Infection and Drug Resistance

Dovepress
open access to scientific and medical research

ORIGINAL RESEARCH

Comparative genomic analysis of multidrug-resistant *Streptococcus pneumoniae* isolates

This article was published in the following Dove Press journal:
Infection and Drug Resistance

Fen Pan[1]
Hong Zhang[1]
Xiaoyan Dong[2]
Weixing Ye[3]
Ping He[4]
Shulin Zhang[4]
Jeff Xianchao Zhu[5]
Nanbert Zhong[1,2,6]

[1]Department of Clinical Laboratory, Shanghai Children's Hospital, Shanghai Jiaotong University, Shanghai, China; [2]Department of Respiratory, Shanghai Children's Hospital, Shanghai Jiaotong University, Shanghai, China; [3]Shanghai Personal Biotechnology Co., Ltd, Shanghai, China; [4]Department of Medical Microbiology and Immunology, Shanghai Jiao Tong University School of Medicine, Shanghai, China; [5]Zhejiang Bioruida Biotechnology co. Ltd, Zhejiang, China; [6]New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY, USA

Correspondence: Hong Zhang
Department of Clinical Laboratory,
Shanghai Children's Hospital, Shanghai
Jiaotong University, No. 24, Lane 1400,
West Beijing Road, Shanghai 200040,
China
Tel/Fax +86 21 6247 4880 ext 82301
Email zhanghong3010@126.com

Nanbert Zhong
New York State Institute for Basic
Research in Developmental Disabilities,
1050 Forest Hill Road, Staten Island, NY
10314, USA
Tel/Fax +1 718 494 5242 ext 4882
Email Nanbert.zhong@opwdd.ny.gov

Introduction: Multidrug resistance in *Streptococcus pneumoniae* has emerged as a serious problem to public health. A further understanding of the genetic diversity in antibiotic-resistant *S. pneumoniae* isolates is needed.

Methods: We conducted whole-genome resequencing for 25 pneumococcal strains isolated from children with different antimicrobial resistance profiles. Comparative analysis focus on detection of single-nucleotide polymorphisms (SNPs) and insertions and deletions (indels) was conducted. Moreover, phylogenetic analysis was applied to investigate the genetic relationship among these strains.

Results: The genome size of the isolates was ~2.1 Mbp, covering >90% of the total estimated size of the reference genome. The overall G+C% content was ~39.5%, and there were 2,200–2,400 open reading frames. All isolates with different drug resistance profiles harbored many indels (range 131–171) and SNPs (range 16,103–28,128). Genetic diversity analysis showed that the variation of different genes were associated with specific antibiotic resistance. Known antibiotic resistance genes (*pbps*, *murMN*, *ciaH*, *rplD*, *sulA*, and *dpr*) were identified, and new genes (*regR*, *argH*, *trkH*, and *PTS-EII*) closely related with antibiotic resistance were found, although these genes were primarily annotated with functions in virulence as well as carbohydrate and amino acid transport and metabolism. Phylogenetic analysis unambiguously indicated that isolates with different antibiotic resistance profiles harbored similar genetic backgrounds. One isolate, 14-LC.ER1025, showed a much weaker phylogenetic relationship with the other isolates, possibly caused by genomic variation.

Conclusion: In this study, although pneumococcal isolates had similar genetic backgrounds, strains were diverse at the genomic level. These strains exhibited distinct variations in their indel and SNP compositions associated with drug resistance.

Keywords: *Streptococcus pneumoniae*, antimicrobial resistance, whole-genome sequencing, insertions/deletions, SNPs, phylogenetic analysis

Introduction

Streptococcus pneumoniae causes a series of severe infective disorders ranging from noninvasive infections (otitis media and pneumonia) to invasive infections (meningitis and sepsis), which could become life-threatening. Worldwide, *S. pneumoniae* is responsible for ~582,000–926,000 deaths annually in children aged 1–59 months.[1] Antimicrobial agents are recommended as the first choice for treating these infections; however, the appearance of drug-resistant *S. pneumoniae*, especially multidrug-resistant (MDR) strains, decreases the effectiveness of antibiotics, and a number of such strains have been reported clinically. A study reported by the Asian

http://dx.doi.org/10.2147/IDR.S147858

Network for Surveillance of Resistant Pathogens (ANSORP) illustrated that the overall rate of MDR in pneumococcal isolates was 59.3% (59.4% and 57.5% in nonmeningeal and meningeal isolates, respectively), with the highest MDR rate being 83.3% in China.[2] Several factors contribute to this phenomenon. Some studies have shown that drug-resistant isolates can be created via the transmission of already resistant strains (primary resistance) or by selection of resistance-conferring mutations via inadequate therapy (secondary resistance).[3] Especially, horizontal transmission through the transfer of genetic material between bacteria has the greatest effect on drug resistance transmission. As a transformable strain, the mechanisms of MDR *S. pneumoniae* involve different genes associated with different antibiotics. Pneumococcal resistance to β-lactams is mainly caused by mutations in penicillin-binding proteins (PBPs; *pbp1a*, *pbp2x*, and *pbp2b*) which have a decreased affinity for the antibiotic as well as other non-PBP mechanisms, including mutations in *ciaH* and *cpoA* as well as encoding histidine protein kinase and putative glycosyltransferase, respectively.[4,5] Simultaneously, pneumococcal strains with resistance to macrolides are detected to be associated with two major mechanisms: target modification by ribosomal methylation encoded by the *ermB* gene and drug efflux encoded by the *mefA* gene.[4] However, owing to the lack of comprehensive understanding, complicated resistance mechanisms of *S. pneumoniae* require investigation.

Over the past few decades, the methods in place to investigate the genetic characteristics of drug-resistant *S. pneumoniae* are restricted to PCR-related molecular approaches such as nest PCR, real time-PCR, and multilocus sequence typing (MLST).[6] Although these methods have significantly enhanced our knowledge of *S. pneumoniae*, their discriminatory power for identifying potential drug resistance mechanisms is limited. However, in recent years, new genome-scale technologies have started to make an important contribution to our comprehension of the biology and molecular epidemiology of *S. pneumoniae* as an important bacterial pathogen.[7]

High-throughput whole-genome sequencing (WGS) – a powerful tool based on next-generation sequencing – offers new opportunities for studying pathogenic bacteria and is now used widely to investigate pathogenesis and drug resistance mechanisms in a wide range of pathogens.[8–10] Supplementary to WGS, is whole-genome resequencing (WGRS) – a technique whereby a genome is sequenced and compared with the genome of a previously sequenced reference strain to identify sequence polymorphisms between the two sequences, such as single-nucleotide polymorphisms (SNPs), gene acquisition, recombination, loss as well as insertions and deletions (indels), that are potentially important in epidemiological analyses, drug resistance detection, and phylogenetic investigations.[11] Although there are several pneumococcal genomes in the public domain, WGRS is rarely used for investigation of drug resistance of *S. pneumonia* isolated in China. In this study, we conducted WGRS to analyze 25 *S. pneumoniae* pneumococcal isolates and determine the genetic diversity including SNPs and indels among them.

Methods

Microbiology methods

Twenty-five pneumococcal strains were part of the routine hospital laboratory procedure and were selected according to their similar resistance patterns. They were isolated from children <14 years of age who were diagnosed with pneumonia and had received antibiotic therapy before the samples collected at the Shanghai Children's Hospital, Shanghai, People's Republic of China. Written informed consent was obtained from the patients' guardians on behalf of the children enrolled in this study. All strains were identified on the basis of typical characteristics (ie, colony morphology, alpha hemolysis, and gram-positive diplococcus), after which the optochin sensitivity (Oxoid, Hampshire, UK) and bile solubility tests were used for final confirmation.[12] The strains were stored at −80°C in 40% sterilized glycerin–bouillon medium until required.

The minimal inhibitory concentrations of penicillin (PEN), cefuroxime (CXM), ceftriaxone (CRO), erythromycin, azithromycin, levofloxacin (LEV), moxifloxacin (MXF), vancomycin (VAN), and sulfamethoxazole/trimethoprim (SXT) were determined by the broth microdilution method.[13] The breakpoints used for interpretation were those recommended by the Clinical and Laboratory Standards Institute (CLSI) in 2016.[13] In addition, all isolates were investigated by MLST analysis as described previously.[14]

DNA extraction and WGRS

Chromosomal DNA was isolated from overnight cultures of the isolates previously grown on agar plates supplemented with 5% sheep blood using a TIANamp bacteria DNA kit (Tiangen, Beijing, People's Republic of China) according to the manufacturer's instructions. The extracted DNA was dissolved in TE buffer (10 mM Tris-HCl and 1 mM EDTA, pH 8.0), and its concentration and purity were measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). The total DNA quantity in each sample was ≥10 µg.

Next, the genomic DNA was fragmented in a Covaris instrument (Woburn, MA, USA) to an average size of 250–300 nucleotides. Library preparation was done using TruSeq Nano DNA Sample Preparation Kit (Illumina, San Diego, CA, USA), and the libraries were then sequenced using an Illumina HiSeq 2000 instrument (125-bp paired-end reads), and the individual samples were barcoded using Illumina's Multiplex Sample Preparation Oligonucleotide Kit. Sequence-read adapters were removed using AdapterRemoval (version 2.1.7).[15] Low-quality bases were trimmed using a 5-bp sliding window, cutting once the average PHRED quality scores to below 20. After trimming, if either read from a pair of reads were <50 bp in length, then both reads were discarded. The sequence data files have been deposited in the National Center for Biotechnology Information's Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/).

Variant analysis

High-quality sequence reads were mapped to the *S. pneumoniae* R6 strain reference genome (GenBank accession no. ASM704v1) using BWA (version 0.7.12-r1039).[16] The alignments were improved using the Picard package (http://sourceforge.net/projects/picard/) with the following two commands: the "FixMateInformation" command was used to ensure that all mate-pair information was in sync between each read and its mate pair, and the "MarkDuplicates" command was used to mark potential PCR duplicates. Where multiple read pairs had identical external coordinates, only the pair with the best mapping quality was retained; the others were marked as PCR duplicates. We then undertook a local realignment of the mapped reads around indels using the GATK package in two steps: the "RealignerTargetCreator" command was used to determine suspicious intervals that probably need realigning, and the "IndelRealigner" command was used for realignment of such intervals. After alignment, we carried out variant calling using the Bayesian approach as implemented in the GATK package (https://software.broadinstitute.org/gatk/). The variants were further filtered according to the following criteria: RMS mapping quality of ≥25, site quality score of ≥30, variant confidence/quality by depth of ≥2, ≥16 reads covering each site with eight reads mapping to each strand, and the reads covering a major variant were at least five times greater than that of the minor variant. Sites that failed these criteria in any strain were removed from the analysis.

Genome assembly and comparative analysis

The high-quality reads were de novo assembled with Newbler 2.9 (version 20130529_1641) (http://www.454.com/products/analysis-software/). Gaps inside the scaffold were closed with GapCloser (version 1.12) (http://soap.genomics.org.cn/soapdenovo.html). All newly sequenced genomes were ordered on the basis of the reference genome with Mauve (version 2.3.1).[17] For each genome, regions that were absent in the reference genome were extracted with custom Perl script, and were then searched against the SwissProt database using blastx (version 2.2.26).[18] Possible genes associated with drug resistance were identified by the following criteria: identity ≥45%, e-value <1e-6, and coverage ≥70%.

Phylogenetic analysis

A SNP supermatrix for each strain with reference alleles was constructed using custom Perl scripts. A neighbor-joining phylogeny was generated using the Kimura 2-parameter model of nucleotide substitution in MEGA 6.02.[19] For each method, 1,000 bootstrap replicates provided support for nodes on the tree. The phylogenies derived from each method were congruent. We conducted principal component analysis on the SNPs using EIGENSOFT 5.02, and the eigenvectors were obtained from the covariance matrix using the R function.

Results

Epidemiological data and bioinformatics analysis of the pneumococcal isolates

The detailed epidemiological data from the selected *S. pneumoniae* clinical isolates are summarized in Table 1. Antimicrobial susceptibility tests revealed that 21 strains were resistant to one or more antibiotics, whereas four isolates were susceptible to all the antibiotics we tested. None were resistant to LEV, MXF, and VAN. Most pneumococci (72%, 18/25) were defined as MDR *S. pneumoniae*. Additionally, MLST analysis identified 19 sequence types (STs), and different serotypes had different STs.

The WGS statistics for the 25 isolates are shown in Table S1. The sequencing read assembly revealed that the pneumococcal isolate sequences were ~2.1 Mbp in size, covering >90% of the total estimated size of the genome. The overall G+C% content was ~39.5% and the genomes contained 2,200–2,400 open reading frames. The scaffold range among the isolates was 46–86. The largest scaffold of each isolates

Table 1 Epidemiological data from *S. pneumoniae* strains isolated from children

Sample ID	Age	Gender	Year of isolation	MIC value of isolates (μg/mL)						Drug resistance profiles*	Serotype	Sequence types	Sequence depth
				PEN	CXM	CRO	ERY	AZM	SXT				
14LC.ER1023	10m	Male	2012	8	8	4	≥16	≥16	8/152	PEN-CXM-CRO-MAC-SXT	19F	271	127.89
14LC.ER1024	3m	Female	2012	8	8	8	≥16	≥16	8/152	PEN-CXM-CRO-MAC-SXT	19F	1,464	119.24
14LC.ER1025	2y	Male	2012	8	8	8	≥16	≥16	8/152	PEN-CXM-CRO-MAC-SXT	6A	2,971	103.09
14LC.ER1026	2y	Male	2012	2	8	4	≥16	≥16	8/152	CXM-CRO-MAC-SXT	19F	271	117.43
14LC.ER1027	3y	Female	2012	2	8	4	≥16	≥16	8/152	CXM-CRO-MAC-SXT	6B	744	118.7
14LC.ER1028	6y	Male	2013	2	16	4	≥16	≥16	8/152	CXM-CRO-MAC-SXT	19F	271	125.03
14LC.ER1029	1y	Male	2012	2	8	4	≥16	≥16	0.25/4.75	CXM-CRO-MAC	14	876	128.97
14LC.ER1030	11m	Male	2012	2	8	4	≥16	≥16	0.25/4.75	CXM-CRO-MAC	19A	3,111	114.82
14LC.ER1031	5m	Female	2012	0.25	16	4	≥16	≥16	0.125/2.38	CXM-CRO-MAC	14	876	130.6
14LC.ER1032	10m	Male	2012	1	4	1	≥16	≥16	8/152	CXM-MAC-SXT	19F	271	119.63
14LC.ER1033	5m	Male	2012	1	4	1	≥16	≥16	8/152	CXM-MAC-SXT	6A	9,789	116.8
14LC.ER1034	6m	Male	2013	1	4	1	≥16	≥16	8/152	CXM-MAC-SXT	19A	320	113.15
14LC.ER1035	1y	Female	2012	1	8	0.5	≥16	≥16	0.25/4.75	CXM-MAC	19F	236	115.43
14LC.ER1036	3y	Male	2012	1	8	0.5	≥16	≥16	0.25/4.75	CXM-MAC	14	876	123.41
14LC.ER1037	2y	Male	2013	1	4	0.5	≥16	≥16	0.25/4.75	CXM-MAC	14	876	111.11
14LC.ER1038	4y	Male	2012	0.5	1	0.125	≥16	≥16	8/152	MAC-SXT	19F	2,754	78.74
14LC.ER1039	6y	Male	2013	1	1	1	≥16	≥16	4/76	MAC-SXT	19F	8,250	120.77
14LC.ER1040	2y	Male	2012	1	1	1	≥16	≥16	4/76	MAC-SXT	23F	6,325	119.97
14LC.ER1041	7m	Male	2012	0.125	0.125	0.125	≥16	≥16	0.25/4.75	MAC	19F	7,751	130.99
14LC.ER1042	4y	Male	2012	0.125	0.125	0.125	≥16	≥16	0.25/4.75	MAC	6A	855	116.59
14LC.ER1043	1y	Male	2013	0.125	0.125	0.125	≥16	≥16	0.25/4.75	MAC	7F	3,545	127.01
14LC.ER1044	4y	Male	2012	0.125	0.125	0.125	0.125	0.125	0.25/4.75	None	19A	416	121.17
14LC.ER1045	1m	Male	2012	0.5	0.5	0.5	0.125	0.125	0.25/4.75	None	19F	4,662	109.32
14LC.ER1046	4y	Female	2012	0.5	0.5	0.5	0.125	0.125	0.25/4.75	None	6B	3,173	106.37
14LC.ER1047	19m	Female	2013	0.125	0.125	0.125	0.125	0.125	0.125/2.38	None	6A	180	136.59

Abbreviations: *PEN, penicillin; CXM, cefuroxime; CRO, ceftriaxone; MAC, macrolides (erythromycin and azithromycin) and clindamycin; SXT, sulfamethoxazole–trimethoprim AZM, azithromycin; ERY, erythromycin; MIC, minimal inhibitory concentrations; m months; y, years.

was 91,339–310,534 bp, and the N50 scaffold length was 41,305–82,460 bp. Furthermore, we observed that similar numbers of indels (range 131–171) were present in isolates exhibiting different levels and profiles of drug resistance. The number of SNPs representing nonsynonymous/synonymous mutations in coding regions and nonsynonymous/synonymous mutations in noncoding regions ranged from 16,103 to 28,128 among these isolates. The WGRS for 25 clinical isolates (14LC.ER1023–14LC.ER1047) have been deposited in GenBank under accession numbers: MCGY00000000, NAOM00000000, NAOL00000000, NAOK00000000, NAOJ00000000, NAOI00000000, NAOH00000000, NAOG00000000, NAOF00000000, NAOE00000000, NAOD00000000, NAOC00000000, NAOB00000000, NAOA00000000, NANZ00000000, NANY00000000, NANX00000000, NANW00000000, NANV00000000, NANU00000000, NANT00000000, NANS00000000, NANR00000000, NANQ00000000, and NANP00000000, respectively.

Correlation analysis of pneumococcal isolates at the genome level

Sequence alignments of the pneumococcal isolates against *S. pneumoniae* R6 revealed that the mapping rates were 70.71–87.93% and the read depth range was 703–1,930. To further assess the relationships of the isolates, we conducted a phylogenetic analysis based on the SNPs identified in their whole-genome alignments. The phylogenetic tree revealed unambiguously that isolates with different antibiotic resistance profiles harbored similar genetic backgrounds (Figures 1 and 2); furthermore, the STs of the isolates (Table 1) were consistent with this finding. However, one isolate (14-LC.ER1025) showed a dissimilar genetic relationship to the other isolates, possibly as a result of genomic variation.

Compared with the *S. pneumoniae* R6 reference strain, the strain-specific regions in the isolates varied from 146,809 to 303,600 bp in size, which is proportionately 7.3–13.8% of the reference strain. Next, we conducted a functional analysis of these specific regions using the SwissProt database and

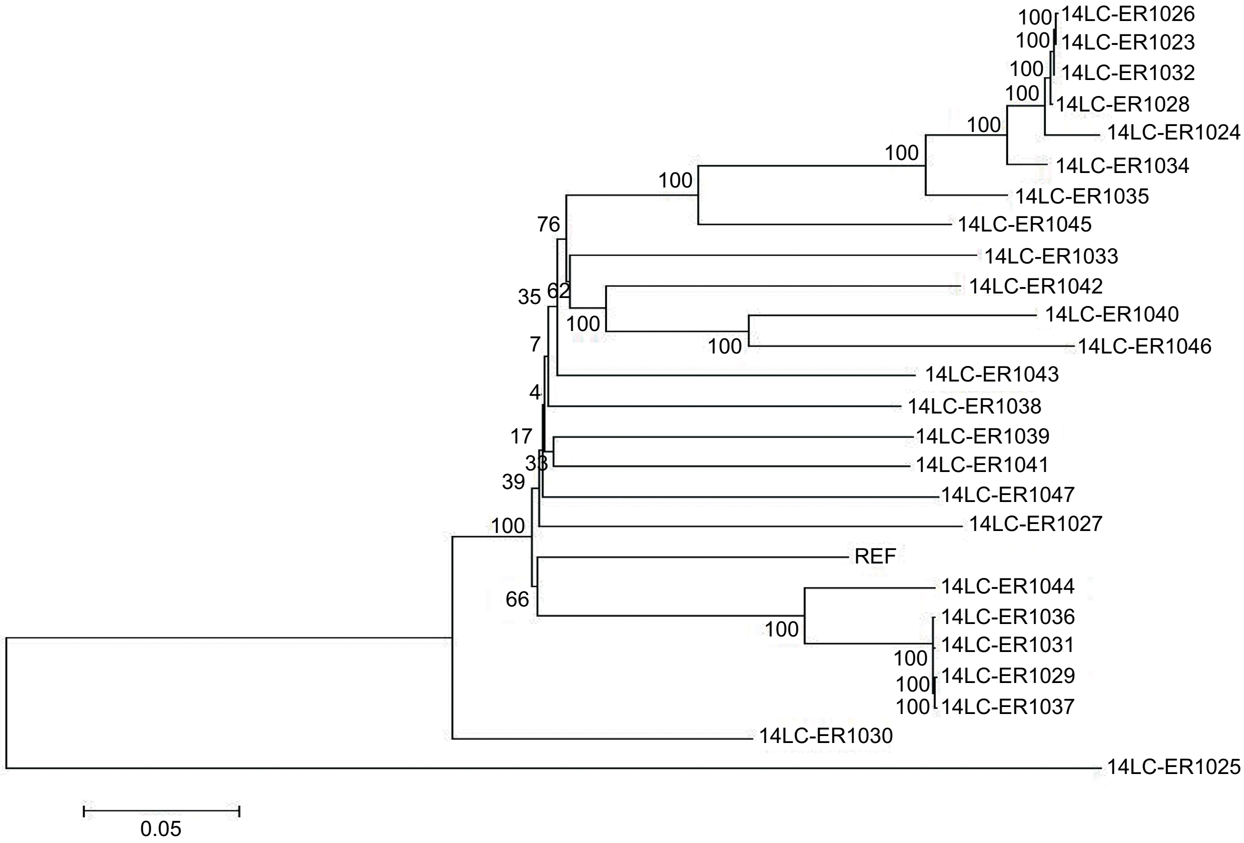

Figure 1 Phylogenetic relationships of *S. pneumoniae* isolates based on single-nucleotide polymorphisms from whole DNA sequences.

Notes: A tree representing the isolates. The branch lengths show the evolutionary distances among the isolates. The bootstrap values of the nodes represent the reliability of a branch being formed by all isolates in this branch, and values >70% are considered to be reliable. Resistant profiles: 14LC.ER1023-25, PEN-CXM-CRO-MAC–SXT; 14LC.ER1026-28, CXM-CRO-MAC-SXT; 14LC.ER1029-31, CXM-CRO-MAC; 14LC.ER1032-34, CXM-MAC-SXT; 14LC.ER1035-37, CXM-MAC; 14LC.ER1038-40, MAC-SXT; 14LC.ER1041-43, MAC; and 14LC.ER1044-47, NONE.

Abbreviations: PEN, penicillin; CXM, cefuroxime; CRO, ceftriaxone; MAC, macrolides (erythromycin and azithromycin) and clindamycin; SXT, sulfamethoxazole–trimethoprim.

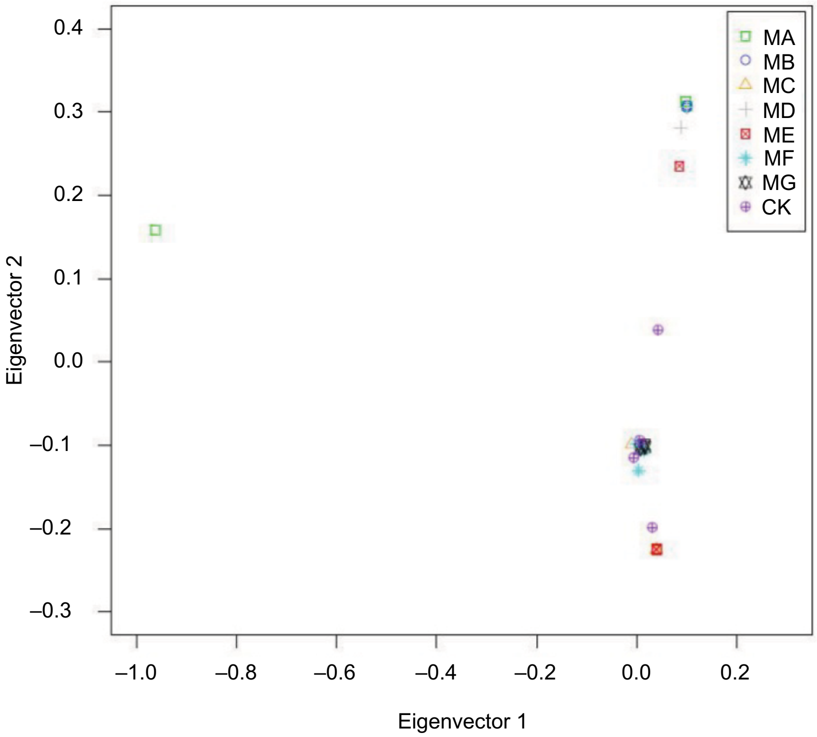

Figure 2 Principle component analysis of the whole genomes of 25 pneumococcal isolates.

Notes: MA, PEN-CXM-CRO-MAC–SXT; MB, CXM-CRO-MAC-SXT; MC, CXM-CRO-MAC; MD, CXM-MAC-SXT; ME, CXM-MAC; MF, MAC-SXT; MG, MAC; and CK, NONE.

Abbreviations: PEN, penicillin; CXM, cefuroxime; CRO, ceftriaxone; MAC, macrolides (erythromycin and azithromycin) and clindamycin; SXT, sulfamethoxazole–trimethoprim.

found that they were located in 76 genes involved in functions related to virulence, protein metabolism, and antimicrobial resistance (Table S2). This suggests that the sequences share similarities and that horizontal genetic transfer may have occurred during their evolution.

Genetic variations potentially associated with drug resistance

Analysis of genetic variation across the whole genomes of the isolates revealed that indels and SNPs occurred in both antimicrobial-susceptible and antimicrobial-resistant strains. To find variations potentially associated with drug resistance, Indels and SNPs were compared between antimicrobial-susceptible and antimicrobial-resistant strains.

Excluding the variations in indels from the susceptible isolates, we analyzed the relationships between indel variations and antibiotic resistance (Table S3). The antibiotic-related genes with high mutation rates (≥0.38) are depicted in Table 2. The numbers of indel changes associated with PEN, CXM, and CRO β-lactams were 77, 237, and 139, respectively. Furthermore, 27 mutations in 21 genes (eg, *dpr*, *dagA*, *zmpB*, *dfr*, *axe1*, *dctA*, *pspC*, *sulA*, and other genes encoding hypothetical protein) showed high mutation rates and shared close relationships with CXM resistance. At the same time, 465 mutations were identified that may be related to macrolide and lincosamide resistance, and 16 mutations in five genes (*lacE*, *ugd*, *pspC*, *pepB*, and a gene encoding a hypothetical protein) were associated with this resistance. There were 253 mutations in the SXT-resistant isolates, and 14 associated with six genes (*dpr*, *pspC*, *dfr*, *dctA*, *sulA*, and a gene encoding a hypothetical protein) had high mutation rates.

Further analysis of the SNPs associated with drug resistance is shown in Table S4, and the genes closely associated with antibiotic resistance are depicted in Table 3. The number of SNPs associated with β-lactam resistance in PEN (3480), CXM (7707), and CRO (5214) involved 1,524, 1,494, and 1,189 genes, respectively. Specifically, 35 mutations in seven genes (*sulA*, *gdhA*, *hom*, *xpt*, *bglA*, *regR*, and a gene encoding a hypothetical protein) showed high mutation rates, and these were closely related to CXM resistance. Moreover, 14,380 mutations were identified that may be related to macrolide and lincosamide resistance, and *trkH* and *PTS-EII* genes were associated with this resistance mechanism. The number of mutations related to SXT resistance was 8,012, involving 1,524 genes, and six of them (*ugd*, *dfr*, *sulB*, *dpr*,

Table 2 Indel changes closely associated with antibiotic resistance in *S. pneumoniae*

Gene symbol	Gene description	Mutations	Potential antibiotics
spr0752	Degenerate transposase (orf1)	748200delAGAAAGA	CXM/MAC
spr0583	Hypothetical protein	599264delGTC/599275delTT/insAACTG	CXM/MAC
spr1091	Hypothetical protein	1086256delTGGC	CXM
spr1403	Hypothetical protein	1380664del TCTTTACCG/1384939del TCCCCTTTTTCACCTTTA	CXM
dpr	DNA-binding protein starved cells-like peroxide resistance protein	1413639delAC/1413632insAG	CXM/SXT
spr1640	Hypothetical protein	1613222delT	CXM
spr1683	UDP-galactose 4-epimerase, truncation	1655609delG	CXM
spr1809	Hypothetical protein	1779791delC	CXM
spr0325	Hypothetical protein	324414insA	CXM
dagA	D-alanine glycine permease	369933del TGCAGC	CXM
zmpB	Zinc metalloprotease	592772ins AAAGGAACAGCAGTC	CXM
spr1079	Degenerate transposase (orf1)	1075949delA	CXM/SXT
spr0109	Hypothetical protein	118002del TGCACTAAACGT	CXM/SXT
spr1403	Hypothetical protein	1385226delACCTTGGGCTCCTTTTTC	CXM
dfr	Dihydrofolate reductase	1412921delGC/1412928insGG	CXM/SXT
spr0138	Hypothetical protein	147352del AACCTTGGGA	CXM/SXT
axe1	Xylan esterase 1	1519850insT	CXM
dctA	Dicarboxylate/amino acid: cation (Na+ or H+) symporter	1575810delA	CXM/SXT
spr0155	Hypothetical protein	163854ins TATAAATAT	CXM/SXT
pspC	Choline-binding protein A	1988359insTGGAGC	CXM
sulA	Dihydropteroate synthase	268192insCGGCCG	CXM/SXT
spr0397	Conserved hypothetical protein	394265insT	MAC
spr1448	Degenerate transposase (orf1)	1433311delA	MAC
spr1873	Conserved hypothetical protein	1843758ins AATAATTGT	MAC/SXT
spr0096	Hypothetical protein	104220del CTCCAGTAGCAGAAA	MAC
lacE	PTS system, lactose-specific IIBC component	107325delTGC/1070335insAAC/1070342insTT	MAC
spr1403	Hypothetical protein	1380459 AACCGCTGGTACTGGTTT	MAC
ugd	UDP-glucose dehydrogenase	148566delT	MAC
pspC	Choline-binding protein A	1988947delTTCTGCTTTTTT	MAC
spr0658	Hypothetical protein	662369insA	MAC
pepB	Group B oligopeptidase	872648insAAAGAAATG	MAC
pspC	Choline-binding protein A	1988905insGTTTTCAAC/1989311delTCTTTC	SXT
spr0312	Degenerate transposase	312730insA	SXT

Abbreviations: CXM, cefuroxime; MAC, macrolides (erythromycin and azithromycin) and clindamycin; SXT, sulfamethoxazole–trimethoprim; del, deletion; ins, insertion; UDP, uridine diphosphate; PTS, phosphotransferase; NADP, nucleoside diphosphate.

dctA, and *pspC*) were closely associated with SXT resistance. An overview of the genetic diversity among the isolates is shown in Figure 3.

Antibiotic resistance-associated regions previously reported in *S. pneumoniae*

We also detected known antibiotic resistance-associated genes (Figure 4A–C). Mutations in *pbp2a*, *pbp2b*, *pbpX*, *murMN*, and *ciaH* genes were mainly confined to PEN resistance, whereas mutations in *pbp2a*, *pbp2b*, *pbp1b*, *murMN*, and *ciaH* genes were associated with CXM and CRO resistance. Additionally, macrolide-resistant isolates were found to carry *ermB* and *mefA* genes, which encode ribosomal methylase and drug-efflux proteins, respectively. Most macrolide-resistant *S. pneumoniae* were also resistant to tetracycline and harbor the *tetM* gene. We found that only mutations in one gene (*rplD*) conferred macrolide resistance, whereas mutations in *dfr* and *sulA* genes were closely related to SXT resistance. These characteristics were not detected among the antibiotic-susceptible isolates.

Discussion

The emergence of WGS as a powerful replacement for traditional molecular technologies has enabled highly detailed molecular epidemiology studies to be conducted on the genomic diversity of bacterial pathogens. As a comparative tool for the bacterial genome, WGRS can confirm a suspected disease outbreak of clinical drug-resistant isolates (eg, MRSA, MDR-tuberculosis).[20,21] In addition to tracking drug-resistant bacteria, WGRS allows the discovery of specific

Table 3 SNP changes associated with antibiotic resistance in *S. pneumoniae*

Gene symbol	Gene description	Mutations	Consequence	Potential antibiotics
pbpX	Penicillin-binding protein 2X	4	NS/MNP	PEN/CXM/MAC
murF	UDP-*N*-acetylmuramoylalanine-D-glutamyl-lysine-D-alanyl-D-alanine ligase	1	NS	PEN
ddl	D-alanine-D-alanine ligase	3	NS/MNP	PEN
rpmG	50S Ribosomal protein L33	1	NS	PEN
pbp2a	Penicillin-binding protein 2a	29	NS/MNP	PEN
sulA	Dihydropteroate synthase	4	NS	CXM/MAC
gdhA	NADP-specific glutamate dehydrogenase	1	NS	CXM
hom	Homoserine dehydrogenase	1	NS	CXM
xpt	Xanthine phosphoribosyltransferase	1	NS	CXM
bglA	6-phospho-beta-glucosidase	1	NS	CXM
regR	Transcription regulator, member of GalR family	10	NS/MNP	CXM
pbpA	Penicillin-binding protein 1A	6	NS/MNP	CXM
zmpB	Zinc metalloprotease	1	NS	CXM
pepT	Aminotripeptidase; tripeptidase	3	NS	CXM
gcp	Secreted metalloendopeptidase Gcp	2	NS/MNP	CXM
kdgK	2-keto-3-deoxygluconate kinase	11	NS/MNP	CXM
gno	5-keto-D-gluconate 5-reductase	2	NS/MNP	CXM
ugl	Unsaturated glucuronyl hydrolase	2	NS/MNP	CXM
recU	Recombination protein U	2	NS/MNP	CXM
dagA	D-alanine glycine permease	1	NS	CXM
galT	Galactose-1-phosphate uridylyltransferase	1	NS	CXM
trkH	Trk transporter membrane-spanning protein-K+ transport	1	NS	MAC
pbp2b	Penicillin-binding protein 2B	5	NS	MAC
PTS-EII	Phosphotransferase system sugar-specific EII component	4	NS	CXM/MAC
adhP	Alcohol dehydrogenase, propanol-preferring	1	NS	MAC
ropA	Trigger factor	1	NS	MAC
bglH	6-phospho-beta-glucosidase	1	NS	MAC
miaA	tRNA Isopentenylpyrophosphate transferase	3	NS/MNP	MAC
divIB	Cell division protein DivIB	1	NS	MAC
truA	tRNA pseudouridine synthase A	1	NS	MAC
pspC	Choline-binding protein A	3	NS	MAC/SXT
ugd	UDP-glucose dehydrogenase	1	NS	SXT
dfr	Dihydrofolate reductase	7	NS/MNP	SXT
sulB	Dihydrofolate synthetase	2	NS/MNP	SXT
dpr	DNA-binding protein starved cells-like peroxide resistance protein	1	NS	SXT
dctA	Dicarboxylate/amino acid:cation (Na+ or H+) symporter	1	NS	SXT

Abbreviations: NS, nonsynonymous; MNP, more than or equal to two bases mutations PEN, penicillin; CXM, cefuroxime; MAC, macrolides (erythromycin and azithromycin) and clindamycin; SXT, sulfamethoxazole–trimethoprim SNP, single-nucleotide polymorphism.

drug resistance mechanisms, such as motifs in mobile genetic elements such as transposons, and the mechanisms underlying gene transfer.[22] The major challenge for WGRS is predicting the patterns of evolution involving genetic diversity in SNPs and indels.[23] In this study, we selected 25 strains with differing antibiotic resistance patterns from different clinical isolates obtained from pediatric samples for WGRS analysis.

Phylogenetic analysis of the 25 isolates found that strains with the same antibiotic resistance profiles harbored different genetic backgrounds and vice versa. That is to say, the genetic changes we observed may not be independent of the drug resistance phenotype. Composite effects of selection, mutation, and recombination can force organisms toward bacterial evolution. Spontaneously, horizontal acquisition of transposons conferring drug resistance may promote the adaptation of organism to an environment with high drug resistance.[24] The dynamics of these mobile elements might be attributed to a clinically important drug resistance shift in *S. pneumoniae*. We found that macrolide resistance strains have acquired *ermB* and *mefA* genes, which are carried on transposons such as Tn916 and Tn5251 which have been previously reported in other articles.[25,26]

To gain insight into the genomic basis for antibiotic resistance traits, we analyzed the genomes of pneumococcal isolates with MDR resistance. As reported previously, MDR isolates display a series of resistance mechanisms including modification of drug targets, inactivation of therapeutic agents, and overexpression of efflux pumps.[27] Efflux pumps

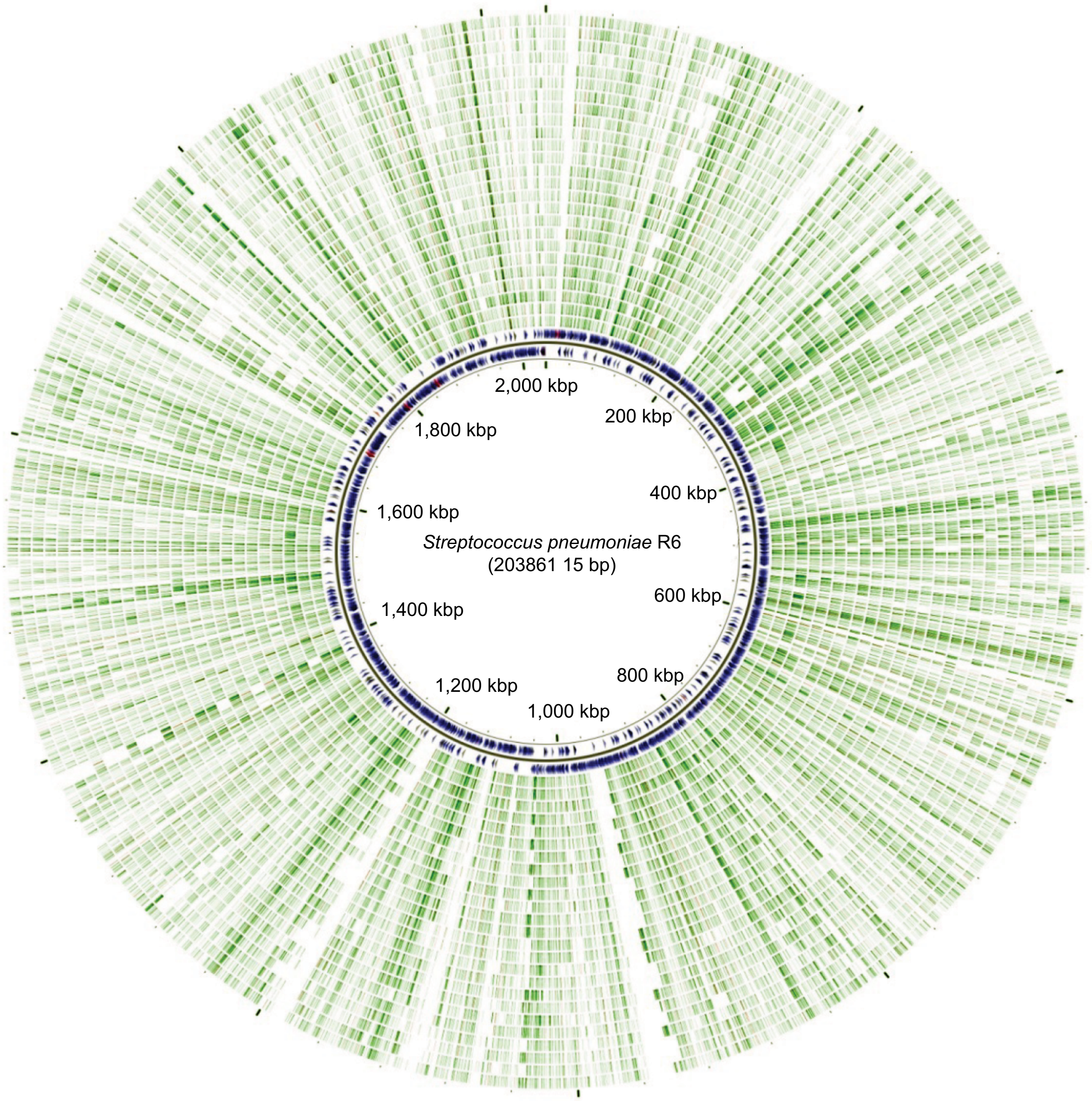

Figure 3 Indel and SNP variation in the complete genome of *S. pneumoniae* R6 compared with 25 *S. pneumoniae* clinical isolates.

Notes: The distribution of nonsynonymous mutations (SNPs and indels) in the genome sequences of 25 isolates compared with the *S. pneumoniae* R6 reference strain is shown. Protein-coding genes on the sense and anti-sense strands are shown in rings II and III (from inside to outside). The arrow directions indicate the translational directions. The distribution of nonsynonymous mutations in the genomes of 25 isolates (from isolate 14LC-ER1023 to 14LC-ER1047) are shown from ring IV to ring XXVII. Green bars represent nonsynonymous SNPs; red bars represent nonsynonymous indels.

Abbreviation: SNP, single-nucleotide polymorphisms.

that contribute to the MDR phenomenon are divided into the ATP-binding cassette (ABC) family of transporters, major facilitator superfamily proteins, small multidrug resistance family proteins, resistance-nodulation-division family proteins and, multidrug and toxic compound extrusion family proteins.[28] In this study, we found that alterations in the ABC-membrane spanning domain efflux pump were observed in the MDR isolates, and this may have led to the MDR strains.

Comparative analysis of all strains showed that mutations in the form of SNPs and indels occurred frequently among different strains of *S. pneumoniae.* Specially, in our search for genome regions with high mutation rates in the isolates, our analysis linked diversity in indels and SNPs in genes (eg, *dpr*, *dagA*, *zmpB*, *pspC*, *adhP*, *truA*, *bglH*) with resistance against several antibiotics, although the gene annotation indicated that these genes are mainly involved in virulence as well as carbohydrate and amino acid transport and metabolism. Furthermore, other genetic changes involved in lone antibiotic resistance were always detected. A large number of mutations in PBPs (eg, *pbp1b*, *pbp2a*, *pbp2b*, and *pbpX*) have

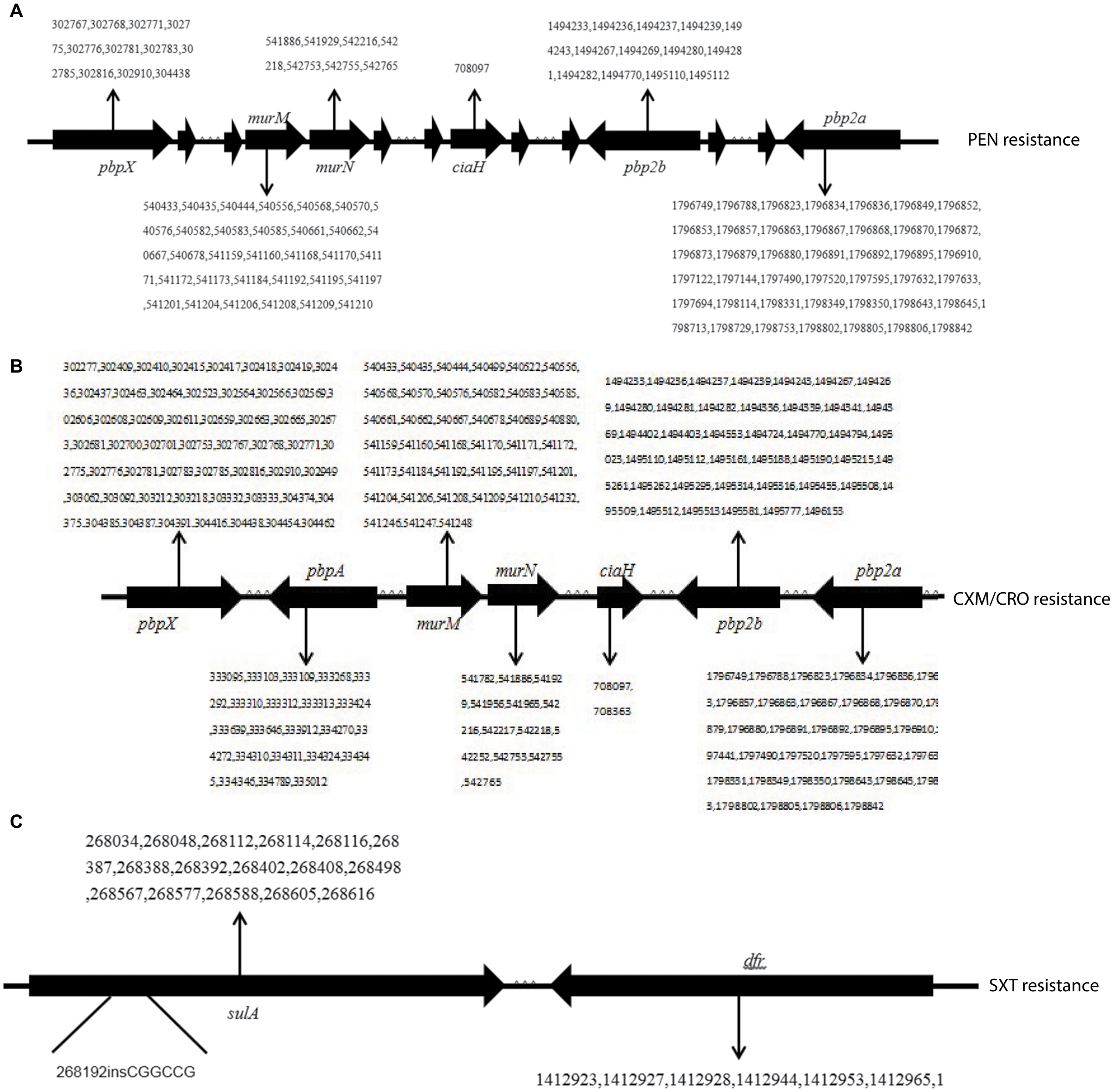

Figure 4 (**A**) SNPs changes in PBPs associated with penicillin resistance in *S. pneumoniae.* (**B**) SNPs changes in PBPs associated with cephalosporins resistance in *S. pneumoniae.* (**C**) SNP changes associated with sulfamethoxazole–trimethoprim resistance in *S. pneumonia.*

Abbreviations: SNPs, single-nucleotide polymorphisms; PBPs, penicillin-binding proteins; PEN, penicillin; CXM, cefuroxime; CRO, ceftriaxone; SXT, sulfamethoxazole–trimethoprim.

been identified, and these are reported to reduce the binding affinities for β-lactam antibiotics, resulting in resistance.[29,30] Moreover, other resistance determinants such as MurMN and CiaRH affecting peptidoglycan structure are known to influence β-lactam resistance.[31,32] Among the indels observed in the resistant isolates in this study, *sulA* is characterized by a 6-bp insertion, resulting in the insertion of two amino acids (Arg and Pro) in dihydropteroate synthase, thereby conferring resistance to SXT.[33] Concurrently, other studies have reported nonsynonymous SNPs in the dihydrofolate reductase encoding *dfr* gene in SXT isolates.[34] Furthermore, it was found that mutations in the ion transporter membrane-spanning protein encoding *trkH* gene were responsible for increased antibiotic susceptibility in the aminoglycoside-resistant population.[35] Moreover, the same study found mutations in a gene close to *trkH*, namely *PTS-EII*, which suggests that an operon in this region regulating the expression of the structural genes exists. However, whether these changes are necessarily related to drug resistance needs further analysis.

This study has several limitations. First, the sample size was fairly small: only 25 clinical isolates were used, and this may have affected the genetic divergence analysis. Second, the strains were isolated over 2 years, and were from different patients with different genetic backgrounds. Last, we did not conduct any experiments to confirm that the genetic diversity in indels and SNPs were associated with antibiotic resistance. In all, further investigation of the function of some of the genes identified in this study is essential.

In conclusion, by undertaking WGRS and comparative genomic analysis of 25 pneumococcal isolates, we have shown that although pneumococcal isolates were similar on genetic background, strains were diverse at the genomic level. These strains exhibited distinct variations in their indel and SNP compositions, which are potentially correlated with drug resistance. Here, we have simply reported the genomic variation present in the isolates. Further in-depth investigations of these variations and their associations with antibiotic resistance mechanisms are required.

Acknowledgments

The authors thank the microbiologists and technical staff from Shanghai Children's Hospital for isolating the *S. pneumoniae* strain. The authors also thank Shanghai Personal Biotechnology Company for providing bioinformatics analysis of these isolates. This work was supported by the Shanghai Municipal Commission of Health and Family Planning (grant no. 2015ZB0203) and Shanghai Shenkang Hospital Development Center (grant no. SHDC22014012).

Disclosure

All authors report no potential conflicts of interest.

References

1. World Health Organization. Pneumococcal conjugate vaccine for childhood immunization: WHO position paper. *Wkly Epidemiol Rec*. 2007;82(12):93–104.
2. Kim SH, Song JH, Chung DR, et al; ANSORP Study Group. Changing trends in antimicrobial resistance and serotypes of *Streptococcus pneumoniae* isolates in Asian countries: an Asian Network for Surveillance of Resistant Pathogens (ANSORP) study. *Antimicrob Agents Chemother*. 2012;56(3):1418–1426.
3. Merker M, Kohl TA, Roetzer A, et al. Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium* tuberculosis Beijing strains in patients. *PLoS One*. 2013;8(12):e82551.
4. Cherazard R, Epstein M, Doan TL, Salim T, Bharti S, Smith MA. Antimicrobial resistant *Streptococcus pneumoniae*: prevalence, mechanisms, and clinical implications. *Am J Ther*. 2017;24(3):e361–e369.
5. Schweizer I, Blättner S, Maurer P, et al. New aspects of the interplay between penicillin binding proteins, murM, and the two-component system CiaRH of penicillin-resistant *Streptococcus pneumoniae* Serotype 19A isolates from Hungary. *Antimicrob Agents Chemother*. 2017;61(7):e00414–e00417.
6. Rayner RE, Savill J, Hafner LM, Huygens F. Genotyping *Streptococcus pneumoniae*. *Future Microbiol*. 2015;10(4):653–664.
7. Andam CP, Hanage WP. Mechanisms of genome evolution of Streptococcus. *Infect Genet Evol*. 2015;33:334–342.
8. Côrtes MF, Costa MO, Lima NC, et al. Complete genome sequence of community-associated methicillin-resistant *Staphylococcus aureus* (strain USA400-0051), a prototype of the USA400 clone. *Mem Inst Oswaldo Cruz*. 2017;112(11):790–792.
9. Jakobsson HE, Salvà-Serra F, Thorell K, et al. Draft genome sequences of six strains of *Streptococcus pneumoniae* from serotypes 5, 6A, 6B, 18C, 19A, and 23F. *Genome Announc*. 2017;5(14):e00125–e00117.
10. Gardner KAJA, Osawa M, Erickson HP. Whole genome re-sequencing to identify suppressor mutations of mutant and foreign *Escherichia coli* FtsZ. *PLoS One*. 2017;12(4):e0176643.
11. Loman NJ, Constantinidou C, Chan JZ, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol*. 2012;10(9):599–606.
12. Wessels E, Schelfaut JJ, Bernards AT, Claas EC. Evaluation of several biochemical and molecular techniques for identification of *Streptococcus pneumoniae* and *Streptococcus pseudopneumoniae* and their detection in respiratory samples. *J Clin Microbiol*. 2012;50:1171–1177.
13. Clinical and Laboratory Standards Institute (CLSI). *Performance Standards for Antimicrobial Susceptibility Testing; 26th Informational Supplement*. CLSI document M100-26. Wayne, PA: Clinical and Laboratory Standards Institute; 2016.
14. Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology*. 1998;144(Pt 11):3049–3060.
15. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016;9:88.
16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
17. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;14(7):1394–1403.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–410.
19. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Bio Evol*. 2013;30(12):2725–2729.
20. Köser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med*. 2012;366(24):2267–2275.
21. Liu F, Hu Y, Wang Q, et al. Comparative genomic analysis of *Mycobacterium* tuberculosis clinical isolates. *BMC Genomics*. 2014;15:469.
22. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. *Pathology*. 2015;47(3):199–210.
23. Relman DA. Microbial genomics and infectious diseases. *N Engl J Med*. 2011;365(4):347–357.
24. Ding F, Tang P, Hsu MH, et al. Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. *BMC Genomics*. 2009;10:158.
25. Korona-Glowniak I, Siwiec R, Malm A. Resistance determinants and their association with different transposons in the antibiotic-resistant *Streptococcus pneumoniae*. *Biomed Res Int*. 2015;2015:836496.
26. Iannelli F, Santoro F, Oggioni MR, Pozzi G. Nucleotide sequence analysis of integrative conjugative element Tn5253 of *Streptococcus pneumoniae*. *Antimicrob Agents Chemother*. 2014;58(2):1235–1239.

27. Miller WR, Munita JM, Arias CA. Mechanisms of antibiotic resistance in enterococci. *Expert Rev Anti Infect Ther*. 2014;12(10):1221–1236.
28. Blair JM, Richmond GE, Piddock LJ. Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. *Future Microbiol*. 2014;9(10):1165–1177.
29. Sanbongi Y, Ida T, Ishikawa M, et al. Complete sequences of six penicillin-binding protein genes from 40 *Streptococcus pneumoniae* clinical isolates collected in Japan. *Antimicrob Agents Chemother*. 2004;48(6):2244–2250.
30. Chaguza C, Cornick JE, Everett DB. Mechanisms and impact of genetic recombination in the evolution of *Streptococcus pneumoniae*. *Comput Struct Biotechnol J*. 2015;13:241–247.
31. Davey L, Halperin SA, Lee SF. Mutation of the *Streptococcus gordonii* thiol-disulfide oxidoreductase SdbA leads to enhanced biofilm formation mediated by the CiaRH two-component signaling system. *PLoS One*. 2016;11(11):e0166656.
32. Metcalf BJ, Chochua S, Gertz RE Jr, et al; Active Bacterial Core Surveillance Team. Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. *Clin Microbiol Infect*. 2016;22(12):1002.e1–1002.e8.
33. Haasum Y, Ström K, Wehelie R, et al. Amino acid repetitions in the dihydropteroate synthase of *Streptococcus pneumoniae* lead to sulfonamide resistance with limited effects on substrate K(m). *Antimicrob Agents Chemother*. 2001;45(3):805–809.
34. Maskell JP, Sefton AM, Hall LM. Multiple mutations modulate the function of dihydrofolate reductase in trimethoprim-resistant *Streptococcus pneumoniae*. *Antimicrob Agents Chemother*. 2001;45(4):1104–1108.
35. Oz T, Guvenek A, Yildiz S, et al. Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution. *Mol Biol Evol*. 2014;31(9):2387–2401.