JIDC | THE JOURNAL OF INFECTION IN DEVELOPING COUNTRIES

Case Report

Genome analysis of a MDR *Streptococcus pneumoniae* 23F serotype causing meningoencephalitis in a 10-months refugee infant

Tamara Salloum[1], Elie Tannous[1], Samar Merheb-Ghoussoub[2], Elie Ghoussoub[3], Sima Tokajian[1]

[1] *Department of Natural Sciences, School of Arts and Sciences, Lebanese American University, Byblos, Lebanon*
[2] *Pierre and Marie-Curie University, Paris, France*
[3] *Paris Descartes University, Paris, France*

Abstract

Purpose: *Streptococcus pneumoniae* is an important human pathogen causing invasive pneumococcal diseases (IPD). The re-emergence of eradicated *S. pneumoniae*-associated meningoencephalitis in Lebanon is a major point of concern.

Methods: We aimed at conducting a comparative genome analysis of a multi-drug resistant *S. pneumoniae*, LAU-23F, linked to meningoencephalitis and fatality in a 10-months Syrian refugee infant in Lebanon, and 24 related publically available genome sequences. Serotype, capsular genes, MLST, SNPs, phylogenetic relatedness and repertoire of resistance genes were investigated. Genes encoding penicillin binding proteins (PBPs) were examined for mosaicity. Virulence factors were screened for SNPs as compared to reference strains.

Results: The isolate belonged to ST-277 and was of serotype 23F. It showed an intermediate resistance to ciprofloxacin, cefuroxime and penicillin and carried multiple components of different efflux pumps. Gene mosaicity was observed in *pbp2x*, it was also distinct from other penicillin-resistant strains; *pbp1a* and *pbp2b* appeared to be conserved between LAU-23F and the reference strain SP49. The arrangement of capsular gene loci was similar to ATCC 700669 though polymorphism was detected in the *cpsABCD* region, believed to be conserved among different *Streptococcus* species. Amplitude of virulence factors was detected showing varying degrees of conservation compared to reference strains. Observed zones of high heterogeneity were associated with phage encoded regions.

Conclusions: The fine levels of diversity throughout the genome could account for the pronounced invasiveness of this isolate. The genomics-based methods used support the importance of implementing WGS in routine clinical diagnostics and surveillance of streptococcal diseases.

Key words: Whole-genome sequencing; *S. pneumoniae;* 23F serotype; ST-277; SNPs; virulence.

J Infect Dev Ctries 2018; 12(3):196-203. doi:10.3855/jidc.10180

(Received 15 January 2018 – Accepted 26 February 2018)

Introduction

α-hemolytic *Streptococcus pneumoniae* (pneumococcus) is a major cause of bacterial meningitis in infants (< 24 months) and pneumonia in elderly [1]. Mortality rates of untreated bacterial meningitis approach 100 % and can still occur with optimal therapy [2].

The rapid intercontinental spread of drug-resistant *S. pneumoniae* is alarming [2]. Mass population movements, inadequate shelter and hygiene conditions and interrupted vaccination programs increase the risk and contribute to the spread of communicable diseases [3]. Syria is experiencing a protracted political and socioeconomic crisis accompanied with mass population displacement, with access to health care being severely restricted by security factors and/or poor living condition within refugee camps [4]. According to the Lebanese ministry of public health, 18 and 33 cases of *S. pneumoniae* associated meningitis were reported in 2014 and 2016, respectively. In 2016, *S. pneumoniae* associated meningitis made 7.2% out of the total meningitis cases reported, out of these 15 cases affected the 0-4 year's old age category and at least six were fatal [5].

Fluoroquinolone and penicillin resistant *S. pneumoniae* were previously described in Lebanon [6-8]. Moreover, resistance to β-lactams through modifications in the penicillin-binding proteins (PBPs) is a hallmark of adaptability in pneumococci [6].

The capsule is the major virulence factor (VF) in *S. pneumoniae*. It is made out of polysaccharides covering the cell wall and acts as the principal anti-phagocytic and defensive component that prevents access of the leukocytes to the cell wall [9]. The genes coding for the capsule biosynthesis in *S. pneumoniae* were previously designated as *cps* [10] or *cap* [11]. All *cps* loci in *S. pneumoniae*, except for type 37, are located between *dexB* and *aliA* co-transcribed from a promoter located

upstream of the first *cps* gene, also referred to as *wzg* or *cpsA* [12].

The capsular polysaccharides are used for identification and serotyping, with 97 different serotypes of capsulated pneumococci being already described [13]. These are targeted by either 23-valent or 13-valent pneumococcal vaccines [9]. Serotypes 6, 14, 18, 19 and 23 are the predominant disease causing types [9]. Serotype 23F is among the 23 serotypes that account for approximately 90% of invasive pneumococcal infections [14]. Capsule switching to alternative types such as that of serotype 23F to 14, 19A or 19F has been previously reported [14].

The use of whole-genome sequencing (WGS) permits the identification of proteins involved in pathogenesis and that are exploited as potential vaccine targets [13]. In this case report, we perform the genome sequencing and comparative genome analysis of *S. pneumoniae* ST-277 serotype 23F causing fatal meningoencephalitis in a 10-month Syrian refugee infant hospitalized in Lebanon with 24 publically available genomes sequences. This is the first report of WGS analysis of *S. pneumoniae* in Lebanon and Syria, which emphasizes on the importance of implementing WGS in routine clinical diagnostics and surveillance of streptococcal disease.

Methodology

Ethics Statement

The bacterial sample was collected and stored as part of routine clinical care. Ethical approval was not necessary as the patient remained anonymous.

Case Description

The 10-months old female patient living in Halat, Lebanon, was admitted to the pediatric intensive care unit (PICU) at the Centre Hospitalier Universitaire Notre Dame de Secours hospital on the 17/05/2016 for high-grade fever and seizure. She has received three doses against hepatitis B, three doses to diphteria-tetanus-pertussis and oral poliovirus vaccines. She was not vaccinated against *Pneumococcus* or Rotavirus. The history of the illness goes back to five days prior to the presentation when the child developed spikes of fever reaching 40°C, responsive to antipyretics. She also developed watery, non-bloody diarrhea, non-mucoid, two episodes per day and multiple episodes of projectile non-bloody, non-bilious vomiting for which she received Domperidone, oral rehydration salts and cefpodoxime (Third-generation cephalosporin). Twenty-four hours prior to hospitalization, somnolence was recorded with reduced poor oral (PO) intake, complicated by focal tonico-clonic seizures of the left body side. Upon arrival to the ER, the patient was somnolent, lethargic, Glasgow Coma scale 10/15, with bulging anterior fontanel and normal deep tendon reflexes all over extremities. Her vital signs were SaO_2:100%, heart rate (HR):157/minute, blood pressure (BP):110/70. Urgent brain CT scan with and without contrast showed signs of diffuse meningoencephalitis. Full sepsis workup on the first and second days of hospitalization results are summarized in Table 1.

C-reactive protein (CRP) was 290 mg/mL on 17/05/2016, 250 mg/mL on 18/05/2016 and 35 mg/mL on 20/05/2016. Lumbar puncture with cerebro-spinal fluid (CSF) analysis showed a very thick, turbid CSF, WBC: 5,280 (N: 90%), glucose: 1 mmol/L, presence of pus and presence of bacteria in chains and pairs. The soluble antigen test and CSF culture were positive for *S. pneumonia* with heavy growth. A repeat CSF culture taken on 27/05/2016 was negative.

The patient was intubated and sedated on continuous Midazolam, Epanutin IVL Q 12 hours for status epilectus (tonico-clonic seizures, nystagmus and right upper extremity clonic movements). Ceftriaxone was given to the child (100mg/Kg/Day Q 12 hours) from 18/05/2016 to 22/05/2016, Vancomycin (20mg/Kg/Dose Q 6 hours) from 18/05/2016 to 08/06/2016 (21 days) and Dexamethasone Q 6 hours for four days prior to the first dose of antibiotics. The patient died shortly after treatment.

Table 1. Results of the full sepsis workup of the patient at the first and second days of administration to the hospital.

Test	Day 1	Day 2
Hemoglobin (Hb)	9.6 mg/dL	9.2 mg/dL
Hematocrit (hct)	28.7%	28.3%
Mean corpuscular volume (MCV)	73 fl	72 fl
Red cell distribution width (RDW)	17.1%	16.5%
Platelets (PLTs)	351,000	184,000
White blood cells (WBCs)	9,400	17,200
Monocytes (N)	67%	76%
Lymphocytes (L)	32%	21%

Antibiotic Susceptibility Testing

Antimicrobial susceptibility testing was performed using the disk diffusion method against 24 antibiotics. Results were interpreted according to the EUCAST breakpoints as updated in 2015 (http://www.eucast.org) [15].

DNA Isolation & Genome Sequencing

DNA extraction was performed using the Nucleospin Kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. Genomic DNA (gDNA) was used as input for library preparation using the Illumina TruSeq DNA library preparation kit (Illumina, San Diego, CA, USA). The library was multiplexed, clustered, and sequenced on an Illumina MiSeq with paired-end 500 cycles protocol to read a length of 250 bp.

Multilocus Sequence Typing

Multilocus sequence typing (MLST) was performed through the amplification of seven housekeeping genes (*aroE*, *gdh*, *gki*, *recP*, *spi*, *xpt* and *ddl*) as previously described [16].

Genome Analysis

Genome assembly was performed *de novo* using A5 with default parameters [17]. tRNAs and rRNAs were detected using ARAGORN v1.2.36 [18] and the RNAmmer Prediction Server 1.2 [19], respectively. MLST 1.8 server was used to determine the ST [20]. The genome was subjected to BLAST analysis against Resfinder 3.0 [21] the Antibiotic Resistance Database (ARDB) [22] and the Comprehensive Antibiotic Resistance Database (CARD) [23]. PHASTER detected putative phage sequences in the genome [24]. Capsular serotype was determined using the automated WGS-based serotyping bioinformatics tool, PneumoCaT (Pneumococcal Capsule Typing) [25]. The presence of genomic islands (GIs) was determined using IslandViewer 4 [26]. The sequences of *gyrA* and *parC* were extracted and compared with reference sequences for *parC* (accession no. Z67739) [27], *gyrA* (accession no. U49087) [28]. Genes encoding capsular loci and major VFs were extracted and aligned with reference genes using BioNumerics v7.6.1 beta software (Applied Maths, Sint-Martens-Latem, Belgium). The obtained alignment was used to screen and locate single nucleotide polymorphism (SNPs).

SNPs-based Phylogenetic Analysis

The sequenced genome was aligned against the chromosome of SP49 using the Burrows–Wheeler Aligner (BWA). SNP-calling was performed by mapping the genome of LAU-23F to 24 genome sequences of *S. pneumoniae* strains retrieved from NCBI using the BioNumerics v7.6.1 beta software (Applied Maths, Sint-Martens-Latem, Belgium). A neighbor-joining (NJ) tree of categorical differences was drawn in BioNumerics by using synonymous mutations from the filtered wgSNPs data as input.

Comparative Genome Analysis

The obtained scaffolds were ordered and oriented with Mauve [29] using SP49 genome as a reference. Genomes were aligned using ProgressiveMauve to identify conserved segments known as Locally Collinear Blocks (LCBs) and search for heterogeneity zones [29]. The Average Nucleotide Identity (ANI) online tool (http://enve-omics.ce.gatech.edu/ani/) was used to calculate the similarity between LAU-23F and closely related strains by using default settings [30]. CGView was used to visualize the circular genome blast hits of LAU-23F against SP49 [31].

Nucleotide sequence accession number

The datasets generated during the current study are available in the DDBJ/ENA/GenBank repository under the accession number MLQV00000000.1.

Results

Genome Annotation

The obtained 2,058,652 bp genome consisted of 54 contigs, 39.7% GC, 48 tRNAs and six rRNAs. The isolate belonged to ST-277 and was of capsular serotype 23F.

Antibiotic Resistance

LAU-23F had an intermediate resistance to ciprofloxacin, cefuroxime and penicillin.

Multiple antibiotic resistance determinants such as components of efflux pumps (*patB, imrC, imrD, pmrA, taeA, mefE,*) associated with resistance to ciprofloxacin and norfloxacin, *mfd* conferring resistance to fluoroquinolones, along with *pmrE* conferring resistance to polymyxin and *ileS* resistance to mupirocin were detected.

Alignment of *pbp2x* in LAU-23F, SP49, TIGR4, R6, ATCC 700669 and Hungary 19A revealed similarities ranging between 88.2-92.9%. *pbp2x* in LAU-23F was most closely related to SP49 (90.2% similarity). In total, 220 SNPs were detected between *pbp2x* in LAU-23F and SP49 and 232 SNPs between *pbp2x* in LAU-23F and R6, the penicillin sensitive strain. *pbp2b* in LAU-23F was 100% similar to that of

SP49 and 88.9-96.1% similar to the other isolates. Similar results were obtained for *pbp1a* . No mutations were detected in *gyrA*; S132N and I462V amino acid substitutions were found in *parC*.

Virulence Factors

Capsule

LAU-23F and ATCC 700669, both of serotype 23F, had an identical capsular gene arrangement. *cpsABCD* was examined for polymorphism. *cpsA* in LAU-23F was most closely related to ATCC 700669 and most distantly to TIGR4. *cpsA* had 162 SNPs, 70 in *cpsB*, 179 in *cpsC*, and 210 in *cpsD* as compared to the reference strains (Figure 1).

Adherence

LAU-23F did not express neither the type I pilus (PI-1) nor the type II (PI-2). One single A→G change was observed in *pavA* between LAU-23F and SP49 at position 856 bp, while ten SNPs were detected in *pavA* between LAU-23F and ATCC 700669.

PLY

Only Four SNPs in *ply* were detected between LAU-23F and ATCC 700669. Similarly, four distinct SNPs were found in *ply* between LAU-23F and SP49.

LPXTG-anchored proteins

hlyA hyaluronidase in LAU-23F was 99.3% similar to *hlyA* in TIGR4. *hlyA* in SP49, Hungary ß19A-6 and ATCC 700669 appeared to be significantly different from the latter two strains. In total 24 SNPs were found in *hlyA* between LAU-23F and TIGR4. No deletions or premature stop codons were detected. All strains carried *nanA*. *nanA* in ATCC 700669, Hungary19A-6, SP49 and LAU-23F appeared to be 99% similar and clustered separately from *nanA* in TIGR4. Although significant variation was observed in *prtA* in TIGR4, Hungary19A-6 and ATCC 700669 (% similarity < 55%), *prtA* in LAU-23F and SP49 were 100% identical.

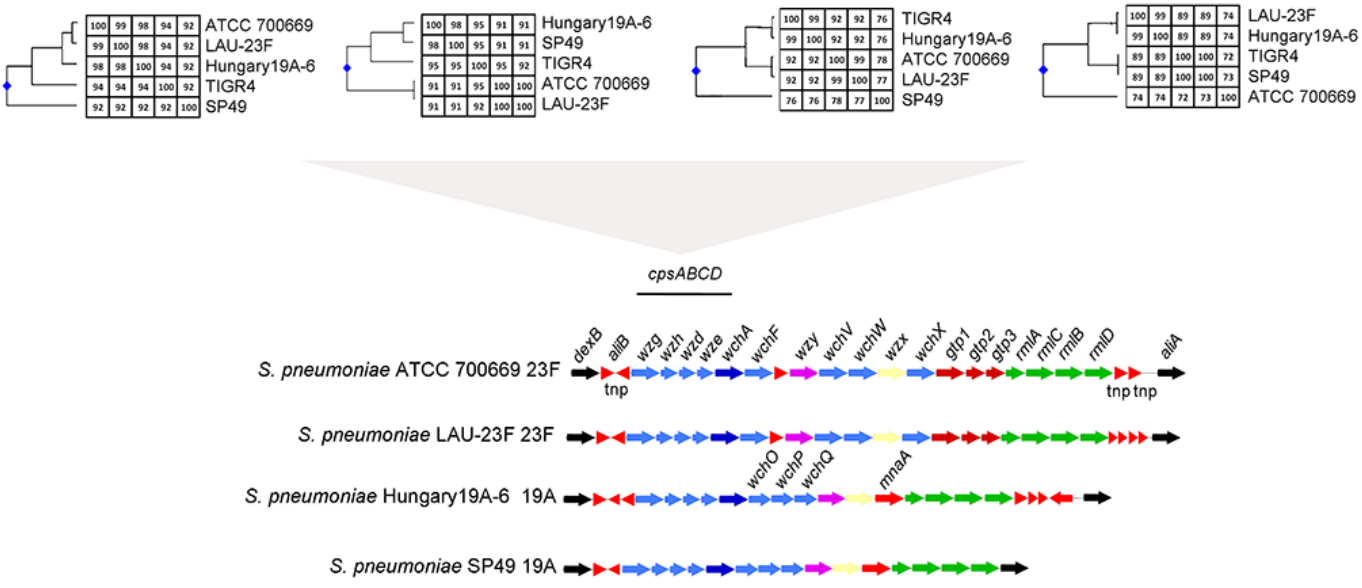
Lipoproteins

psaA in LAU-23F, ATCC 700669, Hungary19A-6, SP49 and TIGR4 were 99.5-100% identical. *psaA* in LAU-23F was more closely related to SP49. Also, SlrA and PpmA, lipoprotein peptidyl prolyl isomerases were 99.3-100% similar between all isolates. *slrA* and *ppmA* in LAU-23F was more closely related to SP49 then to others. Iron compound ABC uptake transporter proteins were also highly conserved in LAU-23F and reference strains.

Choline-Binding Proteins (CBPs)

Ten CBPs were present in LAU-23F and ATCC 700669. *lytA* in LAU-23F was only 87.2% identical with SP49 and 120 SNPs detected. *lytB* in LAU-23F and Hungary19A-6 were 100% identical, while *lytB* in ATCC 700669 was only 66% similar to the other two isolates. *lytB* in TIGR4 was more genetically divergent and thus was not included in the comparison. *lytC* was 99.1-100% identical in LAU-23F and all references, except in SP49 showing only 28% similarity. *cbpE* in

Figure 1. Capsule biosynthesis genes in LAU-23F, ATCC 700669, SP49 and Hungary19A-6. Genes were colored as previously described by Bentley *et al.* (2006) [12]. UPGMA phylogenetic relatedness and percentage similarity of the *cpsABCD* genes is indicated.



LAU23-F was 99% identical to that of Hungary19A-6 while ATCC 700669 and SP49 appeared to be only 57% and 27% similar, respectively.

Phages & Genomic Islands

LAU-23F carried four incomplete phages: *Streptococcus* phage phiARI0460-1, *Pseudoalteromonas* phage PH1, *Enterobacteria* phage phi92 and *Salmonella* phage 118970_sal4. Similarly, SP49 carried an intact *Streptococcus* phage phiARI0460-1 and *Lactococcus* 63301 phage and an incomplete *Ralstonia solanacearum* phage.

In total, 24 genomic islands (GIs) were detected ranging in size between 4,050 bp to 31,602 bp. The majority were associated with the ABC transporter system and iron uptake or carried phage related proteins (Table S1; Figure S1).

Comparative Genome Analysis

Comparative genome analysis by ANI [30] and circular genome mapping with SP49 resulted in 99.73% identity (Figure 2).

Phylogenetic analysis based on wgSNPs separated the 24 *S. pneumoniae* isolates downloaded from NCBI into five clusters (Figure 3; Table S2). LAU-23F was most closely related to SP49 (Serotype 19A; ST-277; highly susceptible to macrolides) isolated in Germany [32].

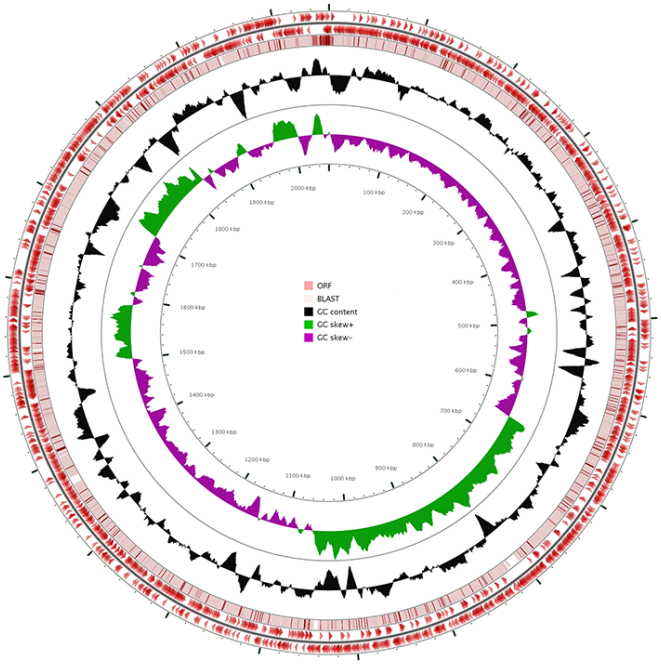
Mauve alignment of the isolates belonging to clade IV identified nine LCBs with few inversions between LAU-23F and SP49 but none with Hungary19A-6 (Figure S2). Accordingly, most of the genomes included in the analysis appeared homogeneous with matching DNA segments. However, using genome alignment on BioNumerics, we detected at least three distinctive non-matching regions designated as 'high heterogeneity zones' (HHZ1; HHZ2 and HHZ3). HHZ1 (23,500-98,000 bp) and HHZ2 (1,452,800-1,494,000 bp) encoded phage-related proteins, while HHZ3 (1,671,600-1,720,000 bp) encoded cell wall surface anchoring proteins. Other smaller regions of non-similarity were related either to phages or to capsular polysaccharide proteins (Figure 4).

Function-based genome comparison of LAU-23F with SP49 using RAST revealed five unique regions in LAU-23F corresponding to carbohydrate and lipid metabolism, and cell wall components (Table S3).

Discussion

S. pneumoniae serotype 23F is amongst the most common causes of IPD [3]. In this case report, the genome sequencing and comparative genome analysis of an emerging *S. pneumoniae* ST-277 serotype 23F causing fatal meningoencephalitis in a 10-months Syrian refugee infant hospitalized in Lebanon is presented.

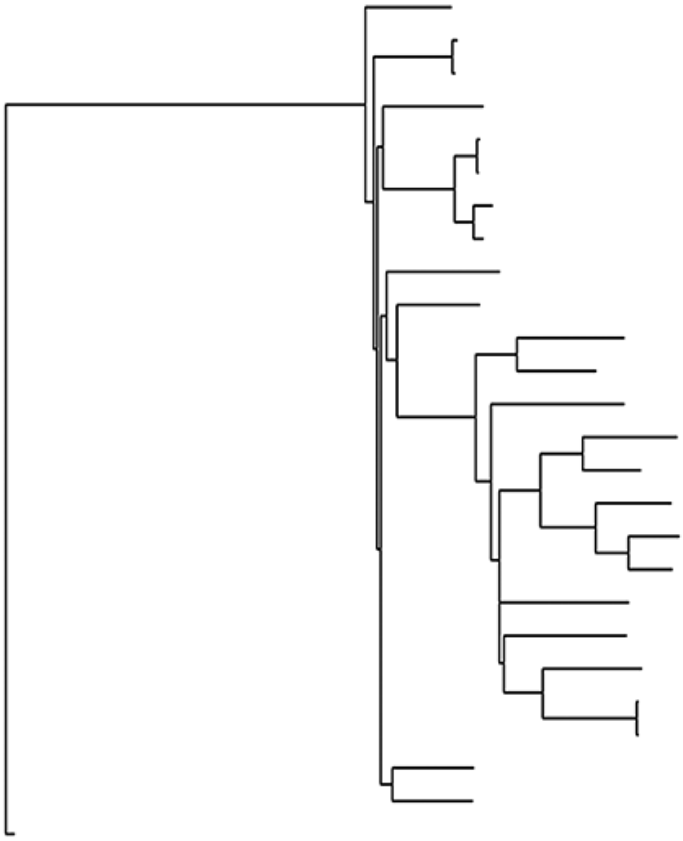
Figure 2. Circular genome representation of LAU-23F as compared to SP49 reference genome. Starting from the outermost ring the feature rings depict: LAU-23F coding DNA sequence (CDS) on the forward strand (red), LAU-23F CDS on reverse strand (red), SP49 Blast results (pink) representing the positions covered by the BLASTN alignment, G+C content (black), G+C positive skew (green) and G+C negative skew (purple). The image was created by CGview Server V 1 [30].



LAU-23F showed an intermediate resistance to ciprofloxacin, cefuroxime and penicillin. Fluoroquinolone and penicillin resistant *S. pneumoniae* were previously described in Lebanon [6-8]. However, with the recent increase in the number of refugees and lack of vaccination among children, streptococcal infections are on the rise [3]. Resistance to β-lactams in *S. pneumoniae* occurs through mutations in the genes coding for the penicillin binding proteins (PBPs) that play an essential role in the bacterial cell wall synthesis [8]. As in LAU-23F, β-lactam-resistant clinical isolates are almost always resistant to both penicillin and cephalosporins [33]. In LAU-23F, the detected low-affinity PBP2× mutants and PBP1a variants explain resistance to ß-lactams [8].

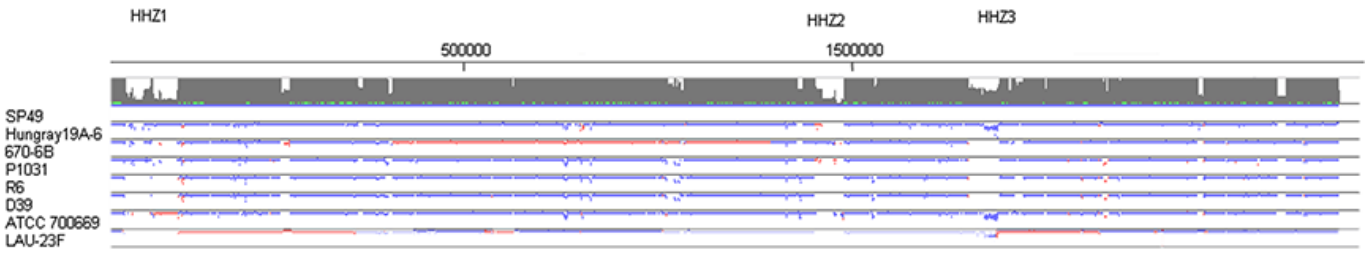
Fluoroquinolone resistance in *S. pneumoniae* has been less frequently documented compared to that for penicillin, tetracycline and chloramphenicol [34].

Figure 3. Neighbor-joining (NJ) tree of wgSNP calling performed by mapping the paired-end reads of LAU-23F and 24 *S. pneumoniae* genomes downloaded from NCBI. The sequences data were aligned against the chromosome of the SP49 reference genome using the Burrows –Wheeler Aligner (BWA). *S. pyogenes* M1 GAS was used as an outgroup to root the tree.



Name	ST	Clade	Serotype	Accession #
D141	353	I		CP016633.2
R6	595	II	2	NC_003098
D39	595		2	NC_008533
670-6B	90	III	6B	NC_014498
TCH8431/19A	320		19A	NC_014251
SP61	2432		19A	CP018137.1
ST556	1392		19F	CP003357.2
Taiwan19F-14	236		19F	NC_012469
SPNA45	6934	IV	3	HE983624.1
P1031	303		1	NC_012467
CGSP14	15		14	NC_010582
INV200	9		14	FQ312029.1
NCTC7465	615		1	LN831051.1
SP49	277		19A	CP018136.1
LAU-23F	277		23F	MLQV00000000
Hungary19A-6	268		19A	NC_010380
Hu15	226		19A	CP020551.1
Hu17	226		19A	CP020549.1
70585	289		5	CP000918.1
TIGR4	205		4	NC_003028
ATCC 700669	81		23F	NC_011900
JJA	66		14	CP000919.1
G54	63	V	19F	NC_011072
AP200	62		11A	NC_014494
S. pyogenes M1 GAS				AE004092.2

Figure 4. Whole-genome alignment of LAU-23F and seven reference genomes (SP49, Hungary19A-6,670-6B, P1031, R6, D39 and ATCC 700669) constructed using BioNumerics v7.6.1 beta software. Stretches mapping on the second sequence in direct orientation are indicated with a blue arrow, those mapping on the second sequence in inverted orientation are marked with a red arrow.



The first case of fluoroquinolone-resistant *S. pneumoniae* in Lebanon was reported in 2006 [6]. LAU-23F showed an intermediate resistance to fluoroquinolones (ciprofloxacin). Two mechanisms contribute to fluoroquinolone resistance in pneumococci: target alteration and active efflux. The first mechanism results from mutations in the quinolone resistance-determining region (QRDR) encoding *gyrA* and *parC* subunits. No mutations however, were detected in *gyrA* in LAU-23F, whereas S132N and I462V amino acid substitutions were seen in *parC*. Mutations in a single target enzyme could be precursors of resistance [35]. The second mechanism contributes to fluoroquinolone resistance through the overexpression of the ABC transporter genes *patA* and *patB*, both found in the LAU-23F genome [36].

Genome plasticity is favored by the capacity of the pneumococcus to obtain DNA through natural competence and by genome rearrangements through mobile components in its genome [14]. Zones of high heterogeneity (HHZ) detected in LAU-23F indicate that the gain and loss of genetic components through phages are essential and are linked to strain-specific differences [37]. Phages could encode toxins or other VFs that contribute to the pathogenesis in *S. pneumoniae* [37].

Due to its antiphagocytic activity, the polysaccharide capsule is probably the most important VF in pneumococci [13]. LAU-23F and ATCC 700669, both of serotype 23F had an identical capsular gene arrangement. Extensive inter-strain sequence variations in the *cps* promoter (the intergenic sequence between *dexB* and *cpsA*) were linked to variations in capsule production, adhesion to host epithelial cells, evasion of phagocytosis and virulence [38]. Interestingly, we noticed that SNPs frequently occur in the *cpsABCD* region suggesting that this region should not be overlooked when studying capsule variations.

Another major pneumococcal VF is *ply* encoding pneumolysin that was linked to brain damage and deafness in pneumococcal meningitis [39]. Similarly to the other reference genomes used, LAU-23F also carried *ply*. Only four SNPs were detected in this region when comparing LAU-23F to ATCC 700669 and LAU-23F to SP49.

At least, 17 putative LPXTG-anchored proteins were found in the pneumococcal genome [38]. Hyaluronidase, encoded by *hlyA* and secreted by most clinical isolates, favors inflammation during pneumococcal infections [38]. *hlyA* in LAU-23F showed to be 99.3% similar to *hlyA* TIGR4, and different from the one detected in the other reference strains (SP49, Hungary 19A-6 and ATCC 700669). No deletions or premature stop codons were detected suggesting that the hyaluronidase is functional and hence can be linked to rapid disease onset [40].

The number of CBPs varies between *S. pneumoniae* strains, with ten and 15 CBPs being detected in R6, LAU-23F, and ATCC 700669 and 15 in TIGR4 [40]. Several CBPs are associated with virulence, including four cell wall hydrolytic enzymes (LytA, LytB, LytC, and CbpE). *lytA* in LAU-23F appeared to be dissimilar from the references with 87.2% identity with SP49 with 120 SNPs. On the other hand, *lytC* was 99.1-100% identical in LAU-23F and all references, except in SP49 showing only 28% similarity. *lytB, lytC* and *cbpE* in TIGR4 were much more genetically divergent and thus was not included in the comparison.

PavA mediates adherence to the host endothelial cells by binding to fibronectin [38]. PavA was present in LAU-23F and reference sequences. One single substitution was observed between LAU-23F and SP49, while ten SNPs were detected in *pavA* between LAU-23F and ATCC 700669.

The high level of similarity observed between LAU-23F and SP49, though collected from different geographical locations, suggests evolutionary relatedness between the two strains and an intercontinental spread of this clone [14]. wgSNPs-based phylogenetic analysis of the 24 genomes revealed five major lineages. However, as previously reported, the distribution of the isolates did not significantly correlate with capsular serotype or site of isolation [32]. Clade III showed no correlation between the STs and serotypes. Interestingly, the *ddl* MLST locus was the one that varied within the isolates that clustered in this clade. This was previously observed in other *S. pneumoniae* serotypes [41]. The *ddl* locus closely aligns to *pbp2b*, which was acquired by horizontal gene transfer during the evolution of penicillin resistance, and accordingly not recommended to be used in MLST based phylogenetic analysis [32]. LAU-23F clustered in clade IV and as such was more closely related to serotype 19A with single or double locus ST variants than to serotype 23F ATCC 700669 isolate. It has been previously reported that large-scale recombination events have been observed in serotype 23F favoring a shift in capsular serotype from 23F to 19A under vaccine pressure [14], possibly explaining the observed phylogenetic relatedness.

Conclusion

The variations detected in the *cpsABCD* capsular regions and among the studied VFs suggest that a constant selection in pneumococci occurs during

carriage to escape the immune system which could have an impact over the disease outcome [33]. As the data on serotypes associated with IPD from the regions with the highest disease burden are limited, the importance of surveillance to monitor and control spread cannot be undermined. Thus, the routine use of WGS appears to be a low-cost, rapid, high resolution technique driving the future of clinical diagnostics and surveillance.

References

1. Gonçalves VM, Takagi M, Lima RB, Massaldi H, Giordano RC, Tanizaki MM (2003) Purification of capsular polysaccharide from *Streptococcus pneumoniae* serotype 23F by a procedure suitable for scale-up. Biotechnol Appl Biochem 37: 283-287.
2. Wen HY, Chou ML, Lin KL, Kao PF, Chen JF (2001) Recurrence of pneumococcal meningitis due to primary spontaneous cerebrospinal fluid fistulas. Chang Gung Med J 24: 724-728.
3. World Health Organization (2015) WHO-UNHCR-UNICEF Joint technical guidance: general principles of vaccination of refugees, asylum-seekers and migrants in the WHO European Region. Available: https://reliefweb.int/sites/reliefweb.int/files/resources/EuropeVaccinationPosition_WHO-UNHCR-UNICEFNov.pdf. Accessed 21 March 2018.
4. Mipatrini D, Stefanelli P, Severoni S, Rezza G (2017) Vaccinations in migrants and refugees: a challenge for European health systems. A systematic review of current scientific evidence. Pathog Glob Health 111: 59-68.
5. Republic of Lebanon, Ministry of Public Health (2017) Epidemiology Surveillance Unit, Meningitis Surveillance. Available: http://www.moph.gov.lb/userfiles/files/Esu_data/Esu_currentyear/Meningitis2017.htm. Accessed 21 March 2018.
6. Naba MR, Araj GF, Baban TA, Tabbarah ZA, Awar GN, Kanj SS (2010) Emergence of fluoroquinolone-resistant *Streptococcus pneumoniae* in Lebanon: a report of three cases. J Infect Public Health 3: 113-117.
7. Araj GF, Bey HA, Itani LY, Kanj SS (1999) Drug-resistant *Streptococcus pneumoniae* in the Lebanon: implications for presumptive therapy. Int J Antimicrob Agents 12: 349-354.
8. Uwaydah M, Mokhbat JE, Karam-Sarkis D, Baroud-Nassif R, Rohban (2006) Penicillin-resistant *Streptococcus pneumoniae* in Lebanon: the first nationwide study. Int J Antimicrob Agents 27: 242-246.
9. Henriques-Normark B, Tuomanen EI (2013) The pneumococcus: epidemiology, microbiology, and pathogenesis. Cold Spring Harb Perspect Med 3: a010215.
10. Dillard JP, Yother J (1994) Genetic and molecular characterization of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 3. Mol Microbiol 12: 959-972.
11. Llull D, Muñoz R, López R, García E (1999) A single gene (*tts*) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide. J Exp Med 190: 241-252.
12. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G (2006) Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet 2: e31.
13. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, Konradsen HB, Nahm MH (2015) Pneumococcal capsules and their types: past, present, and future. Clin Microbiol Rev 28: 871-899.
14. Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, Mitchell AM, Quail MA, Andrew PW, Parkhill J, Bentley SD (2009) Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumonia* Spain 23F ST81. J Bacteriol 191: 1480-1489.
15. European Committee on Antimicrobial Susceptibility Testing EUCAST (2015) Breakpoint tables for interpretation of MICs and zone diameters Version 5.0. Available: http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_8.0_Breakpoint_Tables.pdf. Accessed: 21 March 2018.
16. Jolley KA, Maiden MC (2010) BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11: 595.
17. Tritt A, Eisen JA, Facciotti MT, Darling AE (2012) An integrated pipeline for de novo assembly of microbial genomes. PloS One 7: e42304.
18. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32: 11-16.
19. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35: 3100-3108.
20. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Pontén TS, Ussery DW, Aarestrup FM, Lund O (2012) Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50: 1355-1361.
21. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV (2012) Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67: 2640-2644.
22. Liu B, Pop M (2009) ARDB—Antibiotic Resistance Genes Database. Nucleic Acids Res 37: 443-447.
23. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S (2017) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45: 566-573.
24. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44: 16-21.
25. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, Fry NK (2016) Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. Peer J 4: e2477.
26. Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL, Brinkman FS (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45: 30-35.
27. Pan XS, Fisher LM (1996) Cloning and characterization of the *parC* and *parE* genes of *Streptococcus pneumoniae* encoding DNA topoisomerase IV: role in fluoroquinolone resistance. J Bacteriol 178: 4060-4069.
28. Tankovic J, Perichon B, Duval J, Courvalin P (1996) Contribution of mutations in *gyrA* and *parC* genes to

fluoroquinolone resistance of mutants obtained in vivo and in vitro. Antimicrob Agents Chemother 40: 2505-2510.
29. Darling AE, Mau B, Perna NT (2010) Progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. PloS one 5: e11147.
30. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57: 81-91.
31. Grant JR, Stothard P (2014) The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36: 181-184.
32. Donner J, Bunk B, Schober I, Spröer C, Bergmann S, Jarek M, Overmann J, Wagner-Döbler I (2017) Complete genome sequences of three multidrug-resistant clinical isolates of *Streptococcus pneumoniae* serotype 19A with different susceptibilities to the Myxobacterial metabolite carolacton. Genome Announc 5: e01641-16.
33. Grebe T, Hakenbeck R (1996) Penicillin-binding proteins 2b and 2x of *Streptococcus pneumoniae* are primary resistance determinants for different classes of beta-lactam antibiotics. Antimicrob Agents Chemother 40: 829-834.
34. Jones RN, Sader HS, Mendes RE, Flamm RK (2013) Update on antimicrobial susceptibility trends among *Streptococcus pneumoniae* in the United States: report of ceftaroline activity from the SENTRY Antimicrobial Surveillance Program (1998–2011). Diagn Microbiol Infect Dis 75: 107-109.
35. Lim S, Bast D, McGeer A, de Azavedo J, Low DE (2003) Antimicrobial susceptibility breakpoints and first-step *parC* mutations in *Streptococcus pneumoniae*: redefining fluoroquinolone resistance. Emerg Infect Dis 9: 833.
36. Baylay AJ, Ivens A, Piddock LJ (2015) A novel gene amplification causes upregulation of the PatAB ABC transporter and fluoroquinolone resistance in *Streptococcus pneumoniae*. Antimicrob Agents Chemother 59: 3098-3108.
37. Dobrindt U, Hacker J (2001) Whole genome plasticity in pathogenic bacteria. Curr Opin Microbiol 4: 550-557.
38. Mitchell AM, Mitchell TJ (2010) *Streptococcus pneumoniae*: virulence factors and variation. Clin Microbiol Infect 16: 411-418.
39. Benton KA, Paton JC, Briles DE (1997) The hemolytic and complement-activating properties of pneumolysin do not contribute individually to virulence in a pneumococcal bacteremia model. Microb Pathog 23: 201-209.
40. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, Durkin AS (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 293: 498-506.
41. Sandgren A, Sjöström K, Liljequist BO, Christensson B, Samuelsson A, Kronvall G, Normark BH (2004) Effect of clonal and serotype-specific properties on the invasive capacity of *Streptococcus pneumoniae*. J Infect Dis 189: 785-796.

Corresponding author

Dr. Sima Tokajian,
Department of Natural Sciences, School of Arts and Sciences, Lebanese American University, Blat, Byblos, Lebanon, P.O. Box 36.
Phone: +961-9-547254
Fax: +961-9-944851
E-mail: stokajian@lau.edu.lb

Conflict of interests: No conflict of interests is declared.

Annex – Supplementary items

Table S1. List of genomic islands predicted in LAU-23F.

GI	Island start	Island end	Length	Gene start	Gene end	Product	Function
GI-I	88757	96774	8017	88757	89128	FIG01115961: hypothetical protein	
				89136	89348	hypothetical protein	
				89722	89847	hypothetical protein	
				90229	91287	FIG01114070: hypothetical protein	
				91469	92425	FIG01114918: hypothetical protein	
				92802	93176	FIG01115961: hypothetical protein	
				93384	94286	FIG01114534: hypothetical protein	
				94342	94674	FIG01114964: hypothetical protein	
				94671	95399	FIG01114515: hypothetical protein	
				96643	96774	FIG01119909: hypothetical protein	
GI-II	111614	118005	6391	111614	112168	Exopolysaccharide biosynthesis glycosyltransferase EpsF (EC 2.4.1.-)	
				112295	112756	Exopolysaccharide biosynthesis glycosyltransferase EpsF (EC 2.4.1.-)	
				112858	113655	Beta-1,3-glucosyltransferase	ABC Transporter and metabolism
				113742	115358	ABC transporter ATP-binding/membrane spanning protein - multidrug resistance	
				115823	115945	Asparagine synthetase [glutamine-hydrolyzing] (EC 6.3.5.4)	
				116070	116819	FIG01114613: hypothetical protein	
				116983	118005	UDP-glucose dehydrogenase (EC 1.1.1.22)	
GI-III	151113	162260	11147	151113	152627	ATP-dependent DNA helicase recG (EC 3.6.1.-)	
				152800	153513	Phage transcriptional repressor	
				153714	153851	conserved hypothetical protein - phage associated	
				153911	154537	prophage ps3 protein 13	
				154966	155361	FIG01115536: hypothetical protein	
				155358	155558	Glycerate kinase (EC 2.7.1.31)	
				155560	155775	FIG01116276: hypothetical protein	Phage related proteins
				155772	156113	FIG01114419: hypothetical protein	
				156106	156387	FIG01114073: hypothetical protein	
				156522	157388	Phage protein	
				157441	158910	DNA primase, phage associated	
				159273	159440	FIG01114328: hypothetical protein	
				159524	160054	FIG01115461: hypothetical protein	
				161841	162260	Phage protein	

GI-IV	151113	157388	6275	151113	152627	ATP-dependent DNA helicase recG (EC 3.6.1.-)	Phage related proteins
				152800	153513	Phage transcriptional repressor	
				153714	153851	conserved hypothetical protein - phage associated	
				153911	154537	prophage ps3 protein 13	
				154966	155361	FIG01115536: hypothetical protein	
				155358	155558	Glycerate kinase (EC 2.7.1.31)	
				155560	155775	FIG01116276: hypothetical protein	
				155772	156113	FIG01114419: hypothetical protein	
				156106	156387	FIG01114073: hypothetical protein	
				156522	157388	Phage protein	
GI-V	202864	207086	4222	202864	203247	Two-component sensor kinase, associated with ferric iron transporter, SPy1061 homolog	Iron Transport
				203580	204356	Pyruvate formate-lyase activating enzyme (EC 1.97.1.4)	
				204478	205224	Transcriptional regulator, DeoR family	
				205237	206217	transcriptional regulator, putative	
				206412	206732	PTS system, cellobiose-specific IIA component (EC 2.7.1.69)	
				206778	207086	PTS system, cellobiose-specific IIB component (EC 2.7.1.69)	
GI-VI	261769	267764	5995	261769	261951	Mobile element protein	Sugar metabolism
				262107	262445	Glutamate synthase [NADPH] large chain (EC 1.4.1.13)	
				262635	264071	6-phospho-beta-glucosidase (EC 3.2.1.86)	
				264123	264362	FIG01114066: hypothetical protein	
				264492	264806	PTS system, cellobiose-specific IIB component (EC 2.7.1.69)	
				264930	266903	Transcriptional antiterminator of lichenan operon, BglG family	
				266913	267227	PTS system, cellobiose-specific IIA component (EC 2.7.1.69)	
				267258	267764	FIG01114705: hypothetical protein	
GI-VII	308894	317721	8827	308894	309010	hypothetical protein	Polysaccharide biosynthesis
				309139	310332	Polysaccharide biosynthesis protein CpsH(V)	
				310336	311307	Polysaccharide biosynthesis glycosyl transferase CpsO	
				311314	312546	hypothetical protein	
				312548	313933	Polysaccharide biosynthesis protein CpsM(V)	
				313941	315089	CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase (EC 2.7.8.12)	
				315125	316153	Glycerol-1-phosphate dehydrogenase [NAD(P)] (EC 1.1.1.261)	
				316169	316873	nucleotidyl transferase family protein	

				316888	317721	Hypothetical NagD-like phosphatase	
GI-VIII	435571	441272	5701	435571	435708	hypothetical protein	
				437582	438304	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	
				438868	439665	Mobile element protein	
				439722	441272	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	
GI-IX	445432	477034	31602	445432	445560	hypothetical protein	
				445761	446198	FIG01115427: hypothetical protein	
				446362	447396	Heat-inducible transcription repressor HrcA	
				447423	447947	Heat shock protein GrpE	
				448427	450250	Chaperone protein DnaK	
				450252	450398	hypothetical protein	
				451009	452145	Chaperone protein DnaJ	
				452174	452305	Substrate-specific component BioY of biotin ECF transporter	
				452467	452754	Hypothetical protein in cluster with Ecs transporter (in Streptococci)	
				452764	453174	Histidine triad (HIT) nucleotide-binding protein, similarity with At5g48545 and yeast YDL125C (HNT1)	
				453242	453973	ABC transporter, ATP-binding protein EcsA	
				453970	455019	ABC transporter, permease protein EscB	
				455187	455516	BlpT protein, fusion	
				455792	456130	BlpS protein	
				456135	456872	Response regulator of the competence regulon ComE	ABC Transporter system
				456886	458226	Histidine kinase of the competence regulon ComD	
				458483	459844	Competence-stimulating peptide ABC transporter permease protein ComB	
				459855	461210	Competence-stimulating peptide ABC transporter ATP-binding protein ComA	
				461219	461902	Competence-stimulating peptide ABC transporter ATP-binding protein ComA	
				462183	462413	Bacteriocin BlpU	
				462648	462764	hypothetical protein	
				462832	462996	FIG01116395: hypothetical protein	
				462993	463397	Bacteriocin immunity protein BlpL	
				463595	463783	Mobile element protein	
				464011	464400	Mobile element protein	
				464550	464804	Bacteriocin-like peptide M BlpM	
				464820	465023	Bacteriocin-like peptide N BlpN	
				465267	465416	Bacteriocin-like peptide O BlpO	
				465398	465517	hypothetical protein	
				465520	465639	hypothetical protein	

				466425	466808	Bactriocin immunity protein BlpX
				466860	467549	Bactriocin immunity protein BlpY
				467591	467824	BlpZ protein, fusion
				467975	468586	ABC transporter ATP binding protein - unknown substrate
				468810	469541	aminoglycoside phosphotransferase family protein
				469538	470173	tRNA (guanine46-N7-)-methyltransferase (EC 2.1.1.33)
				470228	470779	FIG000325: clustered with transcription termination protein NusA
				470823	471959	Transcription termination protein NusA
				471981	472274	COG2740: Predicted nucleic-acid-binding protein implicated in transcription termination
				472267	472566	ribosomal protein L7Ae family protein
				472583	475375	Translation initiation factor 2
				475626	475976	Ribosome-binding factor A
				476255	477034	FIG01115101: hypothetical protein
GI-X	488877	493754	4877	488877	489089	Mobile element protein
				489170	489331	DNA/RNA HELICASE (DEAD/DEAH BOX FAMILY)
				489631	490470	Beta-glucoside bgl operon antiterminator, BglG family
				490488	492326	PTS system, beta-glucoside-specific IIB component (EC 2.7.1.69) / PTS system, beta-glucoside-specific
				492339	493754	6-phospho-beta-glucosidase (EC 3.2.1.86)
GI-XI	774093	778143	4050	774093	774635	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)
				774632	775750	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)
				775810	776046	Prevent host death protein, Phd antitoxin
				776043	776456	Death on curing protein, Doc toxin
				776574	777539	Tyrosine recombinase XerC
				777505	778143	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)
GI-XII	789424	797450	8026	789424	790929	Pyruvate kinase (EC 2.7.1.40)
				791974	792120	FIG01114840: hypothetical protein
				792664	793536	Uncharacterized secreted protein associated with spyDAC
				793643	794026	Mobile element protein
				794169	794480	FIG01114246: hypothetical protein
				794609	795199	FIG01114816: hypothetical protein
				795329	796144	ABC transporter, substrate-binding protein

				796141	797046	possible permease	
				797047	797178	hypothetical protein	
				797292	797450	Substrate-specific component BioY of biotin ECF transporter	
GI-XIII	866875	872412	5537	866875	867465	DNA polymerase III epsilon subunit (EC 2.7.7.7)	
				867474	869237	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)	
				869451	869780	Mobile element protein	
				869959	870132	Mobile element protein	
				870260	870460	FIG01114961: hypothetical protein	
				870502	870906	FIG01115324: hypothetical protein	
				871137	871844	Cytochrome c-type biogenesis protein CcdA (DsbD analog)	
				871855	872412	Thiol:disulfide oxidoreductase related to ResA	
GI-XIV	901784	917080	15296	901784	901930	FIG01114400: hypothetical protein	
				902323	903348	Iron compound ABC uptake transporter substrate-binding protein PiaA	
				903338	904345	Iron compound ABC uptake transporter permease protein PiaB	
				904345	905352	Iron compound ABC uptake transporter permease protein PiaC	
				905352	906146	Iron compound ABC uptake transporter ATP-binding protein PiaD	
				907018	907332	FIG01115356: hypothetical protein	
				907642	909030	mobilization/transfer protein	
				909715	910242	FIG01114922: hypothetical protein	
				911685	912365	FIG01116881: hypothetical protein	
				912984	913145	FIG01115587: hypothetical protein	Phage related proteins & Iron uptake
				913330	913740	Phage integrase (Site-specific recombinase)	
				913712	914917	Phage integrase (Site-specific recombinase)	
				915185	915751	FIG01114410: hypothetical protein	
				916072	916260	FIG01115371: hypothetical protein	
				916322	917080	FIG01114267: hypothetical protein	
				901784	901930	FIG01114400: hypothetical protein	
				902323	903348	Iron compound ABC uptake transporter substrate-binding protein PiaA	
				903338	904345	Iron compound ABC uptake transporter permease protein PiaB	
				904345	905352	Iron compound ABC uptake transporter permease protein PiaC	
				905352	906146	Iron compound ABC uptake transporter ATP-binding protein PiaD	

				907018	907332	FIG01115356: hypothetical protein	
				907642	909030	mobilization/transfer protein	
				909715	910242	FIG01114922: hypothetical protein	
				911685	912365	FIG01116881: hypothetical protein	
				912984	913145	FIG01115587: hypothetical protein	
				913330	913740	Phage integrase (Site-specific recombinase)	
				913712	914917	Phage integrase (Site-specific recombinase)	
				915185	915751	FIG01114410: hypothetical protein	
				916072	916260	FIG01115371: hypothetical protein	
				916322	917080	FIG01114267: hypothetical protein	
GI-XV	1142689	1147667	4978	1142689	1142859	FIG01114804: hypothetical protein	ABC Transporter System
				1142872	1143318	ABC transporter ATP-bindng protein - possibly multidrug efflux, truncation	
				1143876	1144016	Mobile element protein	
				1144313	1144534	FIG01114768: hypothetical protein	
				1145119	1145622	Mobile element protein	
				1145631	1146470	Mobile element protein	
				1146777	1146980	Methionine ABC transporter ATP-binding protein	
				1147056	1147667	V-type ATP synthase subunit D (EC 3.6.3.14)	
GI-XVI	1150914	1156622	5708	1150914	1151234	V-type ATP synthase subunit F (EC 3.6.3.14)	
				1151224	1152231	V-type ATP synthase subunit C (EC 3.6.3.14)	
				1152296	1152877	V-type ATP synthase subunit E (EC 3.6.3.14)	
				1152911	1153387	V-type ATP synthase subunit K (EC 3.6.3.14)	
				1153403	1155394	V-type ATP synthase subunit I (EC 3.6.3.14)	
				1155384	1155707	V-type ATP synthase subunit G (EC 3.6.3.14)	
				1155732	1156622	N-acetylmannosamine kinase (EC 2.7.1.60)	
GI-XVII	1272784	1281802	9018	1271810	1272787	oxidoreductase, Gfo/Idh/MocA family	
				1272784	1273866	ATP-dependent RNA helicase YfmL	
				1273930	1274109	Mobile element protein	
				1274174	1274416	Mobile element protein	
				1274448	1274954	Mobile element protein	
				1275338	1276534	Translation elongation factor Tu	
				1276877	1276999	hypothetical protein	
				1277445	1278314	Glycerol uptake facilitator protein	
				1278339	1278461	hypothetical protein	
				1278538	1279215	Cell wall surface anchor family protein	
				1279173	1280252	FIG01114480: hypothetical protein	

				1280363	1280593	FIG01115938: hypothetical protein	
				1280957	1281490	Mobile element protein	
				1281477	1281611	Mobile element protein	
				1281659	1281802	TnpB	
GI-XVIII	1557017	1572156	15139	1557017	1557547	Nucleotide sugar synthetase-like protein	
				1557576	1558451	Beta-1,3-glucosyltransferase	
				1558491	1559369	Maltose O-acetyltransferase (EC 2.3.1.79)	
				1559366	1561450	Beta-1,3-glucosyltransferase	
				1561570	1562916	DNA polymerase III subunits gamma and tau (EC 2.7.7.7)	
				1562950	1563948	DNA polymerase III subunits gamma and tau (EC 2.7.7.7)	
				1565008	1565874	Putative peptidoglycan bound protein (LPXTG motif) Lmo1799 homolog	
				1567457	1567918	Degenerative transposase	
				1568296	1568676	Transcriptional regulator, MarR family	
				1568775	1568996	FIG01114254: hypothetical protein	
				1569013	1569327	Thioredoxin	
				1569633	1570166	NADPH dependent preQ0 reductase (EC 1.7.1.13)	
				1570228	1570575	Substrate-specific component BioY of biotin ECF transporter	
				1570600	1571268	Aquaporin Z	
				1571379	1571933	hypothetical protein	
				1572001	1572156	putative membrane protein	
GI-XIX	1575695	1580313	4618	1575695	1576714	FIG01118513: hypothetical protein	
				1576765	1579008	putative serine protease	
				1579010	1580044	AAA superfamily ATPase	
				1580155	1580313	MutT/nudix family protein; 7,8-dihydro-8-oxoguanine-triphosphatase	
				1580294	1581382	FIG01114441: hypothetical protein	
GI-XX	1617591	1627329	9738	1617591	1617704	hypothetical protein	
				1617808	1618518	MgtC/SapB family protein	
				1618521	1620212	Ferric iron ABC transporter, permease protein	
				1620224	1621234	ABC transporter ATP-binding protein	
				1621338	1622405	ABC transporter, substrate-binding protein	ABC Transporter System
				1622430	1623089	FIG01113986: hypothetical protein	
				1623802	1624098	FIG01114460: hypothetical protein	
				1624082	1626208	cell wall surface anchor family protein	
				1627153	1627329	FIG01115726: hypothetical protein	
GI-XXI	1635529	1641329	5800	1635529	1635702	hypothetical protein	
				1636335	1636916	Xanthine phosphoribosyltransferase (EC 2.4.2.22)	
				1636916	1638178	Xanthine permease	

				1638265	1639131	Type II restriction enzyme MjaIII (EC 3.1.21.4)	
				1639118	1639888	DNA modification methyltransferase (EC 2.1.1.-)	
				1639914	1640675	Methyl-directed repair DNA adenine methylase (EC 2.1.1.72)	
				1640925	1641329	Phenylacetic acid degradation protein PaaD, thioesterase	
GI-XXII	1716047	1724949	8902	1716047	1716178	Mobile element protein	
				1716183	1716446	Mobile element protein	
				1716923	1717756	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)	
				1717775	1718836	ABC transporter, substrate-binding protein	
				1718853	1719857	ABC transporter ATP-binding protein	ABC Transporter System
				1719869	1721563	Ferric iron ABC transporter, permease protein	
				1721565	1722266	Mg(2+) transport ATPase protein C	
				1722313	1722912	HAD superfamily hydrolase	
				1722922	1724949	ABC-type sugar transport system, periplasmic component	
GI-XXIII	1851059	1862010	10951	1851059	1851520	unknown	
				1851712	1851825	hypothetical protein	
				1852053	1852811	FIG01116235: hypothetical protein	
				1852841	1853059	hypothetical protein	
				1853065	1853748	FtsK/SpoIIIE family	
				1853906	1854295	unknown	
				1854462	1855235	FIG01119803: hypothetical protein	
				1855333	1856115	Rep protein	
				1856129	1856332	FIG01116431: hypothetical protein	
				1856336	1857469	Integrase	
				1857510	1857947	Guanosine-3',5'-bis(Diphosphate) 3'-pyrophosphohydrolase (EC 3.1.7.2)	
				1858163	1858399	RelB/StbD replicon stabilization protein (antitoxin to RelE/StbE)	
				1858392	1859123	serine/threonine phosphatase, putative	
				1861576	1862010	Cytidine/deoxycytidylate deaminase family protein	
GI-XXIV	1923667	1929134	5467	1923667	1924599	Transketolase, C-terminal section (EC 2.2.1.1)	
				1924596	1925453	Transketolase, N-terminal section (EC 2.2.1.1)	
				1925457	1926803	FIG00732228: membrane protein	
				1926816	1927100	FIG00629163: hypothetical protein	
				1927104	1929134	Predicted galactitol operon regulator (Transcriptional antiterminator), BglG family / PTS system, ma	
		Max	31602				
		Min	4050				
		Average	9305.15				

Table S2. ST, accession number and related information of the samples used for comparative genome analysis.

#	Name	ST	*aroe*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	Clade	Notes	Country of isolation	Accession #
1	D141	353								I			CP016633.2
2	R6	595								II	serotype 2 avirulent strain		NC_003098
3	D39	595									serotype 2 laboratory strain		NC_008533
4	670-6B	90	5	6	1	2	3	3	4		serotype 6 B ,multidrug resistant especially to ciprofloxacin and penicillin G		NC_014498
5	TCH8431/19A	320	4	16	19	15	6	20	1		serotype 19A sensitive to penicillin G		NC_014251
6	SP61	2432	4	16	19	15	6	202	1	III	serotype 19A is highly susceptible to macrolide		CP018137.1
7	ST556	1392	15	16	19	15	6	20	156		serotype 19F multidrug resistant especially to ciprofloxacin and penicillin G		CP003357.2
8	Taiwan19F-14	236	15	16	19	15	6	20	26		serotype 19F multidrug resistant especially to ciprofloxacin and penicillin G	Taiwan	NC_012469
9	SPNA45	6934	10	9	4	12	287	426	470		serotype 3 sensitive to penicillin		HE983624.1
10	P1031	303	10	5	4	1	7	19	9		serotype 1 cause invasive pneumococcal disease		NC_012467
11	CGSP14	15	1	5	4	5	5	3	8		serotype 14 multidrug resistant especially to penicillin and trimethoprim-sulfamethoxazole	Taiwan	NC_010582
12	INV200	9	1	5	4	5	5	1	8		serotype 14 expresses resistance to a variety of antimicrobial agents, including penicillin, erythromycin, and ceftriaxone.		FQ312029.1
13	NCTC7465	615	10	31	4	1	6	4	94		serotype 1 cause invasive pneumococcal disease		LN831051.1
14	SP49	277	7	13	8	6	6	12	8		serotype 19A is highly susceptible to macrolide	Germany	CP018136.1
15	LAU-23F	277	7	13	8	6	6	12	8		serotype 23F	Lebanon	
16	Hungary19A-6	268	7	13	42	6	10	6	56	IV	serotype 19A; Penicillin resistant		NC_010380
17	Hu15	226	7	13	42	6	10	6	14		serotype 19A sequence type 226 clinical isolates from Hungary with high-level beta-lactam resistance and penicillin sensitive phenotype		CP020551.1
18	Hu17	226	7	13	42	6	10	6	14				CP020549.1
19	70585	289	16	12	9	1	41	33	33		serotype 5	Bangladesh	CP000918.1
20	TIGR4	205	10	5	4	5	13	10	18		serotype 4 synthesizes a functional lytC lysozyme	Norway	NC_003028
21	ATCC 700669	81	4	4	2	4	4	1	1		serotype 23F penicillin - resistant is the consequence of alterations of 3 majors PBP(pbp1a,pbp2b,pbp2x)		NC_011900
22	JJA	66	2	8	2	4	6	1	1		serotype 14 expresses resistance to a variety of antimicrobial agents, including penicillin, erythromycin, and ceftriaxone	Brazil	CP000919.1
23	G54	63								V	serotype 19 F multidrug resistant especially to		NC_011072

24	AP200	62	ciprofloxacin and penicillin G serotype 11A sensitive to penicillin G	NC_014494
25	*S. pyogenes* M1 GAS			AE004092.2

Table S3. List of the unique genes in LAU-23F and SP49 reference strain.

#	Isolate	Subsystem	Role
1	SP49	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	Isochorismatase (EC 3.3.2.1)
2	SP49	CMP-N-acetylneuraminate Biosynthesis	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14)
3	SP49	Teichoic and lipoteichoic acids biosynthesis	N-acetylmannosaminyltransferase (EC 2.4.1.187)
4	SP49	CBSS-393131.3.peg.612	Putative metallopeptidase (Zinc) SprT family
5	SP49	DNA-replication	DNA replication protein DnaC
6	SP49	Phage capsid proteins	Phage capsid and scaffold
7	SP49	Phage capsid proteins	Phage capsid protein
8	SP49	Phage capsid proteins	Phage head maturation protease
9	SP49	Phage capsid proteins	Phage major capsid protein
10	SP49	Phage capsid proteins	Phage minor capsid protein
11	SP49	Phage lysis modules	Phage endolysin
12	SP49	Phage lysis modules	Phage holin
13	SP49	Phage packaging machinery	Phage DNA binding protein
14	SP49	Phage packaging machinery	Phage portal protein
15	SP49	Phage packaging machinery	Phage terminase, large subunit
16	SP49	Phage packaging machinery	Phage DNA binding protein
17	SP49	Phage packaging machinery	Phage portal protein
18	SP49	Phage packaging machinery	Phage terminase, large subunit
19	SP49	Phage packaging machinery	Phage terminase, small subunit
20	SP49	Phage replication	DNA helicase, phage-associated
21	SP49	Phage replication	DNA primase/helicase, phage-associated
22	SP49	Phage replication	Phage replication protein
23	SP49	Phage replication	Single stranded DNA-binding protein, phage-associated
24	SP49	Phage tail proteins	Phage major tail protein
25	SP49	Phage tail proteins	Phage major tail shaft protein
26	SP49	Phage tail proteins	Phage tail assembly
27	SP49	Phage tail proteins	Phage tail length tape-measure protein
28	SP49	Phosphate metabolism	Alkaline phosphatase (EC 3.1.3.1)
29	SP49	Ribosome LSU bacterial	LSU ribosomal protein L36p
30	SP49	Lanthionine Synthetases	Lanthionine biosynthesis protein LanL
31	LAU-23F	Fructooligosaccharides(FOS) and raffinose utilization	Sucrose-6-phosphate hydrolase (EC 3.2.1.26)
32	LAU-23F	Glycogen metabolism	Predicted glycogen debranching enzyme (pullulanase-like, but lacking signal peptide)
33	LAU-23F	Rhamnose containing glycans	Alpha-D-GlcNAc alpha-1,2-L-rhamnosyltransferase (EC 2.4.1.-)
34	LAU-23F	Teichoic and lipoteichoic acids biosynthesis	CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase (EC 2.7.8.12)
35	LAU-23F	Glycerolipid and Glycerophospholipid Metabolism in Bacteria	Glycerol-1-phosphate dehydrogenase [NAD(P)] (EC 1.1.1.261)

Figure S1. Distribution of genomic islands on the LAU-23F chromosome.

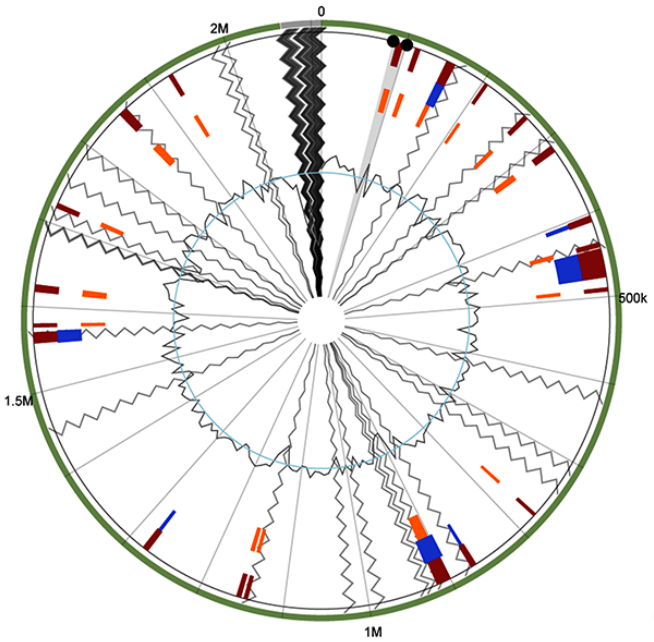


Figure S2. Genome alignment of LAU-23F with six reference genomes (SP49, Hungary19A-6,670-6B, P1031, R6 and D39) using the Mauve tool. Regions located outside blocks did not match to homologous regions. The average level of conservation is illustrated by the height of the similarity profile. White areas are unique regions in a genome.