Multilocus sequence typing of *Mycoplasma agalactiae*

L. McAuliffe,¹ F. Gosney,¹ M. Hlusek,¹ M. L. de Garnica,² J. Spargser,³ M. Kargl,³ R. Rosengarten,³ R. D. Ayling,¹ R. A. J. Nicholas¹ and R. J. Ellis¹

Correspondence

R. J. Ellis

r.ellis@vla.defra.gsi.gov.uk

¹Veterinary Laboratories Agency (Weybridge), Woodham Lane, Addlestone, Surrey KT15 3NB, UK

²Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León, Spain

³Institute of Bacteriology, Mycology and Hygiene, University of Veterinary Medicine, A-1210 Vienna, Austria

Mycoplasma agalactiae is the main cause of contagious agalactia, a serious disease of sheep and goats, which has major clinical and economic impacts. We have developed a multilocus sequence typing (MLST) scheme using the sequenced genomes of the *M. agalactiae* strains PG2 and 5632. An MLST scheme based on the genes *gltX*, *metS*, *gyrB*, *tufA* and *dnaA* was designed and in total 3468 bp of sequence were analysed for each strain. MLST offers a highly discriminatory typing method for *M. agalactiae* and was capable of subdividing 53 strains into 17 distinct sequence types, largely according to geographical origin. MLST detected unexpected diversity in recent isolates from Spain, identifying two novel outliers, and enabled typing of novel Mongolian isolates for the first time. Genetic diversity in the sequenced regions was largely due to mutation, with recombination playing a much smaller role. A web-accessible database has been set up for this MLST scheme for *M. agalactiae*: <http://pubmlst.org/magalactiae/>. MLST offers a robust, objective molecular epidemiological tool for *M. agalactiae* that enables interlaboratory comparison of data.

Received 15 November 2010

Accepted 28 February 2011

INTRODUCTION

Known for nearly 200 years, contagious agalactia is primarily a disease of dairy sheep and goats characterized by mastitis, arthritis and keratoconjunctivitis (OIE, 2003). It is mainly caused by *Mycoplasma agalactiae* but in recent years, *M. capricolum* subsp. *capricolum*, *M. mycoides* subsp. *capri* and *M. putrefaciens* have also been isolated in many countries from goats with mastitis, arthritis and, occasionally, respiratory disease. The clinical signs of the infections are sufficiently similar to those of contagious agalactia for the Office International des Epizooties to include them as causes of this listed disease (OIE, 2003).

M. agalactiae is prevalent worldwide but causes particular problems around the Mediterranean basin, where it has a major clinical and economic impact on the small-ruminant dairy industry (Bergonier *et al.*, 1997). In most cases, infected hosts spontaneously recover from acute clinical signs within a few weeks but develop a chronic infection accompanied by shedding of *M. agalactiae* in milk and/or other body secretions for years without presenting any clinical signs; these asymptomatic carriers can transmit the

bacteria to other susceptible animals and cause acute disease (Bergonier *et al.*, 1997).

As *M. agalactiae* shows differing prevalence across the world, and as it is currently absent from some countries, notably the UK, there is a pressing need for molecular epidemiological techniques which enable a high degree of strain differentiation, allowing the tracing of the source of disease outbreaks. Early studies showed that *M. agalactiae* was largely homogeneous, with PFGE analysis revealing no variation in over 80 Italian isolates although some degree of antigenic variation was evident (Solsona *et al.*, 1996; Tola *et al.*, 1996, 1997). Insertion sequence analysis of *M. agalactiae* isolates of wide geographical origin found some degree of variation, with some isolates containing and others lacking the IS element ISMag, although subtyping isolates was not possible (Pilo *et al.*, 2003).

Recently the genome of the *M. agalactiae* type strain PG2 has been sequenced (Sirand-Pugnet *et al.*, 2007), making it highly amenable to analysis using sequence-based typing methods such as variable number of tandem repeats (VNTR) analysis and multilocus sequence typing (MLST). We have previously described the identification and characterization of VNTRs within the *M. agalactiae* PG2 genome, the selection of four VNTRs which showed most

Abbreviations: HR, homologous recombination; MLST, multilocus sequence typing; ST, sequence type; VNTR, variable number of tandem repeats.

intraspecific variation, and their use as a molecular epidemiological tool for the analysis of 88 *M. agalactiae* strains (McAuliffe *et al.*, 2008). VNTR analysis revealed unexpected diversity among *M. agalactiae* strains, with 14 different profiles seen (McAuliffe *et al.*, 2008).

MLST utilizes the sequences of internal fragments of several housekeeping genes for unambiguous characterization of bacterial isolates. For each gene, different sequences are assigned as distinct alleles and, for each isolate, the alleles at each of the loci define the allelic profile or sequence type (ST) (reviewed by Maiden, 2006).

Although the number of nucleotide differences between alleles is ignored when assigning alleles for epidemiological purposes, the embedded sequence data can provide information about the population structure of the species and the relevant contribution of mutation and recombination in the genetic diversity seen (Feil *et al.*, 2001; Maiden, 2006). In addition to its use for determining evolutionary relationships between bacteria, MLST offers many advantages over other molecular typing schemes as it is portable and unambiguous, and as data are shared via the internet the results are directly comparable between laboratories around the world.

Originally developed for *Neisseria meningitidis* (Maiden *et al.*, 1998), MLST has subsequently been applied to numerous bacterial species, including the porcine pathogen *Mycoplasma hyopneumoniae* (Mayor *et al.*, 2008). The present study describes the development and application of a novel MLST scheme to *M. agalactiae* and a comparison of MLST with a previously developed VNTR scheme.

METHODS

Mycoplasma culture and DNA extraction. *M. agalactiae* isolates (as listed in Table 3: see Results) were stored at -80°C and grown in 3 ml aliquots of Eaton's broth medium (Nicholas & Baker, 1998) for

24 h. A loopful of the culture was then plated on Eaton's solid medium and incubated at 37°C , 5% CO_2 for between 48 and 72 h. Single colonies were picked and grown in 3 ml Eaton's broth until growth was detected and then aliquots were frozen in 15% (v/v) glycerol at -80°C until required.

Confirmatory tests. DNA from each strain was extracted from 1.5 ml of culture using a Genelute gDNA extraction kit (Sigma), following the manufacturer's instruction. To confirm the identity of the isolates as *M. agalactiae*, PCR based on the *uvrC* gene (Subramaniam *et al.*, 1998) and 16S rRNA gene PCR DGGE (McAuliffe *et al.*, 2005) were performed.

Primer design and PCRs. Genes were chosen for MLST based on the sequenced genome of *M. agalactiae* type strain PG2 (Sirand-Pugnet *et al.*, 2007). The gene targets and primer sequences are listed in Table 1. Gene targets were chosen so that they were distributed evenly within the genome as calculated by the distance from the origin of replication. Initially primers (primer set 1) were chosen so that they amplified the whole of the gene. Initial screens looked at genes *gyrB*, *tpi*, *gltX*, *recA*, *uvrC*, *tufA*, *dnaA* and *metS*. Primers were chosen so that they had an optimum annealing temperature of 55°C . All PCRs were carried out using the cycling conditions of denaturation at 94°C for 5 min followed by 30 cycles of 95°C for 1 min, 55°C for 60 s and 72°C for 1 min, and a final extension step of 72°C for 10 min; samples were kept at 4°C until analysis. For the PCR, 1 μl extracted DNA was added as a template to 49 μl of a reaction mixture containing 10 mM Tris/HCl (pH 9.0), 1.5 mM MgCl_2 , 50 mM KCl, 0.1% Triton X-100, 0.2 mM each deoxynucleoside triphosphate, 0.5 mM each primer and 0.5 U Taqgold (Applied Biosystems). Samples were checked for correct amplification by electrophoresis on a 2% agarose gel followed by visualization with ethidium bromide under UV illumination.

Initial MLST studies on a subset of eight strains using primer set 1 gave long reads encompassing most of the whole coding region for *dnaA*, *gltX*, *gyrB*, *metS*, *tufA*, *tpi*, *recA* and *uvrC*. Analysis of sequences showed some variation in *dnaA* between 300 and 900 bp, in *gltX* between 350 and 950 bp, in *gyrB* between 1000 and 1300 bp, in *metS* between 500 and 1000 bp and in *tufA* between 300 and 1100 bp. No variation was seen in *tpi*, *recA* or *uvrC*. Based on these sequences, a new set of internal primers was designed to ensure that high-quality forward and reverse sequence was obtained over the variable regions

Table 1. Primers based on *M. agalactiae* PG2 genome used for initial screen (1), for subsequent MLST analysis (2) and alternative primers designed based on *M. agalactiae* 5632 genome (3)

Gene target	Forward	Reverse	Product size (bp)
<i>gyrB</i> (1)	tagcaaaacaccctcaaact	attttcaggattcattgtgg	1968
<i>gltX</i> (1)	tagtggtgctagaactgct	tgctcaatgtaagttgtgc	1392
<i>tpi</i> (1)	ccgtcagaattgacagaaat	tagaaaagccaatccacaat	777
<i>tufA</i> (1)	attttgaccgtagcaaagaa	gaccacgtgaacgtctgtt	1191
<i>dnaA</i> (1)	aacccaaactcacaatatg	attgctctgtttcttttcg	1401
<i>uvrC</i> (1)	tgttttgtatgttgcaaaag	tagcaacattaaatggcaca	1716
<i>metS</i> (1)	gcatcagaagattcattg	ctgctgctgctaataaggat	1550
<i>recA</i> (1)	tgctcaatgtaagttgtgc	agtgatgttcttccttgat	993
<i>gyrB</i> (2)	caatacacatcaaccttcca	agcagagttaccttcgaca	519
<i>gltX</i> (2)	gccttgctcacaactcttat	atagttgcttaagcgcaaac	792
<i>tufA</i> (2)	gaacatgattactggtgctg	acggccttttcaaattcta	781
<i>dnaA</i> (2)	taacgtaaccccaactcac	cataattcaggcgctcatctt	762
<i>metS</i> (2)	gcctgtaaatagcccttct	ttttaacaaaatcaagctg	756
<i>gltX</i> (3)	ggcgaagaagtcagttaga	gagcagtttctgaaggctta	1080

Table 2. Details of the five loci used in the *M. agalactiae* typing scheme, including genetic variation, where *H* is genetic diversity, and d_N/d_S is the ratio of non-synonymous to synonymous mutations

Locus	Gene function	Start position (in PG2 genome)	Seq. size (bp)	Nucleotide position within gene	No. of alleles	<i>H</i>	d_N/d_S
<i>dnaA</i>	Chromosomal replication initiation protein	1	672	238–909	8	0.5035	0.0193
<i>metS</i>	Methionyl-tRNA synthetase	237325	687	451–1137	6	0.2227	0.0508
<i>tufA</i>	Elongation factor Tu	382492	681	331–1011	5	0.3404	0.0231
<i>gltX</i>	Glutamyl-tRNA synthetase	669536	699	373–1071	5	0.2227	0.0729
<i>gyrB</i>	DNA gyrase subunit B	854497	420	901–1320	4	0.6016	0.0197

in *gyrB*, *gltX*, *metS*, *tufA* and *dnaA*. These primers (primer set 2) are shown in Table 1. It was found the set 2 *gltX* primers that were designed based on the *M. agalactiae* PG2 genome did not yield amplicons for two strains (L9 and 281F03) and no sequence data could be produced. Therefore additional primers were designed based on the *M. agalactiae* 5632 genome (Nouvel *et al.*, 2010). These primers are also given in Table 1 (set 3); they provided high-quality forward and reverse sequence in the region of interest for both L9 and 281F03.

PCR purification and DNA sequencing. PCR products were purified using a Qiaquick PCR purification kit (Qiagen), quantified by running on 2% agarose E gels with an E gel quantitative ladder (Invitrogen). Sequencing reactions were then carried out using the BigDye v3.1 kit (Applied Biosystems) and samples were run using the ABI Prism 3730 Genetic Analyzer. Sequences were analysed and contigs generated and aligned against a reference sequence using SeqScape v2.6 (Applied Biosystems). Any apparent polymorphisms in sequence were confirmed by manually analysing the sequencing traces.

Allele and ST assignment. Allele numbers were assigned to sequences. For each locus, distinct allele sequences were assigned arbitrary allele numbers in the order of identification. Each genotype was therefore designated by five numbers (e.g. 1-3-2-4-1) that constituted an allelic profile or sequence type (ST; e.g. ST-3). The STs were assigned arbitrary numbers in the order of description.

Data analysis. MLST data analysis was performed using the non-redundant database (NRDB) at pubMLST (<http://pubmlst.org/analysis/>) to assign sequence types. Dendrograms to show relatedness between strains based on allelic profiles were constructed using the PHYLIP neighbour-joining methods at <http://pubmlst.org/analysis/>.

Simpson's index of diversity, which is based on the probability that two unrelated strains will be placed into different typing groups, was calculated according to Hunter & Gaston (1988). Simpson's index of diversity ranges from 0.0 to 1.0, where 0.0 means that all strains in a population are of an identical type and 1.0 indicates that all of the strains in a population are different and can be distinguished from one other.

BURST analysis to reveal the relationship of MLST sequence types and to analyse clonal complexes was carried out with eBURST V3 (accessible at <http://eburst.mlst.net>).

The 'standardized' index of association (I_A) was used to test the null hypothesis of linkage equilibrium for multilocus data and therefore determine the relative contribution of mutation and recombination to the diversity seen by MLST. I_A is zero for linkage equilibrium and deviation from this indicates a degree of linkage disequilibrium.

Analysis was done using the LIAN program available at <http://pubmlst.org/analysis/>.

VNTR analysis. VNTR analysis was carried out on new strains not included in our previous study as was described previously (McAuliffe *et al.*, 2008). Four targets were assessed (VNTR 5 and VNTR 19 in predicted lipoproteins, and VNTR 14 and VNTR 17 in intergenic regions). VNTR profiles were recorded as character data using allelic profiles and then dendrograms were constructed using neighbour-joining with PHYLIP hosted at <http://pubmlst.org/analysis/>.

RESULTS

Analysis of sequence variation in housekeeping genes

Genetic variation was seen in the genes *gyrB*, *gltX*, *metS*, *dnaA* and *tufA*. An online database for this MLST scheme has been set up using specially designed software, BIGSdb (Jolley & Maiden, 2010) and is available at <http://pubmlst.org/magalactiae/>. The greatest variation was seen in *metS*, with 36 polymorphisms found giving 6 different alleles. Twenty-three polymorphisms were found in *dnaA* giving 8 different alleles, 11 polymorphisms in *tufA* giving 5 different alleles, 3 polymorphisms in *gltX* giving 5 different alleles and 21 polymorphisms in *gyrB* giving 4 different alleles. The genetic diversity (*H*) was calculated for each locus using LIAN: *dnaA* had a diversity of 0.5035, *gltX* of 0.2227, *gyrB* of 0.3404, *metS* of 0.2227 and *tufA* of 0.6016 (Table 2).

Determination of the index of association

LIAN analysis gave a standardized index of association of 0.2318. Essentially this value of I_A shows that the genetic diversity seen is due to mutation and not recombination and the genetic variation seen is in linkage disequilibrium.

Determination of allelic profiles and sequence types for strains

Sequences were analysed using the NRDB at <http://pubmlst.org/analysis/> and it was found that there were 17 STs among the 51 strains (as summarized in Table 3). ST-5

Table 3. Description of strains used in this study, their ST, allelic profile, BURST type and VNTR profile

Strain	Origin	Year of isolation	BURST group*	ST	Allelic profile					VNTR profile
					<i>dnaA</i>	<i>gltX</i>	<i>gyrB</i>	<i>metS</i>	<i>tufA</i>	
NCTC 10123 (PG2)	USA	1983	CC1	1	1	1	1	1	1	1111
222F03	Greece	2003	CC1	2	1	1	1	2	1	1111
233F03	Greece	2003	CC1	2	1	1	1	2	1	1111
240F03	Greece	2003	CC1	2	1	1	1	2	1	1111
244F03	Sardinia, Italy	2003	CC2	3	1	1	1	2	2	1021
247F03	Sardinia, Italy	2003	CC2	3	1	1	1	2	2	1021
235F03	Greece	2003	AG	4	1	1	2	2	1	1121
259F03	Portugal	2003	CC2	5	1	1	2	2	2	1121
269F03	Portugal	2003	CC2	5	1	1	2	2	2	1121
271F03	Portugal	2003	CC2	5	1	1	2	2	2	1121
289F03	Spain	2003	CC2	5	1	1	2	2	2	0121
1640/98	Sicily, Italy	1998	CC2	5	1	1	2	2	2	1011
14SR97	Macedonia†	1997	CC2	5	1	1	2	2	2	1121
S2	Spain	2010	CC2	5	1	1	2	2	2	1121
I	Spain	2010	CC2	5	1	1	2	2	2	1121
27	Spain	2010	CC2	5	1	1	2	2	2	1121
13	Spain	2010	CC2	5	1	1	2	2	2	1121
1101	Spain	2010	CC2	5	1	1	2	2	2	1121
1801	Spain	2010	CC2	5	1	1	2	2	2	1121
2501	Spain	2010	CC2	5	1	1	2	2	2	1121
0102	Spain	2010	CC2	5	1	1	2	2	2	1121
E	Spain	2010	CC2	5	1	1	2	2	2	1121
23	Spain	2010	CC2	5	1	1	2	2	2	1121
165/99	Sicily, Italy	1999	CC2	5	1	1	2	2	2	1012
250F03	Italy	2003	CC2	5	1	1	2	2	2	1012
256F03	Italy	2003	CC2	5	1	1	2	2	2	1012
243F03	Sardinia, Italy	2003	CC2	5	1	1	2	2	2	1121
287F03	Spain	2003	CC2	5	1	1	2	2	2	1121
501F03	Sicily, Italy	2006	CC2	5	1	1	2	2	2	1121
Lo Faso	Sicily, Italy	2000	CC2	5	1	1	2	2	2	1121
263F03	Portugal	2003	CC2	5	1	1	2	2	2	1121
234F03	Greece	2003	AG	6	1	1	2	4	1	0121
249F03	Sardinia, Italy	2003	CC2	7	3	1	2	2	2	
224F03	Greece	2003	AG	8	1	2	2	3	1	1131
225F03	Greece	2003	AG	8	1	2	2	3	1	1131
227F03	Greece	2003	CC1	9	1	3	1	2	1	1101
401	Mongolia	2006	AG	10	2	1	2	2	1	1122
752	Mongolia	2006	AG	10	2	1	2	2	1	1122
14	Mongolia	2006	AG	10	2	1	2	2	1	1122
334	Mongolia	2006	AG	10	2	1	2	2	1	1122
44	Mongolia	2006	AG	10	2	1	2	2	1	1122
356	Mongolia	2006	AG	11	2	1	2	2	3	1112
564	Mongolia	2006	AG	11	2	1	2	2	3	1112
659	Mongolia	2006	AG	11	2	1	2	2	3	1112
666	Mongolia	2006	AG	11	2	1	2	2	3	1112
281F03	Spain	2003	Singleton	12	4	4	3	5	4	1001
568	Mongolia	2006	AG	13	5	1	2	2	1	1122
572	Mongolia	2006	AG	14	5	1	2	2	3	1112
L9	Gran Canaria, Spain	Unknown	Singleton	15	6	5	4	6	5	2001
S1	Spain	2010	CC2	16	7	1	2	2	2	1121
H	Spain	2010	CC2	16	7	1	2	2	2	1121
S4	Spain	2010	CC2	17	8	1	2	2	2	1121

*AG, ancestral group; CC, clonal complex. †Republic of Macedonia.

Fig. 1. Dendrogram of MLST allelic profiles of isolates as determined using neighbour-joining analysis (scale bar represents number of differences). Pairwise distances were calculated using the unweighted pair group method with arithmetic averages (UPGMA). ST designations are indicated for each branch.

with the allelic profile of 11222 was the most numerically dominant ST and corresponded to 24 strains of European origin. Spanish strains were diverse, representing 5 STs. Greek strains represented 5 STs, Mongolian strains represented 4 STs, and Sardinian strains represented 3 STs. All other Italian strains, the single Macedonian strain tested and Portuguese strains were characterized as ST-5. A dendrogram of the MLST allelic profiles is shown in Fig. 1.

Determination of BURST types

Three clonal complexes were found using BURST analysis. For this study, clonal complexes are defined as a group of multilocus genotypes in which every genotype shares at

least three of the five loci in common with at least one other member of the group. The most likely ancestral genotype was identified by comparing in turn each genotype within a clonal complex with all other genotypes within the clonal complex. The ancestral genotype was defined as the genotype within the clonal complex that differs from the highest number of other genotypes in the clonal complex at only one locus out of five. To put it another way, the ancestral clone is that genotype defining the highest number of single-locus variants, or SLVs. SLVs are identical to the ancestral genotype at four loci, but differ at the fifth. The ancestral group comprised ST-4, which contained a single Greek strain 235F03. SLVs of the ancestral group were ST-6 from Greece and ST-10 and

Fig. 2. BURST analysis of *M. agalactiae* STs. The assigned ancestral genotype is within the central ring (ST-4). The central group is thus considered to be the ancestral group (AG), with single-locus variants (SLVs) and double-locus variants (DLVs) shown in concentric circles. Two additional clonal complexes were identified, CC1 (centred on ST-2) and CC2 (centred on ST-5). ST-2 and ST-5 are both SLVs of the ancestral genotype ST-4.

ST-13 from Mongolia. Double-locus variants (DLVs) of the ancestral group were ST-8 from Greece, ST-11 from Mongolia and ST-14 from Mongolia. Clonal complex 1 was centred on ST-2 from Greece and SLVs were ST-1, the type strain from the USA and ST-9 from Greece. Clonal complex 2 was centred on ST-5, which contained strains of multiple European origins, and SLVs were ST-3 from Sardinia, ST-7 from Sardinia, and ST-16 and ST-17 from Spain. ST-12 and ST-15 (Spain) were not linked to the clonal complex and were singletons (Fig. 2).

VNTR analysis

As was seen previously (McAuliffe *et al.*, 2008) *M. agalactiae* isolates were relatively heterogeneous, with 12 different profiles obtained using VNTR analysis; these are summarized in Fig. 3. Analysis of the 12 strains of *M. agalactiae* from Mongolia produced a unique set of profiles not observed before in any other *M. agalactiae* strains with either a 1112 or 1122 profile given.

Group 1 contained three Greek strains and the Spanish type strain. Groups 2 and 3 comprised five and six Mongolian strains respectively. Group 4 comprised two Greek isolates. Group 5 was a single Sicilian strain. Group 6 contained three Italian strains. Group 7 was a single Spanish isolate. Group 8 also consisted of a single Spanish

isolate, L9. Group 9 contained two Greek strains, and the closely related group, group 10, also contained a (single) Greek strain. Group 11 contained a Spanish strain and a Greek strain, and group 12 was a large indistinguishable group of 23 European strains of multiple origins.

Congruence between MLST and VNTR analysis

MLST was capable of differentiating some strains which were indistinguishable using VNTR analysis. For example, VNTR group 1 could be separated into two groups using MLST analysis, with the type strain found to be distinctly different from all other strains. Strains in VNTR group 6 could also be differentiated using MLST analysis. Similarly, strains found to be identical by VNTR analysis (VNTR group 11, and strains S1, H and S4 in VNTR group 12) could be distinguished using MLST analysis. Conversely, a single strain 1640, which was different from all other strains using VNTR analysis, was found to fall into the large homogeneous European group using MLST analysis. Overall, the methods placed 75 % of strains into equivalent groups.

Determination of Simpson's index of diversity

Simpson's index of diversity was calculated to be 0.7843 for MLST analysis of these strains and 0.7746 for VNTR analysis.

DISCUSSION

To our knowledge this study is the first to undertake MLST analysis of the important small-ruminant pathogen *M. agalactiae*. Of the eight genes initially examined, five (*gyrB*, *gltX*, *metS*, *dnaA* and *tufA*) showed some degree of variation and were selected for inclusion in this MLST scheme. However, the other three genes that were investigated (*recA*, *tpi* and *uvrC*) were completely invariant over the entire region of sequence coverage (3486 bp). This suggests that these loci are under stabilizing selection, that is, the sequence is optimal for the function that it performs. Any other variation is deleterious and therefore it will not persist in the population. Nonetheless, this MLST scheme of five loci (3160 nt) offers a highly discriminatory typing method for *M. agalactiae* and was capable of subdividing 53 strains into 17 distinct sequence types, largely according to geographical origin. MLST detected unexpected diversity in recent isolates from Spain, identifying two novel outliers, and also enabled typing of novel Mongolian isolates for the first time. The analysis of the Mongolian isolates is interesting as they form a distinct branch (Fig. 1) and it can therefore be speculated that there are different lineages prevalent in Asia compared to Europe.

Previous comparative genome sequence analysis has indicated significant genome variation between *M. agalactiae* strains: an additional 95 genes (representing an extra 130 kbp) found in a field strain when compared to the PG2

Fig. 3. Dendrogram of VNTR analysis of *M. agalactiae* strains. VNTR profiles were recorded as character data using allelic profiles and then dendrograms were constructed using neighbour-joining (scale bar represents number of differences). Pair-wise distances were calculated using the unweighted pair group method with arithmetic averages (UPGMA). VNTR groups are designated for each branch.

type strain (Nouvel *et al.*, 2010). However, these genes are clearly non-essential and as such could be considered to be 'accessory genes'. The MLST work here demonstrates that sequence variation also occurs within housekeeping genes, indicating that the core genome is also variable. However, the large majority of variation for all five loci was synonymous (Table 2), meaning that there was a smaller amount of variation in the expressed proteins. Each locus was shown to be under purifying selection by means of a one-tailed Z-test ($P < 0.005$ for all loci), indicating that the coded proteins are essential and conserved for their respective functions. This is not unexpected for an organism with a minimal genome.

MLST analysis detected a large homogeneous group of European strains that were indistinguishable; a similar

finding was also made using VNTR analysis (McAuliffe *et al.*, 2008), despite there being some isolates that differed using the two methods. Large groups such as this may be more successful (see below) or this result may indicate that these strains have been oversampled. It has previously been described how highly successful clones will become widespread in a population (Vos & Didelot, 2009). Therefore these clonal strains may offer a selective advantage and studies should be undertaken to determine if they are significantly more persistent, more virulent or more recalcitrant to antibiotics than other strains.

Two strains (L9 and 281F03 from Spain) were found to be substantially different from all other strains studied and as such were singletons which did not fit into any clonal complex. Some regions from these strains could also not be

amplified using primers based on the PG2 genome and required alternative primers based on the 5632 genome. These strains represent unusual isolates and are likely to be related to the unusual field strains described previously (Nouvel *et al.*, 2010), which possess additional genes, mobile genetic elements, gene families encoding surface proteins and integrative conjugative elements. The impact of these genomic differences on virulence and persistence of strains is worthy of future study.

Analysis was also carried out to determine what the relative contributions of mutation and homologous recombination (HR) were in the genetic diversity seen among *M. agalactiae* strains. It was found that diversity was largely due to mutation, with recombination playing a much smaller role. However, as described by Feil and co-workers (Feil *et al.*, 2001; Feil & Enright, 2004), with the possible exception of *Salmonella enterica* (Selander & Smith, 1990), few bacterial species appear to be truly clonal, such that recombinational exchanges are absent, or so rare that they are observed only in genes under strong diversifying selection. In these species, clones should be stable because diversification at housekeeping loci depends almost entirely on the accumulation of point mutations, but, with increasing ratios of recombination to point mutation, bacterial clones should become increasingly transient. Care must be taken in the interpretation of values of HR, as Maynard Smith *et al.* (1993) first pointed out that the over-representation of closely related, high-frequency (epidemic) clones in a sample will lead to an inflated estimate of clonality of the population as a whole. Oversampling of a single clone in an epidemic population structure will therefore result in an underestimation of HR. Nonetheless, ST-5 was represented by isolates from temporally and spatially separated disease outbreaks, suggesting that this type may be stable and more pathogenic than other potential variants.

Interestingly, MLST has only been carried out on one other mollicute species, *Mycoplasma hyopneumoniae*, so it is difficult to test statistically whether HR is reduced in certain phylogenetic or ecological groups such as the mollicutes. HR was actually found to be relatively high for *M. hyopneumoniae* (Mayor *et al.*, 2008) but as has been shown for other phyla, wide variation in HR rates can occur among species belonging to the same phylum or division (Vos & Didelot, 2009).

This study has also confirmed the utility of VNTR analysis as a rapid screening tool for molecular epidemiology purposes and it gave results that were largely congruent with the MLST analysis. Differences may largely be due to MLST being based on sequences of genes under purifying selection whilst VNTR regions are by their nature likely to be under different levels of selection. As such MLST is likely to be more representative of evolutionary lineages, whilst VNTR could reflect local adaptations. However, MLST was capable of greater discrimination and has the added advantages of also providing meaningful information about population structure and being totally objective and subject to easy inter-laboratory comparison.

ACKNOWLEDGEMENTS

Financial support for this work was from the UK Department for Environment, Food and Rural Affairs (Defra). We would like to thank Keith Jolley at University of Oxford for setting up the web database.

REFERENCES

- Bergonier, D., Bertholet, X. & Poumarat, F. (1997). Contagious agalactia of small ruminants: current knowledge concerning epidemiology, diagnosis and control. *Rev Sci Tech* **16**, 848–873.
- Feil, E. J. & Enright, M. C. (2004). Analyses of clonality and the evolution of bacterial pathogens. *Curr Opin Microbiol* **7**, 308–313.
- Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M. S., Day, N. P., Enright, M. C., Goldstein, R., Hood, D. W., Kalia, A. & other authors (2001). Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci U S A* **98**, 182–187.
- Hunter, P. R. & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* **26**, 2465–2466.
- Jolley, K. A. & Maiden, M. C. (2010). BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**, 595.
- Maiden, M. C. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60**, 561–588.
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K. & other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–3145.
- Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**, 4384–4388.
- Mayor, D., Jores, J., Korczak, B. M. & Kuhnert, P. (2008). Multilocus sequence typing (MLST) of *Mycoplasma hyopneumoniae*: a diverse pathogen with limited clonality. *Vet Microbiol* **127**, 63–72.
- McAuliffe, L., Ellis, R., Lawes, J., Ayling, R. D. & Nicholas, R. A. J. (2005). 16S rDNA PCR and denaturing gradient gel electrophoresis; a single generic test for detecting and differentiating *Mycoplasma* species. *J Med Microbiol* **54**, 731–739.
- McAuliffe, L., Churchward, C. P., Lawes, J. R., Loria, G. R., Ayling, R. D. & Nicholas, R. A. J. (2008). VNTR analysis reveals unexpected genetic diversity within *Mycoplasma agalactiae*, the main causative agent of contagious agalactia. *BMC Microbiol* **8**, 193.
- Nicholas, R. A. J. & Baker, S. E. (1998). Recovery of mycoplasmas from animals. In *Mycoplasma Protocols*, pp. 37–43. Edited by R. J. Miles & R. A. J. Nicholas. Totowa, USA: Humana Press.
- Nouvel, L. X., Sirand-Pugnet, P., Marenda, M. S., Sagné, E., Barbe, V., Mangenot, S., Schenowitz, C., Jacob, D., Barré, A. & other authors (2010). Comparative genomic and proteomic analyses of two *Mycoplasma agalactiae* strains: clues to the macro- and micro-events that are shaping mycoplasma diversity. *BMC Genomics* **11**, 86.
- OIE (2003). Contagious agalactia. In *Manual of Standards for Diagnostic Tests and Vaccines*, pp. 992–999. Paris: World Organisation for Animal Health.
- Pilo, P., Fleury, B., Marenda, M., Frey, J. & Vilei, E. M. (2003). Prevalence and distribution of the insertion element ISMag1 in *Mycoplasma agalactiae*. *Vet Microbiol* **92**, 37–48.

- Selander, R. K. & Smith, N. H. (1990).** Molecular population genetics of *Salmonella*. *Rev Med Microbiol* **1**, 219–228.
- Sirand-Pugnet, P., Lartigue, C., Marends, M., Jacob, D., Barré, A., Barbe, V., Schenowitz, C., Mangenot, S., Couloux, A. & other authors (2007).** Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. *PLoS Genet* **3**, e75.
- Solsona, M., Lambert, M. & Poumarat, F. (1996).** Genomic, protein homogeneity and antigenic variability of *Mycoplasma agalactiae*. *Vet Microbiol* **50**, 45–58.
- Subramaniam, S., Bergonier, D., Poumarat, F., Capaul, S., Schlatter, Y., Nicolet, J. & Frey, J. (1998).** Species identification of *Mycoplasma bovis* and *Mycoplasma agalactiae* based on the *uvrC* genes by PCR. *Mol Cell Probes* **12**, 161–169.
- Tola, S., Idini, G., Manunta, D., Casciano, I., Rocchigiani, A. M., Angioi, A. & Leori, G. (1996).** Comparison of *Mycoplasma agalactiae* isolates by pulsed field gel electrophoresis, SDS-PAGE and immunoblotting. *FEMS Microbiol Lett* **143**, 259–265.
- Tola, S., Angioi, A., Rocchigiani, A. M., Idini, G., Manunta, D., Galleri, G. & Leori, G. (1997).** Detection of *Mycoplasma agalactiae* in sheep milk samples by polymerase chain reaction. *Vet Microbiol* **54**, 17–22.
- Vos, M. & Didelot, X. (2009).** A comparison of homologous recombination rates in bacteria and archaea. *ISME J* **3**, 199–208.