

# Development of a Multilocus Sequence Typing Scheme for Molecular Typing of *Mycoplasma pneumoniae*

Rebecca J. Brown,<sup>a,b</sup> Matthew T. G. Holden,<sup>c</sup> O. Brad Spiller,<sup>a</sup> Victoria J. Chalker<sup>b</sup>

Cardiff University, School of Medicine, Department of Child Health, University Hospital of Wales, Cardiff, United Kingdom<sup>a</sup>; Bacteriology Reference Department, Public Health England, London, United Kingdom<sup>b</sup>; University of St. Andrews, School of Medicine, Medical & Biological Sciences, North Haugh, St. Andrews, United Kingdom<sup>c</sup>

*Mycoplasma pneumoniae* is a major human respiratory pathogen causing both upper and lower respiratory disease in humans of all ages, and it can also result in other serious extrapulmonary sequelae. A multilocus sequence typing (MLST) scheme for *M. pneumoniae* was developed based on the sequences of eight housekeeping genes (*ppa*, *pgm*, *gyrB*, *gmk*, *glyA*, *atpA*, *arcC*, and *adk*) and applied to 55 *M. pneumoniae* clinical isolates and the two type strains M129 and FH. A total of 12 sequence types (STs) resulted for 57 *M. pneumoniae* isolates tested, with a discriminatory index of 0.21 STs per isolate. The MLST loci used in this scheme were shown to be stable in 10 strains following 10 sequential subculture passages. Phylogenetic analysis of concatenated sequences of the eight loci indicated two distinct genetic clusters that were directly linked to multilocus variable-number tandem repeat analysis (MLVA) type. Genetic MLST clustering was confirmed by genomic sequence analysis, indicating that the MLST scheme developed in this study is representative of the genome. Furthermore, this MLST scheme was shown to be more discriminatory than both MLVA and P1 typing for the *M. pneumoniae* isolates examined, providing a method for further and more detailed analysis of observed epidemic peaks of *M. pneumoniae* infection. This scheme is supported by a public Web-based database (<http://pubmlst.org/mpneumoniae>).

*Mycoplasma pneumoniae* is a common cause of community-acquired pneumonia (CAP) transmitted by aerosol or close contact (1). *M. pneumoniae* may cause other serious extrapulmonary sequelae, such as encephalitis (2). The pathogen is found in all age groups, with a higher prevalence in children age 5 to 14 years (3, 4). Admissions to a United Kingdom hospital in patients with CAP that were attributed to *M. pneumoniae* were estimated at 18% in 1982 and 4% in 1999 (5). Major increases and decreases in *M. pneumoniae* infection have occurred periodically in the United Kingdom; historically, epidemics have occurred at approximately 4-year intervals and have lasted 12 to 15 months, concurrent with sporadic infection at a lower level and seasonal peaks from December to February (4, 6). However, globally, peaks of infection have been observed in either summer or autumn, with no obvious explanation for this seasonal variation (7–10).

Typing of clinical isolates by molecular methods is of importance for the understanding of the epidemiology of *M. pneumoniae* infection and for an analysis of endemic outbreaks. It is generally considered that molecular typing of *M. pneumoniae* is hampered by the fact that the pathogen is a genetically homologous species (11). Initial molecular typing targeted the gene encoding the major surface protein (P1) of *M. pneumoniae*. PCR-restriction fragment length polymorphism (PCR-RFLP) analysis of the P1 gene, which encodes a major adhesion, is the most common genotyping method. This enables the separation of isolates into two types, 1 and 2 (11–13). More recent studies utilize the repetitive regions, RepMp2/3 and RepMp4, which can be found in the P1 gene, for molecular typing and have resulted in the identification of an additional subtype and three variants of these subtypes (14, 15). Multilocus variable-number tandem-repeat (VNTR) analysis (MLVA) has also been used, which is based on the variation in the copy number of tandem repeated sequences (VNTRs) found at different loci across the genome. The variation in the copy number of these tandem repeats (TRs) depends on the isolate tested. Initially, 265 strains were grouped into 26 MLVA

types based on five VNTR loci (Mpn1 and Mpn13 to Mpn16), and an additional 18 novel types have since been reported (16–18). However, the Mpn1 locus is unstable in both clinical strains and in laboratory passages, and most of the novel types came from variations in Mpn1; therefore, there is international consensus that this locus should be removed from the typing scheme (19).

Multilocus sequence typing (MLST) was previously attempted for the molecular typing of *M. pneumoniae*; however, due to the homogeneity of the *M. pneumoniae* species, very little polymorphisms were found in the housekeeping genes examined, and it was previously concluded that MLST with housekeeping and structural genes was not useful for molecular typing (22). Only three housekeeping genes were thoroughly examined for polymorphisms across 30 isolates of either P1 type 1, 2, or a variant strain. The other genes selected for analysis were examined against a single representative strain from each subtype. In this study, an MLST scheme was developed with a high discriminatory ability to differentiate *M. pneumoniae* isolates based on sequence polymor-

Received 14 May 2015 Returned for modification 9 June 2015

Accepted 14 July 2015

Accepted manuscript posted online 22 July 2015

Citation Brown RJ, Holden MTG, Spiller OB, Chalker VJ. 2015. Development of a multilocus sequence typing scheme for molecular typing of *Mycoplasma pneumoniae*. *J Clin Microbiol* 53:3195–3203. doi:10.1128/JCM.01301-15.

Editor: N. A. Ledebauer

Address correspondence to Victoria J. Chalker, [vicki.chalker@phe.gov.uk](mailto:vicki.chalker@phe.gov.uk).

O.B.S. and V.J.C. are co-senior authors.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JCM.01301-15>.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.01301-15

phisms within eight housekeeping genes, improving on all published typing methods for *M. pneumoniae*.

## MATERIALS AND METHODS

***M. pneumoniae* strains, culture conditions, and sample preparation.** The strains analyzed in this study are listed in Table 1. Fifty-five *M. pneumoniae* strains were submitted to Public Health England, United Kingdom, for clinical diagnostic purposes, and the two *M. pneumoniae* type strains, FH (NCTC 10119, ATCC 15531) and M129 (ATCC 29342), were obtained from National Collection of Type Cultures (NCTC) (held by Public Health England). All strains were triple cloned on *Mycoplasma* agar (*Mycoplasma* Experience, Surrey, United Kingdom) and confirmed to be *M. pneumoniae* by amplification of the *p1* gene (23).

All strains were subsequently cultured in *Mycoplasma* liquid medium (MLM) (*Mycoplasma* Experience, Surrey, United Kingdom). For genomic sequencing, strains were grown in 100 ml of broth culture, and genomic DNA was extracted using the GenElute bacterial genomic DNA kit (Sigma, Dorset, United Kingdom). PCR amplification was performed on bacterial DNA from a 500- $\mu$ l 4-day culture that was released by boiling lysis (95°C for 10 min) following centrifugation at 17,000  $\times$  g for 10 min, removal of all MLM, and resuspension in 50  $\mu$ l of sterile water.

**Multilocus sequence typing.** Housekeeping genes considered to be conserved in other bacterial species under a low rate of selective pressure were chosen for analysis (Table 2). The locus sequences were selected using the available genome sequences of *M. pneumoniae* FH and M129 (FH GenBank accession no. NC\_017504.1, and M129 GenBank accession no. NC\_000912.1) and the available whole-genome sequences of 35 clinical isolates. Ten genes were included for initial analysis: RecA protein (*recA*), inorganic phosphatase (*ppa*), phosphoglycerate mutase (*pgm*), DNA gyrase subunit B (*gyrB*), guanylate kinase (*gmk*), serine hydroxymethyltransferase (*glyA*), elongation factor P (*efp*), ATP synthase subunit  $\alpha$  (*atpA*), carbamate kinase (*arcC*), and adenylate kinase (*adk*); however, *recA* and *efp* were excluded from the resulting MLST scheme. The locus regions for PCR amplification were selected based on areas of the protein-coding sequences (CDS) containing nucleotide polymorphisms.

PCR utilizing the primers listed in Table 3 was used to amplify the target genes from a further 20 *M. pneumoniae* clinical isolates. Amplification of each of the locus sequences was performed in a DNA thermocycler (Techne Prime, Stone, United Kingdom) in 50- $\mu$ l reaction mixtures containing 1 $\times$  GoTaq Flexi buffer (Promega, Southampton, United Kingdom), 1.5 mM MgCl<sub>2</sub>, 0.2 mM deoxynucleoside triphosphates, 0.5 pmol/ $\mu$ l each primer, 1.56 units of GoTaq DNA polymerase (Promega), and 2.5  $\mu$ l of template DNA. PCRs consisted of an initial denaturation step of 3 min at 94°C, followed by 35 cycles of 60 s at 94°C, 60 s at 60°C, and 60 s at 72°C. A final extension step was maintained for 10 min at 72°C. The primer sequences and PCR product sizes are shown in Table 3. The PCR products were analyzed on 1.5% agarose gels with ethidium bromide visualization. All PCRs were performed in duplicate.

PCR amplicons were purified using a Qiagen miniprep kit (Qiagen, Inc., Hilden, Germany), as per the manufacturer's instructions, and sequenced using the amplification primers, which was performed by MWG Eurofins (Ebersberg, Germany). The sequences obtained from each corresponding forward and reverse primer were assembled and trimmed for double-stranded high-quality sequences. All the sequences obtained for each locus were aligned using Clustal W (Vector NTI, Paisley, United Kingdom), and different allelic types (ATs) (sequences with at least a 1-nucleotide difference) were assigned sequential numbers. The combination of the eight alleles determined the allelic profile of a strain, and each unique allelic profile was designated a unique sequence type (ST). Open reading frame amino acid sequences were identified using the Expasy translation tool (*Mycoplasma* setting [web.expasy.org/translate/]) for each AT. The deduced amino acid sequences were aligned using Clustal W (Vector NTI) for each locus, and synonymous changes were identified.

**MLVA and P1 typing.** MLVA types were determined as described by Dégrange et al. (16), excluding the VNTR locus Mpn1 and using interna-

tional nomenclature consensus (19). P1 types were determined as described by Dumke et al. (15).

**Genomic sequencing.** Genomic sequence data for 35 isolates were obtained using the Illumina Nextera XT sample prep kit (Illumina, Cambridge, United Kingdom) and sequenced on an Illumina HiSeq 2500 platform with TruSeq rapid SBS kits (200 cycles; Illumina) and cBOT for cluster generation (Illumina). Fastq reads were trimmed using Trimmomatic 0.32, with the following parameters: leading, 30; trailing, 30; slidingwindow, 10:30; and minlen, 50 (20). The Illumina reads were assembled to the M129 type strain (GenBank accession no. NC\_000912.1) using SPAdes version 2.5.0 (21) and mapped to M129 using Geneious version 8.0.4. Sequencing yielded at least one contig of between 99,047 bp and 324,397 bp with homology to the M129 type strain (GenBank accession no. NC\_000912.1), passing quality and coverage checks. Identification as *M. pneumoniae* from the genomic data was confirmed with 16S rRNA sequence analysis. The Illumina reads for all isolates were mapped against the reference chromosome of M129 (EMBL accession no. U00089) using SMALT (<http://www.sanger.ac.uk/resources/software/smalt/>) to identify single-nucleotide polymorphisms (SNPs), as previously described (39). Regions of recombination in the whole chromosomes of the isolates were analyzed using Genealogies Unbiased By recomBINations In Nucleotide Sequences (GUBBINS) (40).

**Phylogenetic analysis.** The locus sequences corresponding to each strain were concatenated head-to-tail for diversity analysis. Sequence analyses and tree construction were performed using MEGA 6.0. Neighbor-joining trees were constructed for each individual locus and concatenated sequences using Kimura's two-parameter model (24, 25). Maximum-likelihood trees were constructed for each individual locus using the Jukes-Cantor model of sequence evolution (26). Maximum-likelihood trees were constructed from concatenated sequences of the eight MLST loci using the generalized time-reversible (GTR) model of sequence evolution, with uniform rates of variation (27). Bootstrap analyses with 1,000 replicates were performed for every phylogenetic tree (28). Relatedness between STs was analyzed based on allelic profiles using eBURST version3. Maximum-likelihood trees were constructed from genomic sequences after the removal of areas of recombination. In total, 1,854 SNP sites were identified in comparison to the M129 reference chromosome. Three regions were predicted to contain SNP sites that had arisen by recombination, and these contained 28 SNP sites.

## RESULTS

**MLST of *M. pneumoniae*.** An initial examination of 10 gene targets in the two type strains M129 and FH and the genomic sequences of 35 *M. pneumoniae* clinical isolates identified variation, in the form of SNP differences, in 8 out of the 10 genes. The *recA* and *efp* genes were 100% conserved in all sequences analyzed and were therefore excluded from the MLST scheme. Genomic sequence analysis and additional PCR and sequencing of all eight targets in a further 20 clinical isolates resolved a total of 12 STs. The discriminatory typing ability for *M. pneumoniae* was 0.21 ST per isolate. The number of SNPs observed within each individual locus and the percentage of polymorphic sites are indicated in Table 3, with *pgm* having the highest number of SNPs (10 SNPs) and the highest percentage of polymorphic sites corrected for sequence length (0.93%). The number of alleles per locus ranged from two (*ppa*, *gyrB*, *gmk*, and *arc*) to four (*atpA*) (Table 3). Examination of the Hunter-Gaston diversity index (DI) (which ranges from 0.0 for no diversity to 1.0 for complete diversity) indicated moderate diversity between the STs (DI, 0.784; 95% confidence interval [CI], 0.716 to 0.852), with the greatest individual diversity shown in *pgm* (DI, 0.620; 95% CI, 0.566 to 0.674) and the lowest diversity shown in *arcC* (DI, 0.069; 95% CI, 0.000 to 0.158).

TABLE 1 Description of *M. pneumoniae* strains used in this study, their sequence type, allelic profile, and MLVA and P1 types

Strain	Yr of isolation	Country of isolation	Isolation site	ST	Allelic profile							MLVA type	P1 type <sup>a</sup>	
					<i>ppa</i>	<i>pgm</i>	<i>gyrB</i>	<i>gmk</i>	<i>glyA</i>	<i>atpA</i>	<i>arcC</i>			<i>adk</i>
M129 (ATCC 29342)	1969	USA	Unknown	1	1	2	1	1	1	3	2	1	4572	1
MPN135	1986	USA	Unknown	1	1	2	1	1	1	3	2	1	4572	V1
FH (ATCC 15531)	1944	USA	Sputum	2	2	3	2	2	2	4	1	1	3662	2
MPN007	1978	UK	Throat swab	2	2	3	2	2	2	4	1	1	NT <sup>a</sup>	2
MPN021	1983	UK	Unknown	2	2	3	2	2	2	4	1	1	3662	NT
MPN022	2010	UK	Sputum	2	2	3	2	2	2	4	1	1	3562	2c
MPN023	1983	UK	Sputum	2	2	3	2	2	2	4	1	1	3662	2
MPN101	1978	UK	Unknown	2	2	3	2	2	2	4	1	1	3562	2
MPN102	1981	UK	Brain frontal lobe	2	2	3	2	2	2	4	1	1	3662	2
MPN107	1983	UK	Sputum	2	2	3	2	2	2	4	1	1	3562	2
MPN114	1983	UK	Sputum	2	2	3	2	2	2	4	1	1	3662	NT
MPN117	1982	UK	Sputum	2	2	3	2	2	2	4	1	1	3562	2
MPN119	1982	UK	Sputum	2	2	3	2	2	2	4	1	1	3562	2
MPN121	1983	UK	Sputum	2	2	3	2	2	2	4	1	1	3662	2c
MPN123	1983	UK	Sputum	2	2	3	2	2	2	4	1	1	3662	2
MPN125	1983	UK	Sputum	2	2	3	2	2	2	4	1	1	3562	2
MPN126	1979	UK	Unknown	2	2	3	2	2	2	4	1	1	3662	2
MPN128	1976	USA	Unknown	2	2	3	2	2	2	4	1	1	3662	NT
MPN132	1982	UK	Sputum	2	2	3	2	2	2	4	1	1	3562	2
MPN133	1982	UK	Sputum	2	2	3	2	2	2	4	1	1	3662	2
MPN134	1982	UK	Sputum	2	2	3	2	2	2	4	1	1	3662	2
MPN005	1983	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN006	1982	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	NT
MPN013 <sup>b</sup>	2009	UK	Nose and throat swabs	3	1	2	1	1	1	3	1	1	4572	1
MPN014 <sup>b</sup>	2009	UK	Nose and throat swabs	3	1	2	1	1	1	3	1	1	4572	1
MPN015 <sup>b</sup>	2009	UK	Nose and throat swabs	3	1	2	1	1	1	3	1	1	4572	1
MPN016 <sup>b</sup>	2009	UK	Nose and throat swabs	3	1	2	1	1	1	3	1	1	4572	1
MPN017 <sup>b</sup>	2009	UK	Nose and throat swabs	3	1	2	1	1	1	3	1	1	4572	1
MPN020	1982	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	NT
MPN103	1982	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN105	1983	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN108	1983	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN109	1982	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	NT
MPN113	1967	UK	Unknown	3	1	2	1	1	1	3	1	1	4572	1
MPN116	1982	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN118	1996	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN120	1982	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN122	1982	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN136	1982	UK	Sputum	3	1	2	1	1	1	3	1	1	4572	1
MPN004	1981	UK	Sputum	4	2	1	2	2	2	4	1	1	3662	2
MPN104 <sup>c</sup>	1981	UK	Sputum	4	2	1	2	2	2	4	1	1	3662	2
MPN106 <sup>c</sup>	1981	UK	Sputum	4	2	1	2	2	2	4	1	1	3662	2
MPN110 <sup>c</sup>	1981	UK	Sputum	4	2	1	2	2	2	4	1	1	3662	2
MPN124	1981	UK	Sputum	4	2	1	2	2	2	4	1	1	3662	2
MPN131	1981	UK	Sputum	4	2	1	2	2	2	4	1	1	3662	2
MPN111	1968	UK	Unknown	5	1	2	1	1	1	2	1	1	4572	1
MPN011	1983	UK	Sputum	6	2	3	2	2	2	1	1	1	3662	1
MPN112	1983	UK	Sputum	6	2	3	2	2	2	1	1	1	3662	1
MPN127	1982	UK	Sputum	7	2	3	2	2	2	4	1	2	3662	2
MPN129	1983	UK	Sputum	8	2	3	2	2	2	4	1	3	3662	2
MPN130	1983	UK	Sputum	9	1	2	1	1	1	3	1	4	4572	1
MPN008	1981	UK	Sputum	10	2	1	2	2	2	4	1	2	3662	2
MPN018	1981	UK	Sputum	10	2	1	2	2	2	4	1	2	3662	2
MPN010	1983	UK	Sputum	11	1	2	1	1	3	3	1	1	3662	1
MPN003	1983	UK	Sputum	11	1	2	1	1	3	3	1	1	4572	1
MPN012	1981	UK	Brain cyst	11	1	2	1	1	3	3	1	1	3562	NT
MPN019	1983	UK	Sputum	12	2	2	1	1	3	3	1	4	4572	1

<sup>a</sup> NT, *M. pneumoniae* strain not classified by MLVA/P1 typing.<sup>b</sup> Strains isolated from the same patient.<sup>c</sup> Strains isolated from the same patient.

TABLE 2 MLST loci used in established bacterial MLST schemes also present in *M. pneumoniae*

Bacterial species	Presence of MLST locus <sup>a</sup>									
	<i>recA</i>	<i>ppa</i>	<i>pgm</i>	<i>gyrB</i>	<i>gmk</i>	<i>glyA</i>	<i>efp</i>	<i>atpA</i>	<i>arcC</i>	<i>adk</i>
<i>Bacillus cereus</i>					✓					
<i>Chlamydia trachomatis</i>						✓				
<i>Campylobacter jejuni</i>			✓			✓				
<i>Escherichia coli</i>	✓			✓						✓
<i>Enterococcus faecium</i>								✓		✓
<i>Haemophilus influenzae</i>	✓									✓
<i>Helicobacter pylori</i>		✓					✓	✓		
<i>Moraxella catarrhalis</i>		✓					✓			✓
<i>Neisseria meningitidis</i>			✓							✓
<i>Staphylococcus aureus</i>					✓				✓	
<i>Staphylococcus epidermidis</i>									✓	
<i>Streptococcus suis</i>	✓									
<i>Vibrio vulnificus</i>				✓						
<i>Yersinia pseudotuberculosis</i>										✓

<sup>a</sup> MLST loci were chosen based on the frequency of use in other bacterial MLST schemes (<http://www.mlst.net/>) and the presence of the gene in the published M129 and FH whole genomes.

Neighbor-joining and maximum-likelihood trees constructed from concatenated sequences of the eight loci for the 57 *M. pneumoniae* isolates (Fig. 1) illustrated two genetically distinct clusters, which were confirmed by eBURST examination of relatedness (Fig. 2). The two clusters, designated clonal complex 1 (CC1) and CC2, contained ST1, ST3, ST5, ST9, and ST11, and ST2, ST4, ST6, ST7, ST8, and ST10, respectively. ST12 located distal to the two main clusters; however, phylogenetic analysis revealed closer positioning to CC1. Neighbor-joining and maximum-likelihood trees were constructed for the eight loci individually (data not shown), and the topology of both neighbor-joining and maximum-likelihood trees was consistent for all loci and concatenated sequences.

Five homogenous strains (*M. pneumoniae* MPN13 to MPN17) originating from nose and throat swabs from the same patient with Stevens-Johnson syndrome had identical STs (ST3). Addi-

tionally, two clinical isolates (*M. pneumoniae* MPN104 and MPN106) originating from separate sputum samples from a patient with bronchopneumonia taken 4 days apart also had identical STs (ST4). This indicates a single clonal population responsible for infection in these cases.

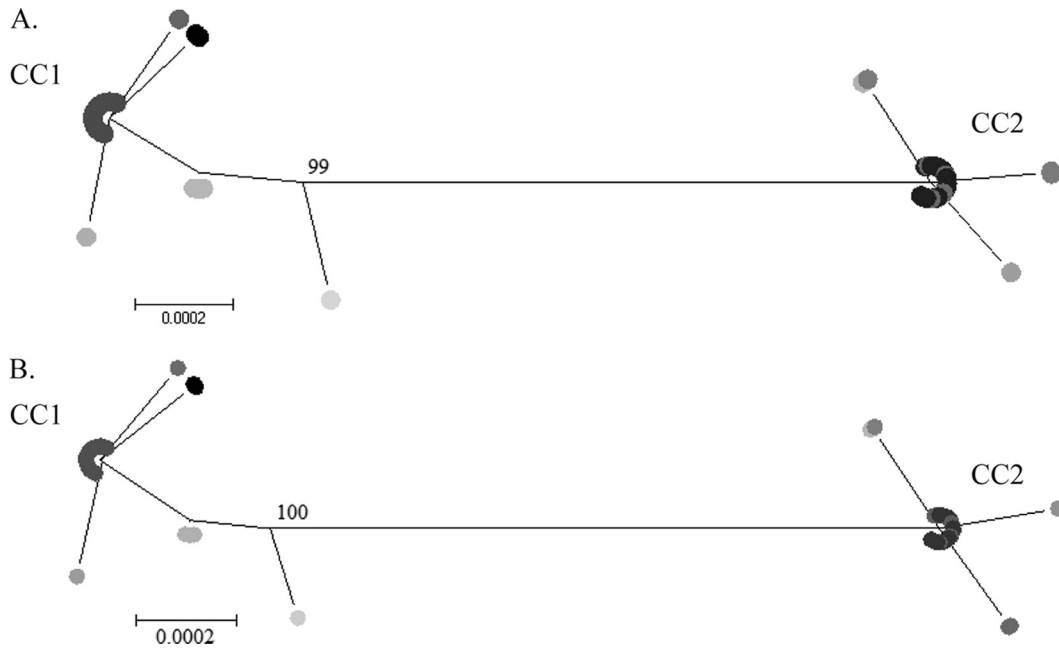
The possibility of synonymous sequence changes (indicating a pressure to conserve amino acid sequence and protein structure) was investigated by comparing the predicted translated sequences for each locus. Analysis of the deduced amino acid sequences of the eight loci for the 57 strains indicated that both synonymous and nonsynonymous SNPs occurred, of which approximately 44% resulted in an amino acid change. Nonsynonymous SNPs are highlighted in Fig. SA2 in the supplemental material. The amino acid sequences for ArcC, Gmk, and GyrB yielded homologous sequences for all ATs, numbering at two ATs for each locus. In comparison, Pgm analysis resulted in the largest number of non-

TABLE 3 Primer pairs developed in this study and variability of the different loci

Name	Direction <sup>a</sup>	Primer sequence (5'–3')	Amplicon size (bp)	MLST locus location	No. of alleles	No. of polymorphic sites	% polymorphic sites	Avg G+C content (%)	Hunter-Gaston diversity index <sup>b</sup>	95% confidence interval
<i>ppa</i>	F	CGCTGACCAAGCCTTCTAC	256	192–440	2	1	0.39	38.4	0.501	0.470–0.533
	R	CACTCCAAACTTTGCACTCCC								
<i>pgm</i>	F	AGCACCTTGCACGATGAAGA	1,072	456–1652	3	10	0.93	43.7	0.620	0.566–0.674
	R	CCTGCGCCTTCGTTAATTGG								
<i>gyrB</i>	F	TTGTCCCGACTTTACCGTG	429	524–952	2	2	0.47	39.9	0.505	0.482–0.528
	R	TGTTTTCGACAGCAAAGCGG								
<i>gmk</i>	F	GAGCGGTGTGGCAAAGTA	394	189–582	2	1	0.25	40.1	0.505	0.482–0.528
	R	TGCATCCTCGTCATTACGCTT								
<i>glyA</i>	F	CAGAGAACTATGTGAGTAGGGACA	676	74–749	3	2	0.30	45.6	0.560	0.493–0.627
	R	TGACAACCCGGAAAGACACC								
<i>atpA</i>	F	GTCGCTGATGGCATTGCTAAG	796	100–895	4	3	0.38	44.8	0.557	0.502–0.612
	R	CCAGTAAACGCGAGTGCAAG								
<i>arcC</i>	F	CCCCATCAAGCCGTGTA	570	304–873	2	1	0.18	45.5	0.069	0.000–0.158
	R	TTGGGCAATAATGCGCGTCT								
<i>adk</i>	F	GTAGCCAACACCACCGGATT	473	70–542	4	3	0.63	47.5	0.199	0.063–0.335
	R	ACGGTGTCTTCGTAAGCGT								

<sup>a</sup> F, forward; R, reverse.

<sup>b</sup> Hunter-Gaston diversity index (DI) ranges from 0.0 for no diversity to 1.0 for complete diversity.



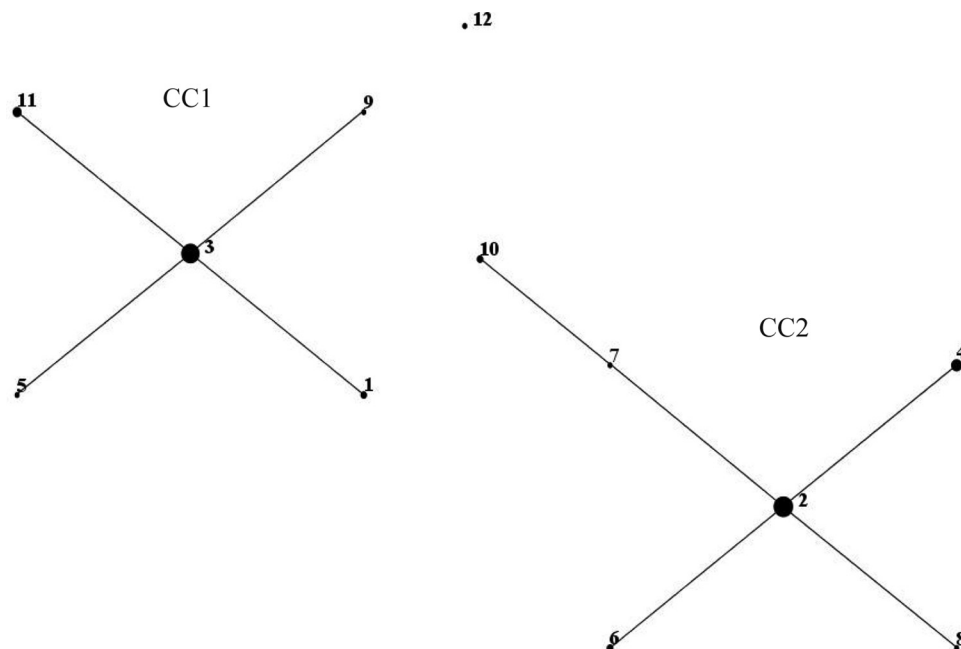
**FIG 1** Phylogenetic trees were constructed based on concatenated sequences of eight housekeeping loci for 12 unique STs using maximum-likelihood (A) and neighbor-joining (B) methods. Bootstrap support values of >70% are shown. STs are indicated by differential shading.

synonymous changes in the amino acid sequence, with four changes in the sequence between three ATs.

The MLST scheme was applied to the published complete genome sequences of *M. pneumoniae* available from NCBI for strains 309 (GenBank accession no. [NC\\_016807.1](#)), M129-B7 (GenBank accession no. [CP003913.2](#)), and M29 (GenBank accession no. [NZ\\_CP008895.1](#)) and to assembled contigs available from NCBI for strains PO1 (GCA\_000319655.1), PI 1428 (GCA\_

000319675.1), and 19294 (GCA\_000387745.1). These strains were determined to belong to ST2, ST1, ST3, ST2, ST1, and ST7, respectively.

The stability of each MLST locus was assessed in 10 *M. pneumoniae* isolates. Isolates were retyped following short-term passage (10 sequential subculture passages) in liquid medium. All loci were found to be completely stable, with no SNPs in comparison to the original isolate.



**FIG 2** eBURST version 3 was used to analyze the 12 unique STs resolved for all 57 *M. pneumoniae* isolates. Two main clonal complexes (CC) were defined. The size of each dot is proportional to the number of isolates included in the analysis for each ST.

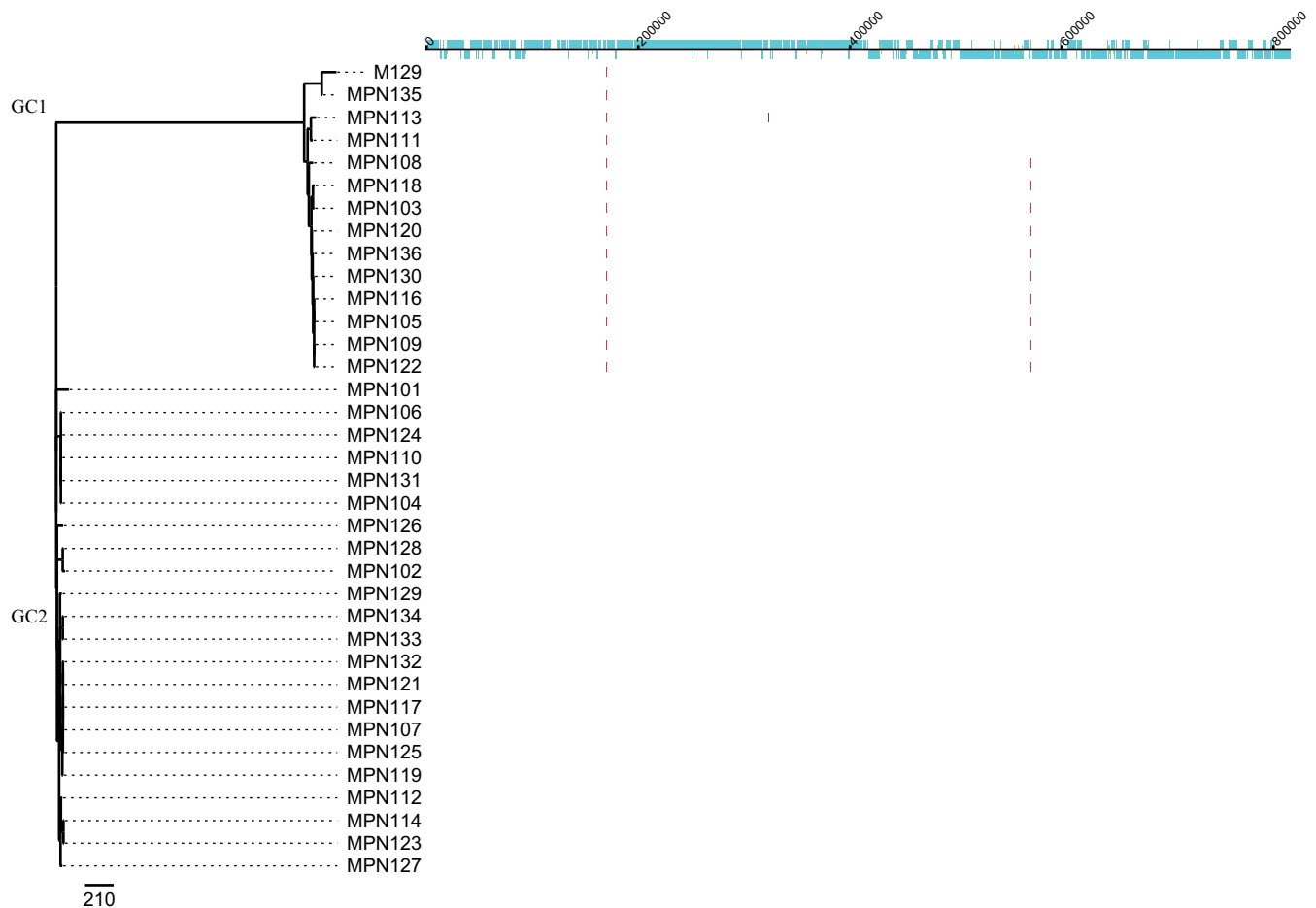


FIG 3 Prediction of recombination in the *M. pneumoniae* isolate chromosomes. Regions of variation in the genomes of the 35 clinical *M. pneumoniae* isolates and the type strain M129, which are predicted to have arisen by homologous recombination, are shown on the right. Red blocks indicate recombination predicted to have occurred on internal nodes, and blue indicates taxon-specific recombination. Isolates are ordered according to the phylogenetic tree displayed on the left. The track along the top of the figure displays the M129 chromosome and annotation, in which protein-coding sequences (CDS) are indicated in light blue.

**Genomic sequence analysis.** Three regions of SNPs were predicted to have arisen by homologous recombination in the chromosomes of the 35 clinical *M. pneumoniae* isolates (Fig. 3), one of which distinguished genomic clade 1 (GC1) from GC2, and the other two occurred within GC1. Area 1 was predicted to occur in all strains in GC1, area 2 in 10 strains, and a single strain, MPN113, had a single additional predicted area of recombination, area 3. Following removal of the predicted areas of recombination, two distinct genetic clades were identified, GC1 and GC2 (Fig. 3). Excellent parity was found using this method and concatenated MLST sequences, with all strains collocating to the corresponding CC and GC.

**Comparison to other typing methods.** There was no obvious link between the ST and the year when the strains were collected, the patient's age, and the sample origin; however, limited numbers of strains were available per year, and for some years, there were no strains. Indeed, multiple STs can be observed in a single year. Furthermore, ST was related to P1 type (Table 1), with the two most common STs, ST2 and ST3, containing strains that were P1 type 2 and P1 type 1, respectively. Additionally, this MLST scheme

was also comparable to MLVA typing. The two major clusters observed, CC1 and CC2, were directly linked to MLVA type: CC1 contained MLVA type 4572, whereas CC2 contained MLVA types 3662 and 3562. Each ST contained only one MLVA type, with the exception of ST2, which contained both 3662 and 3562, and ST11, which contained 4572, 3662, and 3562 (Table 1). Distributions of MLVA type, P1 type, and ST can be observed in Fig. 4, indicating that P1 type 1, MLST ST2, and MLVA types 3662 and 4572 occurred most frequently in the isolates tested.

In the isolates tested in this scheme, MLST was deemed to be more discriminatory than both MLVA typing and P1 typing, resulting in 0.21, 0.05, and 0.07 types per isolate, respectively. This was confirmed by examination of the Hunter-Gaston DI, which indicated higher discriminatory ability for the MLST scheme (DI, 0.784; 95% CI, 0.716 to 0.852) than that of the current MLVA scheme (DI, 0.633; 95% CI, 0.583 to 0.683) and P1 typing (DI, 0.567; 95% CI, 0.510 to 0.625).

**Online database.** An *M. pneumoniae* MLST online database was created for both MLST alleles, profile definitions, and isolate data (29) (<http://pubmlst.org/mpneumoniae>).

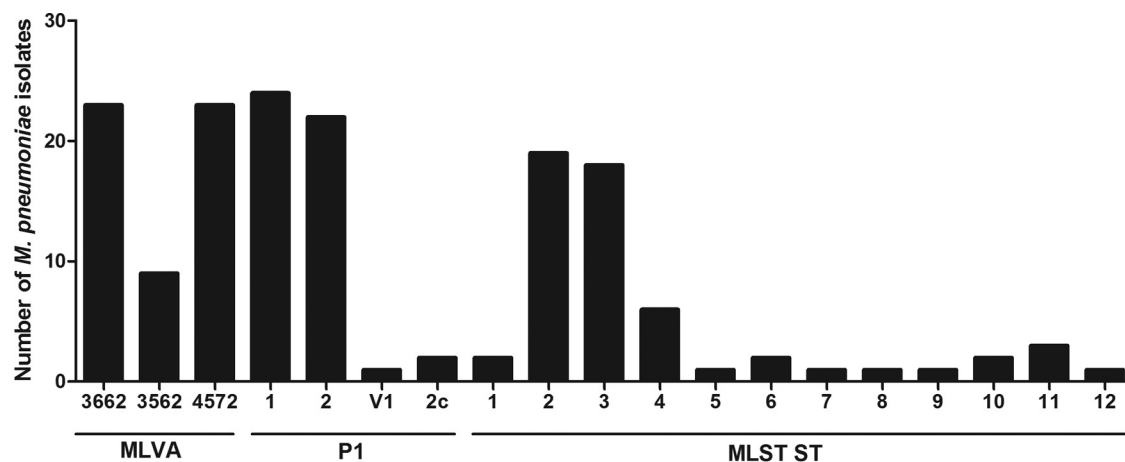


FIG 4 Distribution of MLVA, P1 type, and ST for 57 *M. pneumoniae* isolates (each group defined by lines below).

## DISCUSSION

MLST has been used to genotype several species of bacteria, including several *Mycoplasma* species: *Mycoplasma agalactiae*, *Mycoplasma bovis*, and *Mycoplasma hyorhinis* (30–32). This study has described the successful development of a novel *M. pneumoniae* MLST scheme to allow the characterization of clinical isolates. This scheme was successfully used to discriminate 55 clinical isolates of *M. pneumoniae* from British patients (with the exception of two U.S. isolates) within the reference laboratory collection, from respiratory and extrapulmonary sites, and the two type strains M129 and FH. Eight housekeeping genes were identified as suitable targets for the scheme, and these were used to genotype *M. pneumoniae* isolates by either PCR followed by sequencing or whole-genome sequence analysis. *gyrB* contains a quinolone resistance-determining region (QRDR) with documented *in vitro* mutations at amino acid positions 443, 464, and 483. The clinical use of quinolones may increase selective pressure *in vivo*, resulting in a high mutation rate (33). However, the *gyrB* locus sequence amplified in this MLST scheme is in a different region of the gene from the QRDR and is therefore considered a suitable MLST target. The stability of the eight loci was evaluated *in vitro* and was confirmed before and after 10 repeated passages of 10 strains in liquid medium. However, stability over a larger number of passages in liquid medium and an evaluation of stability using *in vitro* tissue culture were not assessed.

The discriminatory power of this MLST scheme with the eight loci was 0.784 for the collection of 57 isolates. In comparison, the Hunter-Gaston DI of the P1 typing method for the 57 isolates was 0.567, and the DI of the MLVA scheme was 0.633; therefore, this MLST scheme was more discriminatory for the isolates tested. However, it has been shown that the established MLVA method is more discriminatory than P1 typing (16), which was confirmed in this study. The allelic diversity of each of the MLST loci varied significantly at each locus, with the *pgm*, *glyA*, *atpA*, *gyrB*, *gmk*, and *ppa* loci being more discriminatory than the *adk* and *arcC* loci. The association of this set of markers with varied Hunter-Gaston DIs makes this MLST, in theory, more optimal for epidemiological studies than other existing methodologies.

Analysis of *M. pneumoniae* infection at the individual patient level was possible using this scheme. Multiple clinical isolates were available from two of 50 patients: five were from a patient with

Stevens-Johnson syndrome (MPN013 to MPN017), and two were from a patient with bronchopneumonia, taken 4 days apart. In both cases, the ST, MLVA type, and P1 type remained the same, indicating that a single clonal isolate was responsible for infection. Recurring infection or reinfection of *M. pneumoniae* can be determined using this scheme. Recurring infection would have the same ST as the original infection, whereas reinfection with *M. pneumoniae* would likely be a different ST. Genetic MLST instability in isolates might occur; however, in this study, this was not seen over 10 passages.

The eBURST analysis illustrates the relationship of STs on the basis of the number of MLST loci that differ between two STs. Analysis of this population modeling indicates that the two clusters, CC1 and CC2, differed by more than one locus, but within each cluster, the STs did not differ by more than one locus. Within a cluster, this highlights the homogenous nature of the *M. pneumoniae* species; however, a definitive split can be observed between the two clusters in both ST and MLVA type. A possible divergent clade with ST12 from CC1 is also apparent; however, more isolates need to be typed by this method to confirm this observation. Few typing methods have been able to detect significant differences between strains, including one previous attempt to subtype *M. pneumoniae* by MLST with housekeeping and structural genes (12, 15, 22). The previous MLST was determined to be not sufficiently discriminatory to be used for epidemiological purposes. However, the MLST scheme developed in this study was able to discriminate between *M. pneumoniae* isolates and resulted in two genetically distinct clusters, indicating significant differences between strains.

A comparison between genomic sequence analysis after the removal of predicted areas of recombination and phylogenetic analysis of concatenated MLST sequences showed similar topology and the same distinct genetic clustering. This indicates that this MLST scheme is representative of the genome and confirms that *M. pneumoniae* can be subdivided into two distinct genetic lineages.

Typing of clinical *M. pneumoniae* isolates is becoming increasingly important due to the global increase in *M. pneumoniae* infections and the increase in macrolide-resistant strains (34, 35). This scheme provides a more discriminatory method than both the MLVA and P1 typing methods currently in use, allow-

ing further and more detailed analysis of observed epidemic peaks of *M. pneumoniae* infection. Community outbreaks of pneumonia caused by *M. pneumoniae* have been described worldwide (36–38), and it would be interesting to evaluate this MLST scheme in such epidemic situations. The level of discrimination of this typing method and its usefulness in epidemic analysis should be confirmed by comparing outbreak-related strains to a set of control strains that were isolated from a similar time period and geographical area but that are not epidemiologically related. More severe or adverse infections with *M. pneumoniae* are seen in some patients. The reason for this is not clear; however, it can be postulated that this is due to specific microbe pathogenicity (identified through genetic markers) or variance in host susceptibility. This method could assist in determining if this is a strain-specific phenomenon. One advantage of MLST is that it is PCR based and does not require the growth of bacteria, which can be a lengthy process for *M. pneumoniae*. However, there is a large amount of sequencing required for this method, which can be laborious and expensive; therefore, its adaptation for widespread use directly on clinical specimens would be beneficial.

In conclusion, this study presents a robust MLST scheme that has proven discriminatory for *M. pneumoniae*, providing isolate characterization and a higher level of discrimination than MLVA and P1 typing methods. In addition, phylogenetic analysis of both STs and whole-genome sequence data revealed two genetically distinct clusters. Crucially, this scheme for *M. pneumoniae* is also supported by a public Web-based database (<http://pubmlst.org/mpneumoniae>).

#### ACKNOWLEDGMENTS

This work was funded by Public Health England. These studies were supported by funding initiatives by the National Institute for Social Care and Health Research (NISCHR) (research support from the Welsh Government) via the registered research group Microbial and Infection Translational Research Group (MITReG) and Children and Young Persons Research Network (CYPRN).

We declare no conflicts of interest.

#### REFERENCES

1. Waites K, Talkington D. 2004. *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin Microbiol Rev* 17:697–728. <http://dx.doi.org/10.1128/CMR.17.4.697-728.2004>.
2. Meyer Sauter P, Jacobs B, Spuesens E, Jacobs E, Nadal D, Vink C, van Rossum AMC. 2014. Antibody responses to *Mycoplasma pneumoniae*: role in pathogenesis and diagnosis of encephalitis? *PLoS Pathog* 10:e1003983. <http://dx.doi.org/10.1371/journal.ppat.1003983>.
3. Polkowska A, Harjunpää A, Toikkanen S, Lappalainen M, Vuento R, Vuorinen T, Kauppinen J, Flinck H, Lyytikäinen O. 2012. Increased incidence of *Mycoplasma pneumoniae* infection in Finland, 2010–2011. *Euro Surveill* 17:pii=20072. <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20074>.
4. Chalker VJ, Stocki T, Mentasti M, Fleming D, Sadler C, Ellis J, Bermingham A, Harrison TG. 2011. *Mycoplasma pneumoniae* infection in primary care investigated by real-time PCR in England and Wales. *Eur J Clin Microbiol Infect Dis* 30:915–921. <http://dx.doi.org/10.1007/s10096-011-1176-3>.
5. Howard LS, Sillis M, Pasteur MC, Kamath AV, Harrison BD. 2005. Microbiological profile of community-acquired pneumonia in adults over the last 20 years. *J Infect* 50:107–113. <http://dx.doi.org/10.1016/j.jinf.2004.05.003>.
6. Chalker V, Stocki T, Litt D, Bermingham A, Watson J, Fleming D, Harrison T. 2012. Increased detection of *Mycoplasma pneumoniae* infection in children in England and Wales, October 2011 to January 2012. *Euro Surveill* 17:pii=20081. <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20081>.
7. Jacobs E. 2012. *Mycoplasma pneumoniae*: now in the focus of clinicians and epidemiologists. *Euro Surveill* 17:pii=20084. <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20084>.
8. Liu J, Ai H, Xiong Y, Li F, Wen Z, Liu W, Li T, Qin K, Wu J, Liu Y. 2015. Prevalence and correlation of infectious agents in hospitalized children with acute respiratory tract infections in central China. *PLoS One* 10:e0119170. <http://dx.doi.org/10.1371/journal.pone.0119170>.
9. Rastawicki W, Kaluzewski S, Jagielski M, Gierczyński R. 1998. Epidemiology of *Mycoplasma pneumoniae* infections in Poland: 28 years of surveillance in Warsaw 1970–1997. *Euro Surveill* 3:99–100. <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=95>.
10. Tjhih JH, Dorigo-Zetsma JW, Roosendaal R, Van Den Brule AJ, Bestebroer TM, Bartelds AI, Vandenbroucke-Grauls CM. 2000. *Chlamydia pneumoniae* and *Mycoplasma pneumoniae* in children with acute respiratory infection in general practices in The Netherlands. *Scand J Infect Dis* 32:13–17. <http://dx.doi.org/10.1080/00365540050164155>.
11. Cousin-Allery A, Charron A, de Barbeyrac B, Frey G, Skov Jensen J, Renaudin H, Bebear C. 2000. Molecular typing of *Mycoplasma pneumoniae* strains by PCR-based methods and pulsed-field gel electrophoresis. Application to French and Danish isolates. *Epidemiol Infect* 124:103–111.
12. Dorigo-Zetsma JW, Dankert J, Zaat SAJ. 2000. Genotyping of *Mycoplasma pneumoniae* clinical isolates reveals eight P1 subtypes within two genomic groups. *J Clin Microbiol* 38:965–970.
13. Sasaki T, Kenri T, Okazaki N, Iseki M, Yamashita R, Shintani M, Sasaki Y, Yayoshi M. 1996. Epidemiological study of *Mycoplasma pneumoniae* infections in Japan based on PCR-restriction fragment length polymorphism of the P1 cytoadhesin gene. *J Clin Microbiol* 34:447–449.
14. Dumke R, Von Baum H, Luck PC, Jacobs E. 2010. Subtypes and variants of *Mycoplasma pneumoniae*: local and temporal changes in Germany 2003–2006 and absence of a correlation between the genotype in the respiratory tract and the occurrence of genotype-specific antibodies in the sera of infected patients. *Epidemiol Infect* 138:1829–1837. <http://dx.doi.org/10.1017/S0950268810000622>.
15. Dumke R, Lück PC, Noppen C, Schaefer C, von Baum H, Marre R, Jacobs E. 2006. Culture-independent molecular subtyping of *Mycoplasma pneumoniae* in clinical samples. *J Clin Microbiol* 44:2567–2570. <http://dx.doi.org/10.1128/JCM.00495-06>.
16. Dégrange S, Cazanave C, Charron A, Renaudin H, Bebear C, Bebear CM. 2009. Development of multiple-locus variable-number tandem-repeat analysis for molecular typing of *Mycoplasma pneumoniae*. *J Clin Microbiol* 47:914–923. <http://dx.doi.org/10.1128/JCM.01935-08>.
17. Dumke R, Jacobs E. 2011. Culture-independent multi-locus variable-number tandem-repeat analysis (MLVA) of *Mycoplasma pneumoniae*. *J Microbiol Methods* 86:393–396. <http://dx.doi.org/10.1016/j.mimet.2011.06.008>.
18. Zhao F, Liu G, Cao B, Wu J, Gu Y, He L, Meng F, Zhu L, Yin Y, Lv M, Zhang J. 2013. Multiple-locus variable-number tandem-repeat analysis of 201 *Mycoplasma pneumoniae* isolates from Beijing, China, from 2008 to 2011. *J Clin Microbiol* 51:636–639. <http://dx.doi.org/10.1128/JCM.02567-12>.
19. Chalker VJ, Pereyre S, Dumke R, Winchell J, Khosla P, Sun H, Yan C, Vink C, Bébérac C, ESCMID Study Group for *Mycoplasma* Infections. 2015. International *Mycoplasma pneumoniae* typing study: the interpretation of *Mycoplasma pneumoniae* multilocus variable-number tandem-repeat analysis. *New Microbes New Infect* 7:37–40. <http://dx.doi.org/10.1016/j.nmni.2015.05.005>.
20. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
21. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
22. Dumke R, Catrein I, Pirkil E, Herrmann R, Jacobs E. 2003. Subtyping of *Mycoplasma pneumoniae* isolates based on extended genome sequencing and on expression profiles. *Int J Med Microbiol* 292:513–525. <http://dx.doi.org/10.1078/1438-4221-00231>.
23. Pitcher D, Chalker VJ, Sheppard C, George RC, Harrison TG. 2006. Real-time detection of *Mycoplasma pneumoniae* in respiratory samples



- with an internal processing control. *J Med Microbiol* 55:149–155. <http://dx.doi.org/10.1099/jmm.0.46281-0>.
24. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
  25. Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120. <http://dx.doi.org/10.1007/BF01731581>.
  26. Jukes T, Cantor C. 1969. Evolution of protein molecules, p 21–132. *In* Munro H (ed), *Mammalian protein metabolism*. Academic Press, New York, NY.
  27. Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York, NY.
  28. Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791. <http://dx.doi.org/10.2307/2408678>.
  29. Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. <http://dx.doi.org/10.1186/1471-2105-11-595>.
  30. Manso-Silvan L, Dupuy V, Lysnyansky I, Ozdemir U, Thiaucourt F. 2012. Phylogeny and molecular typing of *Mycoplasma agalactiae* and *Mycoplasma bovis* by multilocus sequencing. *Vet Microbiol* 161:104–112. <http://dx.doi.org/10.1016/j.vetmic.2012.07.015>.
  31. Tocqueville V, Ferre S, Nguyen NH, Kempf I, Marois-Crehan C. 2014. Multilocus sequence typing of *Mycoplasma hyorhinis* strains identified by a real-time TaqMan PCR assay. *J Clin Microbiol* 52:1664–1671. <http://dx.doi.org/10.1128/JCM.03437-13>.
  32. Rosales RS, Churchward CP, Schnee C, Sachse K, Lysnyansky I, Catania S, Iob L, Ayling RD, Nicholas RA. 2015. Global multilocus sequence typing analysis of *Mycoplasma bovis* isolates reveals two main population clusters. *J Clin Microbiol* 53:789–794. <http://dx.doi.org/10.1128/JCM.01910-14>.
  33. Gruson D, Pereyre S, Renaudin H, Charron A, Bebear C, Bebear CM. 2005. *In vitro* development of resistance to six and four fluoroquinolones in *Mycoplasma pneumoniae* and *Mycoplasma hominis*, respectively. *Antimicrob Agents Chemother* 49:1190–1193. <http://dx.doi.org/10.1128/AAC.49.3.1190-1193.2005>.
  34. Diaz MH, Benitez AJ, Winchell JM. 2015. Investigations of *Mycoplasma pneumoniae* infections in the United States: trends in molecular typing and macrolide resistance from 2006 to 2013. *J Clin Microbiol* 53:124–130. <http://dx.doi.org/10.1128/JCM.02597-14>.
  35. Zhou Z, Li X, Chen X, Luo F, Pan C, Zheng X, Tan F. 2015. Macrolide-resistant *Mycoplasma pneumoniae* in adults in Zhejiang, China. *Antimicrob Agents Chemother* 59:1048–1051. <http://dx.doi.org/10.1128/AAC.04308-14>.
  36. Chen FQ, Yang YZ, Yu LL, Bi CB. 2015. Prevalence of *Mycoplasma pneumoniae*: a cause for community-acquired infection among pediatric population. *Niger J Clin Pract* 18:354–358. <http://dx.doi.org/10.4103/1119-3077.153247>.
  37. Klement E, Talkington DF, Wasserzug O, Kayouf R, Davidovitch N, Dumke R, Bar-Zeev Y, Ron M, Boxman J, Lanier Thacker W, Wolf D, Lazarovich T, Shemer-Avni Y, Glikman D, Jacobs E, Grotto I, Block C, Nir-Paz R. 2006. Identification of risk factors for infection in an outbreak of *Mycoplasma pneumoniae* respiratory tract disease. *Clin Infect Dis* 43:1239–1245. <http://dx.doi.org/10.1086/508458>.
  38. Meyer Sauter PM, Bleisch B, Voit A, Maurer FP, Rely C, Berger C, Nadal D, Bloemberg GV. 2014. Survey of macrolide-resistant *Mycoplasma pneumoniae* in children with community-acquired pneumonia in Switzerland. *Swiss Med Wkly* 144:w14041.
  39. Hsu LY, Harris S, Chlebowicz M, Lindsay J, Koh TH, Kristnan P, Tan TY, Hon PY, Grubb W, Bentley S, Parkhill J, Peacock S, Holden MTG. 2015. Evolutionary dynamics of methicillin-resistant *Staphylococcus aureus* within a healthcare system. *Genome Biol* 16:81. <http://dx.doi.org/10.1186/s13059-015-0643-z>.
  40. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43:e15. <http://dx.doi.org/10.1093/nar/gku1196>.