Comparative analysis of pSMA198 found in *Streptococcus macedonicus* ACA-DC 198, the first streptococcal plasmid of the pCI305/pWV02 family of theta-replicating replicons

Thomas Plakas [1,2], Rania Anastasiou [1], Marina Georgalaki [1], Ioanna-Areti Asteri [1], Stéphanie Ferreira [3], Philip Supply [3,4], Nikos C. Papandreou [2], Bruno Pot [4], Effie Tsakalidou [1] and Konstantinos Papadimitriou [1,*]

[1]Laboratory of Dairy Research, Department of Food Science and Technology, Agricultural University of Athens, Iera Odos 75, 118 55 Athens, Greece,
[2]Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 157 01, Greece,
[3]Genoscreen, Genomic Platform and R&D, Campus de l'Institut Pasteur, 1 rue du Professeur Calmette, 59000 Lille, France,
[4]Institut Pasteur de Lille, Center for Infection and Immunity of Lille (CIIL), F-59019 Lille, France

Correspondence to: kpapadimitriou@aua.gr

Abstract

Here we analyze pSMA198, the first plasmid isolated from *Streptococcus macedonicus* ACA-DC 198, and we attempt to clarify the route of its original acquisition. Based on the similarity profiles of the plasmid's replication initiation protein (Rep) and its origin of replication (*ori*), pSMA198 was found to be a novel member of the pCI305/pWV02 family of theta-replicating plasmids. The pCI305/pWV02 family consists of plasmids of narrow host range that are mainly found in lactococcal species. Comparative analysis of the pSMA198 revealed a high degree of similarity with plasmids pSK11b, pVF22 and pIL5 over its replication backbone, its mobilization backbone and most of its length, respectively. All these three plasmids have been isolated from *Lactococcus lactis* strains deriving from milk or its products supporting that *S. macedonicus* acquired pSMA198 from the latter species and that this acquisition took place in the dairy environment. Both pSMA198 and the chromosome of *S. macedonicus* exhibit a high degree of pseudogenes, indicating that they must have evolved under the same gene decay processes. Furthermore, we were able to determine chromosomal regions that may have originated from pSMA198, also supporting a long co-existence of the two replicons. In addition, pSMA198 is carried by *S. macedonicus* strains segregated in five different genotypes by pulsed-field gel electrophoresis (PFGE), showing that pSMA198's acquisition is not a recent event. We propose that our overall analysis of pSMA198 points towards the habituation of *S. macedonicus* ACA-DC 198 to the dairy environment.

Results and Discussion

Streptococcus macedonicus ACA-DC 198 carries a novel plasmid of 12,728 bp assigned as pSMA198. The plasmid has a 35.0% G+C content, lower than that of the *S. macedonicus* chromosome (37.6%), indicating that it may have been acquired from another organism. Overall 17 CDSs were annotated on pSMA198 (Fig.1 and Table 1).

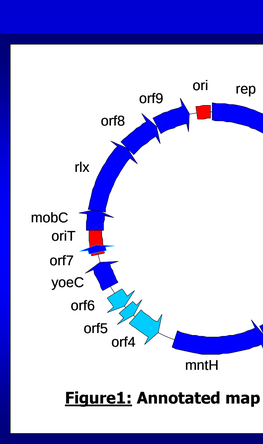

Figure1: Annotated map of pSMA198.

Table 1. Annotated features of pSMA198

locus_tag	gene	size nt	Best WU-Blastn hit (locus or locus_tag/ organism/ identity/ e-value)	Protein function
SMA_p0001	rep	1194	LACR_A06/ Lactococcus lactis subsp. cremoris SK11 plasmid 1/ 87%/ $1.7e^{-196}$	Initiator RepB protein
SMA_p0002	orfX	585	BN193_11490/ Lactococcus raffinolactis 4877/ 91%/ $2.2e^{-101}$	Replication associated protein
SMA_p0003	ydeE	858	ENT_30280 / Enterococcus sp. 7L76 / 99%/ $2.2e^{-184}$	AraC family transcriptional regulator
SMA_p0004	orf1	582	EfmE1039_1841/ Enterococcus faecium E1039/ 99%/ $4.6e^{-121}$	Integral membrane protein
SMA_p0005	yoeC	591	AF179848/ Lactococcus lactis subsp. lactis UC317 pCI305/ 88%/ $1.4e^{-97}$	Integrase/recombinase plasmid associated
SMA_p0006	orf2	459	CAC42047/ Listeria innocua Clip11262 pL1100/ 99%/ $3.6e^{-94}$	Putative pseudo
SMA_p0007	orf3	438	LACR_D31/ Lactococcus lactis subsp. cremoris SK11 plasmid 4/ 99%/ $4.8e^{-90}$	Universal stress protein family
SMA_p0008	mntH	1578	HMPREF0848_00725 / Streptococcus sp. C150/ 99%/ 0.0	Manganese transport protein MntH
SMA_p0009	orf4	480	llmg_pseudo_13/ Lactococcus lactis subsp. cremoris MG1363 pseudogene/ 97%/ $3.2e^{-214}$	Putative pseudo
SMA_p0010	orf5	195	llmg_pseudo_13/ Lactococcus lactis subsp. cremoris MG1363 pseudogene/ 97%/ $3.2e^{-214}$	Putative pseudo
SMA_p0011	orf6	276	llmg_pseudo_13/ Lactococcus lactis subsp. cremoris MG1363 pseudogene/ 97%/ $3.2e^{-214}$	Putative pseudo
SMA_p0012	yoeC	465	LACR_C25/ Lactococcus lactis subsp. cremoris SK11 plasmid 3/ 98%/ $3.5e^{-93}$	Integrase/recombinase plasmid associated
SMA_p0013	orf7	132	pIL7_28/ Lactococcus lactis subsp. lactis IL594 plasmid pIL7/ 84%/ $1.2e^{-13}$	Putative pseudo
SMA_p0014	mobC	366	HMPREF9519_01999/ Enterococcus faecalis TX1346/ 89%/ $5.7e^{-61}$	Mobilisation protein
SMA_p0015	rlx	1233	CI5MOBPRO/ Lactococcus lactis cremoris UC503 pCI528/ 99%/ $2.7e^{-268}$	Mobilization protein
SMA_p0016	orf8	627	ENT_30400 / Enterococcus sp. 7L76/ 96%/ $6.1e^{-124}$	Conserved hypothetical protein
SMA_p0017	orf9	603	BN193_11500/ Lactococcus raffinolactis 4877/ 99%/ $3.4e^{-125}$	Fic family protein

The first gene codes for a replication initiation protein (Rep). The *rep* gene showed 87% identity (e-value $1.7e^{-196}$) with the respective gene found on plasmid 1 of *L. lactis* subsp. *cremoris* SK11. Among the WU-Blast hits of Rep we identified the RepB proteins of the pCI305 and the pWV02 plasmids (78% identity, e-value $7.4e^{-162}$ and 75% identity, e-value $2.6e^{-152}$, respectively) that are the prototypes of the pCI305/pWV02 family of the lactococcal theta-replicating plasmids. Multiple sequence alignment of the top Rep hits including the RepB proteins of pCI305 and pWV02 with ClustalW revealed a high degree of conservation over most of their length (Fig. 2).

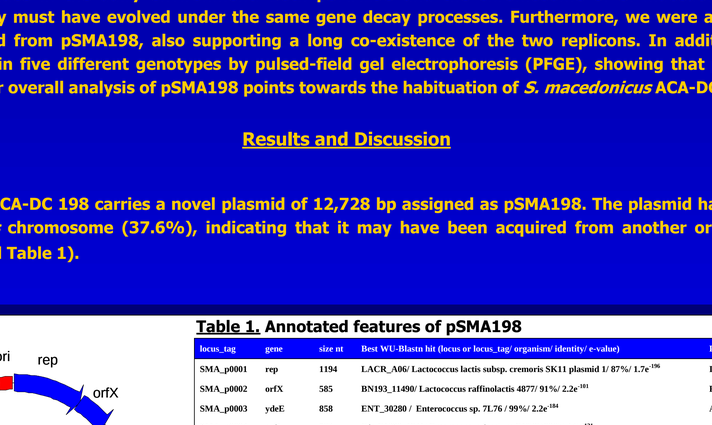

Figure 2: Multiple sequence alignment of RepB sequences relevant to the respective protein of pSMA198 performed with ClustalW. Double-headed arrows indicate the positions of the sequence signatures within the proteins as determined by InterProScan as follows: initiator Rep protein (dashed line), *L. lactis* RepB C-terminal (dotted line) and two consecutive winged helix-turn-helix transcription repressor DNA-binding (solid line).

Subsequently, we looked upstream of the rep gene in an attempt to identify the origin of replication (*ori*) of pSMA198. WU-Blastn similarity searches and multiple sequence alignment directed us towards a pCI305/pWV02 type of *ori* (Fig. 3). Indeed, we determined a segment spanning 242 nucleotides that contains an AT-rich region, three and a half direct repeats (DRs) of 22-bp iterons and two inverted repeats (IRs). This structure includes all the necessary elements for the binding of the required protein complexes, the melting of DNA, the initiation of the replication and the control of PCN. The pattern of the pSMA198 *ori* along with the similarity of its Rep with the lactococcal RepB shows that pSMA198 is certainly a member of the narrow host range pCI305/pWV02 family of replicons, which are normally found in *Lactococcus* species.

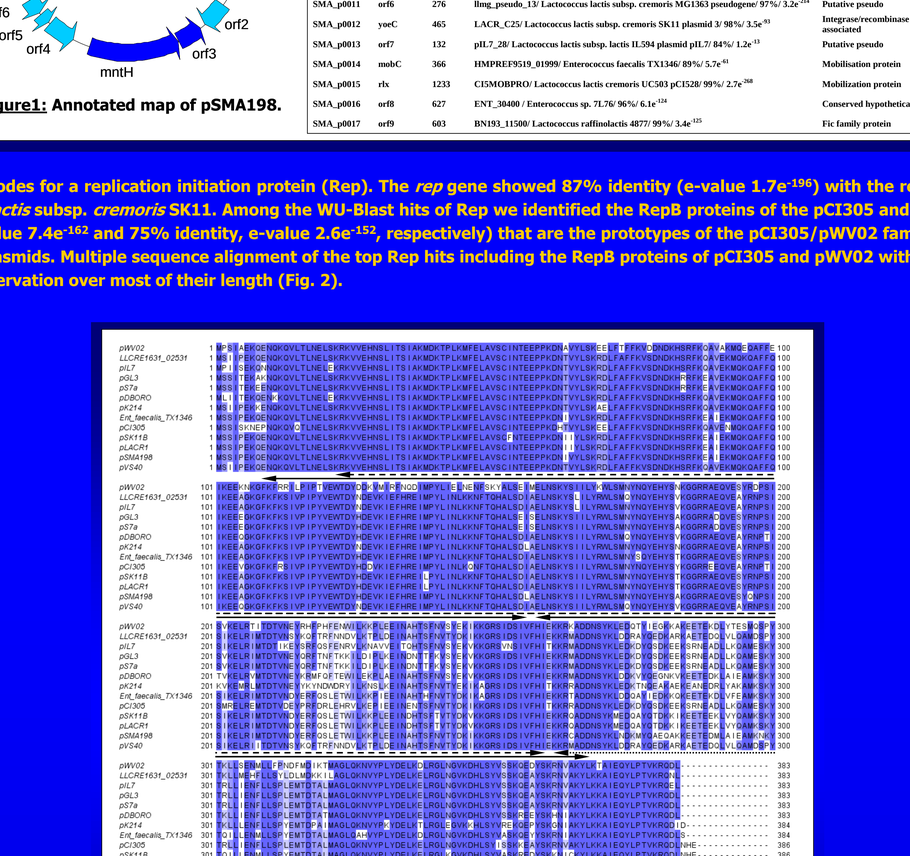

Figure 3: Multiple sequence alignment of the *ori* of pSMA198 and its related plasmids performed using ClustalW. Arrows indicate the position of the AT-rich region, the 22-bp direct repeat (DR) iterons and the two inverted repeats (IR). The promoter (-35, -10 and the RBS) and the start codon of the *rep* gene are underlined.

Acknowledgments

The present work was cofinanced by the European Social Fund and the National resources EPEAEK and YPEPTH through the Thales project.

Even though pSMA198 is not a self-transmissible plasmid, a cis-acting origin for transfer (*oriT*) that would allow its mobilization in the presence of a true conjugative plasmid was also predicted upstream of *mobC* (Fig. 4). The *oriT* exhibited a region of six consecutive IRs and two DRs. In addition, we determined an identical nick site to those proposed previously for plasmids pS7a and pS7b, eight bases after the end of IR3. The nick site is where the mobilization nucleases cleave duplex DNA during transfer. Once more, these structures were highly conserved among several lactococcal plasmids including pCI305.

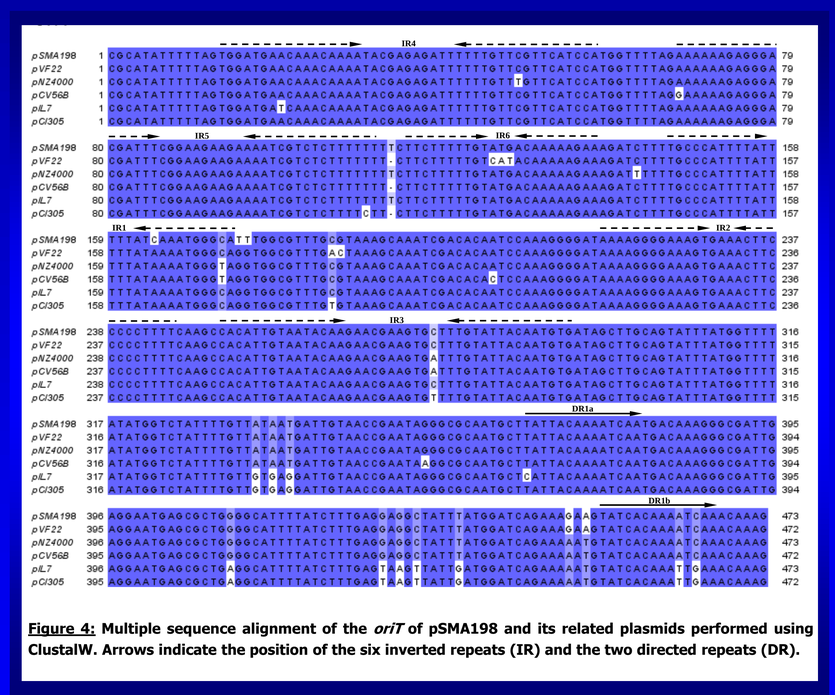

Figure 4: Multiple sequence alignment of the *oriT* of pSMA198 and its related plasmids performed using ClustalW. Arrows indicate the position of the six inverted repeats (IR) and the two directed repeats (DR).

The relation of pSMA198 to other plasmids was further investigated (Fig. 5). pSMA198 showed highest identity over the entire length of its replication (*ori-rep-orfX*) and mobilization (*oriT-mobC-rlx-orf8-orf9*) backbones to plasmids pSK11b (78% identity, e-value $8.4e^{-253}$) and pVF22 (96% identity, e-value 0.0), respectively. Plasmid pSK11b has been isolated from *L. lactis* subsp. *cremoris* SK11, which is a widely used industrial starter in cheese making, and plasmid pVF22 has been isolated from the raw milk cheese strain *L. lactis* subsp. *lactis* biovar. diacetylactis DPC3901. Interestingly, the similarity between pSMA198 and each of the two plasmids mentioned above was basically restricted to the loci under investigation (i.e. the replication or the mobilization backbones) (Fig. 5). This led us to look for the plasmid that would have the highest identity with the complete sequence of pSMA198. The plasmid identified was pIL5 that has been isolated from *L. lactis* subsp. *lactis* IL594, which is also a cheese starter. pIL5 exhibited 97% identity (e-value 0.0) over approximately the three quarters of the pSMA198 sequence (Fig. 5). It should be emphasized that apart from the closest similarity hits mentioned above, the overriding majority of top hits in all similarity searches for the different features annotated on pSMA198 at the protein or nucleotide level originated from *L. lactis* dairy strains. For example, nine out of the ten top hits for the replication backbone derived from strains isolated from milk or its products.

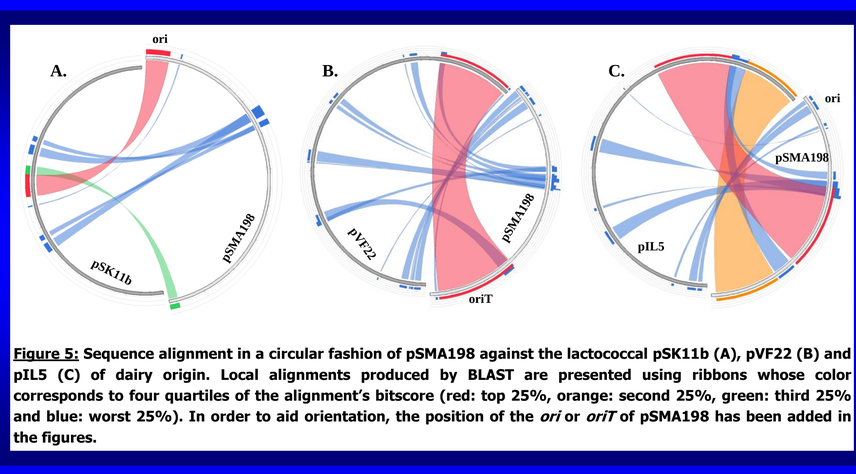

Figure 5: Sequence alignment in a circular fashion of pSMA198 against the lactococcal pSK11b (A), pVF22 (B) and pIL5 (C) of dairy origin. Local alignments produced by BLAST are presented using ribbons whose color corresponds to four quartiles of the alignment's bitscore (red: top 25%, orange: second 25%, green: third 25% and blue: worst 25%). In order to aid orientation, the position of the *ori* or *oriT* of pSMA198 has been added in the figures.

The possibility of genetic exchange between pSMA198 and the chromosome of *S. macedonicus* ACA-DC 198 was also examined. We were led to this hypothesis due to the existence of genes or pseudogenes of elements that would facilitate such an exchange in both the plasmid and the chromosome. In addition, the reduced size of pSMA198 as compared to some of its related plasmids, e.g. pVF22 and pIL5 (12 kb vs. 22 and 23 kb, respectively) indicates loss of genetic material, some of which could have moved to the chromosome. In the attempt to identify such regions, we employed two different strategies. We investigated for the presence of chromosomal genes showing high identity to the genes found on pSMA198 or its related plasmids (Fig. 6).

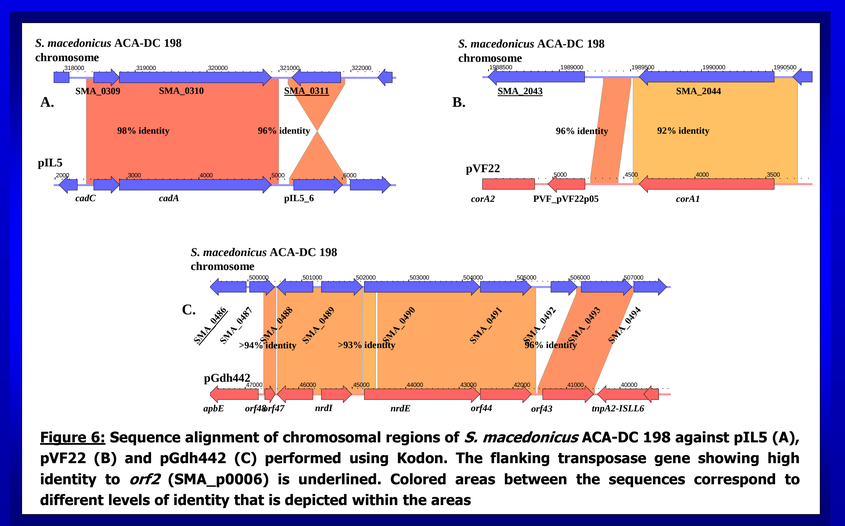

Figure 6: Sequence alignment of chromosomal regions of *S. macedonicus* ACA-DC 198 against pIL5 (A), pVF22 (B) and pGdh442 (C) performed using Kodon. The flanking transposase gene showing high identity to *orf2* (SMA_p0006) is underlined. Colored areas between the sequences correspond to different levels of identity that is depicted within the areas

Our findings demonstrate that pSMA198 is a novel member of the pCI305/pWV02 family of theta-replicating plasmids. The pCI305/pWV02 replicon has been shown to be of narrow host range, mainly replicating in *Lactococcus* spp. pSMA198 is the first streptococcal plasmid to be described within this family. Based on the levels of identity of the pSMA198 replication backbone, which may reflect its evolutionary history, to the respective backbones of other plasmids, our data supports that *S. macedonicus* acquired pSMA198 from *L. lactis*. This exchange took place most probably in the milk environment as the overriding majority of closest related plasmids to pSMA198 (i.e. pSK11b, pVF22 and pIL5) are of dairy origin. This is in agreement with the isolation source of *S. macedonicus* ACA-DC 198 and the fact that *S. macedonicus* strains have dairy products as their primary ecological niche. The acquisition of pSMA198 by *S. macedonicus* ACA-DC 198 seems not to be a recent event. Prominently, the fact that the chromosome of *S. macedonicus* ACA-DC 198 and pSMA198 exhibit a high percentage of pseudogenes, indicates that they may have both evolved under the same gene decay processes. In addition, the potential exchange of genetic material between the chromosome and the plasmid designates a long co-existence of the two replicons. We propose that our overall analysis of pSMA198 points towards the habituation of *S. macedonicus* ACA-DC 198 to the dairy environment.

Bibliography

• Papadimitriou K., S. Ferreira, N. C. Papandreou, E. Mavrogonatou, P. Supply, B. Pot, and E. Tsakalidou (2012) Complete genome sequence of the dairy isolate *Streptococcus macedonicus* ACA-DC 198. J Bacteriol 194:1838-9.

• Papadimitriou K., T. Plakas, R. Anastasiou, S. Ferreira, P. Supply, P. Renault, N. C. Papandreou, B. Pot, and E.Tsakalidou (In preparation) Analysis of the lactococcal plasmid pSMA198 found in *Streptococcus macedonicus* ACA-DC 198 points towards the habituation of the strain to the dairy environment.