Characterization of plasmid pSMA198 found in *Streptococcus macedonicus* ACA-DC 198 supports the relation of the species to the milk environment

Konstantinos Papadimitriou1, Thomas Plakas1, Rania Anastasiou1, Nikos C. Papandreu2, Stavros J. Hamodrakas2, Stéphanie Ferrère3, Philippe Supp4, Pierre Renaud5, Bruno Pot6 and Effie Tsakalidou1

1Laboratory of Dairy Research, Department of Food Science and Technology, Agricultural University of Athens, Iera Odos 75, 118 55 Athens, Greece, 2Department of Cell Biology and Biochemistry, Faculty of Biology, University of Athens, Perama 157 71, Greece, 3Génomique, Génétique, Protéomique et RRR, Campus du Froidefilet, 7 rue de Provence, 59932 Lille, France, 4IMBA, UMR1139, Faculté de Vétérinaire, Jouy-en-Josas, France

**Correspondence to:** k.papadimitriou@aua.gr

**Acknowledgments:**

The present work was financed by the European Social Fund and the National resources SPEAK and YVFTH through the Thales project.

### Abstract

Background: *Streptococcus macedonicus* is an intriguing streptococcal species whose most frequent source of isolation is fermented foods similarly to *Streptococcus thermophilus*. During the genome sequencing of *S. macedonicus* ACA-DC 198 a plasmid was identified.

Objectives: To analyse pSMA198, the plasmid isolated from *S. macedonicus* and its role and utility in the breeding process.

Methods: Similarity searches of nucleotide and protein sequences, comparative analysis of whole plasmid sequences and phylogenetic analysis were performed using the respective bioinformatic tools.

Results and Discussion

*Streptococcus macedonicus* ACA-DC 198 carries a novel plasmid of 12.7 kb assigned as pSMA198. The plasmid has a 36.3% G+C content, lower than that of the *S. macedonicus* chromosomal DNA (37.4%), indicating that they may have been acquired from another organism. Overall 57 CDS were annotated on pSMA198 (Fig. 1 and Table 1).

Even though pSMA198 is not a self-transmissible plasmid, a co-acting origin for transfer (oriT) that would allow its mobilization in the presence of a conjugative plasmid was also predicted upstream of *yoeC* (Fig. 4). The oriT regions show a high content of direct repeats (22 bp) and their DNA. In addition, we determined an identical size to those present previously for the plasmids pVF2 and pVF7, eight bases after the end of the oriT. This site is also to where the mobilization increases clean duplex DNA during transfer. Once more, these structures were highly conserved among several lactococcal plasmids including pCH5.

The relations of pSMA198 to other plasmids was further investigated (Fig. 5). pSMA198 showed highest identity over the entire length of its replication region with the lactococcal plasmid pIL5 (96% identity), and in its other regions with the lactococcal plasmids pVF2 and pVF7. We also found that pSMA198 possesses an AT-rich region (96% identity, orf6, respectively) that are the prototype of the pIL560000 orf22(2), orf23, orf24 (96% identity, 84% and 93% identity, respectively). Among the pSMA198 plasmids, pVF2 and pVF7 plasmids are the closest (87% identity), followed by pIL5 (82% identity) and pVF2, where pSMA198 and pVF2 plasmids are the closest (79% identity, orf6, respectively). Plasmid pVF2 has been isolated from *L. lactis* subsp. cremoris. In this study, the similarity between pSMA198 and each of the two plasmids mentioned above was basically maintained in the last under-transcription (Fig. 5a). The replication region of the most conserved backbones (Fig. 5b). This led us to find that the backbone that would have the highest identity with the complete sequence of the lactococcal plasmid pIL5 (96%) that has been isolated from *L. lactis* subsp. cremoris ACA-DC 198, which is also a closer starter that the closest one recorded. The repetition region has a 97% identity (nucleotide) over approximately three-quarters of the pSMA198 sequence (Fig. 5f). It should be emphasised that from the closest starters, the differences were associated to the additional DNA in pSMA198 and the different backbone was associated to the orf6 in the repeat region and the nucleotide level originated from *L. lactis* dairy strains. For example, nine out of the ten top hits for the backbone backbone derived from strains isolated from milk or its products.

The possibility of genetic exchanges between pSMA198 and the chromosome of *S. macedonicus* ACA-DC 198 was also examined. We were interested in this hypothesis due to the nature of the organism. We thereby sequenced the plasmids pSMA198 and pIL5 to find evidence for the existence of such exchanges, which could be facilitated by scavenging DNA elements that would facilitate these processes. We sequenced the plasmids pSMA198 and pIL5 to find evidence for the existence of such exchanges, which could be facilitated by scavenging DNA elements that would facilitate these processes. The starting point for this approach was the fact that *S. macedonicus* ACA-DC 198 contains a repetitive region that is rich in GC repeats, which could be used as a marker for the identification of such regions, which have been used for different strategies. We investigated the presence of these common regions showing high identity to the genes found on pSMA198 and the related plasmid.

The plasmid pSMA198 is a novel member of the pIL560000/pVF2 family of theta-replicating plasmids. The pIL500000/pVF22 replica has a co-acting origin for transfer (oriT) that would allow its mobilization in the presence of a conjugative plasmid. The plasmids pSMA198 and pVF2 can be considered as part of a new family of lactococcal plasmids.

In addition, the plasmid pSMA198 is capable of co-mobilization in *S. macedonicus* ACA-DC 198 and the fact that it can be transferred by a conjugative mechanism is of technological value. The acquisition of pSMA198 by *S. macedonicus* ACA-DC 198 seems to be a recent event. For this reason, the chromosome of *S. macedonicus* ACA-DC 198 is enriched for genes that were present in the plasmid pSMA198. For this reason, the chromosome of *S. macedonicus* ACA-DC 198 is enriched for genes that were present in the plasmid pSMA198.

Our findings demonstrate that pSMA198 is a novel member of the pIL560000/pVF22 family of theta-replicating plasmids. The pIL500000/pVF22 replica has a co-acting origin for transfer (oriT) that would allow its mobilization in the presence of a conjugative plasmid, which could facilitate these processes. Based on our analysis, we investigated the presence of these common regions showing high identity to the genes found on pSMA198 and the related plasmid. Annotated map of pSMA198 (Fig. 1).

### Table 1.

<table>
<thead>
<tr>
<th>Map of pSMA198</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 1.</strong> Annotated Map of pSMA198.</td>
</tr>
<tr>
<td><strong>Table 2.</strong> Multiple sequence alignment of pSMA198 and its related plasmids.</td>
</tr>
</tbody>
</table>

**Table 2.** Multiple sequence alignment of pSMA198 and its related plasmids. The arrows indicate the position of the atf-rich regions. The 22-bp direct repeat (DR) flankers and the two inverted repeats (IR). The promoter (P) bands and the IRs and the start codons of the gene are underlined.

**Figure 1.** 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).

**Figure 2.** Multiple sequence alignment of pSMA198 and its related plasmids using Blast2. The arrows indicate the position of the atf-rich regions. The 22-bp direct repeat (DR) flankers and the two inverted repeats (IR).

**Figure 3.** Sequence alignment in a circular feature of pSMA198 against the bacteriocin pJK18 (A), pJK22 (B) and pIL5 (C) of origin. Local alignments produced by BLAST are presented using ribbons whose color corresponds to the quartile of the alignment's identity (red: top 25%, orange: second 25%, green: third 25% and yellow: 25%). In order to identify the position of the oriT of pSMA198, it has been added to the figure. (D) Maximum likelihood tree of the pJK18 of origin generated using the Phylogen JJ pipeline.

**Acknowledgments:**

The present work was financed by the European Social Fund and the National resources SPEAK and YVFTH through the Thales project.