Whole-genome sequencing analysis of multidrug-resistant serotype 15A *Streptococcus pneumoniae* in Japan and the emergence of a highly resistant serotype 15A-ST9084 clone

Supplementary Material

Supplementary Methods

Illumina raw read processing

In this study, we multiplexed and sequenced 52 new samples (isolates) on an Illumina NextSeq instrument for 300 cycles (2×150-bp paired end), in addition to the 34 samples (isolates) sequenced in our previous study, which were sequenced using a MiSeq system for 600 cycles (2×300-bp paired-end) (1). All data were processed using Trimmomatic v0.38 (2) and cutadapt v1.18 (3) to remove adapter sequences and perform quality filtering.

Information about *S. pneumoniae* G54, which was used as a reference sequence in this study

S. pneumoniae G54 is serotype 19F and belongs to ST63. According to the NCBI database, this isolate was recovered in 1991 in Italy from respiratory samples, and its complete whole-genome information is available under GenBank accession No. NC_011072.1. The total length of the sequence is 2078953, and the isolate contains a total of 2269 genes and 1986 coding genes. This isolate was previously annotated by Dopazo et.al (4). It has a defective Tn*6002*-like element that is one of the Tn*916*-like elements with a 2849 bp *ermB* insertion between ORF19 and 20.

Mapping and phylogenetic analysis

To create a phylogenetic tree, we used Genealogies Unbiased By recomBinations In Nucleotide Sequences (Gubbins) v1.4.6 (5), which identifies recombination events using an algorithm that iteratively identifies loci containing increased densities of base substitutions while concurrently constructing a phylogeny based on the putative point mutations outside of these regions. To prepare input files for Gubbins with *S. pneumoniae* G54 as an outgroup isolate, we mapped processed reads to the reference genome of *S. pneumoniae* G54 (GenBank accession No. NC_011072.1) using the Burrows-Wheeler Aligner v0.7.17 (6). The methods used for this process were described previously (1). The same process was performed for foreign isolates. First, we downloaded read data from the Sequence Read Archive (SRA) database (http://www.ncbi.nlm.nih.gov/sra/). Then, we checked their serotypes based on a previous study (7) and performed multilocus sequence typing using the reads (8).

We created three phylogenetic trees. For the first tree, we used all global serotype 15A-ST63 isolates, including serotype 15A-CC63 isolates from Japan (Figure S2), with *S. pneumoniae* G54 included as an outgroup isolate. For the second tree, we used only serotype 15A-CC63 isolates from Japan, with *S. pneumoniae* G54 included as an outgroup isolate (Figure 1). Finally, we created a clade Y-specific tree to predict recombination sites that could generate the observed serotype 15A-ST9084 isolate-specific cefotaxime resistance.

Core-genome analysis

To identify serotype 15A-ST9084 isolate-specific genes, we performed a core-genome comparison using all clade Y isolates with get_homologues using standard parameters (9). To create input files for get_homologues, we annotated all samples using Prokka v1.12 with standard parameters (10). Core-genomes of subclade CTRX isolates and pangenomes of the remaining clade I isolates were obtained, and subclade CTRX-specific genes (identified within all of the subclade CTRX isolates but not identified in any remaining clade Y isolates) were identified.

Genome assembly and genome comparison using BLAST+

The trimmed reads sequenced in this study were assembled using SPAdes v3.12.0 (11), with k-mer values ranging from 29 to 101 in careful mode. Foreign isolate reads were also assembled using SPAdes, with standard parameters in careful mode. The quality of the assemblies was evaluated using QUAST v5.0.0 (12), and BLAST+ v2.6.0 was used to identify the presence of target genes (13). The details of these processes were described previously (1).

Dating the origin of the multidrug-resistant serotype 15A-ST9084 clone

To estimate the date of the most recent common ancestor (MRCA) of the serotype 15A-ST9084 clone, we performed Bayesian analysis of molecular sequences using Markov chain Monte Carlo (MCMC) in BEAST v1.10.4 (14). In this analysis, we used all of the clade Y isolates whose alignment of base substitutions occurring outside of putative recombination regions was entered into BEAUti v1.10.4 to create input data for BEAST. The model was selected via a comparison of the marginal likelihood using path sampling for a strict clock and an uncorrelated relaxed clock in a molecular clock model; and for a constant size, an exponential growth and Bayesian skyride in a tree prior model was used. Consequently, we selected an uncorrelated relaxed clock model and an exponential growth model. To have an effective sample size (ESS) greater than 200 for all factors, we set the MCMC length to 2.0×10^8. The median (HPD interval of growth rate) was 0.6209 (0.0345-1.5241), with an ESS of 856.

Accessory genome variation comparison

To elucidate the genetic details of serotype 15A-ST9084 isolates, we analyzed the *cps* locus and the Tn*916* mobile element structure. We analyzed all 86 serotype 15A-CC63 isolates from Japan and the PMEN15A-25 isolate. In the *cps* locus analysis, we obtained

the sequences of the corresponding region by mapping short reads to the previously published reference sequence (15) (NCBI Reference Sequence: CR931663.1) using the Burrows-Wheeler Aligner v0.7.17 (6). After aligning the sequence, a phylogenetic tree was created using RAxML v8.2.10 (16) with a GTRGAMMA DNA substitution model. Node support was assessed by using 500 bootstrap replicates. In the tree, the PMEN15A-25 isolate was used as an outgroup isolate. In the Tn*916* mobile element analysis, we obtained the corresponding sequences from the processed contigs using BLAST+ v2.6.0 (13), and the sequences were compared using ACT v18.0.0 (17) with standard parameters in BLAST+ v2.6.0. The reference of the Tn*916* mobile element was published as NCBI Reference Sequence U09422.1.

Supplementary Results

The amino acid sequence that was specific to subclade CTXR isolates is as follows:

MVFAFHLPPDELITNVIFKEKINSMLKCYIDRLLYVFINPTHFTEKVNLQFYGSFF
SYEFICREVGNILKNKGVKCNLNFFEGKEYL

Supplementary Dataset

Dataset S1. Foreign isolate information used in this study.

Dataset S2. Domestic isolate information analyzed in this study.

Dataset S3. PBP transpeptidase domains that were not present in the US database.

Supplementary Figures

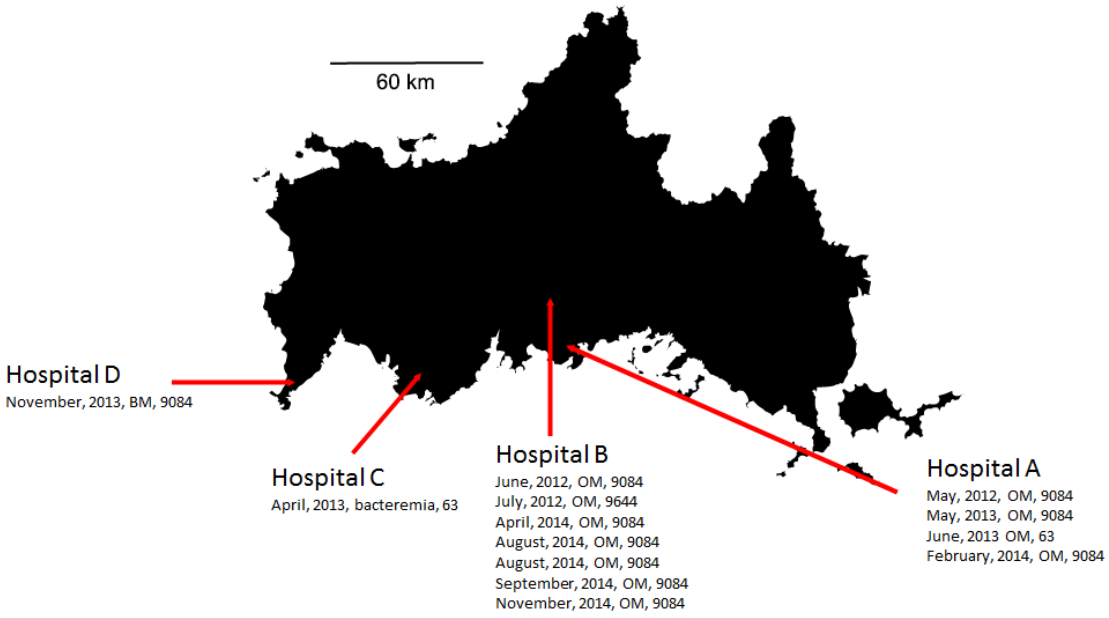

Figure S1. Location of the Yamaguchi prefecture in Japan and all hospitals in the Yamaguchi prefecture where multidrug-resistant serotype 15A isolates were identified. Time (month and year), diagnosis, and sequence type (ST) are indicated in that order. ST9084 and ST9644 were single-locus variants of ST63. BM, bacterial meningitis; OM, otitis media.

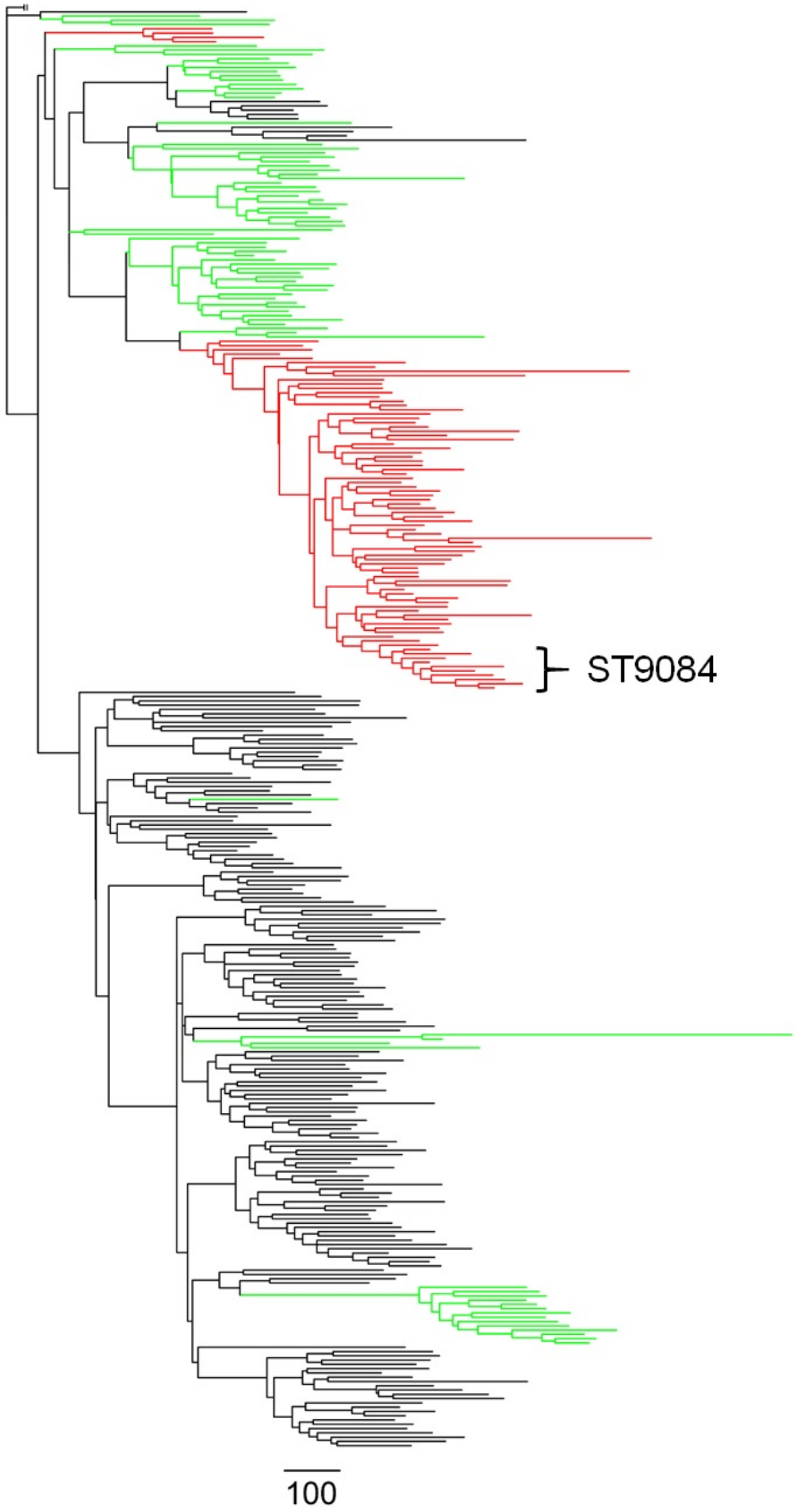
ST9084
100

Figure S2. Phylogenetic tree created by Gubbins using global serotype 15A-CC63 isolates. The branch colors in the tree indicate where the isolates were collected: red, Japan; black, United States; and green, United Kingdom.

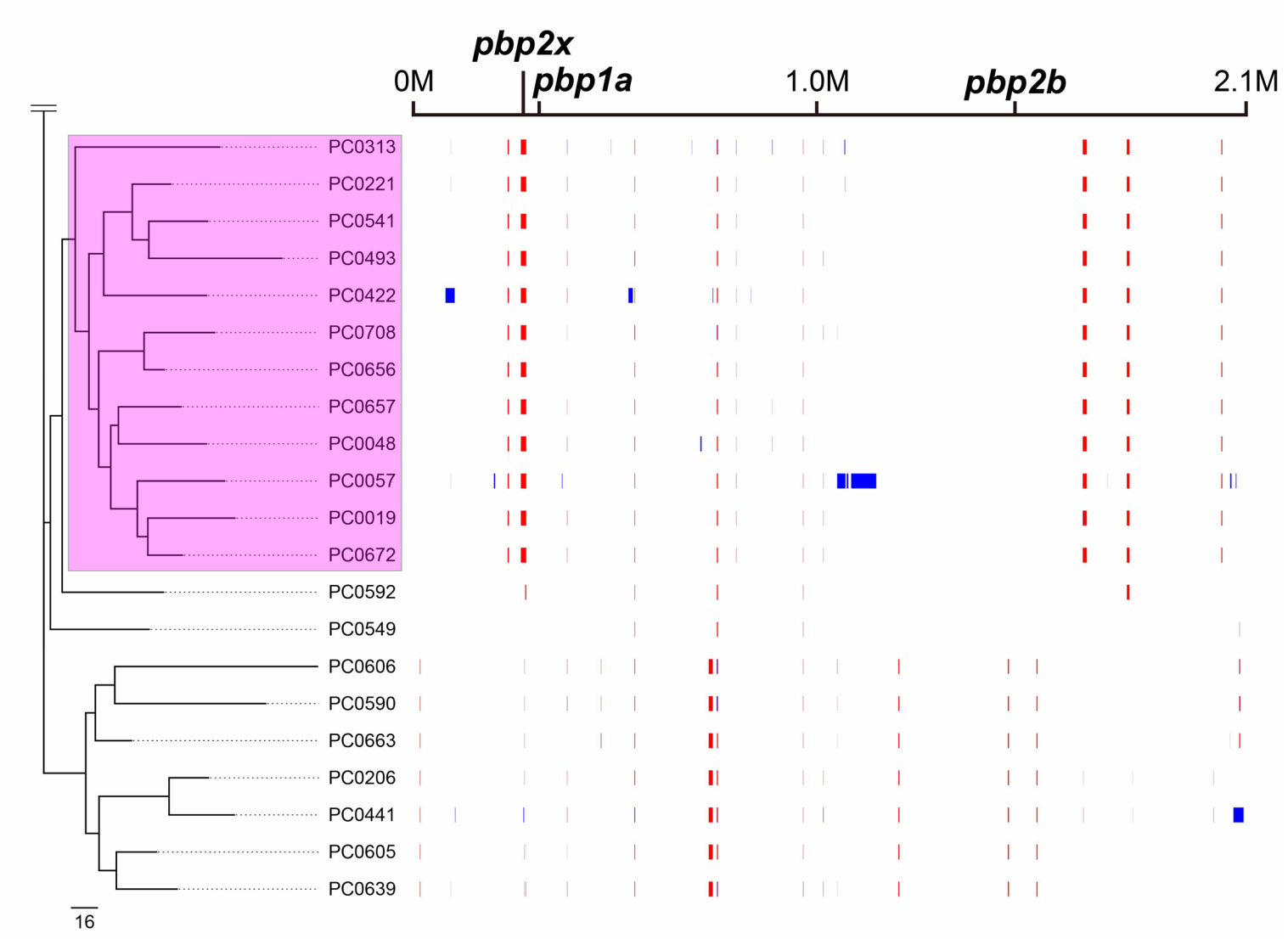
pbp2x
pbp1a
pbp2b
0M
1.0M
2.1M
PC0313
PC0221
PC0541
PC0493
PC0422
PC0708
PC0656
PC0657
PC0048
PC0057
PC0019
PC0672
PC0592
PC0549
PC0606
PC0590
PC0663
PC0206
PC0441
PC0605
PC0639
16

Figure S3. The phylogenetic tree created by Gubbins using all clade Y isolates (Figure 1) generated a multidrug-resistant serotype 15A-ST9084-specific clade (highlighted in pink). All isolates included in the clade, except for PC0313, were ST9084. PC0313 represents the multidrug-resistant serotype 15A-ST63 isolates derived from the Yamaguchi prefecture. The block chart on the right shows the predicted recombination sites for each isolate. Blue blocks are unique to a single isolate, while red blocks are shared by more than one isolate in the same node.

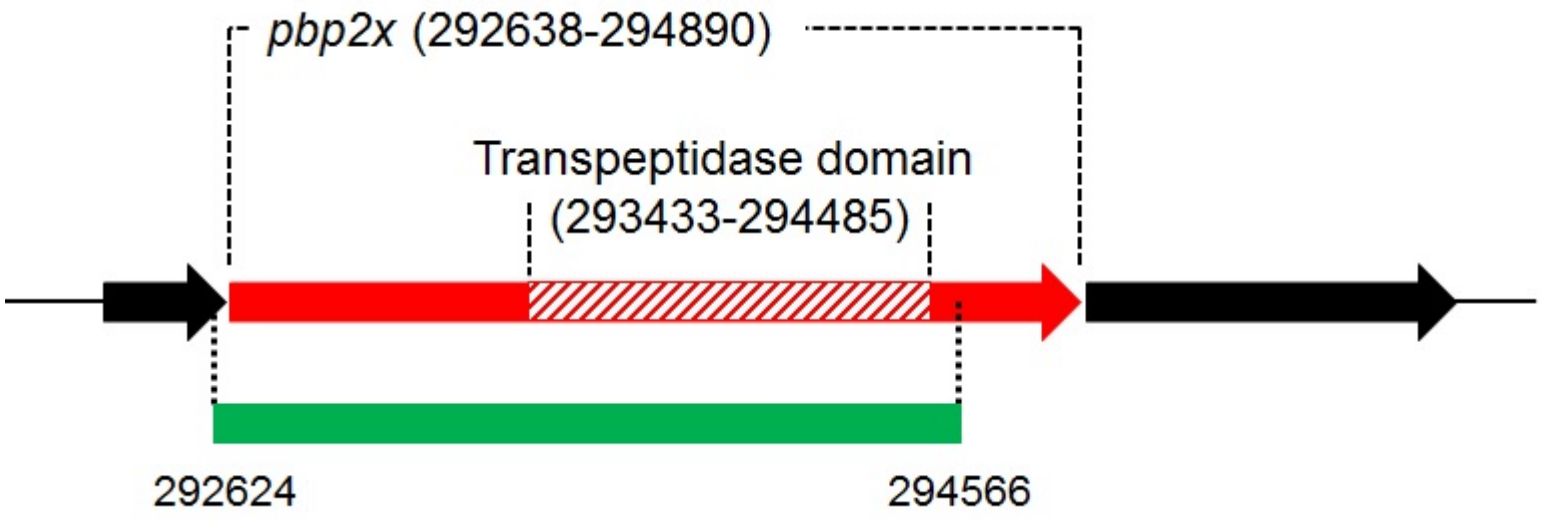

Figure S4. Sketch of the predicted recombination sites overlapping the *pbp2x* region. The green block indicates the region in which the recombination site was observed. The numbers indicate sequence coordinates using *Streptococcus pneumoniae* G54 as a reference sequence (NCBI Reference Sequence: NC_011072.11). This recombination site was shared by all subclade CTXR isolates and was not observed in any of the other clade Y isolates.

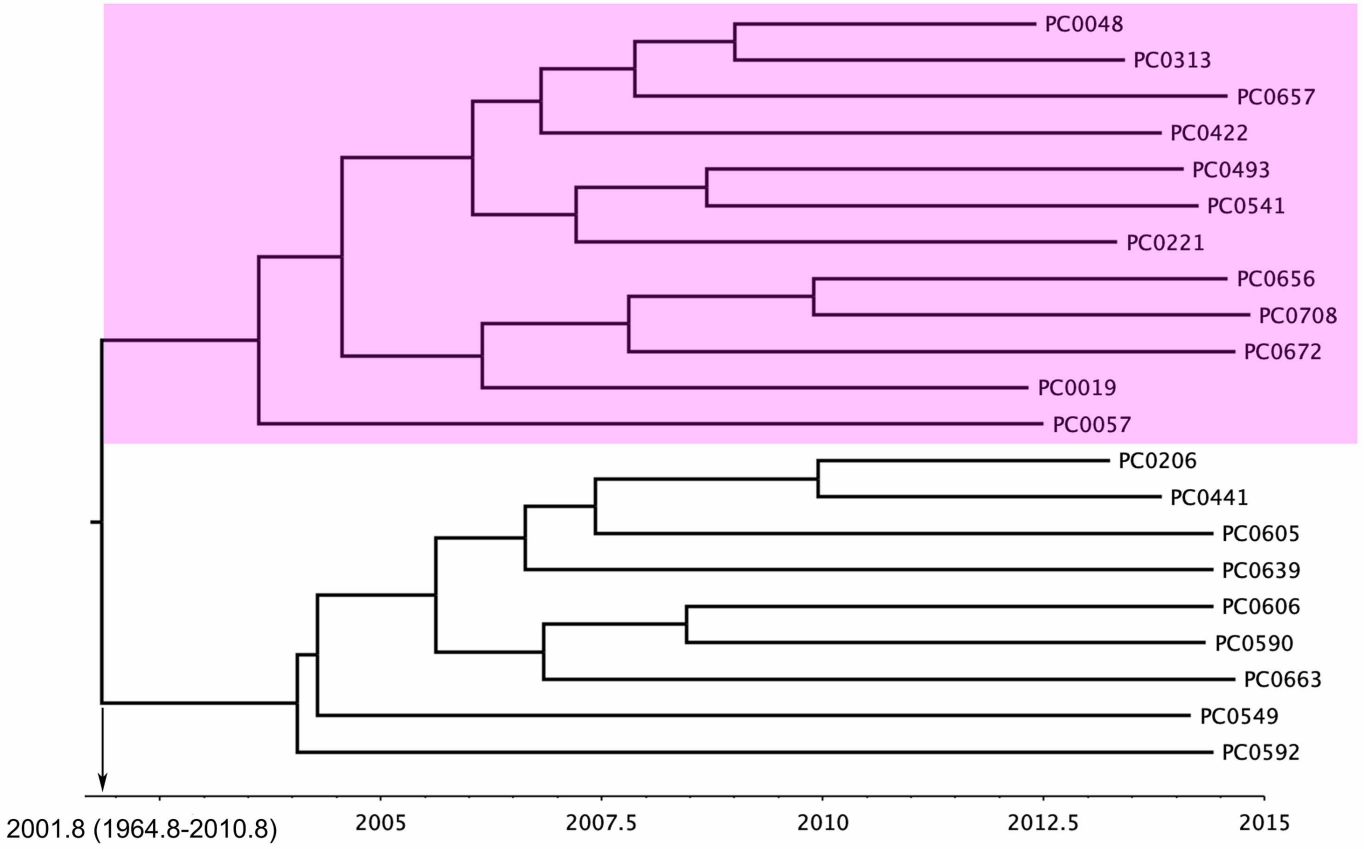
PC0048
PC0313
PC0657
PC0422
PC0493
PC0541
PC0221
PC0656
PC0708
PC0672
PC0019
PC0057
PC0206
PC0441
PC0605
PC0639
PC0606
PC0590
PC0663
PC0549
PC0592
2001.8 (1964.8-2010.8)
2005
2007.5
2010
2012.5
2015

Figure S5. Results of the date estimation for the most recent common ancestor (MRCA) of the multidrug-resistant serotype 15A-ST9084 clone. The pink-colored clade contains multidrug-resistant serotype 15A-ST9084 isolates. The year indicated with an arrow shows the node year when the resistant strains separated from the other strains and the 95% HPD.

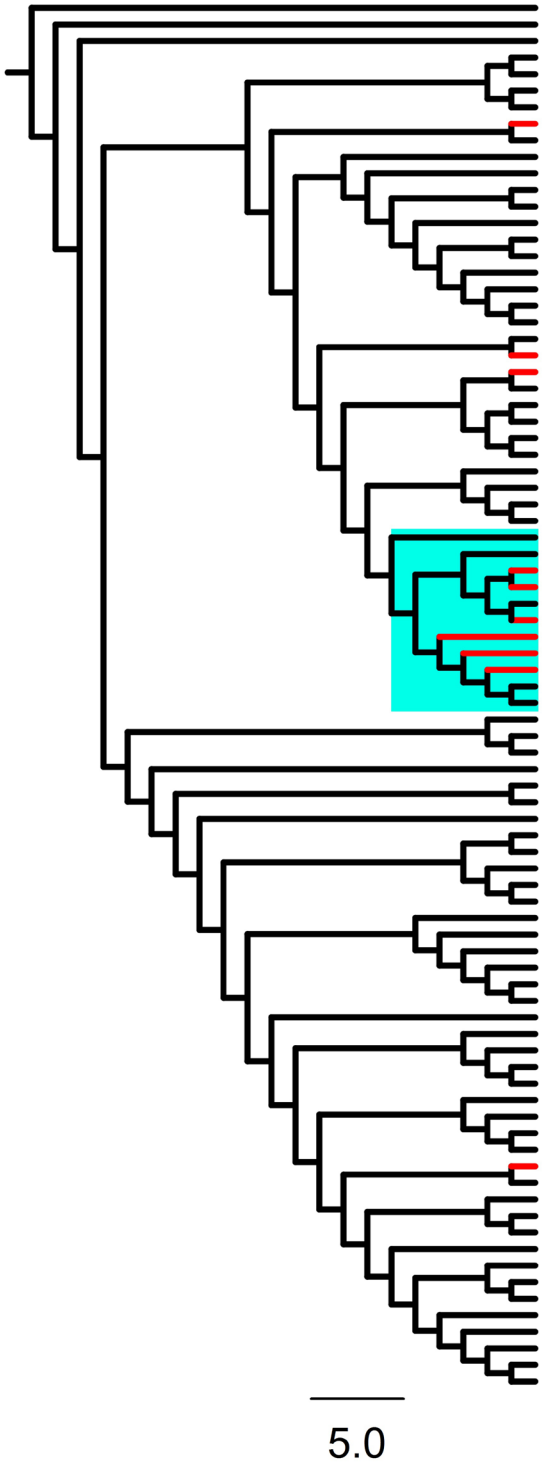
5.0

Figure S6. Variation in the serotype 15A *cps* loci. Isolates with red branches were serotype 15A-ST9084 isolates. All isolates highlighted in light blue were recovered from the Yamaguchi prefecture. Six of 10 serotype 15A-ST9084 isolates were clustered in the highlighted cluster; however, the other four isolates were clustered outside the cluster. This result indicates that the origin of the 15A *cps* loci of the ST9084 isolates was different; thus, the times of acquisition were also different.

Supplementary Tables

Table S1 Results of annotation and BLAST searching of components within recombination sites*.

Annotation	Results of blastp	Sequence ID	BLAST E-value
polC	DNA polymerase III 2C alpha subunit	CIV85673.1	6e-85
hypothetical protein	type II toxin-antitoxin system RelB/DinJ family antitoxin	WP_000041492.1	4e-55
yafQ	RelE/StbE family addiction module toxin	CMX12116.1	5e-61
hypothetical protein	hypothetical protein	WP_000776632.1	1e-25
pepS	aminopeptidase	WP_000243971.1	0.0
hypothetical protein	GlsB/YeaQ/YmgE family stress response membrane protein	WP_000901566.1	9e-44
sorB	PTS system mannose/fructose/N-acetylgalactosamine-transporter subunit IIB	WP_000178620.1	4e-116
agaC	PTS mannose/fructose/sorbose/N-acetylgalactosamine transporter subunit IIC	WP_000026613.1	0.0
manZ	PTS mannose/fructose/sorbose transporter family subunit IID	WP_000148012.1	0.0
hypothetical protein	preprotein translocase subunit YajC	WP_000381736.1	6e-61
hypothetical protein	oligohyaluronate lyase	WP_000657050.1	0.0
kdgR	LacI family DNA-binding transcriptional regulator	WP_078372374.1	0.0
rsmH	ribosomal RNA small subunit methyltransferase H	WP_000159396.1	0.0
hypothetical protein	uncharacterized protein	CGF52136.1	3e-32
ccpA	LacI family DNA-binding transcriptional regulator	WP_078934694.1	0.0
hypothetical protein	hypothetical protein	WP_000656287.1	1e-81
sapB	MgtC/SapB family protein	WP_000127554.1	3e-168
phnV	iron ABC transporter permease	WP_050116136.1	0.0
malK	ABC transporter family protein	EIC49522.1	0.0
hypothetical protein	extracellular solute-binding protein	WP_000738356.1	0.0
hypothetical protein	hypothetical protein	CTF53637.1	6e-160
iga	hypothetical protein	WP_088803941.1	1e-61
sdrD	hypothetical protein SPAR123_2107	EHD26016.1	4e-57

hypothetical protein	spfh domain/band 7 family	CEY43605.1	2e-115

* Annotation was performed by PROKKA with an E-value cutoff of 1e-6. Genes that were not identified as specifics gene by PROKKA were shown as “hypothetical proteins”. A BLAST search was performed with an E-value cutoff of 1e-10.

Table S2 *pbp1a* transpeptidase domain amino acid substitutions that were identified in this study.

		Coordinates of amino acids																											
		3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5
pbp2x	MIC of CTX	7	8	9	9	9	9	0	0	1	1	2	3	4	5	6	7	8	4	4	5	7	7	7	7	7	7	8	8
type	(No. of isolates)	1	2	2	3	5	7	5	8	3	4	1	2	7	8	2	5	7	0	6	0	0	1	4	5	6	7	3	5
1a-13	0.25 (1), 0.5 (39), 1.0 (10), 2.0 (3), 4.0 (12)	S	L	S	I	H	V	D	L	H	V	I	T	T	S	A	K	Y	T	G	P	K	Y	N	T	G	Y	M	V
1a-24	<=0.06 (2), 0.12 (1), 0.25 (4), 0.5 (13), 1.0 (1)	T	I	T	M	N	I	S	V	R	A	L	P	N	D	S	Q	F	S	N	A	N	H	T	S	Q	F	L	A

		Coordinates of amino acids			
		660	609	611	612
pbp2x type	MIC of CTX (No. of isolates)				
1a-13	0.25 (1), 0.5 (39), 1.0 (10), 2.0 (3), 4.0 (12)	I	D	F	L
1a-24	≤0.06 (2), 0.12 (1), 0.25 (4), 0.5 (13), 1.0 (1)	L	N	L	T

Table S3 *pbp2b* transpeptidase domain amino acid substitutions that were identified in this study.

		Coordinates of amino acids											
		4	4	4	4	4	4	4	4	4	5	5	6
pbp2x	MIC of CTX	1	2	3	3	3	4	6	8	9	1	4	2
type	(No. of isolates)	7	7	1	2	5	3	0	1	4	1	3	9
2b-27	≤0.06 (2), 0.12 (1), 0.25 (4), 0.5 (12), 1.0 (1)	S	N	T	Q	G	Q	L	G	S	D	D	E
2b-57	0.5 (1)	P	N	T	Q	D	E	L	G	S	E	D	A
2b-JP1	0.25 (1), 0.5 (39), 1.0 (10), 2.0 (3), 4.0 (12)	P	Y	K	L	G	E	I	E	T	E	N	E

Table S4 *pbp2x* transpeptidase domain amino acid substitutions that may affect the cefotaxime resistance of serotype 15A-CC63 isolates*.

		Coordinates of amino acids																											
		2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	
pbp2x	MIC of CTX	3	6	6	8	1	3	3	4	4	4	5	5	6	6	6	7	7	8	8	8	9	9	0	0	1	4	4	4
type	(No. of isolates)	0	5	8	1	1	8	9	3	6	7	5	8	4	6	9	1	8	2	4	9	3	9	0	1	7	4	7	9
2x-28	≤0.06 (2), 0.12 (1)	T	I	P	Q	D	T	M	M	A	A	G	V	L	I	A	I	E	G	R	S	A	G	M	T	N	N	M	A
2x-JP1	0.25 (1)	T	I	P	Q	D	P	M	M	A	A	G	V	L	I	A	I	E	G	R	S	A	G	M	T	N	N	M	A
2x-43	0.25 (4), 0.5 (52), 1.0 (8)	T	I	T	L	N	A	M	T	S	S	S	Y	F	I	A	T	E	T	G	L	A	G	M	S	K	S	Q	S
2x-112	1.0 (1)	K	L	P	L	N	A	M	T	S	S	S	Y	F	I	A	T	E	T	G	L	A	G	M	S	K	S	Q	S
2x-JP2	1.0 (2)	K	I	T	L	N	A	M	T	S	S	S	Y	F	I	A	T	D	T	G	L	A	G	M	S	S	S	Q	S
2x-JP19	2 (1)	T	L	P	L	N	A	F	M	S	S	S	Y	F	M	V	T	G	T	G	L	A	G	M	S	K	N	Q	S
2x-JP3	2.0 (2), 4.0 (10)	K	L	P	L	N	A	F	M	S	S	S	Y	F	I	A	T	E	T	G	S	A	G	T	S	K	S	Q	S
2x-JP6	4.0 (1)	K	L	P	L	N	A	F	M	S	S	S	Y	F	I	A	T	A	T	G	L	V	A	T	S	K	S	Q	S
2x-147	4.0 (1)	T	I	T	L	N	A	M	T	S	S	S	Y	F	I	A	T	E	T	G	L	A	G	M	S	K	S	Q	S

		Coordinates of amino acids																						
		4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	MIC of CTX	8	8	8	0	0	0	1	1	1	1	2	3	3	3	3	3	4	5	6	6	7	7	7
pbp2x type	(No. of isolates)	3	6	8	1	5	6	0	3	6	7	3	1	5	6	7	8	6	2	3	8	2	4	6
2x-28	≤0.06 (2), 0.12 (1)	L	P	D	N	K	E	V	D	V	M	T	S	A	T	V	N	L	E	T	N	V	S	H
2x-JP1	0.25 (1)	L	P	D	N	K	E	V	D	V	M	T	S	A	T	V	N	L	E	T	N	V	S	H
2x-43	0.25 (4), 0.5 (52), 1.0 (8)	I	T	N	N	K	E	T	N	I	L	L	Y	P	I	I	T	V	Q	V	Y	A	T	N
2x-112	1.0 (1)	I	T	N	N	K	E	T	N	I	L	L	Y	P	I	I	T	V	Q	V	Y	V	T	N
2x-JP2	1.0 (2)	I	T	N	N	K	E	T	N	I	L	L	Y	P	I	I	T	V	Q	V	Y	V	T	N
2x-JP19	2 (1)	I	T	N	K	E	D	T	N	I	L	L	Y	P	I	I	T	V	Q	V	Y	V	T	N
2x-JP3	2.0 (2), 4.0 (10)	I	T	N	N	K	E	T	N	I	L	L	Y	P	I	I	T	V	Q	V	Y	V	T	N
2x-JP6	4.0 (1)	I	T	N	N	K	E	T	N	I	L	L	Y	P	I	I	T	V	Q	V	Y	V	T	N
2x-147	4.0 (1)	I	T	N	N	K	E	T	N	I	L	L	Y	P	I	I	T	V	Q	V	Y	V	T	N

* The coordinate numbers were allocated according to those of the R6 strain PBP2x. CTX, cefotaxime.

References

1. Nakano S, Fujisawa T, Ito Y, Chang B, Matsumura Y, Yamamoto M, et al. Spread of Meropenem-Resistant Streptococcus pneumoniae Serotype 15A-ST63 Clone in Japan, 2012-2014. Emerg Infect Dis. 2018 Feb;24(2):275-83.
2. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 01;30(15):2114-20.
3. Marcel M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
4. Dopazo J, Mendoza A, Herrero J, Caldara F, Humbert Y, Friedli L, et al. Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate. Microb Drug Resist. 2001 Summer;7(2):99-125.

5. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015 Feb 18;43(3):e15.
6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754-60.
7. Metcalf BJ, Gertz RE, Jr., Gladstone RA, Walker H, Sherwood LK, Jackson D, et al. Strain features and distributions in pneumococci from children with invasive disease before and after 13 valent conjugate vaccine implementation in the United States. Clin Microbiol Infect. 2015 Sep 9.
8. Enright MC, Spratt BG. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiology. 1998 Nov;144 (Pt 11):3049-60.
9. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 2013 Dec;79(24):7696-701.
10. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068-9.
11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455-77.
12. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072-5.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10.
14. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007 Nov 8;7:214.

15. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. 2006 2:e31.

16. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014 30:1312-3.

17. Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, et al. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics 2008 24:2672-6.