Whole-Genome Analysis of Mycobacterium avium subsp. paratuberculosis IS900 distribution reveals strain type-specific modalities

Cyril Conde, Marian Price-Carter, Thierry Cochard, Maxime Branger, Karen Stevenson, Richard Whittington, John Bannantine, Franck Biet

Whole-Genome Analysis of *Mycobacterium avium* subsp. *paratuberculosis* IS*900* Insertion Reveals Strain Type-Specific Modalities

Cyril Condé[a], Marian Price-Carter[b], Thierry Cochard[a], Maxime Branger[a], Karen Stevenson[c], Richard Whittington[d], John P. Bannantine[e] and Franck Biet[a]

[a]ISP, INRAE, Université de Tours, UMR 1282, 37380 Nouzilly, France, [b]AgResearch Ltd., Hopkirk Research Institue, Palmerson North, New Zealand, [c]Moredun Research Institute, Penicuik, UK, [d]School of Veterinary Science, The University of Sydney, Sydney, Australia, [e]USDA-ARS, National Animal Disease Center, Ames, USA

Background & Objective

Mycobacterium avium subsp. *paratuberculosis* (Map) is the etiological agent of paratuberculosis or Johne's disease that causes chronic intestinal inflammation in ruminants (1,2). The IS*900* insertion sequence, only present in multiple copies in the chromosome of Map, was used as an epidemiological marker and target for qPCR diagnosis but it biological role remains elusive (3,4). During its transposition, the IS may promote a number of important genetic modifications with consequences in the host-pathogen evolution of the strain (5,6). Thanks to the development of sequencing technologies, the number of available Map genomes is increasing rapidly, which makes it easy to analyze the distribution of IS900 in this slow-growing bacterium.

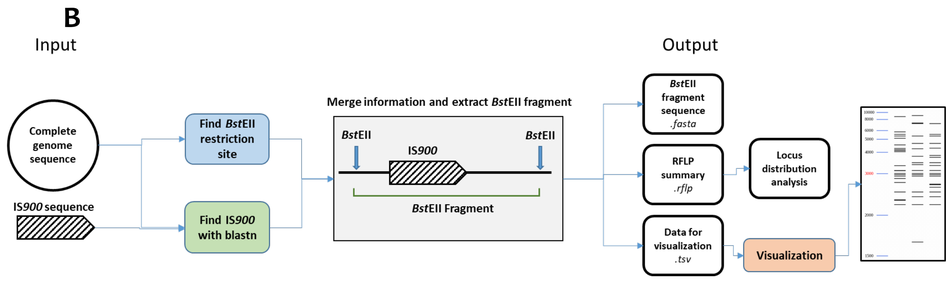
The objective of this study is to use bioinformatics approaches to study the IS*900* distribution in the genome of Map strains. In this study we included the genomes of strains belonging to the three genetic lineages that emerged in Map. These lineages are classified as S-type (including subtype I and III) (ovine) and C-type (bovine) strains. As a secondary goal, a tool will be develop to automate the *in silico* restriction fragment length polymorphism (RFLP) analysis using IS*900*.

Figure 1. Materials and methods

A

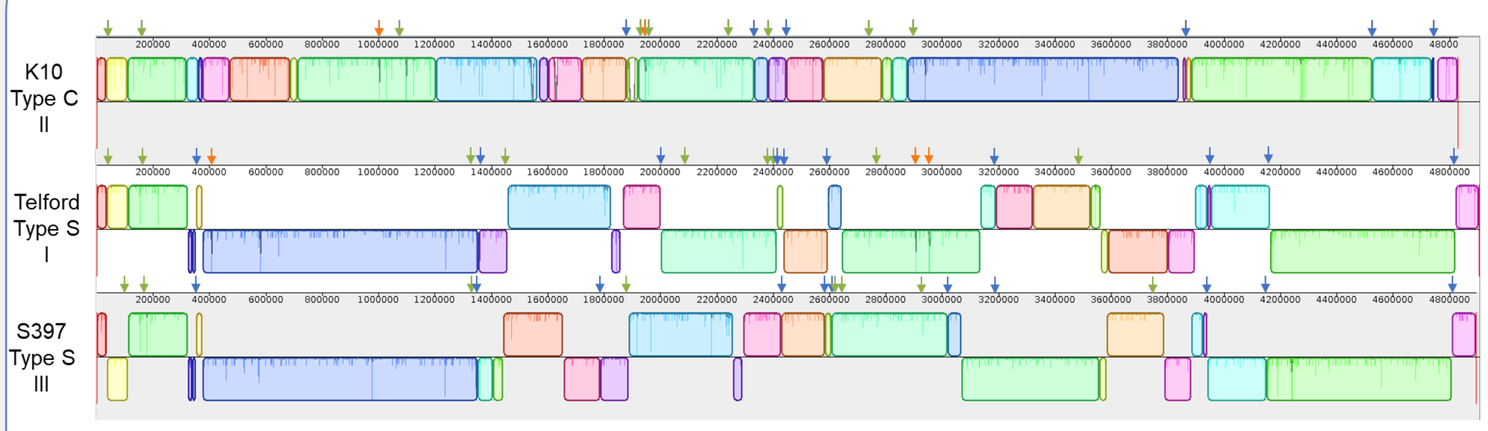
Strain	Telford	K-10	S397
Type	S	C	S
Subtype	I	II	III
IS*900 Bst*EII RFLP profile	S1	R01	A
Number of bands detected *in vitro*	17	15	15
Number of bands detected *in silico*	22	17	19
Number of contigs	1	1	1
Accession number	NZ_CP033688.1	NC_002944.2	NZ_CP053749.1

B



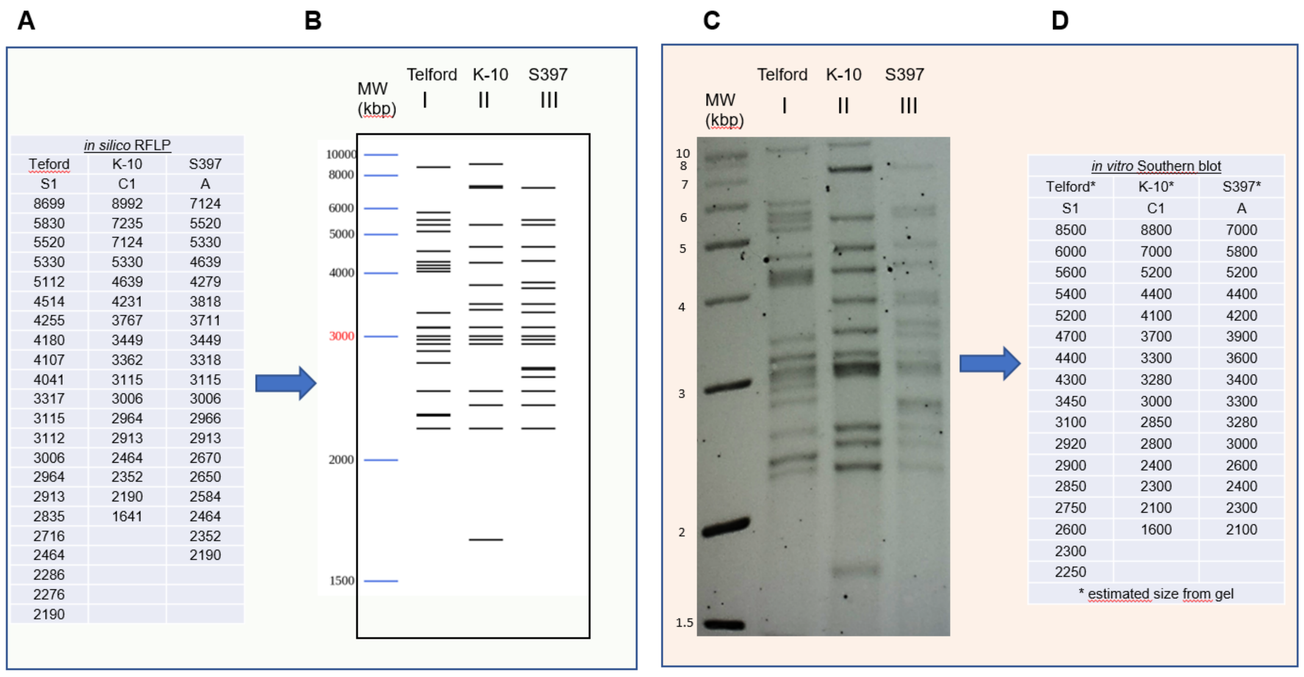
(A) Table 1: details of the strains/genomes used and information on the number of copies of the IS*900*. **(B)** Details of steps performed by the IS*900* RFLP *in silico* pipeline

Figure 2. Distribution of IS*900* copies on Map genomes



Mauve alignment of K-10 (top) (7) with Telford (middle) (8) and S397 (bottom) (9) showing genomic reorganization of the genomes. The colored boxes represent homologous regions present in each genome. Blocks below the centerline indicate regions with inverse orientation. Regions outside the blocks lack homology between the genomes. Within each block there is a similarity profile of the DNA sequences and the white areas indicate sequences specific to a genome. The scale is in base pairs. Orthologous insertions are indicated, by a green arrow, specific insertions are indicated by a orange arrow and conserved loci only in two genomes by a blue arrow.

Figure 3. IS*900* restriction fragment length polymorphism (RFLP) "*in vitro*" versus "*in silico*"



in silico RFLP

Teford	K-10	S397
S1	C1	A
8699	8992	7124
5830	7235	5520
5520	7124	5330
5330	5330	4639
5112	4639	4279
4514	4231	3818
4255	3767	3711
4180	3449	3449
4107	3362	3318
4041	3115	3115
3317	3006	3006
3115	2964	2966
3112	2913	2913
3006	2464	2670
2964	2352	2650
2913	2190	2584
2835	1641	2464
2716		2352
2464		2190
2286		
2276		
2190		

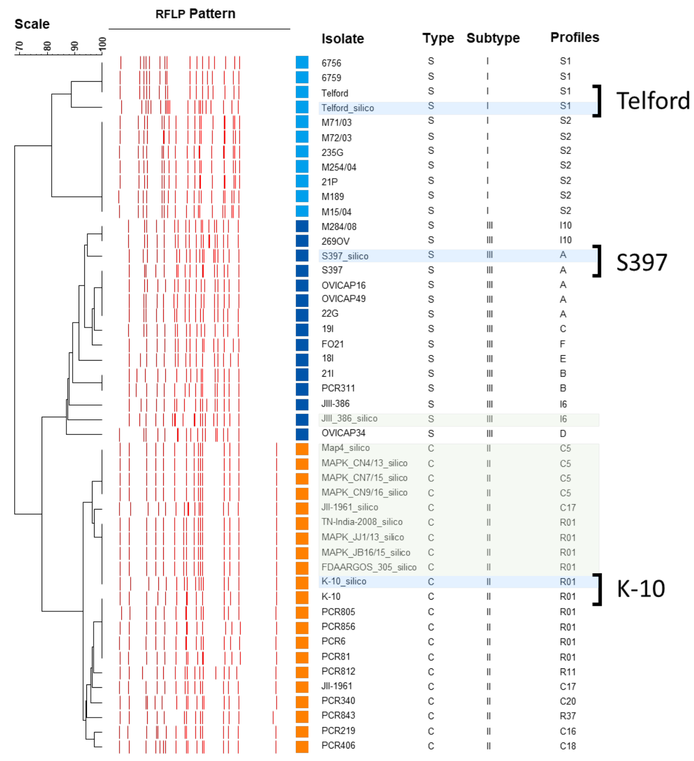
in vitro Southern blot

Telford*	K-10*	S397*
S1	C1	A
8500	8800	7000
6000	7000	5800
5600	5200	5200
5400	4400	4400
5200	4100	4200
4700	3700	3900
4400	3300	3600
4300	3280	3400
3450	3000	3300
3100	2850	3280
2920	2800	3000
2900	2400	2600
2850	2300	2400
2750	2100	2300
2600	1600	2100
2300		
2250		

* estimated size from gel

(A) An *in silico* RFLP analysis was developed using complete genomes. This automated procedure identified the *Bst*EII restriction sites to obtain the exact size of all DNA fragments carrying a copy of the IS*900* sequence. IS*900* RFLP profiles were compared using fragment sizes **(A)** or **(B)** by digital visualization of the restriction fragments separated according to their size by mimicking their migration in agarose gel. **(C)** The IS*900* RFLP profiles obtained by classical Southern blot method and hybridization to IS*900* were used to find the approximate fragment sizes by band analysis software **(D)**

Figure 4. Comparison of digital profiles with the existing profiles in the database



UPGMA dendrogram based on IS*900* RFLP typing using *Bst*EII (10). Profiles obtained *in silico* can be included and compared to those obtained by classical technic (dark brackets)

Results

- The program developed in this study allowed to automatically locate the sequences of the IS*900*, to know their positions and their number.
- Between 16 to 22 copies of the IS*900* were found in the genomes studied.
- Nine insertion site locations were conserved across all genomes studied.
- Analysis of insertion sites reveal that IS*900* rarely insert within coding sequences
- The *in silico* IS900 RFLP profiles can be compared by size or by digital visualization mimicking their migration in agarose gel.

Conclusion & Perspectives

- This study provide a new tool make it possible to automate the IS*900* distribution analysis in Map genome to enrich our knowledge on the dynamics of distribution of this IS for epidemic purpose in link of the evolution within Map species and study the biological implication of the presence of this IS*900*.
- The profiles and digital profiles can be compared to the many existing profiles in the literature using our dedicated application: http://mac-inmv.tours.inra.fr
- More genomes will be analyzed (particularly type S) to determine if there is any correlation between copy number and host lineage.

References : 1. Rathnaiah G, Front Vet Sci 2017 2. Biet F & Boschiroli ML, Res Vet Sci 2014 3. Pavlik I, J Mic Met 1999 4. Bull T, Microbiology 2000 5. Moran NA & Plague GR, 2004 6. Toft & Anderson, 2010 7. Li L, 2005 8. Brauning R, MRA 2019 9. Bannantine JP, MRA 2021 10. Biet F, BMC Microbiology 2012

Link to our published work in frontiers in Microbiology

SCAN ME

Cyril Conde – INRAE
cyril.conde@inrae.fr

Franck Biet – INRAE
franck.biet@inrae.fr