A multi-locus sequence typing scheme for the emerging respiratory pathogen *Mycoplasma amphoriforme*.

Lily Simpson[1], Joshua Macleod[1], Richard S Rowlands[1], Owen B Spiller[2], Charlotte Dixon[1], Jorgen S Jensen[3], Baharak Afshar[4], Helena M B Seth-Smith[5], Victoria J Chalker[6] & Michael L Beeton[1] on behalf of the Study Group for Mycoplasma and Chlamydia Infections (ESGMAC).

Affiliations:

[1]Microbiology and Infection Research Group, Department of Biomedical Sciences, Cardiff Metropolitan University, Cardiff, UK [2]Cardiff University, Division of Infection and Immunity, Department of Medical Microbiology, University Hospital of Wales, Cardiff, UK, [3]Research Unit for Reproductive Microbiology, Statens Serum Institut, Copenhagen, DK-2300, Denmark [4]Zoonotic and Acute Respiratory Section (ZARS), Respiratory and Vaccine Preventable Bacteria Reference Unit (RVPBRU), UK Health Security Agency, London, UK [5]Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland [6]NHS England, UK.

Running headline: *M. amphoriforme* MLST

Corresponding author: Michael L. Beeton. E-mail mbeeton@cardiffmet.ac.uk

Impact Statement:

The MLST scheme described in this study represents a tool for typing *M. amphoriforme* in future studies, allowing for a better understanding of transmission dynamics and the population structure of this emerging respiratory pathogen.

Abstract

Mycoplasma amphoriforme is an emerging respiratory pathogen for which little is known about the population structure or transmission dynamics. In this study, we developed the first multilocus sequence typing (MLST) scheme for *M. amphoriforme* and applied it to a previous genomic data set. The genomes of seven *M. amphoriforme* isolates from the UK and Denmark were sequenced and used to develop the MLST scheme based on loci used for previous *Mycoplasma* MLST schemes. The resulting MLST scheme consisted of four loci (*gyrB*, *atpG*, *uvrA* and *rpoB*) and was applied to 20 previously sequenced genomes obtained from the UK and France/Tunisia. From the 27 sequences examined, 13 sequence types were identified. A phylogenetic tree of concatenated sequences showed a comparable topology to a previously described tree based on whole genome data. Additionally, the MLST scheme corroborated the previous suggestion of possible healthcare-associated transmission of *M. amphoriforme* between two separate patients. The MLST scheme gave a population structure analysis comparable to previous whole-genome-based analyses.

Key Words:

Whole-genome sequencing, pathogen typing, common variable immunodeficiency, emerging pathogen, respiratory tract infection

Introduction:

Mycoplasma amphoriforme was first proposed as an emerging respiratory pathogen in 1999 when isolated from respiratory secretions from a patient with primary antibody deficiency (PAD) (Webster *et al.* 2003). Later studies, with larger sample numbers, determined the prevalence of *M. amphoriforme* to be substantially higher among immunocompromised patients (19%-24%) as compared to immunocompetent patients (5 – 6.3%) (Ling *et al.* 2014; Rehman *et al.* 2021). Patients tend to present with symptoms suggestive of respiratory tract infection, and on rare occasions, *M. amphoriforme* has been identified among patients with sepsis in the absence of other known pathogens (Pereyre *et al.* 2010). Similarly to other mycoplasmas, treatment options are limited due to the lack of a cell wall, and acquired antibiotic resistance has been noted among *M. amphoriforme* isolated in the UK and Denmark (Day *et al.* 2022). Little is known regarding the mechanism of transmission, although past reports have suggested healthcare-associated transmission as well as infection among siblings (Gillespie *et al.* 2014; Katsukawa *et al.* 2016).

Molecular typing of pathogens is essential to understanding population structures for epidemiological studies and gaining insight into transmission dynamics in situations of outbreaks. Multilocus sequence typing (MLST) is one method which has seen substantial uptake since its first application for *Neisseria meningitidis* (Maiden *et al.* 1998). Owing to its ease of application on clinical material via PCR and sequencing of select housekeeping genes, as well as its reproducibility across laboratories and unambiguous results, the highly discriminatory MLST method has been developed for many clinically relevant human mycoplasmas (Brown *et al.* 2015; Jironkin *et al.* 2016; Jolley, Bray and Maiden 2018).

Here, we describe the development of the first typing scheme for the emerging respiratory pathogen *M. amphoriforme*. We then applied this typing method to previously published genome sequences of *M. amphoriforme* to assess its effectiveness in determining possible healthcare-associated transmission of infection.

Materials and methods

Mycoplasma amphoriforme strains and culture conditions.

Seven *M. amphoriforme* isolates, which were described previously by Day *et al.*, (Day *et al.* 2022) were used for the genomic sequencing and subsequent selection of housekeeping genes for the MLST design. These included the A39 type strain (NCTC 11740),(Webster *et al.* 2003; Pitcher *et al.* 2005) A55 and A84 (Pitcher *et al.* 2005). Isolate H04 was isolated from a bronchial alveolar lavage obtained from a 48-year-old male patient admitted on a cardiothoracic ward. Isolate H29 was isolated from a 33-year-old female with common variable immunodeficiency (CVID), recurrent chest infections and receiving a prolonged course of prophylactic azithromycin. No clinical information was available for isolate A70. The final isolate, M5572, was previously isolated in Denmark from the sputum of a patient with bronchitis (Pitcher *et al.* 2005). All isolates were grown statically in Mycoplasma Liquid Medium (MLM) (Mycoplasma Experience, UK) at 37°C under normoxic conditions until signs of a colour change in the medium from orange to yellow, indicating growth of viable organisms.

Whole genome sequencing and assembly

Fifty-millilitre cultures were grown in MLM until a colour change was noted. Cells were collected by centrifugation at 4800 xg for 1 h and resuspended in 400 µL of sterile distilled water for DNA extraction using the Qiagen EZ1 Advanced XL automated extractor utilizing the EZ1 DSP virus kit as per the manufacturer's instructions. Genomic sequencing was undertaken using a Nextera XTv2 library preparation kit with V3 chemistry on an Illumina MiSeq platform. Sequence read QC and assembly into contigs were as described elsewhere (Sands *et al.* 2021). Short reads were trimmed with trimgalore and the resulting reads assessed using fastqc(Andrews 2010). *De novo* genome assembly was performed using Shovill (v0.9.0)(Seemann *et al.*). Additional analysis was performed on isolates using Geneious Prime (v2025.1.2; Biomatters ltd. New Zealand).

MLST design

To assist in identifying suitable genes for the MLST scheme, housekeeping genes that were part of current MLST schemes for other human and animal *Mycoplasma* spp. present on the PubMLST website(Jolley, Bray and Maiden 2018), were compiled and visualised using the ComplexHeatmap package (Gu, Eils and Schlesner 2016). Initially, 11 genes were identified as potential targets for MLST, of which *atpG*, *gyrB*, *rpoB*, and *uvrA* were finally selected. To ensure primers were designed to anneal within conserved regions of each target gene, the full-length sequence for each gene was extracted from the WGS sequence data for each of the seven isolates. These sequences were aligned using Geneious Prime, and PCR primers were designed using the Design New Primer function.

To confirm the specificity of the primers, PCRs were undertaken on DNA isolated from *M. amphoriforme* A39 (NCTC 11740) (Table 1). PCRs included GoTaq G2 Green Master Mix (Promega, UK) with a final $MgCl_2$ concentration of 1.5 mM and 0.2 µM for each of the forward and reverse primers. Thermocycling conditions were as follows: 95°C for 3 minutes, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at the specified temperature in Table 1 for 30 s and extension at 72°C for 1 minute, followed by a final extension at 72°C for 5 min. Amplicons were then run on a 2% w/v agarose gel and visualised with ethidium bromide and a UV transilluminator to confirm the expected product sizes.

Designation of sequence types and drawing of phylogenetic trees.

To determine sequence types (STs) with the new MLST scheme, an online database hosted on PubMLST for *M. amphoriforme* was created https://pubmlst.org/organisms/mycoplasma-amphoriforme. Sequences for the seven isolates from this study, in addition to 20 sequenced from Gillespie *et al.*,(Gillespie *et al.* 2014) available under accession number ERP000340, were uploaded for ST determination. To assess the relatedness between sequences and allow for comparison to previously published trees for *M. amphoriforme*, a phylogenetic tree was generated as follows: Mafft

v7.407 was used to align the four loci across the 27 sequences using default settings (Katoh *et al.* 2002). Seqkit v2.10.0 was used to concatenate the alignments(Shen, Sipos and Zhao 2024). Phylogenetic analysis was performed on the resulting alignment using RAxML-NG v1.2.2 with the GTR+G substitution model, achieving bootstrap convergence after 3700 replicates. The resulting tree was visualized and annotated using iTOL.(Stamatakis 2014; Letunic and Bork 2021).

Results and Discussion

Assessment of MLST schemes for mycoplasma and design of a novel MLST scheme for *M. amphoriforme*.

The PubMLST database was examined for available MLST schemes for *Mycoplasma* spp.. Inclusive of the new *M. amphoriforme* scheme, 15 *Mycoplasma* MLST schemes are now listed on PubMLST, encompassing those associated with human (n = 4) and non-human (n = 11) hosts (Figure 1). A total of 42 loci were noted across the various species, with the number of loci used across species ranging from 3 – 8. Based on the data from the 14 schemes available prior to ours, initially, 11 genes were identified as potential targets for MLST, including ADK, *atpG*, *dnaA*, *fusA*, *gltX*, *gmk*, *gyrB*, *rpoB*, *ung*, *uvrA*, and *valS*; however, only *atpG*, *gyrB*, *rpoB*, and *uvrA* were taken forward for the MLST scheme due to difficulties in identifying suitable conserved regions for annealing across the seven genome sequences initially available and then subsequent PCR amplification. The *uvrA* forward primer was retrospectively amended after assessing the sequences available from Gillespie *et al*.,(Gillespie *et al.* 2014) to include an A/G degenerate base to accommodate SNPs in isolates Ma3663 and Ma4072.

Initial assessment of the MLST scheme

The diversity among sequences was initially examined across the seven isolates sequenced in this study. The number and percentage of polymorphic sites at each locus varied dramatically depending on whether sequences from isolate M5572 were included or omitted: 28 (4.35%) vs 2 (0.31%) for *atpG*, 5 (0.95%) vs 1 (0.19%) for *gyrB*, 16 (3.6%) vs 2 (0.45%) for *rpoB* and 21 (2.81%) vs 1 (0.13%) for *uvrA* (Table 1). The Hunter-Gaston diversity index values indicated that *rpoB* gave the greatest discriminatory power (0.81) followed by *atpG* (0.67), *gyrB* (0.52) and *uvrA* (0.52).

Assigning sequence type to *M. amphoriforme* sequences using the new MLST scheme.

An online database for *M. amphoriforme* ST designation was created at https://pubmlst.org/organisms/mycoplasma-amphoriforme (Jolley, Bray and Maiden 2018). To explore the utility of the new MLST scheme, we determined the sequence types from the sequences generated in this study (7) in addition to 20 additional sequence data sets, which were available through ERP000340 / PRJEB2281 (Table 2). From these data, allelic variants observed were *atpG* (6), *gyrB* (5), *rpoB* (6) and *uvrA* (7), resulting in 13 different sequence types, with ST1 as the most common ST. Phylogenetic analysis of the concatenated sequences identified a relationship between ST and the patient from which the isolate was obtained (Figure 2). Only one ST was isolated from each patient, with the exception of patient 8, for which two separate ST were isolated (ST1 and ST4).

Multilocus sequence typing has been adopted as a valuable method for typing numerous species of bacteria, including many human and animal mycoplasmas. In this study, we successfully developed the first MLST scheme for the emerging respiratory pathogen *M. amphoriforme,* and applied it to available data sets, allowing phylogenetic analysis of highly diverse strains which were not included in previous studies and further demonstrating potential healthcare-associated transmission of isolates (Gillespie *et al.* 2014).

Previous phylogenetic analysis undertaken on the whole genome data sets by Gillespie *et al*. noted three isolates (Ma3663, Ma4072 and Ma4176), were highly diverse relative to the rest of the strains and therefore omitted from the phylogenetic analysis (Gillespie *et al.* 2014). Using our MLST scheme, we were able to include these in the phylogenetic analysis and confirmed this high level of diversity, which was like that seen in M5572. This level of diversity is related to the geographical location of isolation for these *M. amphoriforme*, with Ma3663, Ma4072 and Ma4176 having been isolated in France/Tunisa (Gillespie *et al.* 2014) and M5572, originally isolated in Denmark (Pitcher *et al.* 2005), and the remaining sequences obtained from UK isolates (Gillespie *et al.* 2014; Day *et al.* 2022).

The phylogenetic analysis undertaken by Gillespie *et al*., and the data generated from the four loci MLST designed in this study show strikingly similar topologies of the resulting trees. This result helps to verify the utility of this MLST in describing phylogenetic relationships. This concordance between the WGS data and our MLST is further exemplified by the hypothesised nosocomial transmission between patients 1 and 8, as discussed by Gillespie in their study. Strain IM117-1 was isolated in 2002 from patient 8, whereas isolates 030-1 and 772-1, which were isolated in 2004 and 2005, from patient 8 were more closely related to the isolates M31a, M219a, M279a, A39, 049-1 and 814, recovered from patient 1. It was noted that these two patients had been present at the same clinic on at least two occasions, suggesting transmission of *M. amphoriforme* from patient 1 to patient 8. Our data corroborate these findings as demonstrated by the relationship in the phylogenetic tree, as well as the original isolate IM117-1 from patient 8 being assigned as ST4, with later isolates from this patient being ST1 which is the same ST as that identified from patient 1. With reports of *M. amphoriforme* infection among siblings (Katsukawa *et al.* 2016), our MLST will be important in understanding transmission dynamics and investigating the relationship in future infections.

Only two publications have documented macrolide-resistant *M. amphoriforme* (MRMA). The first case was from a 32-year-old female with a diagnosis of common variable immunodeficiency (CVID) in which two azithromycin-resistant isolates (770-1 and 770-a), as determined by the presence of a macrolide-

resistance mutation in domain V of the 23S rRNA, were recovered in the UK (Gillespie *et al.* 2014). The second isolate, H29, was isolated in the UK from a 33-year-old female with CVID and a history of recurrent chest infections and prolonged use of prophylactic azithromycin. Macrolide resistance in this instance was confirmed via both phenotypic and genotypic methods (Day *et al.* 2022). Interestingly, isolates 770-1, 770-a and H29 all cluster together on the phylogenetic tree and were designated as ST6, along with a azithromycin susceptible isolate A84. In the absence of full clinical details, we hypothesise that the isolates from the work of Gillespie and Day may have been recovered from the same patient at different time points, therefore explaining the association on the tree. Alternatively, this may represent a macrolide-resistant lineage in circulation among this patient group.

A limitation of this study is the small number of isolates and whole genome sequence data sets available to apply to the MLST scheme. As awareness of this pathogen increases, this tool will be available to contextualise the population structure and potentially understand the role of healthcare and household transmission. As with all MLST schemes, only a small subset of the genome is interrogated. Typing methods, such as core-genome MLST (cgMLST) or whole-genome MLST (wgMLST), offer a vastly greater degree of granular detail on the population structure, but this approach often relies on the availability of bacterial isolates to prepare DNA for WGS. A target enrichment-based approach to WGS of *M. amphoriforme* has been developed (Dennis *et al.* 2022), which allows for whole genome sequencing directly from a patient DNA extract without the need for an isolate, but until this methodology becomes readily accessible, the four-gene MLST developed described here represents a more affordable and accessible option for typing *M. amphoriforme* DNA.

Conclusion

In conclusion, we describe the first MLST scheme for the emerging respiratory pathogen, *M. amphoriforme*. This four gene-based typing system will allow for a better understanding of the population structure of *M. amphoriforme* isolated from patients in the future.

Funding: This work was funded by the Applied Microbiology International Summer studentship awarded to L.S and M.L.B. M.L.B also receives funding from the European Society of Clinical Microbiology and Infectious Disease (ESCMID) Study Group Collaboration Grant.

Conflict of interest statement:

Nothing to declare

Data Availability Statement:

The data underlying this article will be shared on reasonable request to the corresponding author.

Author contribution:

M.L.B conceived the study idea. M.L.B, B.A, V.J.C, J.S.J and O.B.S carried out sample preparation for WGS. L.S. undertook the initial assessment and design of the MLST scheme and primers, which were also optimised for use with PCR and sequencing with assistance from R.S.R. and M.L.B. L.S, H.S.-S, J.M, C.D and M.L.B undertook the data analysis. J.M determined the ST via PubMLST. M.L.B, H.S-S, J.M and C.D drafted the manuscript. All authors approved the manuscript for submission.

234

235 References

236 Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

237 Brown RJ, Holden MTG, Spiller OB *et al.* Development of a Multilocus Sequence Typing Scheme for
238 Molecular Typing of Mycoplasma pneumoniae. *J Clin Microbiol* 2015;**53**:3195–203.

239 Day J, Afshar B, Rowlands RS *et al.* Phenotypic and genotypic antimicrobial susceptibility patterns of
240 the emerging human respiratory pathogen Mycoplasma amphoriforme isolated from the UK and
241 Denmark. *J Antimicrob Chemother* 2022;**77**:3126–9.

242 Dennis TPW, Mable BK, Brunelle B *et al.* Target-enrichment sequencing yields valuable genomic data
243 for challenging-to-culture bacteria of public health importance. *Microb Genom* 2022;**8**:836.

244 Gillespie SH, Ling CL, Oravcova K *et al.* Genomic Investigations Unmask Mycoplasma amphoriforme, a
245 New Respiratory Pathogen. 2014, DOI: 10.1093/cid/ciu820.

246 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional
247 genomic data. *Bioinformatics* 2016;**32**:2847–9.

248 Jironkin A, Brown RJ, Underwood A *et al.* Genomic determination of minimum multi-locus sequence
249 typing schemas to represent the genomic phylogeny of Mycoplasma hominis. *BMC Genomics*
250 2016;**17**, DOI: 10.1186/s12864-016-3284-z.

251 Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the
252 PubMLST.org website and their applications. *Wellcome Open Res* 2018;**3**, DOI:
253 10.12688/WELLCOMEOPENRES.14826.1/DOI.

254 Katoh K, Misawa K, Kuma KI *et al.* MAFFT: a novel method for rapid multiple sequence alignment based
255 on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66.

256 Katsukawa C, Asai S, Mizutani K *et al.* Novel Isolation of Mycoplasma amphoriforme from a Pediatric
257 Patient with Protracted Bronchitis in Japan. *Jpn J Infect Dis* 2016;**69**:450–1.

258 Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and
259 annotation. *Nucleic Acids Res* 2021;**49**:W293–6.

260 Ling CL, Oravcova K, Beattie TF *et al.* Tools for detection of Mycoplasma amphoriforme: A primary
261 respiratory pathogen? *J Clin Microbiol* 2014;**52**:1177–81.

Maiden MCJ, Bygraves JA, Feil E *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 1998;**95**:3140–5.

Pereyre S, Renaudin H, Touati A *et al.* Detection and susceptibility testing of Mycoplasma amphoriforme isolates from patients with respiratory tract infections. *Clinical Microbiology and Infection* 2010;**16**:1007–9.

Pitcher DG, Windsor D, Windsor H *et al.* Mycoplasma amphoriforme sp. nov., isolated from a patient with chronic bronchopneumonia. *Int J Syst Evol Microbiol* 2005;**55**:2589–94.

Rehman SU, Day J, Afshar B *et al.* Molecular exploration for Mycoplasma amphoriforme, Mycoplasma fermentans and Ureaplasma spp. in patient samples previously investigated for Mycoplasma pneumoniae infection. *Clinical Microbiology and Infection* 2021;**27**:1697.e1-1697.e5.

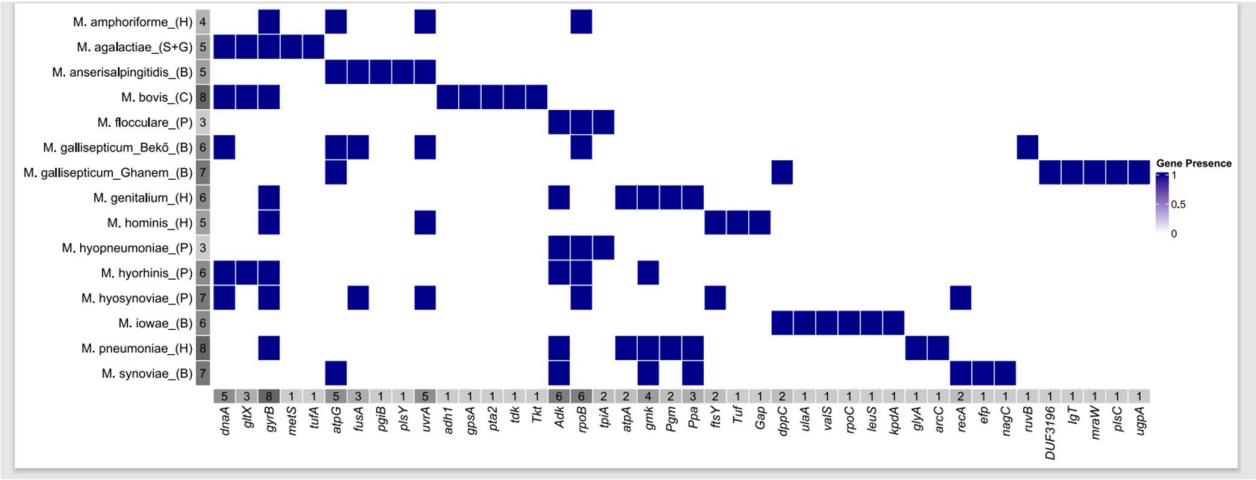
Sands K, Carvalho MJ, Portal E *et al.* Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries. *Nature Microbiology 2021 6:4* 2021;**6**:512–23.

Seemann T, Edwards R, Goncalves da Silva A *et al. De Novo Genome Assembly Was Performed Using Shovill (v0.9.0)*.

Shen W, Sipos B, Zhao L. SeqKit2: A Swiss army knife for sequence and alignment processing. *iMeta* 2024;**3**:e191.

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

Webster D, Windsor H, Ling C *et al.* Chronic bronchitis in immunocompromised patients: Association with a novel Mycoplasma species. *European Journal of Clinical Microbiology and Infectious Diseases* 2003;**22**:530–4.



289

290 Figure 1. *Mycoplasma* MLST schemes currently available via PubMLST. Numbers on the vertical axis
291 correspond to the number of loci in each scheme. Numbers on the horizontal axis correspond to the
292 number of schemes in which this loci is used. Lettering in brackets represents the primary host for
293 each Mycoplasma H = human, S = sheep, G = goat, B = bird, P = pig, and C = cattle.

294

295

296 Figure 2. A tree of *M. amphoriforme* concatenated MLST loci sequences. Numbers in the first column

297 represent sequence type as determined by our MLST scheme, the second column represents patient

298 numbers as per Gillespie et al., 2014., and yellow boxes denote isolates sequenced in this study.

299

300

Table 1. Primers developed and variability between different loci from the seven *M. amphoriforme* sequenced in this project.

Gene name	Direction	Primer Sequence (5′ – 3′)	Annealing temperature	Amplicon size	MLST locus location	Number of alleles	Number of polymorphic sites (%)	Number of polymorphic sites excluding M5572 (%)	Average G+C content %[3]	Hunter-Gaston diversity index
atpG	F	GTAAGTTTCGCCGCATGCTT	60	644	98-741	3	28 (4.35)	2 (0.31)	34.8	0.67
	R	TTCCATTGCGTTTTGCCTGC	60							
gyrB	F	ACACCCCAATCCTGAGTTTGT	60	528	1035-1562	3	5 (0.95)	1 (0.19)	38.1	0.52
	R	CCATCCACATCAGCATCGGT	60							
rpoB	F	TGTTGCACCCAAACAAGTCG	60	444	2400-2843	4	16 (3.6)	2 (0.45)	39	0.81
	R	CCGCGTCAGGTTGTAAATGC	60							
uvrA	F	CCCAGACRATTCAACAGAT	53	747	413-1159	3	21 (2.81)	1 (0.13)	33.9	0.52
	R	CCACACCCTCAACAAYT	53							

Table 2. Description of *M. amphoriforme* allelic profiles and sequence type.

Strain	Country of isolation	Allelic profile				ST
		atpG	*gyrB*	*rpoB*	*uvrA*	
A39 (NCTC 11740)	UK	1	1	1	1	1
030-1	UK	1	1	1	1	1
049-1	UK	1	1	1	1	1
814	UK	1	1	1	1	1
M31a	UK	1	1	1	1	1
M219a	UK	1	1	1	1	1
772-1	UK	1	1	1	1	1
M279a	UK	1	1	1	1	1
m246a	UK	2	1	2	1	2
159-1	UK	2	1	2	1	2
A55	UK	1	1	2	1	3
048a	UK	1	1	2	1	3
24	UK	1	1	2	1	3
IM117-1	UK	1	1	2	2	4
M142A	UK	1	1	3	1	5
770-1	UK	3	1	3	1	6
770-a	UK	3	1	3	1	6
A84	UK	3	1	3	1	6
H29	UK	3	1	3	1	6
M42	UK	3	1	3	1	6
H04	UK	3	1	2	3	7
Ma4526a	France/Tunisia	3	1	4	4	8
M5572	Denmark	4	2	5	5	9
Ma4176	France/Tunisia	5	3	6	6	10
Ma3663	France/Tunisia	3	1	4	7	11
A70	UK	3	4	3	1	12
ma4072	UK	6	5	4	7	13