Single Nucleotide Polymorphisms in the IS900 Sequence of Mycobacterium avium subsp. paratuberculosis Are Strain Type Specific

Elena Castellanos,1,2 Alicia Aranaz,1,2,‡ Lucia de Juan,1,2 Julio Álvarez,1 Sabrina Rodríguez,1,2 Beatriz Romero,1,2 Javier Bezos,1,2 Karen Stevenson,3 Ana Mateos,1,2 and Lucas Domínguez1,2

Centro de Vigilancia Sanitaria Veterinaria (VISAVET), Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040 Madrid, Spain1; Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad Complutense de Madrid, Madrid, Spain2; and Moredun Research Institute, Division of Control of Bacterial Diseases, Pentlands Science Park, Bush Loan, Penicuik, EH26 OPZ Scotland, United Kingdom3

Received 17 March 2009/Accepted 27 April 2009

Insertion sequence IS900 is used as a target for the identification of Mycobacterium avium subsp. paratuberculosis. Previous reports have revealed single nucleotide polymorphisms within IS900. This study, which analyzed the IS900 sequences of a panel of isolates representing M. avium subsp. paratuberculosis strain types I, II, and III, revealed conserved type-specific polymorphisms that could be utilized as a tool for diagnostic and epidemiological purposes.

The insertion sequence IS900 (13) is one of the 20 members of the IS110 family and is considered to be unique to Mycobacterium avium subsp. paratuberculosis, although IS900-like sequences have been found rarely in other environmental mycobacteria (9, 12). IS900 is present in multiple copies (14–18) within the M. avium subsp. paratuberculosis genome (13) and therefore is an ideal target for identification (16, 31). The heterogeneity of M. avium subsp. paratuberculosis isolates based on the number of IS900 copies and their positions within the genome has been exploited for epidemiological purposes. The IS900 sequence has been used in different molecular techniques such as multiplex PCR and restriction fragment length polymorphism analysis with hybridization to IS900, leading to the identification of individual strains and the classification of M. avium subsp. paratuberculosis into strain types (3, 17, 19, 30).

M. avium subsp. paratuberculosis strains have been divided into three clusters named type I (also called sheep [S] strains), type II (also called cattle [C] strains), and type III (also called intermediate strains), although some of the studies include types I and III in the S type (8, 35). This classification into three clusters is based on results from molecular characterization techniques such as restriction fragment length polymorphism analysis with hybridization to IS900 (19), pulsed-field gel electrophoresis (PFGE) (11, 29), PCR-restriction enzyme analysis (PCR-REA) of gyrB (5), PCR-REA of inhA (6), PCR and denaturing gradient gel electrophoresis analyses of MAP1506 (14), PCR sequencing of recF (31), and comparative genomic hybridization analyses (4). The genotypic and phenotypic dissimilarities among types of M. avium subsp. paratuberculosis may be reflected in the differences detected in the progression of paratuberculosis or Johne’s disease among infected herds (33, 34).

Comparisons of IS900 sequences from different M. avium subsp. paratuberculosis isolates with the published genome sequence of M. avium subsp. paratuberculosis strain K-10 have revealed single nucleotide polymorphisms (SNPs) (2, 20, 21, 23, 24, 27, 28, 35, 36) (Table 1).

Of special interest are the results reported by Semret et al. (24), who observed major variance in the IS900 sequences in M. avium subsp. paratuberculosis S strains (classification into type I or III was not indicated in the study) compared to the IS900 sequences in C strains.

The aim of this research was to follow up these findings and to analyze the IS900 sequences from a panel of type I, II, and III M. avium subsp. paratuberculosis isolates to establish the degree of nucleotide identity among them and the relationship between M. avium subsp. paratuberculosis types and IS900 sequence profiles.

For this purpose, we selected 33 isolates of M. avium subsp. paratuberculosis types I, II, and III from separate geographical locations throughout Spain, Scotland, and Denmark (Table 2). Every isolate was cultured and identified as M. avium subsp. paratuberculosis by PCRs directed to the gene F57 (22) (using primers F57-F, 5′-CCCGATAGCTTTTCTCTCCTC-3′, and F57-R, 5′-GATCTCAAGACATGGCGAGGTG-3′ [7]). Additionally, the isolates were typed using a set of PCR-based assays, REA targeting IS1311 (15), PCR analysis of the genomic region described by Collins et al. (DMC) (8), analysis of large sequence polymorphism A20 (LSP^20) (25), and hsp65 sequencing (32) (Table 3). Then the isolates were classified as type I, II, or III by PFGE, PCR-REA of gyrB, or PCR-REA of inhA as described previously (5, 6, 10, 11).
A fragment of 662 bp of the IS900 sequence from each of the isolates was amplified by PCR directed to the 5' end of the insertion sequence, considered to be specific to *M. avium* subsp. *paratuberculosis* (9, 18) and found previously to be polymorphic (24). Primers IS900-F (5' CTTTTTCTGAGGTTG TTCG 3' [24]) and IS900-R (5' CCACCCAGATCGGAA CGTC 3') were used in the amplification reaction, and then amplicons were purified with a QIAquick PCR purification kit

### TABLE 1. Previously described SNPs in the 5' fragments of IS900 sequences relative to the *M. avium* subsp. *paratuberculosis* K-10 genome sequence

| Reference                | Strain type | Geographic origin | Host(s) or source | Molecular technique | SNP                  | Position (bp) | GenBank accession no. |
|--------------------------|-------------|-------------------|-------------------|---------------------|----------------------|---------------|----------------------|
| Bhide et al., 2006 (2)   | –           | –                 | Cattle crossbreed *Bos indicus* | SSCP<sup>a</sup> analysis | No nucleotide (instead of A) | 284           | AY974345             |
| –                        | –           | –                 | Cattle             | SSCP analysis       | No nucleotide (instead of A) | 284           | AY974346             |
| –                        | –           | –                 | Slovak                | SSCP analysis       | C (instead of G)      | 232           | AY974348             |
| –                        | –           | –                 | Slovak                | SSCP analysis       | T (instead of A)      | 243           | AY974348             |
| Pickup et al., 2005 (20) | –           | –                 | United Kingdom River water | Sequencing         | G (instead of A)      | 214<sup>c</sup> | –                    |
| Pickup et al., 2006 (21) | –           | –                 | United Kingdom River water | Sequencing         | G (instead of A)      | 216           | –                    |
| Semret et al., 2006 (24) | S           | –                 | Sheep               | Sequencing          | T/C (with a C small peak) | 169           | –                    |
| Scaru et al., 2007 (23)  | –           | –                 | Humans, Sardinian sheep | Sequencing         | G/A (with an A small peak) | 216           | –                    |
| Sivakumar et al., 2005 (27) | –        | –                 | India               | Sequencing         | No nucleotide (instead of G) | 688           | AY600657             |
| Sivakumar et al., 2009 (36) | –        | –                 | India               | Sequencing         | G (instead of A) | 722           | –                    |
| Willemsen et al., 1999 (36) | –        | –                 | Buffalo            | Sequencing         | G (instead of A) | 899           | –                    |
| Whittington et al., 2001 (35) | S        | Australia         | Sheep              | Sequencing         | No nucleotide (instead of G) | 688           | –                    |

<sup>a</sup> – no information was provided in the reference.

<sup>b</sup> SSCP: single-strand conformational polymorphism.

<sup>c</sup> The SNP corresponds to position 216 in the IS900 sequence.

<sup>d</sup> The SNP was reported to occur at bp 247, but according to the published IS900 sequence of *M. avium* subsp. *paratuberculosis* K-10 (GenBank accession no. AE016958), corresponds to bp 244.

<sup>e</sup> The observed SNP is relative to the IS900 sequence at locus 17 of *M. avium* subsp. *paratuberculosis* K-10.

### TABLE 2. Panel of *M. avium* subsp. *paratuberculosis* isolates used in this study

| Isolate(s) | Strain type<sup>a</sup> | Herd or flock | Host (breed) | Origin<sup>b</sup> |
|------------|--------------------------|---------------|--------------|---------------------|
| 21P        | I                        | Flock A       | Sheep        | Faroe Islands       |
| 235G, 208G, 213G | I                | Flock B       | Sheep        | Shetland            |
| M189       | I                        | Flock C       | Sheep (Finn) | Midlothian          |
| CAM 63     | II                       | Herd A        | Goats (Guadarrama) | Navas del Rey, Madrid |
| CAM 72     | II                       | Herd B        | Goats (Guadarrama) | Becerril de la Sierra, Madrid |
| 464        | II                       | Herd C        | Goats (Guadarrama) | Villa del Prado, Madrid |
| CAM 19     | II                       | Herd D        | Goats (Guadarrama) | Chapinería, Madrid |
| CAM 20     | II                       | Herd E        | Goats (Guadarrama) | Villamantilla, Madrid |
| 574        | II                       | Herd F        | Goats (Murciano-Granadina) | Ciudad Real, Castilla La Mancha |
| 896, 940   | II                       | Herd G        | Cattle (Bullfighting) | Salamanca, Castilla y León |
| CAM 78     | III                      | Herd H        | Goats (Guadarrama) | Sta. María de Alameda, Madrid |
| CAM 40, CAM 38 | III          | Herd I        | Goats (Guadarrama) | Robledo de Chavela, Madrid |
| CAM 86, CAM 87 | III              | Herd J        | Goats (Guadarrama) | Robledo de Chavela, Madrid |
| M107.00175-2 | III             | Herd K        | Goats (Guadarrama) | Hoyo de Manzanares, Madrid |
| M107.00180-2 | III             | Herd L        | Goats (Guadarrama) | Valdemacquita, Madrid |
| 793, 404, 408 | III         | Herd M        | Goats (Murciano-Granadina) | Toledo, Castilla La Mancha |
| 619, 841, 634, M106.00265-2, M106.00286-2 | III | Herd N      | Cattle (Bullfighting) | Albacete, Castilla La Mancha |
| M107.00579-2, M107.00658-2 | III | Herd O        | Cattle (Bullfighting) | Ciudad Real, Castilla La Mancha |
| 734, M105.00897-2 | III | Herd P        | Cattle (Bullfighting) | Salamanca, Castilla y León |
| M106.00198-2 | III | Herd Q        | Cattle (Holstein) | Burgos, Castilla y León |

<sup>a</sup> Every isolate was classified as type I, II, or III by PFGE (10, 11), PCR-REA of *gyrB* (5), or PCR-REA of *inhA* (6).

<sup>b</sup> Geographic distribution: Faroe Islands, Denmark; Shetland, northern Scotland; Midlothian, southeast Scotland; Madrid, central Spain; Castilla La Mancha, south central Spain; and Castilla y León, north central Spain.
(Qiagen, GmbH) and both strands were sequenced by using an ABI Prism 3730 DNA sequencer (Applied Biosystems; CIB Sequencing Facilities, Madrid, Spain). Forward and reverse sequences were aligned with the IS900 elements of the published genome sequence of M. avium subsp. paratuberculosis K-10 (type II strain; GenBank accession no. AE016958; reference sequence no. NC_002944), as suggested by Semret et al. (24). The analysis of the chromatograms showed IS900 SNPs that were homogeneous, conserved, and dependent on the M. avium subsp. paratuberculosis type (Table 3). Every type I ovine isolate showed an SNP at position 216, with a G instead of an A, and no other sequence modifications. This SNP was observed previously in a sample extracted directly from river water (20, 21), in an isolate from a sheep in Australia (35), and in some ovine isolates analyzed by Semret et al. (24). The eight isolates of type II subjected to analysis displayed complete homology to M. avium subsp. paratuberculosis K-10. However, for the type III isolates tested, double peaks of the same size at bp 169 (T/C) and bp 216 (G/A) or a single T peak and a G were noticed. For four of these isolates with ambiguities, another DNA extraction from a single colony was performed with an inoculating needle, but the results were corroborated, ruling out the possibility of contamination by two different strain types. The results of this analysis match previous observations of these SNPs in some of the S isolates tested by Semret et al. (24). This pattern is probably indicative of the presence of a point mutation in some of the copies of IS900 (Table 3) and also confirms the presence of type III isolates outside Spain (10). To our knowledge, this is the first report to establish a correlation between IS900 polymorphisms and strain types I, II, and III. Interestingly, in two previous studies, an irresolvable A/G polymorphism at position 216 (reported initially to be at bp 214) was observed in two samples obtained from a river flow (20, 21), due maybe to the presence of both types I and II in the samples collected. On the other hand, none of our isolates revealed any other SNPs described previously in the literature (2, 23, 27, 28, 36).

The type III strains exhibited polymorphisms identical to those of type I strains, as determined by PCR analysis of DMC, IS311 REA, and LSP^20 and hsp65 sequencing (Table 3), confirming previous observations (1) and showing the genetic

### TABLE 3. Summary of PCR results and IS900 sequencing data obtained for isolates of M. avium subsp. paratuberculosis types I, II, and III

| Isolate | Strain type | Result for: | hsp65 codea | SNP(s)f at bp: |
|---------|-------------|-------------|--------------|---------------|
| 21F     | I           | F57         | S            | 169 G         |
| 23SG    | I           | DMCo        | S            | 169 G         |
| 208G    | I           | IS311b      | S            | 169 G         |
| 213G    | I           | LSPA20      | S            | 169 G         |
| M189    | I           | hsp65       | S            | 169 G         |
| CAM 63  | II          |             | C            | 169 G         |
| CAM 72  | II          |             | C            | 169 G         |
| 464 C   | II          |             | C            | 169 G         |
| CAM 72  | II          |             | S            | 169 G         |
| CAM 20  | II          |             | C            | 169 G         |
| 574     | II          |             | S            | 169 G         |
| 896     | II          |             | S            | 169 G         |
| 940     | II          |             | S            | 169 G         |
| CAM 78  | III         |             | S            | 169 G         |
| CAM 80  | III         |             | S            | 169 G         |
| CAM 82  | III         |             | S            | 169 G         |
| CAM 87  | III         |             | S            | 169 G         |
| MI07.01787-2 | III | IS311b | S | 169 G |
| MI07.04010-2 | III | IS311b | S | 169 G |
| 793     | III         |             | S            | 169 G         |
| 404     | III         |             | S            | 169 G         |
| 619     | III         |             | S            | 169 G         |
| 841     | III         |             | S            | 169 G         |
| 634     | III         |             | S            | 169 G         |
| MI06.00285-2 | III | IS311b | S | 169 G |
| MI06.00286-2 | III | IS311b | S | 169 G |
| MI07.06579-2 | III | IS311b | S | 169 G |
| MI07.06582-2 | III | IS311b | S | 169 G |
| 734     | III         |             | S            | 169 G         |
| MI05.00897-2 | III | IS311b | S | 169 G |
| MI06.01981-2 | III | IS311b | S | 169 G |

### Notes

- PCR analysis of DMC was performed as described by Collins et al. (8). Isolates were identified as S or C strains.
- PCR-REA was performed as described by Marsh et al. (15). Isolates were identified as S or C strains.
- PCR analysis of DMC was performed as described by Collins et al. (8). Isolates were identified as S or C strains.
- The PCR method used was developed by Semret et al. (25).
- SNPs found in the IS900 sequences.
- IS311 data were also confirmed in previous work (29).
- The DNA template was insufficient and did not yield a visible product in the PCR.
similarity of these two strain types. However, this report provides further evidence of additional polymorphic loci that can be used to distinguish between these two strain types. Also, it supports the theory that less genomic divergence exists among type II (bovine) isolates than among type I and III (S) isolates (4, 26, 31).

The stability and conservation of the IS900 sequence drift is reflected at positions 169 and 216 in some copies of IS900 in type I and III strains from herds in different locations. These results are consistent with suggestions in previous studies that M. avium subsp. paratuberculosis strains tend to be clonal (34). The IS900 sequence analysis could be used as a complementary diagnostic tool for epidemiological purposes to study the geographical distribution patterns of the three clusters within the M. avium subsp. paratuberculosis group. On the other hand, from this evidence, it is still not possible to obtain a correlation between pathogen adaptation to environmental factors or virulence pathways and the divergences in IS900 sequences.

Notwithstanding these results, M. avium subsp. paratuberculosis nomenclature and the subdivision of strains into groups are still controversial. Current classifications have been based on the data gathered by several research groups. It would be possible to obtain a correlation between pathogen adaptation to environmental factors or virulence pathways and the sequence drift of IS900, which is reflected at positions 169 and 216 in some copies of IS900.

The IS900 sequence similarity of these two strain types. However, this report provides further evidence of additional polymorphic loci that can be used to distinguish between these two strain types. Also, it supports the theory that less genomic divergence exists among type II (bovine) isolates than among type I and III (S) isolates (4, 26, 31).

The stability and conservation of the IS900 sequence drift is reflected at positions 169 and 216 in some copies of IS900 in type I and III strains from herds in different locations. These results are consistent with suggestions in previous studies that M. avium subsp. paratuberculosis strains tend to be clonal (34). The IS900 sequence analysis could be used as a complementary diagnostic tool for epidemiological purposes to study the geographical distribution patterns of the three clusters within the M. avium subsp. paratuberculosis group. On the other hand, from this evidence, it is still not possible to obtain a correlation between pathogen adaptation to environmental factors or virulence pathways and the divergences in IS900 sequences.

Notwithstanding these results, M. avium subsp. paratuberculosis nomenclature and the subdivision of strains into groups are still controversial. Current classifications have been based on the data gathered by several research groups. It would be possible to obtain a correlation between pathogen adaptation to environmental factors or virulence pathways and the sequence drift of IS900, which is reflected at positions 169 and 216 in some copies of IS900.

The IS900 sequence similarity of these two strain types. However, this report provides further evidence of additional polymorphic loci that can be used to distinguish between these two strain types. Also, it supports the theory that less genomic divergence exists among type II (bovine) isolates than among type I and III (S) isolates (4, 26, 31).

The stability and conservation of the IS900 sequence drift is reflected at positions 169 and 216 in some copies of IS900 in type I and III strains from herds in different locations. These results are consistent with suggestions in previous studies that M. avium subsp. paratuberculosis strains tend to be clonal (34). The IS900 sequence analysis could be used as a complementary diagnostic tool for epidemiological purposes to study the geographical distribution patterns of the three clusters within the M. avium subsp. paratuberculosis group. On the other hand, from this evidence, it is still not possible to obtain a correlation between pathogen adaptation to environmental factors or virulence pathways and the divergences in IS900 sequences.

Notwithstanding these results, M. avium subsp. paratuberculosis nomenclature and the subdivision of strains into groups are still controversial. Current classifications have been based on the data gathered by several research groups. It would be possible to obtain a correlation between pathogen adaptation to environmental factors or virulence pathways and the sequence drift of IS900, which is reflected at positions 169 and 216 in some copies of IS900.

The IS900 sequence similarity of these two strain types. However, this report provides further evidence of additional polymorphic loci that can be used to distinguish between these two strain types. Also, it supports the theory that less genomic divergence exists among type II (bovine) isolates than among type I and III (S) isolates (4, 26, 31).

The stability and conservation of the IS900 sequence drift is reflected at positions 169 and 216 in some copies of IS900 in type I and III strains from herds in different locations. These results are consistent with suggestions in previous studies that M. avium subsp. paratuberculosis strains tend to be clonal (34). The IS900 sequence analysis could be used as a complementary diagnostic tool for epidemiological purposes to study the geographical distribution patterns of the three clusters within the M. avium subsp. paratuberculosis group. On the other hand, from this evidence, it is still not possible to obtain a correlation between pathogen adaptation to environmental factors or virulence pathways and the divergences in IS900 sequences.

Notwithstanding these results, M. avium subsp. paratuberculosis nomenclature and the subdivision of strains into groups are still controversial. Current classifications have been based on the data gathered by several research groups. It would be possible to obtain a correlation between pathogen adaptation to environmental factors or virulence pathways and the sequence drift of IS900, which is reflected at positions 169 and 216 in some copies of IS900.
27. Sivakumar, P., R. N. Tripathi, and N. Singh. 2005. Detection of Mycobacterium avium subsp. paratuberculosis in intestinal and lymph node tissues of water buffaloes (Bubalus bubalis) by PCR and bacterial culture. Vet. Microbiol. 108:263–270.

28. Sohal, J. S., N. Shororan, K. Narayanasamy, V. Brahmachari, S. Singh, and S. Subodh. 2009. Genomic analysis of local isolate of Mycobacterium avium subspecies paratuberculosis. Vet. Microbiol. 134:375–382.

29. Stevenson, K., V. M. Hughes, L. de Juan, N. F. Inglis, F. Wright, and J. M. Sharp. 2002. Molecular characterization of pigmented and nonpigmented isolates of Mycobacterium avium subsp. paratuberculosis. J. Clin. Microbiol. 40:1798–1804.

30. Thibault, V. C., M. Grayon, M. L. Boschiroli, C. Hubbans, P. Overduin, K. Stevenson, M. C. Gutierrez, P. Supply, and F. Biet. 2007. New variable-number tandem-repeat markers for typing Mycobacterium avium subsp. paratuberculosis and M. avium strains: comparison with IS900 and IS1245 restriction fragment length polymorphism typing. J. Clin. Microbiol. 45:2404–2410.

31. Turenne, C. Y., D. M. Collins, D. C. Alexander, and M. A. Behr. 2008. Mycobacterium avium subsp. paratuberculosis and M. avium subsp. avium are independently evolved pathogenic clones of a much broader group of M. avium organisms. J. Bacteriol. 190:2479–2487.

32. Turenne, C. Y., M. Semret, D. V. Cousins, D. M. Collins, and M. A. Behr. 2006. Sequencing of hsp65 distinguishes among subsets of the Mycobacterium avium complex. J. Clin. Microbiol. 44:433–440.

33. Whipple, D., P. Kapke, and C. Vary. 1990. Identification of restriction fragment length polymorphisms in DNA from Mycobacterium paratuberculosis. J. Clin. Microbiol. 28:2561–2564.

34. Whittington, R., I. Marsh, E. Choy, and D. Cousins. 1998. Polymorphisms in IS1311, an insertion sequence common to Mycobacterium avium and M. avium subsp. paratuberculosis, can be used to distinguish between and within these species. Mol. Cell. Probes 12:349–358.

35. Whittington, R. J., C. A. Taragel, S. Ottaway, I. Marsh, J. Seaman, and V. Fridriksdottir. 2001. Molecular epidemiological confirmation and circumstances of occurrence of sheep (S) strains of Mycobacterium avium subsp. paratuberculosis in cases of paratuberculosis in cattle in Australia and sheep and cattle in Iceland. Vet. Microbiol. 79:311–322.

36. Willemsen, P. T. J., D. Bakker, and F. G. van Zijderveld. 1999. An error in the reported IS900 nucleotide sequence affects the proposed expression of ORF2 (hex) and the detection of M. paratuberculosis using the polymerase chain reaction, p. 466–471. In E. J. B. Manning and M. T. Collins (ed.), Proceedings of the 6th International Colloquium on Paratuberculosis. International Association for Paratuberculosis, Madison, WI.