Adaptation of *Streptococcus macedonicus* and *Streptococcus thermophilus* in milk. Common strategies, distinct ways.

Maria Kazou[1], Voula Alexandraki[1], Pierre Renault[2], Bruno Pot[3], Effie Tsakalidou[1] and Konstantinos Papadimitriou[1]

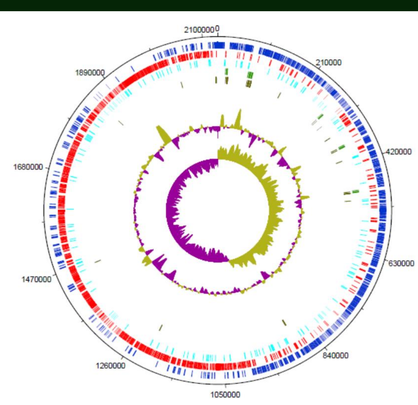
[1]Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Iera Odos 75, 118 55, Athens, Greece.
[2]INRA, UMR1319 Micalis, Jouy-en-Josas, F-78352, France; AgroParisTech, UMR Micalis, Jouy-en-Josas, F-78352, France.
[3]Institut Pasteur de Lille, Center for Infection and Immunity of Lille (CIIL), F-59019, Lille, France; Inserm U1019, F-59019, Lille, France; CNRS UMR8204, F-59019, Lille, France; Univ Lille de Nord France, F-59019, Lille, France.

Abstract

Lactic acid bacteria (LAB) are the most important bacteria in food fermentations. Among the genera included in LAB is the *Streptococcus* genus, containing a number of species that are commensals and opportunistic pathogens. Up to now only one species in the *Streptococcus* genus, *Streptococcus thermophilus*, has been traditionally used as a starter in milk and is considered a domesticated organism. However, *Streptococcus macedonicus*, belonging to the *S. bovis*/*S. equinus* complex (SBSEC), can be also found in milk. In this study, *S. macedonicus* was fully sequenced and a comparative analysis was performed against all other SBSEC members. According to our analysis, *S. macedonicus* missed several genes encoding enzymes for the degradation of complex plant carbohydrates typically met in the genome of *Streptococcus gallolyticus*, indicating a reduced ability to survive in the gastrointestinal tract of herbivores. In addition, two pilus operons necessary for the adhesion of *S. gallolyticus* to the host and the initiation of infection were absent from the *S. macedonicus* genome. Adaptation of *S. macedonicus* to the milk environment was supported by the presence of an extra lactose operon. Finally, the plasmid pSMA198 found in *S. macedonicus* also provided evidence about the habituation of the species to milk. Our findings support adaptation of *S. macedonicus* to the rich in nutrients milk environment characterized by diminished biosynthetic capabilities and loss of pathogenicity-related genes in parallel to gene gain events through horizontal gene transfer, similarly to *S. thermophilus*. Comparative analysis between *S. macedonicus* and *S. thermophilus* though revealed numerous species-specific differences at the genomic and proteomic level, suggesting that common strategies can be used by truly divergent organisms during their evolutionary history

Annotating the genome of *S. macedonicus* ACA-DC 198

- **1st step: initial annotation was performed with the BaSys and the RAST pipelines**
- **2nd step: annotations were manually compiled in one using Kodon software**
- **3rd step: final corrections and quality assessment was performed using GenePRIMP (including predictions for potential pseudogenes)**



2,192 protein coding genes
192 potential pseudogenes
18 rRNA genes
70 tRNA genes

Fig. 1. Genome map of *S. macedonicus* ACA-DC 198

Comparative genomics of *S. macedonicus*

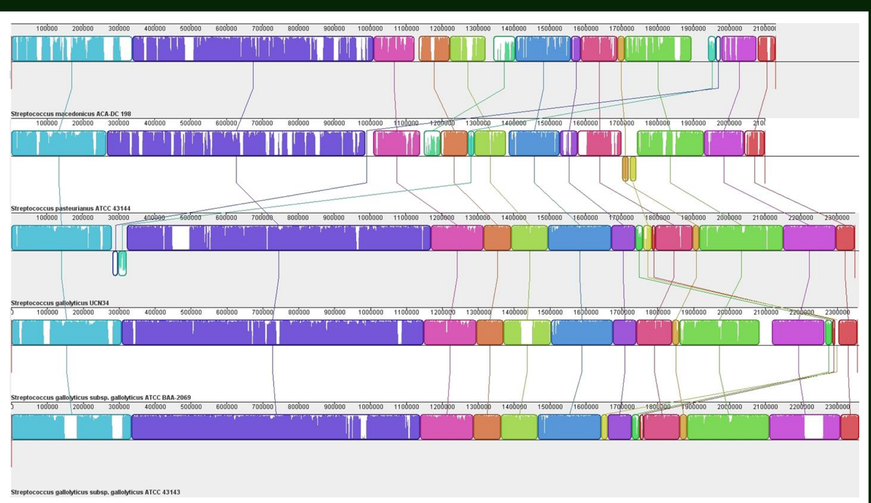


Fig. 2. Complete alignment of the *S. macedonicus* ACA-DC 198 genome against those of related species using local collinear blocks (LCBs)

Full chromosome alignments were performed using local collinear blocks (LCBs) among three SBSEC species. The analysis revealed a mosaic pattern of homology (Fig. 2). Evidently, a significant portion of the genetic information has been overall conserved, since the majority of the LCBs are shared by all species over most of their genome length. It should be noted that numerous strain-specific differences can also be detected. Furthermore, there are LCBs common only among some of the strains, while there are regions divergent enough so as not to be placed within a LCB. These findings indicate that, apart from gene loss through genome decay, gene gain events like lateral gene transfer (LGT) must have played a role during the evolution of the three species. In addition, chromosomal rearrangements seem to have been rather minimal, as the number of "movable" LCBs was low and their length was short. Inclusion of the *S. infantarius* genome in the analysis increased significantly the number of LCBs and reduced drastically the level of conservation among the genomes (data not shown), indicating that this particular genome is fairly different from the rest.

Niche-specific and pathogenicity genes presence/absence

S. gallolyticus UCN 34 locus_tag	gene	function	*S. gallolyticus* ATCC BAA 2069	*S. gallolyticus* ATCC 43143	*S. macedonicus* ACA-DC 198	*S. pasteurianus* ATCC 43144	*S. infantarius* CJ18
gallo_0112	fruA	fructan hydrolase	√	√	-	-	-
gallo_0330	-	beta-1,4-endoglucanase (cellulase)	√	√	-	-	-
gallo_0757	-	a-amylase	√	√	-	-	-
gallo_0162	-	mannase	√	√	-	-	pseudo
gallo_0189	-	endo-beta-1,4-galactanase	√	√	pseudo	√	pseudo
gallo_1577	-	pectate lyase	√	√	√	pseudo	√
gallo_1578	-	pectate lyase	√	√	√	pseudo	√
gallo_1632	amyE	a-amylase	√	√	√	pseudo	√
gallo_0933	tanA	tanins degradation	√	√	-	-	-
gallo_1609	similar to tanA	tanins degradation	√	√	-	-	-
gallo_2106	padC	gallic acid decarboxylation	√	√	√	√	-
gallo_0906		gallic acid decarboxylation	√	√	√	√	-
gallo_0818	bsh	bile salt hydrolase	√	√	pseudo	√	√
gallo_2179	-	accessory pilin (pil1)	√	√	-	-	-
gallo_2178	-	major pilin (pil1)	√	√	-	-	-
gallo_2177	-	sortase C (pil1)	√	√	-	-	-
gallo_1570	-	accessory pilin (pil2)	√	√	-	-	-
gallo_1569	-	major pilin (pil2)	√	√	-	-	-
gallo_1568	-	sortase C (pil2)	√	√	-	-	-
gallo_2040	-	accessory pilin (pil3)	√	√	√	√	√
gallo_2039	-	major pilin (pil3)	√	√	√	√	√
gallo_2038	-	sortase C (pil3)	√	√	√	√	√

Our findings clearly suggest that not only *S. macedonicus*, but also *S. pasteurianus* and *S. infantarius* have deviated from *S. gallolyticus* in their potential to catabolize complex plant carbohydrates and to cope with the harsh environment of the GI tract of herbivores. Furthermore, *in silico* analysis of *S. gallolyticus* has revealed that it contains three pilus gene clusters (*pil1*, *pil2*, *pil3*), which may mediate its binding to the extracellular matrix (ECM), although variations of pilus genes presence/absence within strains have also been reported. Each gene cluster consists of three genes. The first two genes encode two adhesins belonging to the MSCRAMM (microbial surface recognizing adhesive matrix molecules) family, one being the major and one being the minor (or accessory) pilus subunit. Pilus attachment to the peptidoglycan, as well as polymerization of adhesin filaments are catalyzed by a sortase C encoded by the third gene of the cluster. *pil1* and *pil2* loci are absent in *S. macedonicus*, *S. pasteurianus* and *S. infantarius* indicating a diminished tendency to adhere to ECM that could probably influence their ability to colonize host tissues and to produce infections when compared to *S. gallolyticus*.

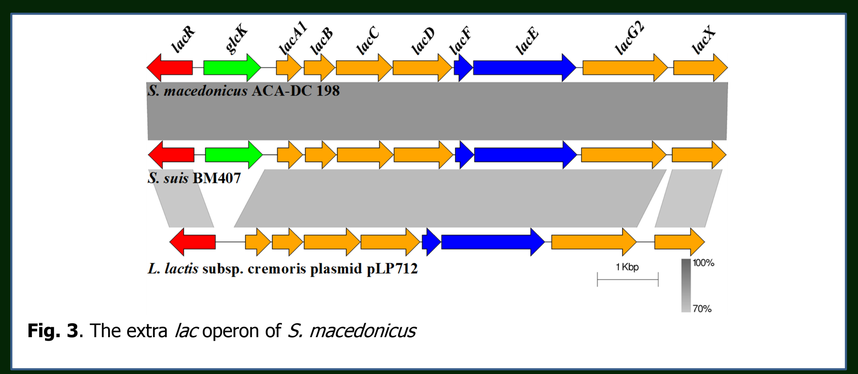


Fig. 3. The extra *lac* operon of *S. macedonicus*

***S. macedonicus* is the only member of the SBSEC that has a second gene cluster for the accumulation of lactose and the degradation of galactose that is also present in *S. suis* and that our analysis indicates that it may originate from *L. lactis* (Fig. 3).**

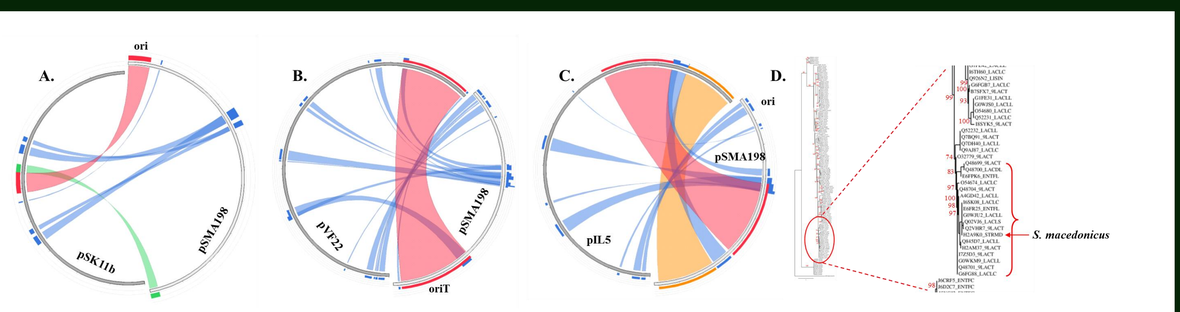


Fig. 4. Sequence alignment in a circular fashion of pSMA198 against the lactococcal pSK11b (A), pVF22 (B) and pIL5 (C) of dairy origin. Local alignments produced by BLAST are presented using ribbons whose color corresponds to four quartiles of the alignment's bitscore (red: top 25%, orange: second 25%, 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. (D) Maximum likelihood tree of the pSMA198 Rep generated using the Phylogeny.fr pipeline

The relation of pSMA198 to other plasmids was further investigated (Fig. 4). 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. 4). 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. 4). 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.

Comparative analysis among *S. thermophilus* and *S. macedonicus* strains

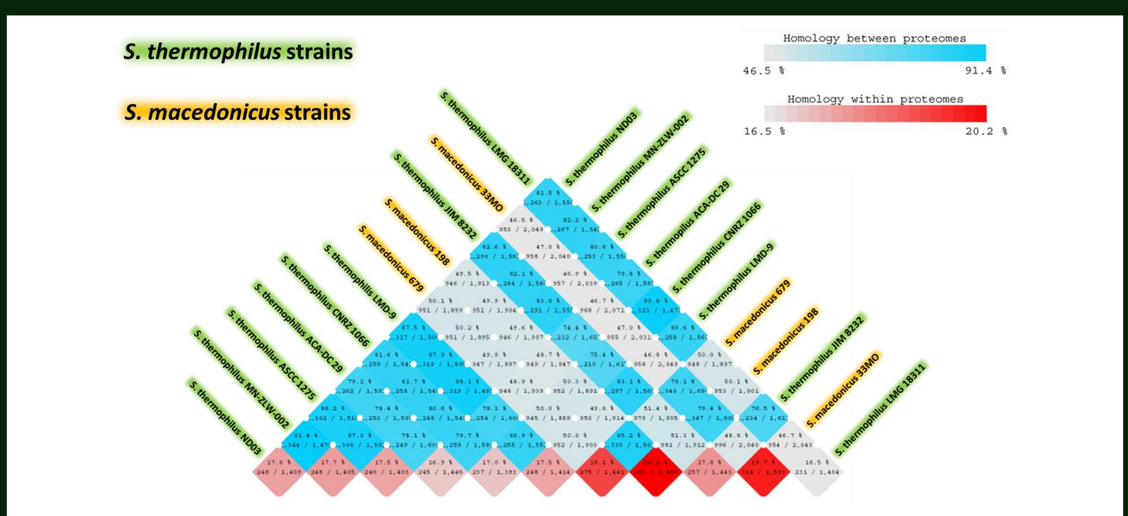


Fig. 5. Comparative proteome analysis among *S. thermophilus* and *S. macedonicus* strains

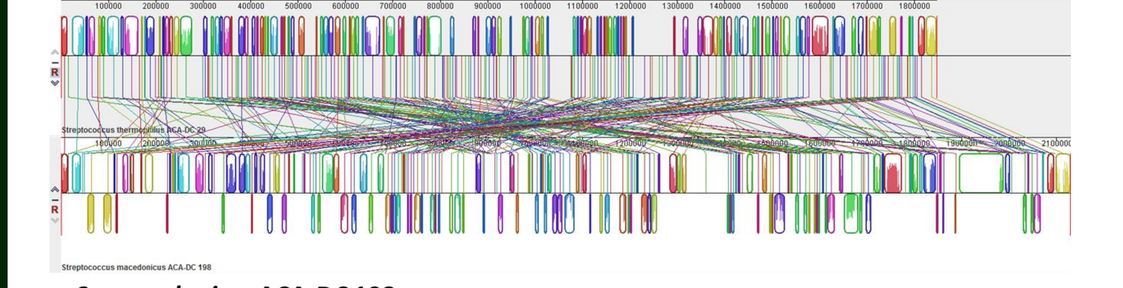


Fig. 6. Comparative genome analysis between *S. thermophilus* ACA-DC 29 and *S. macedonicus* ACA-DC 198

Comparative proteome analysis of the three available *S. macedonicus* strains and all *S. thermophilus* strains with complete genome sequences revealed low degree of conservation among them (Fig. 5). This was also exemplified during the full chromosome alignments of two arbitrary selected strains, i.e. *S. thermophilus* ACA-DC 29 and *S. macedonicus* ACA-DC 198 (Fig. 6).

Conclusions

1. *S. macedonicus* is most probably a separate species from *S. gallolyticus*

2. *In silico* analysis of *S. macedonicus* ACA-DC 198 suggests that:

- **The strain is at the process of adapting to a rich in nutrients environment**
- **It shows a diminished capacity to live and survive in the GI tract of herbivores**
- **It has a diminished pathogenic potential compared to *S. gallolyticus***

3. *S. macedonicus* and *S. thermophilus* demonstrate similar strategies for adapting to the milk environment, but they have followed diverge evolutionary pathways

Bibliography

• Papadimitriou K, Anastasiou R, Mavrogonatou E, Blom J, Papandreou NC, Hamodrakas SJ, Ferreira S, Renault P, Supply P, Pot B, Tsakalidou E. Comparative genomics of the dairy isolate *Streptococcus macedonicus* ACA-DC 198 against related members of the *Streptococcus bovis*/*Streptococcus equinus* complex. BMC Genomics. 2014 Apr 8;15:272.

Acknowledgements

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

Γεωπονικό Πανεπιστήμιο Αθηνών
Agricultural University of Athens

NSRF 2007-2013 programme for development