A clone of the emergent Streptococcus pyogenes emm89 clade responsible for a large outbreak in a post-surgery oncology unit in France

Céline Plainvert, Magalie Longo, Elise Seringe, Benjamin Saintpierre, Elisabeth Sauvage, Laurence Ma, Johann Beghain, Nicolas Dmytruk, Gislène Collobert, Eric Hernandez, et al.

REVISED MMIM-D-18-00037

A clone of the emergent *Streptococcus pyogenes emm89* clade responsible for a large outbreak in a post-surgery oncology unit in France

Céline Plainvert[1,2,3,4], Magalie Longo[1,2,3], Elise Seringe[5], Benjamin Saintpierre[1,2,3], Elisabeth Sauvage[6,7], Laurence Ma[8], Johann Beghain[9], Nicolas Dmytruk[4], Gislène Collobert[4], Eric Hernandez[10], Pascal Astagneau[5,11], Philippe Glaser[6,7], Frédéric Ariey[1,2,3], Claire Poyart[1,2,3,4] and Agnès Fouet[1,2,3]

[1]INSERM, U1016, Institut Cochin, Paris, 75014, France

[2]Université Sorbonne Paris Descartes, Paris, 75014, France

[3]CNRS (UMR 8104), Paris, 75014 France

[4]Assistance Publique Hôpitaux de Paris, Service de Bactériologie, Centre National de Référence des Streptocoques, Groupe Hospitalier Paris Centre Cochin-Hôtel Dieu-Broca, Paris, 75014 France.

[5]Assistance Publique Hôpitaux de Paris, Centre de Prévention des infections associées aux soins, F75014 Paris, France

[6]Institut Pasteur, Unité Ecologie et Evolution de la Résistance aux Antibiotiques, Paris, 75724 France

[7]CNRS UMR3525, Paris, 75724 France

[8]Institut Pasteur, Pôle Biomics, Paris, 75724 France

[9]Institut Pasteur, GGIV unit Paris 75724 France

[10] Centre Médical de Forcilles, 77150 Lesigny, France

[11] : Sorbonne université, Faculté de médecine, APHP, Pitie-Salpétrière, F75013 Paris, France

Correspondence: Claire Poyart, Telephone, 33 1 58 41 15 60; Fax, 33 1 58 41 15 48
claire.poyart@aphp.fr

Keywords (3 - 10): Group A *Streptococcus*; *emm89*; emerging clade; phylogeny; biofilm; bacterium-cell interaction; outbreak

Running title (<50): *emm89* GAS outbreak

Abstract

An outbreak of nosocomial infections due to *Streptococcus pyogenes* (Group A *Streptococcus*; GAS) occurred in a post-surgery oncology unit concerned more than 60 patients and lasted 20 months despite enhanced infection control and prophylaxis measures. All GAS strains were characterized (*emm* genotype, toxin gene profile and pulse-field gel electrophoresis subtype). Selected strains were sequenced and phylogenetic relationship established. Capacity to form biofilm and interaction with human pulmonary epithelial cells and macrophages were determined. Twenty-six GAS strains responsible for invasive infections (II) and 57 for non-II or colonization were isolated from patients (n=66) or healthcare workers (n=13). Seventy strains shared the same molecular markers and 69 the same PFGE pattern; 56 were sequenced. They all belonged to the emerging *emm89* clade 3; all but one were clonal. Whole genome sequencing identified 43 genetic profiles with sporadic mutations in regulatory genes and acquired mutations in two structural genes. Except for two regulatory gene mutants, all strains tested had the same biofilm formation capacity and displayed similar adherence and invasion of pulmonary epithelial cells and phagocytosis and survival in human macrophages. This large outbreak of GAS infection in a post-surgery oncology unit, a setting that contains highly susceptible patients, arose from a strain of the emergent *emm89* clade. No relationship between punctual or acquired mutations, invasive status and strain phenotypic characteristics was found. Noteworthy, the phenotypic characteristics of this clone account for its emergence and its remarkable capacity to elicit outbreaks.

INTRODUCTION

Group A *Streptococcus* (GAS, *Streptococcus pyogenes*) is a human pathogen causing a broad range of diseases, non-invasive and invasive diseases as well as autoimmune sequelae, with an estimated 517,000 deaths per year [1]. These can be community acquired or be the consequence of outbreaks in human settings with high proximity such as schools, military barracks or hospital wards. Residents of the latter are at increased risk of acquiring GAS infections because of underlying medical conditions favorable to streptococcal infections [2].

GAS strains are extremely diverse and currently the most widely used typing-method for GAS strains relies on the 5'end of the *emm* gene sequence that encodes the hypervariable amino-terminus region of the M protein, a surface exposed virulence factor [3]. Among more than 200 *emm*-types defined, the most frequently associated with invasive infections in Europe are *emm1*, *emm28* and *emm89*, with variable distribution worldwide [4, 5]. Epidemiological studies have evidenced the emergence of given genotypes and more precisely of clones as leading cause of disease in Europe or the United States. Such was the case with the *emm1* type and the M1T1 clone in the late 1980s ([6] and references herein) and more recently of the *emm89* type exemplified by the emergent clade or clade 3 variant [7, 8]. The emergence of *emm89* strains, most probably of the same clone, has been described in multiple geographic locations [9-16]. Although *emm89* strain recovery from outbreaks has been described previously [17-19], the newly emerging *emm89* clone has not been decisively implicated in major outbreaks.

In 2012 in a follow-up care and rehabilitation post-surgery oncology unit in France, GAS infections were recorded at an unusual level and, although state-of-the-art prophylaxis measures were taken, this outbreak lasted twenty months. We investigated at the molecular level the cause of this outbreak by analyzing all isolates, be they from invasive or non-invasive diseases or colonization. The strains were genetically characterized and, using

65 biological and date criteria, fifty-six strains were chosen and sequenced. To shed light on
66 possible reasons for this long-lasting outbreak we investigated whether the GAS strains
67 involved had specific properties in terms of biofilm formation capacity, interactions with
68 human lung epithelial cells or human macrophages.
69

MATERIALS AND METHODS

Case definition

Invasive disease is defined through the isolation of GAS from normally sterile sites or the isolation of GAS from non-sterile sites (e.g., sputum, wounds) in the presence of either streptococcal toxic shock syndrome or necrotizing fasciitis and non-invasive disease through the isolation of GAS from a non-sterile site and which developed a clinical syndrome consistent with GAS infection. We defined a carrier as a patient or health care worker with GAS cultured from a non-sterile site and no clinical evidence of infection at the time of culture.

Bacterial strains

All GAS isolates collected during the outbreak were characterized by the French national reference center for streptococci (CNR-Strep) (Table S1). *emm* sequence type was determined by sequencing the variable 5'end of the *emm* gene and comparing sequences with database of the Center for Disease Control and Prevention (http://www.cdc.gov/ncdidod/biotech/strep/doc.htm) [3]. All strains were tested by a multiplex PCR method for the presence of genes encoding the toxins or superantigens SpeA, SpeB, SpeC, Ssa and SmeZ [4]. A prior cluster analysis was carried out by pulsed-field gel electrophoresis (PFGE) with *Sma*I restriction enzyme and interpreted according to Tenover criteria [20]. Other GAS strains collected by the CNR and similarly characterized were used as controls (Table S2). Antibiotic resistance profiles were determined according to the European Committee on Antimicrobial Susceptibility Testing guidelines (http://www.eucast.org).

Bacterial growth conditions

GAS isolates were grown at 37ºC without agitation in Todd-Hewitt broth supplemented with 0.5 % Yeast extract (THY) or in Brain Heart Infusion (BHI) (Difco). Bacteria were collected in mid-log phase, washed once with sterile phosphate-buffered saline (PBS), and diluted to the required inoculum; the number of viable bacteria was determined by counting the colony forming units (CFUs) after plating dilutions on THY agar.

GAS strain sequencing

Chromosomal DNA was extracted using the MasterPure^{TM} Gram-positive DNA purification kit (Tebu-Bio) and sequenced using the Illumina technology, with read length of 100 nt and a more than 100 fold-coverage. Libraries were constructed by using the Illumina TrueSeq kit following the manufacturer's instructions. Illumina short reads were assembled by using the Velvet software [21]. Strain 20120499 considered as the outbreak index case was used as a reference for SNPs calling. Contigs of strains 20120499 were ordered by aligning them to the complete genome sequence of strain MGAS315 (*emm1*, accession number NC_004070.1, no *emm89* complete sequence being available at the time of this analysis) using the Mauve software [22]. Each strain was analyzed for SNPs, small indels and deletions compared to strain 20120499 by using the BRESEQ software (pmid 24838886), which allowed the identification of errors in the Velvet assembly. Sequence specific to each strain was analyzed by assembling unmapped reads with Velvet. Accession numbers pending.

Reconstruction of large insertion from fastq files.

A Perl software named iSeGWalker was newly developed to accomplish a *de novo* genome reconstruction from the reads file in fastq format [23]. The search step is an exact-matching reads selection using a regular expression (seed) and a simple Perl script, reading each

sequence one by one. Once all matching reads of the index case genome (20120456) had been selected, a consensus sequence was determined; then, a new seed, composed by the 30 last consecutive nucleotides, was obtained and a new search was performed.

Phylogeny analysis

Phylogenetic analyses were performed after alignment of all strains against the new reconstructed sequence as reference. The evolutionary history was inferred using the Neighbor-Joining method [24]. The evolutionary distances were computed using the Maximum Composite Likelihood method [25] and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The differences in the composition bias among sequences were considered in evolutionary comparisons [26]. All positions containing gaps and missing data were removed for each sequence pair. Evolutionary analyses were conducted in MEGA7 [27].

PCR amplification

The DNA sequence surrounding the *has* operon was amplified using the oligonucleotides Has-M89_F CTGTGCCACTAAGACTCTTT and Has-M89_R CGGACAGGTGCAGTTGG-TTT using the GoTaq® Green Master Mix with an elongation time of 4 min.

Biofilm formation assay

Overnight cultures grown in BHI, for optimal abiotic biofilm formation as described in [28], were diluted to an optical density at 600 nm (OD_{600}) of 0.06 and 180 µL were distributed in quadruplet in 96-well plates. After 24 h static growth at 37°C, one plate was used to record the OD_{595} after resuspension while the other one was treated with crystal violet as described by Köller et al. [28] except that the aqueous 1% SDS solution was replaced by a 80/20

ethanol/acetone solution. Biofilm formation capacity was defined as the ratio of the OD_{595} elicited by crystal violet and the OD_{595} due to growth on the control plate. The experiment was carried out at least in triplicate.

Growth kinetics.

Overnight cultures grown in THY were diluted 1:100 in fresh THY and 150 µL were distributed in quadruplet in 96-well plates. The plates were incubated at 37°C in a Thermo Scientific Multiskan™ GO microplate spectrophotometer. OD_{600} was measured every 10 min after agitation.

Human lung epithelial cell culture and adhesion and invasion assays

A549 cells were grown in DMEM-Glutamax 10% Fetal Calf Serum (FCS) in 24-well plates during 2 days, washed with PBS and infected with mid-logarithmic GAS resuspended in RPMI cell-culture medium at a multiplicity of infection (MOI) of 1 bacterium / cell. For adhesion assays, the plates were centrifuged 5 min at 1000 rpm and incubated 1 h at 4°C; the non-adherent extracellular bacteria were then eliminated removing the culture medium and washing three times with sterile PBS. For internalization assays, the plates were further incubated 45 min at 37°C, 5% CO_2, and the adherent extracellular bacteria were subsequently killed by incubation with fresh medium containing 50 U/mL penicillin/ and 50 µg/mL streptomycin for 30 min. After washing with PBS A549 cells were lysed with 1 mL cold sterile distilled water and serial dilutions were plated on THY plates and the number of CFUs was determined after 24-48 hours growth at 37ºC. For all experiments, 4 or 5 independent assays in duplicate were carried out for each bacterial isolate.

Human macrophage culture and infection assays

To differentiate the THP-1 into macrophages, 2.10^5 to 4.10^5 cells were inoculated in RPMI GluMax supplemented with 10% FCS and 200 nM PMA (Phorbol 12-myristate 13-acetate) (Sigma, P1585). After 3 days at 37°C, 5% CO_2, the medium was replaced with RPMI GluMax supplemented with 10% FCS and the incubation extended for 5 days. Cells were then infected with mid-logarithmic GAS resuspended in RPMI medium at a MOI of 1. The phagocytosis and survival assays were carried out as previously described [29].

Statistical analyses.

Significance of differences in biofilm formation capacities were assessed by Student's t-test. That of cellular adhesion, invasion, phagocytosis and survival were assessed by Repeated measures Anova and Turkey's multiple comparison test, GraphPad software.

RESULTS

Outbreak description

From 16th April through 7th May 2012, 4 cases of invasive GAS infections were recorded in a 100-bed post-surgery oncology unit (Fig.1). To limit secondary cases, epidemiological investigations were started among the patients and healthcare workers to determine their status against GAS. Although appropriate hygiene measures were taken this outbreak lasted until 23rd December 2013 concerned 66 cases and ended after mass chemoprophylaxis by azithromycin 500 mg *per os* once daily for 3 days and enhanced infection control such as mask wear, hand hygiene and the implementation of appropriate disinfection measures for tracheostomy cannula.

Almost all the patients had surgery for an oral cancer and were in second line of treatment with radiotherapy. Men accounted for 83% (55/66) of the patients and the mean age were 61 year-old (range 24-87) and 67.5 year-old (52-83) for men and women, respectively (Table S1 and data not shown).

During the outbreak, 26 cases of GAS invasive infections were identified, 13 bacteremia, 11 pneumonias and 2 necrotizing fasciitis, among which 5 (19.2%) died. During the epidemiological surveillance, 57 non-invasive GAS strains were isolated; 44 from patients and 13 from healthcare workers with direct patient-care responsibilities (Table S1).

GAS patients' strains, were isolated from blood cultures (n=13), oropharynx (n=9), wounds (n=13), gastrostomy sites (n=8), tracheostomy cannula (n=8), respiratory specimen (n=18), and lymphadenectomy (n=1). All thirteen GAS-positive cultures from healthcare workers were from the oropharynx. A total of 83 GAS strains were thus isolated between April 2012 and December 2013 including 26 invasive, 38 non-invasive, and 19 from colonization. An invasive respiratory device was present in 13/26 (50%) patients with an invasive infection and in only 12/44 (27%) patients colonized or without an invasive infection, p=0.049.

209 **Genetic characterization of the GAS strains**

210 All strains were genotyped (Table S1). All invasive but one (25/26) and 51 out of 57 non-
211 invasive or colonization strains shared the same *emm89* genotype. Sixty-nine/76 (91%)
212 *emm89* GAS strains, be they invasive, non-invasive or colonization strains, were
213 indistinguishable by PFGE displaying an 89-A4 pattern according to our nomenclature (Table
214 S1 and data not shown). Patterns of the seven *emm89* remaining strains showed a maximum
215 of four differences. All 76 *emm89* strains displayed a low-level resistance to aminoglycosides
216 (kanamycin and gentamicin) and were susceptible to β-lactams, vancomycin, macrolides and
217 tetracycline except two non-89-A4 strains that were resistant to tetracycline due to the
218 presence of *tet*(M) determinant located on Tn*916* related conjugative transposon.
219 Thus, whereas non-89-A4 PFGE and non-*emm89* strains were occasionally isolated, 69/83
220 GAS strains isolated throughout the considered period shared the same *emm89* genotype 89-
221 A4 PFGE pattern. This argued for an outbreak origin of these strains; the other strains were
222 considered outside the outbreak. To further analyze these *emm89* 89-A4 strains, a total of 56
223 were randomly chosen for sequencing among a sub-collection selected according to date and
224 clinical criteria as follows: a balanced distribution, 22/26 and 34/50 isolated in 2012 and 2013,
225 respectively, including the first, index case in 2012, and last of each year, and 9 and 11
226 invasive infections from 2012 and 2013, respectively. For sake of comparison, we added an
227 unrelated *emm89* invasive strain isolated in 2012 (20120456) as well as three clinical isolates
228 isolated at the same post-surgery oncology unit but in 2014 (20141462, 20141463 and
229 20140469; Table S2).
230
231 **Sequence analysis and phylogeny of the GAS outbreak isolates**
232 The alignment of three fastq files, including that of the index case strain (20120499), to H293
233 *emm89* genomic sequence (NC_002737.2), and the use of the Tablet software [30], led to the

identification of 5 areas of possible chromosomal rearrangement sites based on a relative increase in the diversity of these loci. We then reconstructed the new reference genome (see Materials and Methods) (1 747 907 base pairs) and realigned all the fastq files on it. The genomes are 99 % identical on a 99% coverage to that of MGAS27061 (accession number NZ_CP013840.1), indicating that these *emm89* clinical isolates belong to the new *emm89* clade 3, lacking the *has* operon and harboring the same mutation in the *slo-nga* promoter region [13, 31]. To get insight into the emergence of this new clade in France we sought, in a collection of *emm89* French strains, when the *has* operon loss could have occurred. The DNA fragment encompassing the *has* operon was amplified from 15 unrelated strains received by the CNR-Strep during the last 15 years (data not shown). In three of these strains, including one from 2003 the *has* operon was absent, suggesting the presence of the new clade in France already at that time.

For phylogenetic analyses we used the whole genome alignment of 61 strains: we combined genomic data obtained from 56 isolates of the 2012-2013 outbreak, one *emm89* control strain independent of this outbreak (20120256), three strains isolated in 2014 at a time-related distance from the outbreak, and the reference MGAS27061 an *emm*89 clade 3 strain [7, 8] (Fig.2; Table S2). This tree (Fig.2) clearly shows the monophyletic character of the 2012/2013 outbreak, well separated from the isolates collected at the same site but in time away from the outbreak (2014 strains). The control strain 20120456 is phylogenetically more distant to the index case strain (42 SNPs compared to the 20120499 index strain) than all other strains. Furthermore, the fact that the reference strain MGAS26761 is positioned between the branches 2012/2013 and 2014, reinforces the independence of the two events and strongly suggests the extinction of the outbreak episode of 2012/2013 (Fig. 2).

The tree covering the 55 outbreak strains (Fig.3) presents three separate subclones linked to their date of isolation. Of the three main subclones, two are composed solely of 2012 strains

(subclones 1 and 2 2012). Among strains belonging to subclone 3 2013, two, 20121368 and 20121198, were the last isolated in 2012, in fact during the last trimester of 2012.

Also they have lost the phage harboring *speC*, however loss of a phage being compatible with a convergence event. The strain 20130948 appears to have a quite peculiar evolutionary history and lies apart from all subclones. Despite these exceptions this topography is highly compatible with an epidemic signature: most strains deriving from a unique colonization event occurring in 2012. The *emm89* outbreak subclone was termed *emm89*-F.

Forty-three genetic profiles were defined (Table S3). At the gene level, four mutations stabilized over the outbreak period; two occurred in intergenic regions and the two others in genes involved in bacterial metabolism, the oxaloacetate decarboxylase beta chain (ODC), found in 36 strains from 8th September 2012 onward, and the shikimate 5−dehydrogenase I alpha (S5D), found in 33 strains from 25th January 2013 onward. Ten non-stabilized mutations were also found in the gene encoding the serum opacity factor (SOF) among 29 strains. Six non-fixed mutations were also found among 8 strains in the gene encoding the CovS regulatory protein. No mutation or group of mutations could be related to the strain invasive status.

Phenotypic characterization of the *emm89*-F strains

Isolates of the *emm89*-F clone were characterized for phenotypes relevant to the clinical settings of this outbreak, *i*. *e*. growth capacities, abiotic biofilm formation, interactions with lung epithelial cells and macrophages. Strains were chosen to cover all the branches of the genetic tree and to, altogether, harbor all mutation profiles (Fig. 3, Table S1).

Biofilm formation

As mentioned a significant correlation between the occurrence of invasive GAS infections and the presence of a medical respiratory device was observed among infected patients. (Table S1). The recognition that biofilm formation by GAS is involved in persistence (for review [32]) and that biofilm formation varies among strains within *emm* genotypes [33] prompted us to assay the capacity of these clinical isolates to form biofilm on abiotic surface (Fig. 4). All strains were able to form biofilms and most of them in similar amounts; no difference was observed between strains sampled from tracheostomy cannula and from other sites. The stronger biofilm producers are mutated in either the global positive regulator gene *mga*, yielding values slightly above those of the other strains, or, and more impressively, in *covS,* the sensor of the CovRS two component system known to negatively regulate the expression of multiple virulence genes [34].

We also assessed whether GAS biofilm formation capacities increased during the outbreak. To that aim we calculated a trend line, after removing the strains mutated in regulatory genes (Fig. S1). It slightly raises ($y = 0.0362\ x + 8.8$) but displays no great change in biofilm formation throughout the outbreak.

To assess whether the outbreak isolates were weak or robust biofilm formers, their capacity to form biofilms was compared to that of other unrelated clinical isolates (Fig. 4, Table S2). Biofilm formation being dependent on the CovRS regulatory system, we included only *covR covS* wild-type isolates from other studies and calculated the mean value for the outbreak isolates excluding all regulatory mutants [35]. The clinical isolates collected during the outbreak formed substantial more biofilms compared to isolates from other *emm* types, ($p = 0.001$). In contrast, no difference was observed when comparison was established with other *emm89* isolates, also belonging to the emergent clade 3 [7, 8], (data not shown) indicating that this may be a specific trait to the emerging *emm89* clones.

In conclusion these results show that *emm89* outbreak isolates are more robust biofilm formers than non-*emm89* isolates; this property could account and contribute for the current emergence and spread of this clone.

Adhesion and invasion into lung epithelial cells

The pathologies observed during the outbreak such as respiratory infections prompted us to analyze the adhesion and invasion of selected clinical isolates into human A549 lung epithelial cells. Four strains were selected, in addition to the index case (IC; 120499), using different criteria to ensure a representative sampling: dispersion throughout the genetic tree and the presence of different mutations, SOF (120811), ODC (120747), S5D (130922) and the regulator CovS (120619) (Fig. 3 and Fig. S1 and Table S1). We first assessed that these strains displayed no growth difference in laboratory conditions (data not shown).

Adhesion capacities of these strains were determined and compared to that of an unrelated *emm89* strain previously studied [29]. No significant difference in the adhesion capacity of these strains was observed (Table 1). In contrast, a difference was noted when invasion was compared: the invasion of the *covS* mutant isolate was lower than that of all the other strains and was significantly lower than that of the *emm89* control strain (Table 1). This suggests that these clinical isolates possessed no specific properties in their interaction with pulmonary cells compared to other unrelated GAS *emm89* clinical isolates and did not acquire any, through their mutations detected.

Phagocytosis and survival within macrophages

Macrophages are critical host defense cells. We have previously shown that *emm89* isolates were more phagocytozed by murine macrophages than non-*emm89* strains, that they were differentially phagocytozed depending on their invasive status and that there was a variability

within the invasive and non-invasive strains [29]. Consequently, we analyzed the behavior of the outbreak clinical isolates towards uptake and survival in the physiologically relevant THP-1 human macrophages and compared it to that of the *emm89* control strain (Fig. 5). Except for the *covS* mutant strain, which seemed less phagocytozed than all other strains and was significantly less phagocytozed than the SOF and S5D strains, all outbreak isolates and the control strain were similarly phagocytozed by THP1 macrophages (Fig. 5A).

We then tested the survival of these isolates in THP-1 (Fig. 5B). Only the *covS* mutant strain survived, at T2, less than other strains, and significantly less than the index isolate.

This indicates that the outbreak strains and the *emm89* control isolate behaved similarly towards the host defense elicited by the human macrophages.

DISCUSSION

This is, to our knowledge, one of the largest outbreaks of GAS infection in a single health care facility among highly vulnerable patients with oral cancer. Here we describe a remarkably long outbreak linked to the persistence and spread of a clone in both patients and healthcare workers. Almost all isolates from patients with GAS infection and those from healthcare personnel were subtype *emm89*, suggesting that one GAS strain was circulating throughout the healthcare facility. However, we could not determine whether a patient or an infected colonized healthcare worker initiated the outbreak since investigations began three weeks after notification to the authorities of the four cases of invasive infections. In France, *emm89* is now among the top three leading *emm* types, accounting for 19 % of all invasive infections in 2017 and the first one (23 %) for GAS clustered cases since 2011 (https://cnr-strep.fr/index.php/bilans-dactivite/rapports-dactivite).

Using whole-genome sequencing we demonstrated that this clone belongs to the emerging *emm89* clade 3 [7, 8, 14]. Interestingly, most but not all isolates could be phylogenetically grouped on the basis of their sampling year (Fig. 2). The exceptions suggest that some clones reappeared after a few months during which they were asymptomatically carried. Several data report that this *emm89* emerged in many countries in a closely similar time frame during the first decade of the 2000s, accompanied by a swift from clade 1 to clade 2 and subsequent emergence of clade 3 [12].

The extent of this outbreak and the case-fatality rate of 19 % among patients with invasive disease prompted us to shed light on the advantages brought by this specific clone; we studied phenotypic traits such as biofilm formation and interaction with human pulmonary cells or macrophages. The similar high level of biofilm formation displayed by the outbreak and other *emm89* clones compared to other *emm*-type strains could be a consequence of all strains belonging to the emerging *emm89* clade. Indeed, *emm89* strains were not described as high

biofilm producer by Koller *et al*. [28], and this could be linked to their strain panel, sampled between 2001 and 2006, that is before the full emergence of this clone [7, 28]. Here we show that the *emm89*-F clone, which is uncapsulated, is a robust biofilm former confirming that capsule is dispensable for abiotic biofilm formation [35, 36]. The *emm89*-F *mga* mutants, Mga being a global transcriptional activator including of the M and M-like encoding genes, yielded slightly higher biofilm production than the wild-type *emm89*-F isolates. M and M-like proteins most often contribute to biofilm formation, which is nevertheless *emm*-type dependent [35-37]; a weaker M and M-like protein production in an *emm89* background may have, as in the *emm2* background, little influence on biofilm formation [37]. Furthermore, a variation in the abundance of Mga-controlled products, other than the M and M-like proteins, could account for the biofilm increase. CovRS is a two-component system that controls, mainly repressing, the expression of 15 % of the GAS genome. Several *covS*-mutant strains produce less, and others more biofilm than their wild-type counterparts, suggesting that the consequences of mutations in *covS* on biofilm formation are strain-dependent [35, 38]. The *emm89*-F *covS* mutant isolates were particularly robust biofilm formers supporting the strain-dependent contribution of CovS to biofilm formation and demonstrating that CovRS-controlled products other than the capsule are involved in biofilm formation.

Several reports suggested that acapsular *emm89* clade-associated GAS strains may have acquired potential for long-term colonization [7, 13]. Moreover, the increased expression of SLO and NADase by this clade could enhance internalization and intracellular bacterial survival in epithelial cells, theoretically providing protection from natural antimicrobial peptides and antibiotic treatment. Adhesion and invasion experiments on human epithelial lung cells performed on a representative sampling did not reveal major difference among strains except for the *covS* strain for which, if adhesion was similar to other strains, its capacity to invade cells was affected. Furthermore, there was a tendency for *covS* strain to be

less phagocytozed and to survive less than the others in the macrophage. Altogether this meant a disadvantage for this strain for which epithelial cells and macrophages would not act as a reservoir and promote a carrier state. This result is consistent with the report that *covS* strains are badly transmitted [39], thus accounting for sporadic apparition of *covS* mutations during this outbreak.

In this outbreak the majority of the patients had undergone surgery associated with radiotherapy for oral cancer and belongs to a disadvantaged and highly vulnerable population. These features, could explain the very high rate of transmission in fragile patients where promiscuity was highly problematic.

In conclusion, this report adds France to the growing list of European countries and North America where, by whole genome sequencing, *emm89* clade 3 strains have been demonstrated to an emergent clone eliciting GAS infections. It strengthens the fact that the spreading of this clone is worldwide and emphasizes the importance of epidemiological monitoring by reference centers and health authorities. It confirms the relevance, power and robustness of the whole genome sequencing approach for epidemiological analysis of major outbreaks and its routine use for surveillance.

Conflict of interest statement

The other authors declare no competing financial interests.

Acknowledgements

This work was supported by Santé Publique France, INSERM, CNRS, Université Paris Descartes and by the High Council for Scientific and Technological Cooperation between France-Israel "Complexity in Biology" program.

Authorship/Contributor

Conceived and designed the study: Céline Plainvert, Agnès Fouet and Claire Poyart. Provided clinical and epidemiological data Elise Seringe, Eric Hernandez and Pascal Astagneau. Performed the bacteriological analyses Magalie Longo, Nicolas Dmytruk and Gislaine Collobert. Performed biofilms and phagocytosis analyses Magalie Longo. Performed Whole genome sequencing of bacterial strains Benjamin Saintpierre, Elisabeth Sauvage, Laurence Ma and Johann Beghain. Analysed genomes data Philippe Glaser, Frédéric Ariey and Agnès Fouet. Wrote the manuscript Claire Poyart, Céline Plainvert and Agnès Fouet.

REFERENCES

1 Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. *Lancet Infect Dis*. 2005; **5**: 685-694.

2 Deutscher M, Schillie S, Gould C, et al. Investigation of a group A streptococcal outbreak among residents of a long-term acute care hospital. *Clin Infect Dis*. 2011; **52**: 988-994.

3 Beall B, Facklam R, Thompson T. Sequencing *emm*-specific pcr products for routine and accurate typing of group A streptococci. *J Clin Microbiol*. 1996; **34**: 953-958.

4 Plainvert C, Doloy A, Loubinoux J, et al. Invasive group a streptococcal infections in adults, france (2006–2010). *Clin Microbiol Infect*. 2012: 702-710.

5 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**: 611-616.

6 Nasser W, Beres SB, Olsen RJ, et al. Evolutionary pathway to increased virulence and epidemic group a *Streptococcus* disease derived from 3,615 genome sequences. *Proc Natl Acad Sci U S A*. 2014; **111**: E1768-1776.

7 Turner CE, Abbott J, Lamagni T, et al. Emergence of a new highly successful acapsular group a *Streptococcus* clade of genotype *emm89* in the united kingdom. *MBio*. 2015; **6**: e00622.

8 Zhu L, Olsen RJ, Nasser W, de la Riva Morales I, Musser JM. Trading capsule for increased cytotoxin production: Contribution to virulence of a newly emerged clade of *emm89 Streptococcus pyogenes*. *MBio*. 2015; **6**: e01378-01315.

9 Friaes A, Machado MP, Pato C, Carrico J, Melo-Cristino J, Ramirez M. Emergence of the same successful clade among distinct populations of *emm89 Streptococcus pyogenes* in multiple geographic regions. *MBio*. 2015; **6**: e01780-01715.

10 Musser JM, Zhu L, Olsen RJ, Nasser W. Musser et al. Reply to "emergence of the same successful clade among distinct populations of *emm89 streptococcus pyogenes* in multiple geographic regions". *MBio*. 2015; **6**: e01838-01815.

11 Turner CE, Lamagni T, Holden MT, et al. Turner et al. Reply to "emergence of the same successful clade among distinct populations of *emm89 Streptococcus pyogenes* in multiple geographic regions". *MBio*. 2015; **6**: e01883-01815.

12 Latronico F, Nasser W, Puhakainen K, et al. Genomic characteristics behind the spread of bacteremic group a *Streptococcus* type emm89 in finland, 2004-2014. *J Infect Dis*. 2016; **214**: 1987-1995.

13 Beres SB, Kachroo P, Nasser W, et al. Transcriptome remodeling contributes to epidemic disease caused by the human pathogen *Streptococcus pyogenes*. *MBio*. 2016; **7**.

14 Beres SB, Olsen RJ, Ojeda Saavedra M, et al. Genome sequence analysis of *emm89 streptococcus pyogenes* strains causing infections in scotland, 2010-2016. *J Med Microbiol*. 2017.

15 Chochua S, Metcalf BJ, Li Z, et al. Population and whole genome sequence based characterization of invasive group a streptococci recovered in the united states during 2015. *MBio*. 2017; **8**.

16 Teatero S, Coleman BL, Beres SB, et al. Rapid emergence of a new clone impacts the population at risk and increases the incidence of type *emm89* group a *Streptococcus* invasive disease. *Open Forum Infect Dis*. 2017; **4**: ofx042.

17 Falkenhorst G, Bagdonaite J, Lisby M, et al. Outbreak of group a streptococcal throat infection: Don't forget to ask about food. *Epidemiol Infect*. 2008; **136**: 1165-1171.

18 Thigpen MC, Thomas DM, Gloss D, et al. Nursing home outbreak of invasive group a streptococcal infections caused by 2 distinct strains. *Infect Control Hosp Epidemiol*. 2007; **28**: 68-74.

19 Yang P, Peng X, Yang J, Dong X, Zhang M, Wang Q. A probable food-borne outbreak of pharyngitis after a massive rainstorm in beijing, caused by emm89 group a *streptococcus* rarely found in china. *Int J Infect Dis*. 2013; **17**: e471.

20 Tenover FC, Arbeit RD, Goering RV, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: Criteria for bacterial strain typing. *J Clin Microbiol*. 1995; **33**: 2233-2239.

21 Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de bruijn graphs. *Genome Res*. 2008; **18**: 821-829.

22 Darling AC, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004; **14**: 1394-1403.

23 Karadjian G, Hassanin A, Saintpierre B, et al. Highly rearranged mitochondrial genome in nycteria parasites (haemosporidia) from bats. *Proc Natl Acad Sci U S A*. 2016; **113**: 9834-9839.

493 24 Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing
494 phylogenetic trees. *Mol Biol Evol*. 1987; **4**: 406-425.
495 25 Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using
496 the neighbor-joining method. *Proc Natl Acad Sci U S A*. 2004; **101**: 11030-11035.
497 26 Tamura K, Kumar S. Evolutionary distance estimation under heterogeneous
498 substitution pattern among lineages. *Mol Biol Evol*. 2002; **19**: 1727-1736.
499 27 Kumar S, Stecher G, Tamura K. Mega7: Molecular evolutionary genetics analysis
500 version 7.0 for bigger datasets. *Mol Biol Evol*. 2016; **33**: 1870-1874.
501 28 Koller T, Manetti AG, Kreikemeyer B, et al. Typing of the pilus-protein-encoding fct
502 region and biofilm formation as novel parameters in epidemiological investigations of
503 *Streptococcus pyogenes* isolates from various infection sites. *J Med Microbiol*. 2010;
504 **59**: 442-452.
505 29 Dinis M, Plainvert C, Kovarik P, Longo M, Fouet A, Poyart C. The innate immune
506 response elicited by group a *Streptococcus* is highly variable among clinical isolates
507 and correlates with the *emm* type. *PLoS One*. 2014; **9**: e101464.
508 30 Milne I, Stephen G, Bayer M, et al. Using tablet for visual exploration of second-
509 generation sequencing data. *Brief Bioinform*. 2013; **14**: 193-202.
510 31 Watanabe S, Sasahara T, Arai N, et al. Complete genome sequence of *Streptococcus*
511 *pyogenes* strain jmub1235 isolated from an acute phlegmonous gastritis patient.
512 *Genome Announc*. 2016; **4**.
513 32 Young C, Holder RC, Dubois L, Reid SD. *Streptococcus pyogenes* biofilm.
514 2016/02/12 edn 2016.
515 33 Lembke C, Podbielski A, Hidalgo-Grass C, Jonas L, Hanski E, Kreikemeyer B.
516 Characterization of biofilm formation by clinically relevant serotypes of group a
517 streptococci. *Appl Environ Microbiol*. 2006; **72**: 2864-2875.
518 34 Churchward G. The two faces of janus: Virulence gene regulation by covr/s in group a
519 streptococci. *Mol Microbiol*. 2007; **64**: 34-41.
520 35 Cho KH, Caparon MG. Patterns of virulence gene expression differ between biofilm
521 and tissue communities of *streptococcus pyogenes*. *Mol Microbiol*. 2005; **57**: 1545-
522 1556.
523 36 Fiedler T, Koller T, Kreikemeyer B. *Streptococcus pyogenes* biofilms-formation,
524 biology, and clinical relevance. *Front Cell Infect Microbiol*. 2015; **5**: 15.

525 37 Courtney HS, Ofek I, Penfound T, et al. Relationship between expression of the
526 family of m proteins and lipoteichoic acid to hydrophobicity and biofilm formation in
527 *Streptococcus pyogenes*. *PLoS One*. 2009; **4**: e4166.
528 38 Sugareva V, Arlt R, Fiedler T, Riani C, Podbielski A, Kreikemeyer B. Serotype- and
529 strain- dependent contribution of the sensor kinase covs of the covrs two-component
530 system to *Streptococcus pyogenes* pathogenesis. *BMC Microbiol*. 2010; **10**.
531 39 Alam FM, Turner CE, Smith K, Wiles S, Sriskandan S. Inactivation of the covr/s
532 virulence regulator impairs infection in an improved murine model of *Streptococcus*
533 *pyogenes* naso-pharyngeal infection. *PLoS One*. 2013; **8**: e61655.
534
535
536

FIGURE LEGENDS

Figure 1. History of GAS isolates. Each isolate is indicated by a symbol; patient invasive strain, plain black circle; patient non-invasive or colonizing strain, open circle; healthcare worker, grey circle; the symbols of sequenced strains are bordered by a red line; sampling months are indicted below.

Figure 2. Evolutionary relationships of taxa of the sequenced isolates.

The optimal tree with the sum of branch length = 1.78211648 is shown. Only bootstrap greater than 50% are shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The analysis involved 66 nucleotide sequences. All positions containing gaps and missing data were eliminated including the region of the 25 kb insertion. There were a total of 206 positions (informative SNP) in the final dataset. 20120456 C in bold red case corresponds to the control strain isolated independently of the outbreak and the strain of the index case is written in black bold case.

Figure 3. Evolutionary relationships of *emm89* outbreak isolates.

The optimal tree with the sum of branch length = 0.00002589 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The analysis involved 55 nucleotide sequences. All positions containing gaps and missing data were eliminated including the SpeC harboring 25 kb insertion, absent from strains 20121198 and 20121368. There were a total of 1716368 positions in the final dataset. Sub-clone 1 (blue) and 2 (red) for 2012 are encircled and identified. 12, 13, isolates sampled in 2012 and 2013 localized in the 2013 (green) and 2012 clades, respectively; Δ, absence of the 25 kb *speC* harboring phage; IC, index case, SOF, ODC, S5D, CovS, isolates with mutations in a fibronectin binding-protein, oxaloacetate

562 decarboxylase beta chain (EC 4.1.1.3), shikimate 5-dehydrogenase I alpha (EC 1.1.1.25), the
563 sensor CovS, respectively.
564
565 **Figure 4.** The outbreak *emm89*-F isolates form substantial biofilms compared to other non-
566 *emm89* GAS strains. (A) Comparison of biofilm formation of isolates from the epidemic
567 clone and of strains from other *emm*-types (Table S2). For each strain, the mean value of its
568 biofilm-forming capacity is indicated. It was obtained by measuring to the ratio between the
569 crystal violet elicited OD_{595} and that of the bacterial growth, of three to six experiments. The
570 bar corresponds to the mean value within each group. The mutants in regulatory genes were
571 not included in the calculation of the mean values but are nevertheless represented on the
572 figure. Grey closed circles 2012 strains; black closed circles, 2013 strains; symbols for the
573 five representative strains for which other phenotypes have been analyzed are highlighted;
574 larger closed circle, IC strain (CNR120499); larger open circle, CovS strain (CNR120619);
575 closed square, ODC strain (CNR120747); closed downward pointing triangle, SOF strain
576 (CNR120811; closed diamond, S5B strain (CNR130922); also open square, the *emm89*
577 control strain; blue circles, *covS* mutants; mauve circles, *mga* mutant strains. (Student's t-test,
578 ***$p \leq 0.001$). (B) A tendency curve was calculated using the means shown in Fig. 3A, $y =$
579 $0.0362\ x + 8.8$ excluding the mutants in regulatory genes. The sampling year is indicated
580 below. Symbols same as in Fig. 1.
581
582 **Figure 5.** Phagocytosis (A) and survival (B) of GAS *emm89*-F isolates in human
583 macrophages. THP-1 cells were infected with *emm89*-F strains or an *emm89* previously
584 characterized control strain (Table S2, [29]) as described in Material and Methods and results
585 are expressed as the percentages of (A) bacterial CFUs recovered after 30 min antibiotic
586 treatment relative to the initial inoculum and (B) the percentage of phagocytozed bacteria that

587 survived. The results represent the mean ± SD of 5 independent experiments carried out in
588 triplicate for each isolate, with significance levels indicated between a given strain and the
589 *covS* mutant strain types (*p ≤ 0.05; **p ≤ 0.01).
590
591 **Figure S1.** The outbreak *emm89*-F isolates form substantial biofilms compared to other non-
592 *emm89* GAS strains. A tendency curve was calculated using the means shown in Fig. 4, y =
593 0.0362 x + 8.8 excluding the mutants in regulatory genes. The sampling year is indicated
594 below. Symbols are the same as in Fig. 1.

Table 1. Adhesion and invasion of human epithelial lung cells (A549) by *emm89*-F isolates

	IC	ODC	SOF	S5D	CovS	M89
Adhesion	40.20 +/-4.6	57.48 +/- 17.2	39.40 +/-13.5	75.67 +/- 21.9	41.06 +/-3.2	51.12 +/- 13.1
Invasion	6.34 +/- 3.4	8.64 +/-4.3	8.40 +/- 4.4	9.44 +/-5.0	0.38 +/-0.26	14.31 +/-6.3**

Results are expressed as the percentages of, for the adhesion, bacterial CFUs recovered after three PBS washing relative to the initial inoculum and, for the invasion, bacterial CFUs recovered after 30 min antibiotic treatment relative to the adherent CFUs. The results represent the mean ± SD of 3 and 5 independent experiments, respectively, carried out in triplicate for each isolate, with significance levels indicated between a given strain and the *covS* mutant strain types (**$p \leq 0.01$)

Figure 1

Click here to download Figure Fig.1.tiff

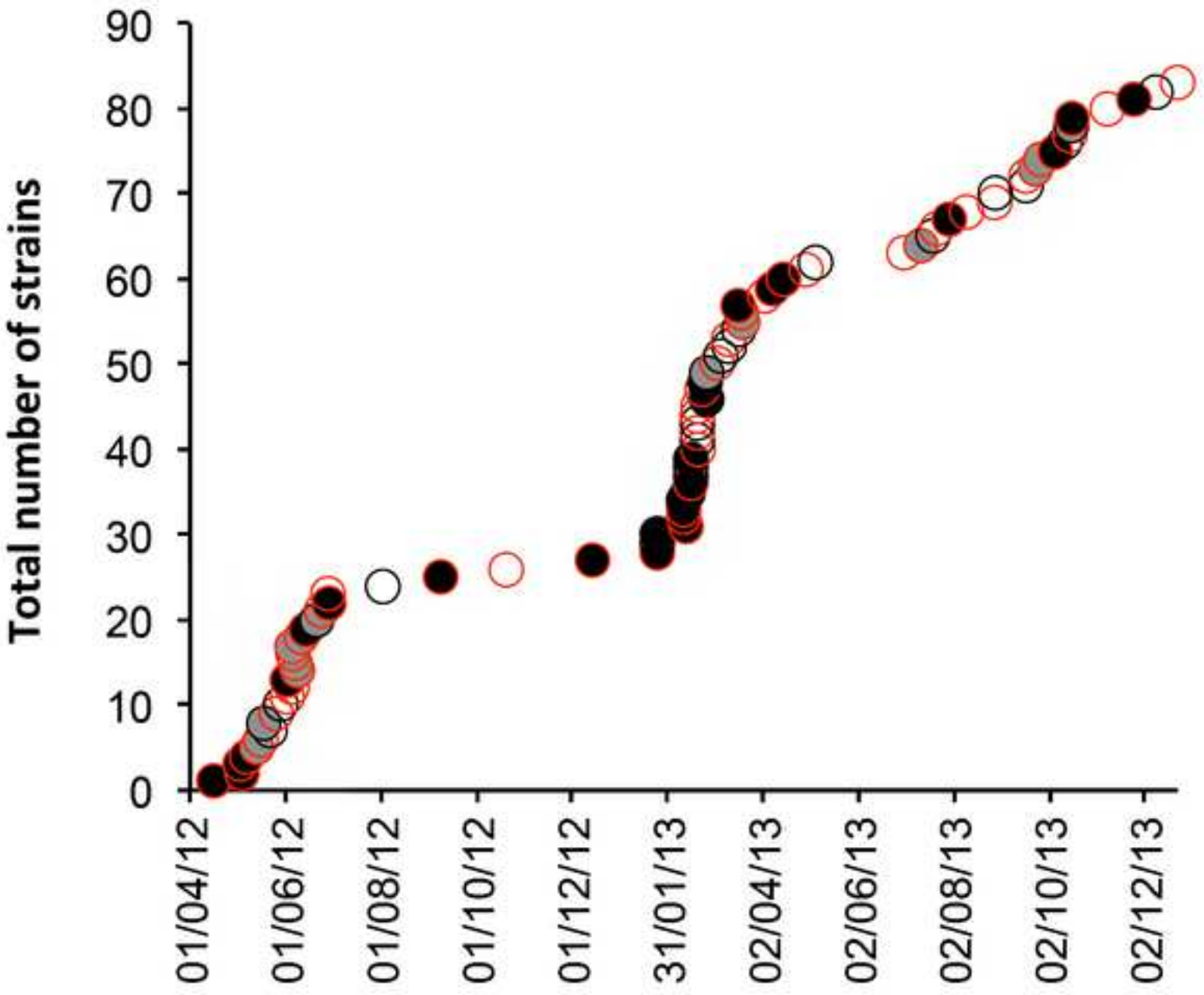

Figure 2

Click here to download Figure FIG. 2.tiff

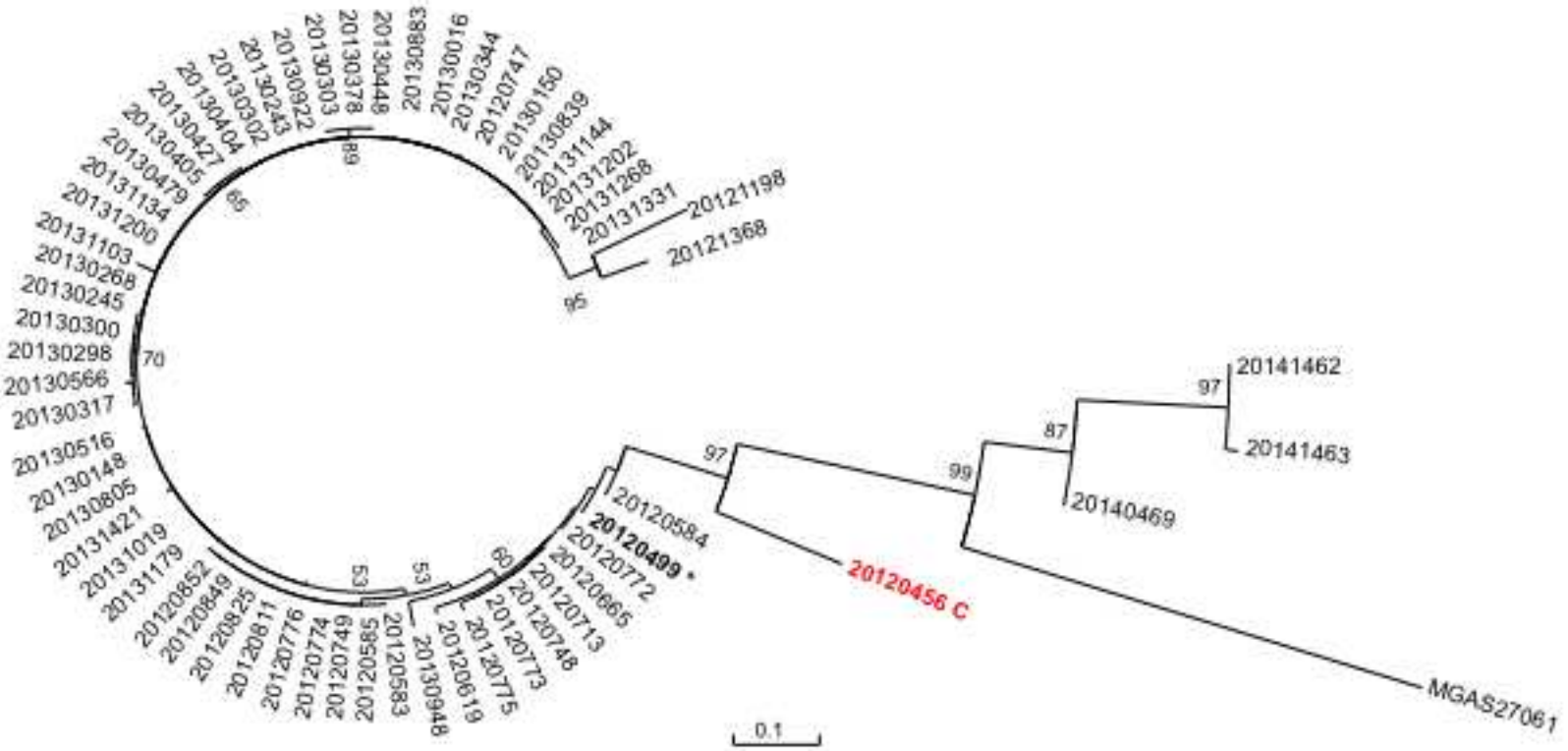

Figure 3

Click here to download Figure Fig.3.tiff

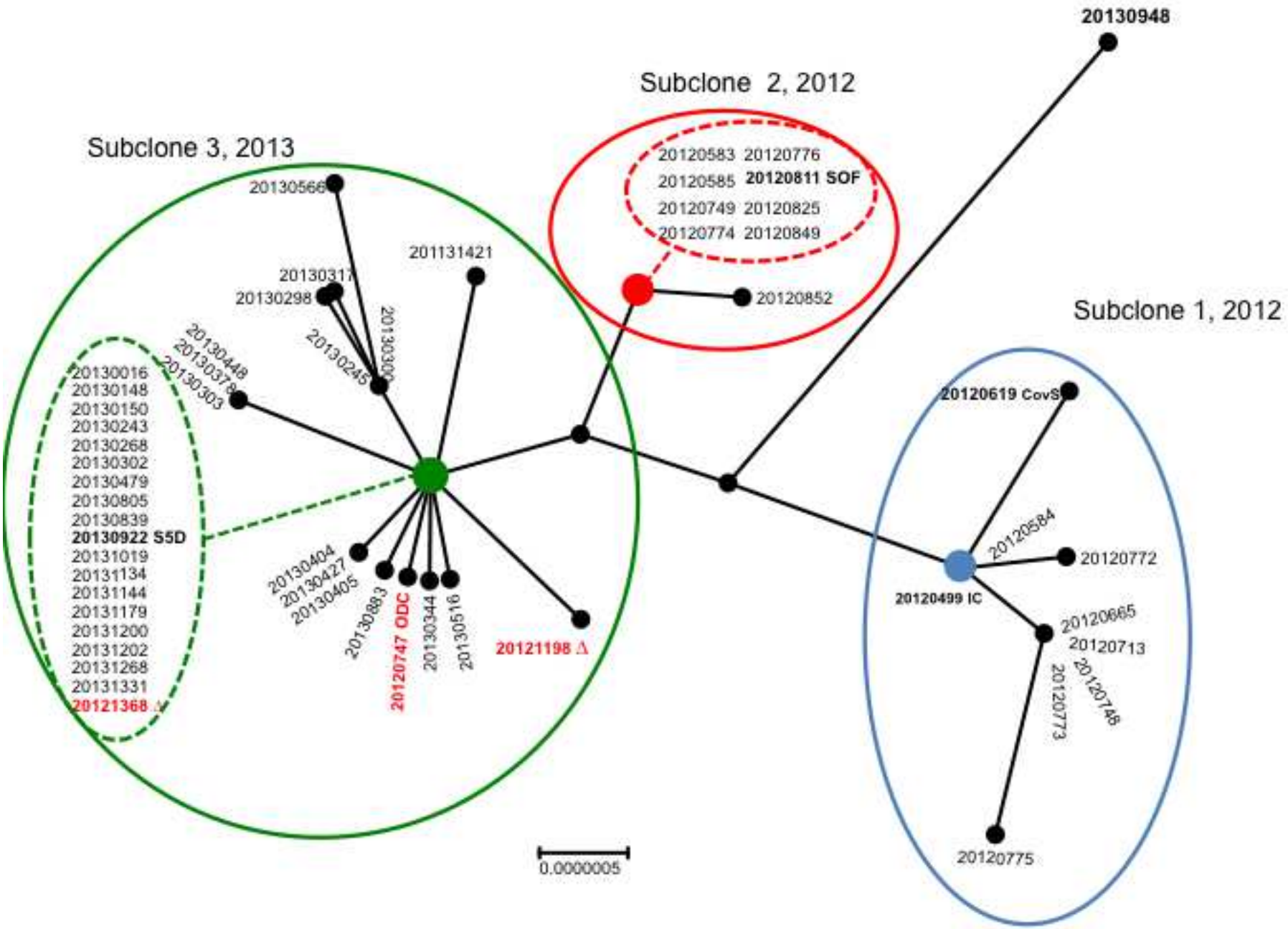

Figure 4

Click here to download Figure Fig.4.tiff

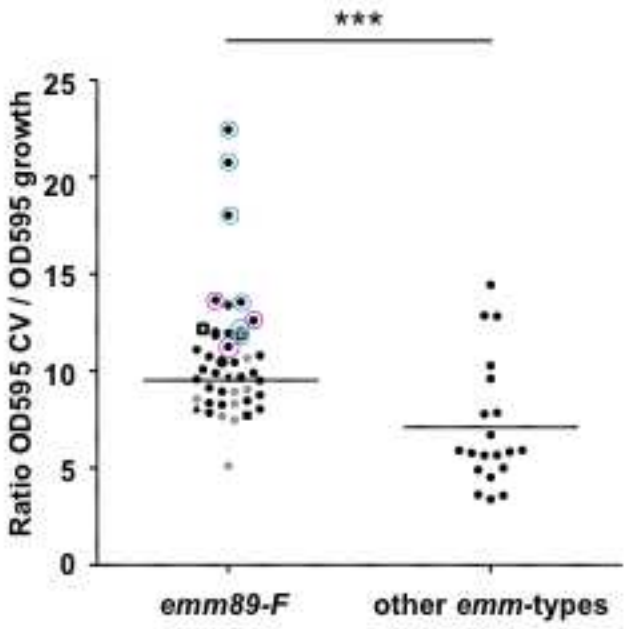

Figure 5

Click here to download Figure Fig.5.tiff

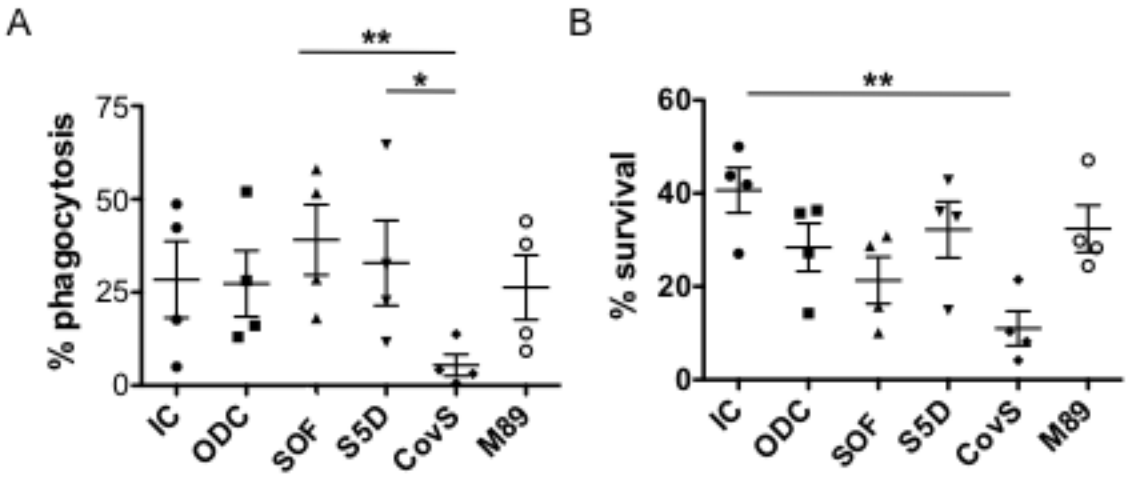

table S1

Table S1. Clinical and microbiological characteristics of *Streptococcus pyogenes* strains isolated during the outbreak

Strain	Isolation Date	Age	Sexe	Anatomical origin / Patient initials [a]	Infection status	Death	Tracheostomy cannula	HCW	Molecular markers	*emm* genotype	PFGE pattern	ATB[b]	WGS	biofilm / cellular analysis [c]
20120499	16/04/12	58	M	Blood culture	Invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	IC
20120583	04/05/12	65	M	Blood culture	Invasive		+		*speB*, *spec*, *smeZ*	*emm89*	89-A4	S	+	+
20120584	03/05/12	64	F	Blood culture	Invasive	+			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20120585	07/05/12	56	M	Blood culture	Invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20120619	14/05/12	27	F	Pharynx	Colonization			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	CovS
20120665	16/05/12	53	M	Gastrostomy site	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	
20120686	22/05/12	35	M	Pharynx	Non-invasive				*speB*, *smeZ*	*emm89*	89-A	S		
20120691	18/05/12	30	F	Pharynx	Non-invasive			+	*speA*, *speB*, *speC*, *smeZ*	*emm6*		S		+
20120713	26/05/12	65	M	Pharynx	Colonization				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20120727	29/05/12	68	M	Sputum	Colonization				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		
20120747	02/06/12	55	M	Pharynx	Colonization		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	ODC
20120748	05/06/12	64	M	Pharynx	Colonization				*speB*, *speC*, *smeZ*	*emm89*	89-A14	TcR	+	
20120749	02/06/12	61	M	Protected specimen brush	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20120772	09/06/12	38	F	Pharynx	Non-invasive			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	
20120773	07/06/12	60	M	Pharynx	Colonization				*speB*, *speC*, *smeZ*	*emm89*	89-A21	TcR	+	
20120774	06/06/12	64	F	Gastrostomy site	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	
20120775	05/06/12	37	F	Pharynx	Non-invasive			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	
20120776	11/06/12	27	F	Pharynx	Non-invasive			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20120811	14/06/12	38	M	Blood culture	Invasive	+			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	FBP
20120824	21/06/12	42	F	Pharynx	Colonization			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		
20120825	24/06/12	52	M	Sputum	Colonization		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20120849	29/06/12	58	M	Protected specimen brush	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	
20120852	28/06/12	56	M	Tracheostomy cannula	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20121026	02/08/12	81	M	Wound	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		
20121198	08/09/12	64	M	Blood culture	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A	S	+	+
20121368	20/10/12	55	M	Sputum	Non-invasive				*speB*, *smeZ*	*emm89*	89-A	S	+	+
20130016	14/12/12	64	M	Protected specimen brush	Invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20130148	25/01/13	63	M	Protected specimen brush	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20130149	25/01/13	49	M	Adenopathy discharge	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		
20130150	25/01/13	73	M	Blood culture	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20130243	12/02/13	60	M	Blood culture	Invasive	+			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20130244	11/02/13	50	M	Protected specimen brush	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		
20130245	11/02/13	64	M	Gastrostomy site	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20130246	11/02/13	50	M	Protected specimen	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		

				brush											
20130267	13/02/13	65	F	DHN/CE	Invasive				*speB*, *smeZ*	*emm104*	104-A	TcR			
20130268	15/02/13	83	F	DHN	Invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130269	15/02/13	58	M	Tracheostomy cannula	Colonization		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S			
20130270	15/02/13	52	F	Pharynx	Non-invasive				*speB*, *smeZ*	*emm104*	104-A	TcR			
20130271	15/02/13	68	M	Protected specimen brush	Invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S			
20130298	20/02/13	66	F	Tracheostomy cannula	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130299	19/02/13	84	M	Wound	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S			
20130300	19/02/13	49	M	Wound	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130301	19/02/13	54	M	Wound	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S			
20130302	19/02/13	65	M	Pharynx	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130303	20/02/13	60	M	Tracheostomy cannula/**BP**	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130316	25/02/13	72	M	Protected specimen brush	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S			
20130317	22/02/13	59	M	Tracheostomy cannula	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130318	24/02/13	69	M	Protected specimen brush	Invasive	+	+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S			
20130319	25/02/13	23	F	Pharynx	Non-invasive			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S			
20130344	04/03/13	73	F	Wound	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130365	06/03/13	74	f	Wound	Colonization				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S			
20130377	12/03/13	24	m	Pharynx	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm75*		S			
20130378	11/03/13	60	m	Tracheostomy cannula/**BP**	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130403	18/03/13	66	F	Wound/CE	Colonization				*speB*, *smeZ*	*emm104*		TcR			
20130404	20/03/13	22	M	Pharynx	Non-invasive			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130405	19/03/13	21	F	Pharynx	Non-invasive			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130427	17/03/13	87	M	Protected specimen brush	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130448	03/04/13	61	M	Tracheostomy cannula/**BP**	Non-invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+		
20130479	08/04/13	78	M	Blood culture	Invasive		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130516	15/04/13	53	M	Protected specimen brush	Invasive	+	+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130566	30/04/13	61	M	Wound	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130587	06/05/13	69	M	Wound	Non-invasive				*speB*, *speC*, *ssa*, *smeZ*	*emm4*		S			
20130805	01/07/13	65	M	Sputum	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130839	12/07/13	28	F	Pharynx/**GV**	Non-invasive			+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130882	20/07/13	80	M	Gastrostomy site	Non-invasive		+		*speB*, *smeZ*	*emm89*	89-A	S			
20130883	22/07/13	60	M	Sputum	Colonization		+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20130922	30/07/13	74	M	Blood culture	Invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	S5D	
20130948	10/08/13	NA	M	Gastrostomy site	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+	
20131019	28/08/13	56	F	Gastrostomy site	Non-invasive				*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+		
20131020	29/08/13	75	M	Wound	Non-invasive				*speB*, *smeZ*	*emm89*	89-A	S			

20131102	17/09/13	65	M	Wound	Non-invasive			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		
20131103	17/09/13	73	M	Gastrostomy site	Non-invasive			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20131134	23/09/13	28	F	Pharynx/**GV**	Colonization		+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20131144	27/09/13	26	F	Pharynx	Non-invasive		+	*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20131179	06/10/13	61	M	Blood culture/**BP**	Invasive			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20131188	12/10/13	63	M	Sputum	Colonization			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		
20131200	15/10/13	54	M	Pharynx	Colonization			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+
20131201	16/10/13	36	F	Pharynx	Colonization		+	*speB*, *speC*, *ssa*, *smeZ*	*emm87*		S		
20131202	17/10/13	74	F	Blood culture	Invasive			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	
20131268	08/11/13	54	M	Gastrostomy site	Colonization			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	
20131331	25/11/13	55	M	Blood culture	Invasive			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	
20131380	09/12/13	73	F	Sputum	Colonization			*speB*, *speC*, *smeZ*	*emm89*	89-A4	S		
20131421	23/12/13	41	M	Tracheostomy cannula	Colonization	+		*speB*, *speC*, *smeZ*	*emm89*	89-A4	S	+	+

[a] The patient's initials, in bold, was added when more than one strain was collected from the same individual

[b] S, susceptible to all antibiotic tested; TcR, tetracycline resistant

[c] + = biofilm formation capacity studied; initials refer to the strains whose interactions with A549 and THP1 have been further studied

Table S2. Control strains used

CNR Strain number	**Anatomical origin**	***emm* genotype**	**Molecular markers** [a]	**Antibiotic resistance** [b]
20120456	Blood culture	*emm89*	*speB*, *speC*	S
20141462	Bronchial aspiration	*emm89*	*speB*, *speC*	S
20141463	Wound	*emm89*	*speB*, *speC*	S
201410469	Wound	*emm89*	*speB*, *speC*	S
20040562	Blood culture	*emm1*	*speA*, *speB*, *speJ*	S
20070928	Blood culture	*emm1*	*speA*, *speB*, *speC*	S
20071056	Blood culture	*emm1*	*speA*, *speB*	S
20080059	Blood culture	*emm1*	*speA*, *speB*	S
20080166	Blood culture	*emm1*	*speA*, *speB*	S
20080176	Blood culture	*emm1*	*speA*, *speB*	S
20080304	Blood culture	*emm1*	*speA*, *speB*	S
20090096	Blood culture	*emm1*	*speA*, *speB*	S
20090301	Blood culture	*emm1*	*speA*, *speB*	S
20080406	Blood culture	*emm101.2*	*speB*	Tc^R
20090408	Blood culture	*emm2*	*speB*, *speC*	S
20090207	Blood culture	*emm2*	*speB*, *speC*	S
20060663	Knee hygroma	*emm28*	*speA*, *speB*	S
20071042	Blood culture	*emm28*	*speA*, *speB*	Kan^R, Str^R, Ery^R, Cli^R, Tc^R
20090413	Blood culture	*emm28*	*speB*, *speC*	S
20090146	Blood culture	*emm28*	*speB*, *speC*	S
20071057	Vagina	*emm28*	*speB*, *speC*	Kan^R, Str^R, Ery^R, Cli^R
20090450	Vagina	*emm28*	*speB*, *speC*	S
20090153	Blood culture	*emm3*	*speA*, *speB*	S
20080308	Blood culture	*emm90,2*	*speB*, *speC*	Tc^R

[a] The presence of *smeZ* was not assayed.
[b] S, susceptible to all antibiotics tested. ^{R}, resistant; Cli, clindamycin; Ery, erythromycin; Kan, kanamycin; Str, streptomycin; Tc, tetracycline

Table S3. Relevant mutations occuring in the sequenced isolates

position			4334	7957	9031
type of mutation			G → A	C → A	22 bp x 2
mutation			G16D (GGC → GAC)	D525Y (GAT → TAT)	duplication
transcription orientation of the gene			→	←	→
Strain name	**mutation profile** [a]	**biofilm / cellular analysis** [b]	hypothetical protein	Dipeptide-binding ABC transporter, periplasmic substrate-binding component DppA (TC 3.A.1.5.2)	Fibronectin-binding protein
20120499	AP	IC			
20120583	AL	Yes			
20120584	AM	Yes			
20120585	E	Yes			
20120619	G	CovS			
20120665	AK				
20120713	AK	Yes			
20120747	AG	ODC			
20120748	AK				
20120749	U	Yes			**X**
20120772	AP				
20120773	AK				
20120774	AP				
20120775	AK				
20120776	AP	Yes			
20120811	AN	FBP			
20120825	T	Yes			**X**
20120849	AP				
20120852	AJ	Yes			
20121198	K	Yes			
20121368	V	Yes			
20130016	AF	Yes			
20130148	AE	Yes			
20130150	J	Yes			
20130243	AO	Yes			
20130245	R	Yes			
20130268	AO				
20130298	Q	Yes			
20130300	S	Yes			
20130302	AD	Yes			
20130303	I	Yes			
20130317	P	Yes			
20130344	L	Yes			
20130378	H	Yes			
20130404	Z	Yes			
20130405	Y	Yes			
20130427	Z	Yes			
20130448	I				
20130479	AH	Yes			
20130516	W	Yes	**X**		

20130566	D	Yes			
20130805	AQ	Yes			
20130839	AH	Yes			
20130883	AB	Yes			
20130922	X	S5D			
20130948	F	Yes			
20131019	AH				
20131103	-	Yes			
20131134	C	Yes			
20131144	A	Yes			
20131179	B	Yes			
20131200	AO	Yes			
20131202	A	Yes			
20131268	A				
20131331	A				
20131421	AA	Yes			

[a] The profile is not indicated for the strains outside the epidemic clone

[b] yes, biofilm formation assayed; strains further analyzed are indicated by their acronymes

13953	14733	21044	21074
Δ1 bp	C → T	11 bp x 2	C → A
coding (189/1284 nt)	A142T (GCC → ACC)	duplication	A38D (GCC → GAC)
←	←	→	→
FIG009439: Cytosolic protein containing multiple CBS domains	L-xylulose 5-phosphate 3-epimerase (EC 5.1.3.-)	Transmembrane histidine kinase CovS	Transmembrane histidine kinase CovS
		X	
X			
			X
			X
			X

21763	21964	22055	22168
Δ2 bp	C → T	A → G	Δ5 bp
coding (802-803/1503 nt)	Q335* (CAA → TAA)	D365G (GAT → GGT)	coding (1207-1211/1503 nt)
→	→	→	→
Transmembrane histidine kinase CovS	Transmembrane histidine kinase CovS	Transmembrane histidine kinase CovS	Transmembrane histidine kinase CovS
			X
		X	
X			

	X		

22538	37183	37290	50763
C → T	C → T	G → A	C → A
T147I (A<u>C</u>T → A<u>T</u>T)	S79F (T<u>C</u>C → T<u>T</u>C)	A115T (<u>G</u>CT → <u>A</u>CT)	D107Y (<u>G</u>AT → <u>T</u>AT)
→	→	→	←
Dihydroorotate dehydrogenase, catalytic subunit (EC 1.3.3.1)	Topoisomerase IV subunit A (EC 5.99.1.-)	Topoisomerase IV subunit A (EC 5.99.1.-)	Response regulator of the competence regulon ComE
X			
			X

		X	

51299	60020	61356	61564
C → T	22 bp x 2	G → A	C → A
H97Y (CAT → TAT)	duplication	A403T (GCG → ACG)	A472D (GCT → GAT)
→	→	→	→
Pyruvate formate-lyase activating enzyme (EC 1.97.1.4)	Fibronectin-binding protein	Fibronectin-binding protein	Fibronectin-binding protein
		X	
			X
			X
			X
			X
			X
	X		
X			

	X	**X**	

61557	62662	62702	62763	62796
Δ1 bp	C → T	G → T	G → A	A → G
coding (1408/2967 nt)	A838V (G<u>C</u>A → G<u>T</u>A)	E851D (GA<u>G</u> → GA<u>T</u>)	V872I (<u>G</u>TC → <u>A</u>TC)	I883V (<u>A</u>TC → <u>G</u>TC)
→	→	→	→	→
Fibronectin-binding protein	Fibronectin-binding protein	Fibronectin-binding protein	Fibronectin-binding protein	Fibronectin-binding protein
			X	
		X		**X**
	X			
			X	**X**
			X	
X				
X				

			X	
			X	

62819	62850	62853	70969	75410
T → G	G → A	A → G	C → T	G → T
D890E (GA<u>T</u> → GA<u>G</u>)	G901S (<u>G</u>GT → <u>A</u>GT)	M902V (<u>A</u>TG → <u>G</u>TG)	T10I (A<u>C</u>A → A<u>T</u>A)	G278V (G<u>G</u>C → G<u>T</u>C)
→	→	→	→	→
Fibronectin-binding protein	Fibronectin-binding protein	Fibronectin-binding protein	Thymidine kinase (EC 2.7.1.21)	M protein trans-acting positive regulator (Mga)
	X	**X**		
	X	**X**		
X	**X**	**X**		
	X	**X**		
	X	**X**		
	X	**X**		
X	**X**	**X**		
	X	**X**		**X**
	X	**X**		
	X	**X**		
	X	**X**		

	X	X		
X				
	X	X		
	X	X		
			X	
X	X	X		
	X	X		

75758	78460	97446	109944
A → G	G → A	T → C	A → G
E394G (GAA → GGA)	D178N (GAT → AAT)	N193S (AAT → AGT)	D274G (GAC → GGC)
→	→	←	→
M protein trans-acting positive regulator (Mga)	Methionyl-tRNA synthetase (EC 6.1.1.10)	ATP synthase beta chain (EC 3.6.3.14)	Oxaloacetate decarboxylase beta chain (EC 4.1.1.3)
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X

X			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
		X	X

136599	145637	194925	249724
Δ26,400 bp	Δ1 bp	T → A	Δ1 bp
	coding (612/636 nt)	S296C (AGC → TGC)	coding (814/873 nt)
[] –	→	←	←
34 genes deletion of the phage containaing *speC*	Guanylate kinase (EC 2.7.4.8)	6-phosphofructokinase (EC 2.7.1.11)	Shikimate 5-dehydrogenase I alpha (EC 1.1.1.25)
X			
X			
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
			X
		X	**X**
		X	**X**
		X	**X**
			X
			X
			X

			X
			X
			X
			X
			X
			X
			X
			X
	X		X
	X		X
	X		X
			X
	X		X
	X		X
	X		X
			X

Fig.S1

Click here to download Electronic Supplementary Material Fig.S1.tiff