome size was 1.9 Mb. Open reading frames were predicted using the Prodigal tool (http://prodigal.ornl.gov) with default parameters. The prediction of protein function was performed by searching against GenBank database using the BLASTP algorithm [10]. Functional classification of gene families was researched using COGnitor against the COG (Clusters of Orthologous Groups) database [11]. The tRNAs and rRNAs were detected using tRNAscan-SE v.1.21 and RNAmmer v.1.2 [12], respectively. The presence or absence of plasmids was verified by searching the gene annotation for any plasmid-related gene and by mapping all contigs against previously published *Streptococcus* sp. plasmid sequences.

To identify putative orthologues and estimate the pan/core-genome composition, comparative genomic analysis was carried out between the two *S. pyogenes* strains G773 and MGAS10270 using bidirectional Best Blast from the BLASTClust algorithm [13], and then specific genes were checked by tBLASTN. Visual inspections were also performed to distinguish genes that are either complete (coverage $\geq$80% of the longest orthologue), split (gene with at least two coding DNA sequences) or fragmented (coverage <80% of the longest orthologue). Before performing these analyses, the MGAS10270 genome was downloaded from GenBank database and then subjected to open reading frame scanning as was done for strain G773. Open reading frame sequences were also submitted to the Rapid Annotation Using Subsystem Technology (RAST) [13] for annotation.

The CRISPRFinder algorithm was used to identify putative CRISPR (clustered regularly interspaced short palindromic repeat) loci [14]. The prediction of genes coding resistance to antibiotics was performed by BLASTP against the ARG-ANNOT database using an e-value of 10^{-5} [15]. These putative antibiotic resistance genes were further verified by a BLASTn search against GenBank.

Nucleotide sequence accession numbers

The genome sequence from *S. pyogenes* strain G773 was deposited in GenBank under Accession numbers CVUH01000001 to CVUH01000022.

Results

General genomic features

The draft genome sequence of *S. pyogenes* strain G773 consisted of 22 scaffolds after assembly and finishing. No putative plasmid sequence was detected. The chromosome size, G+C

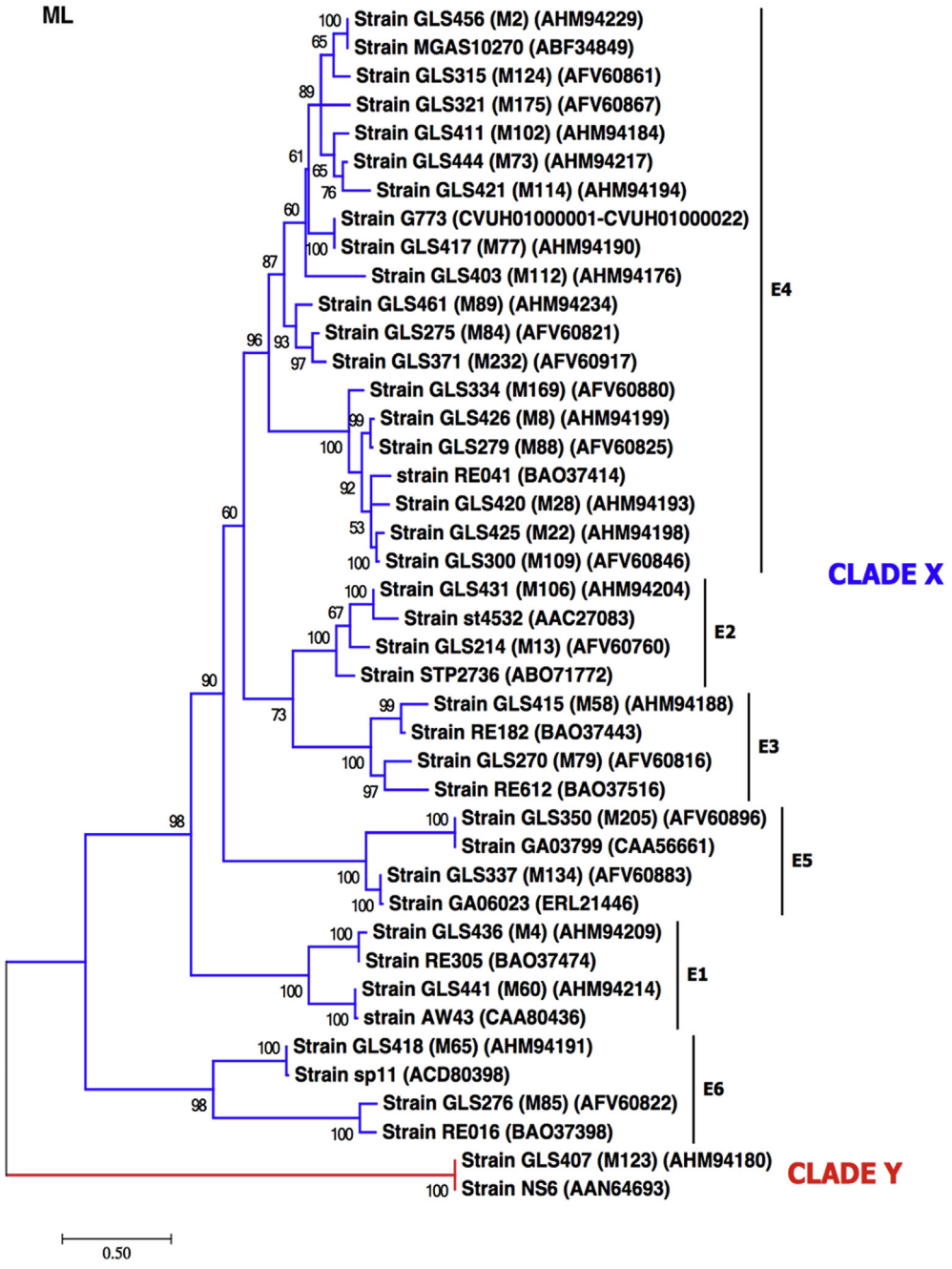

FIG. 1. Phylogenetic maximum likelihood tree of *Streptococcus pyogenes* strains using M proteins. Percentages under the branches correspond to bootstrap values.

and coding DNA sequence contents were 1 948 513 bp, 38.3% and 1886, respectively. A total of 82 RNA genes were identified (one complete rRNA operon, six other 5S rRNAs, 72 tRNAs and one tmRNA). Of the 1886 predicted coding DNA sequences, 1458 (77.3%) were assigned a putative function and 428 (22.7%) were annotated as hypothetical proteins. A total of 1457 (74.03%) genes were assigned a COG functional category.

Phylogenetic analyses

Both ML and NJ-based phylogenetic trees provided similar topologies. Strain G773 belonged to the emm-cluster E4 within the X clade described by Sanderson-Smith *et al.* [6] and is closely related to *S. pyogenes* strains RE041 and MGAS10270 (see Supplementary material, Fig. S1). This result was confirmed by a second phylogenetic analysis focusing on members of the

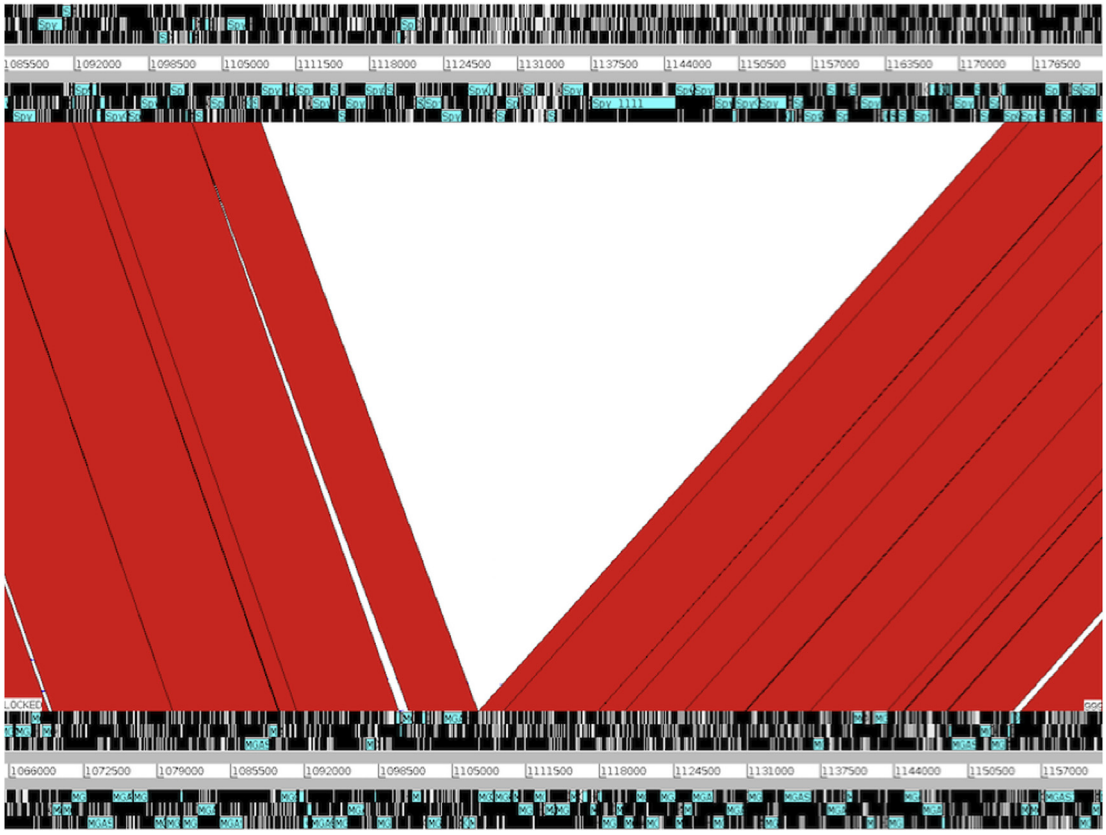

FIG. 2. Example of a specific region identified in the genome of *Streptococcus pyogenes* strain G773 (above, scaffold6) compared with the genome of the strain MGAS10270 (below), using the Artemis Comparion Tool.

emm-cluster E4 (Fig. 1). As only the genome from strain MGAS10270 was sequenced at the time of our study (GenBank Accession number CP000260), we used it for comparative genomic analysis.

Genomic analysis of *S. pyogenes* strains G773 and MGAS10270

The genomic comparison identified a pangenome of 2046 genes and a core genome of 1731 genes. Strains G733 and MGAS10270 harboured 160 and 155 specific genes, respectively (Fig. 2 and Table 1; see Supplementary material, Fig. S3). However, when compared with GenBank, all specific genes had at least one orthologue in at least one *S. pyogenes* strain.

The COG functional classification of the 160 strain G773-specific genes revealed that 17 (10.6%) were involved in transcription, 16 (10%) in replication, 8 (5%) in defensive mechanisms, 6 (3.75%) in metabolism and 4 (2.5%) in intracellular trafficking and secretion. In addition, 12 genes (7.5%) were classified as poorly characterized (Table 1). Moreover, among these G773-specific genes, five were annotated as bacteriocin-type signal sequence, ABC-type bacteriocin transporter, bacteriocin secretion accessory protein, class IIb bacteriocin lactobin

TABLE 1. **COG categories of genes either specific or absent from *Streptococcus pyogenes* strain G773**

		Strain G773-specific genes		Genes absent in strain G773	
Functional category	**Letter**	**Genes**	**%**	**Genes**	**%**
Cell cycle control	D	0	0	0	0
Cell wall/membrane/envelope biogenesis	M	2	1.25	7	4.52
Cell motility	N	0	0	2	1.29
Post-translational modification	O	0	0	0	0
Signal transduction mechanisms	T	1	0.62	2	1.29
Intracellular trafficking and secretion	U	4	2.50	3	1.94
Defence mechanisms	V	8	5	6	3.87
Chromatin structure and dynamics	B	0	0	0	0
Translation	J	4	2.5	1	0.64
Transcription	K	17	10.62	12	7.75
Replication	L	16	10	10	6.45
Energy production and conversion	C	0	0	0	0
Amino acid transport and metabolism	E	0	0	1	0.64
Nucleotide transport and metabolism	F	0	0	2	1.29
Carbohydrate transport and metabolism	G	1	0.62	0	0
Coenzyme transport and metabolism	H	1	0.62	7	4.52
Lipid transport and metabolism	I	2	1.25	0	0
Inorganic ion transport and metabolism	P	2	1.25	1	0.64
Secondary metabolites biosynthesis	Q	0	0	1	0.64
General function prediction only	R	3	1.87	6	3.87
Function unknown	S	12	7.5	11	7.10
Total in COGs		73	45.6	72	46.45
Not in COGs		87	54.4	83	53.55
Total		160	100	155	100

COG, Clusters of Orthologous Groups.

TABLE 2. *In silico* prediction of antibiotic resistance genes in genome of the strain G773

Class	Best match	Length (amino acids)	Related spp.	Similarity (%)	Coverage (%)	Accession numbers
Tetracyclin	TetO	640	*Streptococcus pneumoniae*	100	99	CAQ76847
MLS	Erm	244	*Streptococcus pyogenes*	99	99	WP_019108019

A/cerein or bacteriocin (see Supplementary material, Table S1). No CRISPR region was found in the genome of strain G773. In contrast, two genes involved in resistance to tetracycline and erythromycin were detected in this genome (Table 2). These antibiotic resistances were previously found in *S. pyogenes* in the integrative and conjugative element 2905 (ICEsp2905, Table 2) [16]. Following careful verification, strain G773 possessed a complete ICEsp2905 sequence (Table 3). Table 3 shows that the genome from strain G773 is the only *S. pyogenes* genome containing a complete ICEsp2905 sequence.

Discussion

Streptococcus pyogenes is the most frequent bacterial agent of acute pharyngitis and is also responsible for a variety of cutaneous and systemic infections [17]. Among those, *S. pyogenes* may cause endocarditis, even in people without any pre-existing risk factors [18]. However, the incidence of *S. pyogenes* endocarditis has declined significantly since the introduction of antibiotics, largely due to the systematic treatment of local pyogenic infections preventing bacteraemia and infection of the endocardium [17]. Currently, endocarditis caused by *S. pyogenes* is rare in any age group, representing <5% of all documented cases, but it is often severe and its onset is almost always acute [3,18].

The key factor in the pathogenicity of *S. pyogenes* is the M protein [19]. This surface-exposed antigenic protein is involved in the adhesion of the bacterium to human tissues and in preventing phagocytosis. It has also been used widely to study the bacterium's diversity [19]. The phylogenetic analysis using the *emm* gene classified strain G773 within the X clade. In this cluster, the M protein is able to bind to C4b-binding protein (C4BP) and immunoglobulin [6]. In addition, *S. pyogenes* can produce several exotoxins that have the potential to damage the host tissues either directly or through the stimulation of cytokine production [17].

TABLE 3. BLAST results of the integrative and conjugative element sp2905

Species	Coverage (%)	E-value	Identity (%)	Accession numbers
Streptococcus pyogenes strain G773	100	0.0	99	CVUH01000001
Streptococcus anginosus strain C238	68	0.0	92	CP003861
Streptococcus pyogenes strain HKU360	57	0.0	89	CP009612
Streptococcus agalactiae strain GBS6	56	0.0	89	CP007572
Peptoclostridium difficile strain CD630DERM	51	0.0	86	LN614756
Clostridium difficile 630	51	0.0	86	AM180355
Streptococcus constellatus subsp. *pharyngis* C1050	51	0.0	86	CP003859
Streptococcus anginosus strain SA1	50	0.0	83	CP007573
Streptococcus agalactiae strain GBS1-NY	49	0.0	89	CP007570
Streptococcus dysgalactiae subsp. *equisimilis* RE378	49	0.0	85	AP011114

By comparing the genome of strain G773 to those of other closely related strains and species, we observed that the former was the only one to exhibit a complete ICEsp2905 sequence. The ICEsp2905 carried genes coding tetracycline- and erythromycin-resistance [16], its presence in a genome has been proven to confer resistance to these antibiotics [16]. This was the case in strain G773.

In addition, strain G773 exhibited three genes encoding bacteriocins and two genes encoding an ABC-type bacteriocin transporter and bacteriocin signal sequence (see Supplementary material, Table S1). Bacteriocins are proteins with bactericidal activity principally against strains of species that are closely related to the producer bacterium [20]. They are known to be produced by several *S. pyogenes* strains [21]. The production of bacteriocins can be favourable for colonizing the host by eliminating other bacterial species that share the same environment [22]. In contrast, strain G773 was devoid of the CRISPR-Cas system (see Supplementary material, Table S2). CRISPR-Cas systems are considered as a defence mechanism of bacteria [23]. They confer adaptive immunity against exogenous elements in many bacteria [24] and may limit the horizontal transfer of genes including insertion prophages or plasmids [23].

Overall, the genomic analysis of *S. pyogenes* strain G773, which caused an acute endocarditis in a child, possessed a combination of various pathogenesis factors that may explain its particular virulence.

Conflict of Interest

The authors have no conflict of interest to disclose.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.nmni.2017.01.007.

References

[1] Burkert T, Watanakunakorn C. Group A streptococcus endocarditis: report of five cases and review of literature. J Infect 1991;23:307–16.

[2] Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, et al. Disease manifestations and pathogenic mechanisms of group A *Streptococcus*. Clin Microbiol Rev 2014;27:264–301.

[3] Yeşilkaya A, Azap ÖK, Pirat B, Gültekin B, Arslan H. A rare cause of endocarditis: *Streptococcus pyogenes*. Balk Med J 2012;29:331–3.

[4] McMillan DJ, Drèze P-A, Vu T, Bessen DE, Guglielmini J, Steer AC, et al. Updated model of group A *Streptococcus* M proteins based on a comprehensive worldwide study. Clin Microbiol Infect 2013;19:E222–9.

[5] Bae SY, Kim JS, Kwon J-A, Yoon S-Y, Lim CS, Lee KN, et al. Phenotypes and genotypes of macrolide-resistant *Streptococcus pyogenes* isolated in Seoul, Korea. J Med Microbiol 2007;56:229–35.

[6] Sanderson-Smith M, Oliveira DMPD, Guglielmini J, McMillan DJ, Vu T, Holien JK, et al. A systematic and functional classification of *Streptococcus pyogenes* that serves as a new tool for molecular typing and vaccine development. J Infect Dis 2014;210:1325–38.

[7] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7.

[8] Tritt A, Eisen JA, Facciotti MT, Darling AE. An integrated pipeline for de novo assembly of microbial genomes. PLoS One 2012;7.

[9] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.

[10] Altschul SF. BLAST AlgorithmeLS. Chichester: John Wiley & Sons, Ltd; 2001.

[11] Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 2000;28:33–6.

[12] Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.

[13] Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 2008;9:75.

[14] Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 2007;35:W52–7.

[15] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212–20.

[16] Giovanetti E, Brenciani A, Tiberi E, Bacciaglia A, Varaldo PE. ICESp2905, the erm(TR)-tet(O) element of *Streptococcus pyogenes*, is formed by two independent integrative and conjugative elements. Antimicrob Agents Chemother 2012;56:591–4.

[17] Weidman DR, Al-Hashami H, Morris SK. Two cases and a review of *Streptococcus pyogenes* endocarditis in children. BMC Pediatr 2014;14:227.

[18] Branch J, Suganami Y, Kitagawa I, Stein GH, Tanaka E. A rare case of group A streptococcal endocarditis with absence of valvular vegetation. Intern Med Tokyo Jpn 2010;49:1657–61.

[19] Bidet P, Bonacorsi S. Facteurs de pathogénicité de *Streptococcus pyogenes*. Arch Pédiatrie 2014;21(Suppl. 2):S54–61.

[20] Jack RW, Tagg JR. Factors affecting production of the group A streptococcus bacteriocin SA-FF22. J Med Microbiol 1992;36:132–8.

[21] Tagg JR, Bannister LV. "Fingerprinting" β-haemolytic streptococci by their production of and sensitivity to bacteriocine-like inhibitors. J Med Microbiol 1979;12:397–411.

[22] Lux T, Nuhn M, Hakenbeck R, Reichmann P. Diversity of bacteriocins and activity spectrum in *Streptococcus pneumoniae*. J Bacteriol 2007;189:7741–51.

[23] Nozawa T, Furukawa N, Aikawa C, Watanabe T, Haobam B, Kurokawa K, et al. CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. PLoS One 2011;6.

[24] Louwen R, Staals RHJ, Endtz HP, van Baarlen P, van der Oost J. The role of CRISPR-Cas systems in virulence of pathogenic bacteria. Microbiol Mol Biol Rev MMBR 2014;78:74–88.