Acquisition through Horizontal Gene Transfer of Plasmid pSMA198 by *Streptococcus macedonicus* ACA-DC 198 Points towards the Dairy Origin of the Species

Konstantinos Papadimitriou¹*, Rania Anastasiou¹, Eleni Maistrou¹, Thomas Plakas¹,², Nikos C. Papandreu², Stavros J. Hamodrakas², Stéphanie Ferreira³, Philip Supply³,⁴,⁵,⁶,⁷, Pierre Renault⁸,⁹, Bruno Pot⁴,⁵,⁶,⁷, Effie Tsakalidou¹

¹Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Iera Odos 75, 118 55, Athens, Greece, ²Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, 157 01, Athens, Greece, ³Genoscreen, Service of Research, Development and Innovation in Health and Environment, Campus de l’Institut Pasteur, 1 rue du Professeur Calmette, 59000, Lille, France, ⁴Institut Pasteur de Lille, Center for Infection and Immunity of Lille (CIIL), F-59019, Lille, France, ⁵Inserm UMR8204, F-59019, Lille, France, ⁶CNRS UMR1319 Micalis, Jouy-en-Josas, F-78352, France, ⁷Univ Lille Nord France, F-59019, Lille, France, ⁸INRA, UMR1319 Micalis, Jouy-en-Josas, F-78352, France, ⁹AgroParisTech, UMR Micalis, Jouy-en-Josas, F-78352, France

* kpapadimitriou@aua.gr

Abstract

**Background**

*Streptococcus macedonicus* is an intriguing streptococcal species whose most frequent source of isolation is fermented foods similarly to *Streptococcus thermophilus*. However, *S. macedonicus* is closely related to commensal opportunistic pathogens of the *Streptococcus bovis/Streptococcus equinus* complex.

**Methodology/Principal Findings**

We analyzed the pSMA198 plasmid isolated from the dairy strain *Streptococcus macedonicus* ACA-DC 198 in order to provide novel clues about the main ecological niche of this bacterium. pSMA198 belongs to the narrow host range pCI305/pWV02 family found primarily in lactococci and to the best of our knowledge it is the first such plasmid to be reported in streptococci. Comparative analysis of the pSMA198 sequence revealed a high degree of similarity with plasmids isolated from *Lactococcus lactis* strains deriving from milk or its products. Phylogenetic analysis of the pSMA198 Rep showed that the vast majority of closely related proteins derive from lactococcal dairy isolates. Additionally, cloning of the pSMA198 ori in *L. lactis* revealed a 100% stability of replication over 100 generations. Both pSMA198 and the chromosome of *S. macedonicus* exhibit a high percentage of potential pseudogenes, indicating that they have co-evolved under the same gene decay processes. We identified chromosomal regions in *S. macedonicus* that may have originated from pSMA198, also supporting a long co-existence of the two replicons. pSMA198 was also found in divergent biotypes of *S. macedonicus* and in strains isolated from dispersed...
geographic locations (e.g. Greece and Switzerland) showing that pSMA198’s acquisition is not a recent event.

Conclusions/Significance
Here we propose that *S. macedonicus* acquired plasmid pSMA198 from *L. lactis* via an ancestral genetic exchange event that took place most probably in milk or dairy products. We provide important evidence that point towards the dairy origin of this species.

Introduction
Lactic acid bacteria (LAB) form the most important group of microorganisms used in food fermentations. Several LAB species have a long history of unproblematic use as starters and they are thus considered as safe including a number of probiotic strains that may provide the consumers with health benefits [1]. However, food-related LAB are often phylogenetically related to commensal LAB that may be opportunistic or even invasive pathogens. The most striking example of this situation is probably the genus *Streptococcus*, in which only one species, *Streptococcus thermophilus* is considered safe. Since centuries, *S. thermophilus* is a dairy starter frequently consumed by humans world-wide. We now know that this species is phylogenetically related to notorious streptococcal pathogens such as Group A and Group B streptococci, as well as to *Streptococcus pneumoniae*. Studies of the *S. thermophilus* genome, however, have revealed extensive gene decay, leading to the loss of streptococcal pathogenic determinants during its adaptation to milk [2,3]. In addition, the species was shown to have acquired genes from species like *Lactobacillus delbrueckii* subsp. *bulgaricus* with whom it often shares a symbiotic relationship in the dairy environment [4]. These attributes along with the long empiric experience of its safe use support the benign nature of *S. thermophilus* [5].

Apart from *S. thermophilus*, certain members of the *Streptococcus bovis/Streptococcus equinus* complex (SBSEC), i.e. *Streptococcus macedonicus* and *Streptococcus infantarius*, have also been associated with fermented foods [6,7]. After its original description *S. macedonicus* has been identified in a wide range of dairies around the world and current studies indicate that milk and its products represent its primary ecological niche [6]. Similarly to *S. macedonicus*, strains of *S. infantarius* have been isolated from the food environment [8,9]. Both species are clearly related to the other members of the SBSEC, such as the commensals *Streptococcus galloylgyticus* and *Streptococcus pasteurianus* that are associated with endocarditis, bacteremia and colon cancer [10,11]. *Streptococcus infantarius* has also been implicated with human disease [12], while there are only few studies suggesting the presence of *S. macedonicus* in cases of endocarditis [13,14]. Analysis of the recently sequenced genomes of *S. macedonicus* ACA-DC 198 (isolated from traditional Greek Kasseri cheese) and of *S. infantarius* CJ18 (isolated from suusac, a type of traditionally Kenyan fermented camel milk) has facilitated the assessment of the adaptation of these strains to milk [12,15]. Both strains present traits of adaptation to the dairy environment. For example, *S. macedonicus* appears to have evolved under selective pressures that favored genome decay and gene loss processes, resembling the evolution pattern of *S. thermophilus* [16]. The strain has lost important genes present in the rumen bacterium *S. galloylgyticus* that are involved in complex plant carbohydrate catabolism and in the detoxification of toxic substances met in the gastrointestinal tract of herbivores (e.g. tannins and gallic acid), thus indicating deviation from this environment. Furthermore, the strain owns a second
gene cluster for lactose catabolism unique within the SBSEC, a proteolytic system capable of degrading milk proteins and enhanced resistance to phages like that of known dairy starters. Analogous dairy adaptations were described for *S. infantarius* CJ18 [17]. Perhaps the most convincing evidence suggesting milk as the ecological niche of the two strains is the detection of interactions through horizontal gene transfer (HGT) with the dairy *Lactococcus lactis* and *S. thermophilus*. Such HGT events seem to be specific for *S. macedonicus* and *S. infantarius* and to be absent from *S. galloyticus* and *S. pasteurianus*. Events like these in the evolutionary history of the two species suggest a longstanding relationship with milk and fermented dairy products, even though their actual presence in these niches was only recently confirmed [6,9]. Among the most common HGT mechanisms is conjugation which involves the transfer of DNA by plasmids through plasmid encoded cell-to-cell junctions [18]. Conjugation may also result in the transfer of plasmids that are not conjugative but they carry the necessary mobilization genes.

In this study we attempted to clarify the route of the original acquisition of pSMA198, the first plasmid isolated from *S. macedonicus* ACA-DC 198. Acquisition of pSMA198 by *S. macedonicus* may represent an apparent HGT event in the evolutionary history of the strain. We assessed the hypothesis that the plasmid was originally acquired by *S. macedonicus* from *L. lactis* and that this acquisition took place in the dairy environment. We also tested the replication efficiency of pSMA198 in *L. lactis*. Finally, we investigated whether lactococcal sequences found in the chromosome of *S. macedonicus* ACA-DC 198 could have originated from plasmid pSMA198. Our findings point towards the dairy origin of *S. macedonicus* ACA-DC 198.

### Materials and Methods

**Bacterial strains and culture conditions**

*Streptococcus macedonicus* ACA-DC 198 was originally isolated from traditional Greek Kasseri cheese [19]. The strain was routinely cultured in M17 (Oxoid Ltd., Basingstoke, Hampshire, United Kingdom) supplemented with either 1% (w/v) lactose (LM17) or 0.5% (w/v) glucose (GM17). *Lactococcus lactis* subsp. *cremoris* MG1363 was grown in GM17. Stocks of these bacteria were stored at −80°C in cryovials filled with 1 ml of GM17 supplemented with 20% (v/v) glycerol. *Escherichia coli* HST08 strain (Stellar Competent Cells, Clontech Inc., Mountain View, CA) was grown in Luria Bertani (LB) medium under aerobic conditions. Antibiotics and other supplements were added in the media at appropriate concentrations as indicated below.

**Sequencing and annotation of pSMA198**

The sequence of pSMA198 was identified during the genome sequencing of *S. macedonicus* ACA-DC 198 as described previously [15]. In brief, total DNA of *S. macedonicus* was subjected to pyrosequencing according to the manufacturer’s instructions on a 454 GS-FLX Titanium sequencer (Roche Diagnostics, Basel, Switzerland) (Genoscreen, Lille, France). A round of shotgun pyrosequencing was followed by a 3-kb paired-end pyrosequencing. The acquired sequences were assembled using Newbler into 7 scaffolds, 5 of which could be subsequently aligned against a *NheI* genomic optical map of *S. macedonicus* (OpGen Technologies, Inc., Madison, WI). Further analysis supported that one of the two unaligned scaffolds was a plasmid. The complete sequence of pSMA198 was finally acquired at a >100X coverage through hybrid assembly of the pyrosequencing reads with reads obtained after an additional round of Hiseq2000 (Illumina, San Diego, CA) sequencing (Fidelity Systems, Inc., Gaithersburg, MD) (accession number: HE613570).
Gene prediction and annotation of the plasmid was performed using two ab initio predictor pipelines, RAST [20] and BaSys [21]. We also employed manual WU-Blast similarity searches for the detailed functional annotation of deduced proteins [22]. Pseudogenes were considered on the basis of presence or absence of ribosome-binding sites (RBS), of the coding sequence length and of frame shifts resulting in split genes. The origin of replication (ori) and the origin of mobilization (oriT) were identified based on previously determined sequences in plasmids related to pSMA198.

Nucleotide and protein analysis and phylogenetic analysis

In addition to WU-BlastN and WU-BlastP searches, nucleotide and protein sequences were aligned using ClustalW [23]. Miscellaneous features within the alignments (i.e. direct repeats, inverted repeats and other important sequences) were determined manually based on the available literature. We used InterProScan [24] to detect motifs within protein sequences. For phylogenetic analysis multiple sequence alignments were curated with Gblocks [25] and maximum likelihood trees were generated with the PhyML algorithm [26] as implemented in the Phylogeny.fr pipeline [27]. Branch support values were calculated with the $\chi^2$ parametric approximate likelihood-ratio test (aLRT) [28]. Graphical representation of the phylogenetic tree was carried out with TreeDyn [29]. Comparative analysis of pSMA198 to other plasmids was performed using Circoletto [30] and Kodon software (Applied Maths N.V., Sint-Martens-Latem, Belgium).

Construction of an *E. coli*—*S. macedonicus* shuttle vector based on the pSMA198 replicon

Primers used for the construction of the *E. coli*—*S. macedonicus* shuttle vector are presented in Table 1. Total DNA was extracted using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, St Louis, MO) and plasmids were isolated with the NucleoSpin Plasmid kit (Macherey-Nagel GmbH and Co. KG, Düren, Germany). PCR products were purified from

| Primer | Sequence* |
|--------|-----------|
| pr2Fn  | GAGCTCGGTACCCGGGGATCATGGACGAGCCACTCGTATC |
| pr2R   | CAGGTCGACTCTAGAGGATCTAGGACATGGATGAGTGA |
| ermF1  | TGATCTCTCTCTGATGGTCAGCAGCTGAGGATAATGAGTGCA |
| ermR1  | TTACGCAATGGTCAGCTGAGGATGAGTGAC |
| repF   | GAATCAACGCCCAGTGATC |
| repR   | TACGTCGACACCGTTTTC |
| tufAF  | GGTAGTGTGAGAATGAGGATGTA |
| tufAR  | TAACCCAGGTCTCCTAATCATTCCGACAC |
| mcdMF  | CGGAATTCAGTTCTTCTAAG |
| mcdMR  | GCTTCACAAATAAGCGGTTC |
| reb1F  | GGACGGGAGATTAGCTCTATTT |
| reb1R  | CTGGTCCTGCAAAATTTCAC |
| mobb1F | ATGGGTCTGATTTGGAAGG |
| mobb1R | CACCTAATCAAGAAGCAGTGC |
| M13F   | GTAAACGACGGCCAGT |
| M13R   | CAGGAAACAGCTATGAC |

*aAll primers were designed during this study except from primers tufAF/tufAR that were designed in [55] and primers M13F/M13R that hybridize with the pUC19 backbone. doi:10.1371/journal.pone.0116337.t001
agarose gels using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel GmbH). The entire pSMA198 replication backbone (ori—rep—orfX) was cloned into the pre-linearized pUC19 plasmid using the ligation-free In-Fusion HD Cloning Kit (Clontech Inc.) according to the manufacturer’s instructions. In brief, the replication backbone of pSMA198 was amplified by PCR using primers pr2Fn/pr2R designed to hybridize at the ends of both the replication backbone of pSMA198 and the linearized pUC19 vector. The PCR product was gel purified and then incubated along with pUC19 in the In-Fusion cloning reaction for 15 min at 50°C. Five microliters of the reaction was used to transform E.coli Stellar chemically competent cells. The cells were plated on LB agar plates containing 100 μg/ml ampicillin. Forty microliters of a 20 mg/ml X-gal solution was spread on the surface of the plates for blue—white selection. Colony PCR was performed on white colonies to verify cloning of the right insert. Two colonies were selected for plasmid isolation. The erythromycin resistance gene was amplified by PCR from plasmid pGh9:IS3 [31] using primers ermF1/ermR1. The plasmids containing the pSMA198 replicon were digested with the PstI enzyme (New England Biolabs Inc., Beverly, MA) and the erythromycin resistance gene was cloned in this position. The cloning reactions were used to transform E. coli Stellar cells. Transformed cells were plated on LB agar containing 200 μg/ml erythromycin. Colony PCR was employed to screen for clones containing both inserts. Two plasmids were sequenced in the entire region spanning the hybridization positions of M13F and M13R primers to verify that no mutations were introduced in the two inserts during the cloning procedures. The sequence of the plasmid assigned as pORI198 was deposited in EMBL database under accession number HG974440.

Cloning pORI198 in L. lactis subsp. cremoris MG1363 and assessing the stability of the plasmid in this host

Electrocompetent cells of L. lactis MG1363 were prepared as described previously [32]. One microgram of pORI198 was used during electroporation. Lactococcus lactis cells were plated on GM17 agar containing 1 μg/ml erythromycin at 37°C. Colonies were screened for the pORI198 plasmid using primers M13F/M13R. One positive colony was grown overnight in 5 ml GM17 broth containing 1 μg/ml erythromycin and cells were stored at −80°C as described above. Lactococcus lactis containing the pORI198 plasmid was subcultured once in GM17 broth containing 1 μg/ml erythromycin. Cells were washed twice in Ringer solution to remove any traces of erythromycin and then inoculated in GM17 broth without erythromycin. The strain was subcultured every 12 h so as to multiply for approximately 10 generations within this timeframe. The strain was propagated for a total of 100 generations and every 20 generations samples were serially diluted and plated on GM17 agar containing or not 1 μg/ml erythromycin to assess the stability of the plasmid. Three independent experiments were performed. Differences between the numbers of cells growing in the presence or absence of erythromycin were assessed with Student’s t-test for \( p < 0.05 \).

Quantification of the pSMA198 or pORI198 plasmid copy number (PCN) in the relevant hosts and assessment of the presence of pSMA198 in strains of S. macedonicus

All primers used to quantify the PCN of the two plasmids are presented in Table 1. The PCN of pSMA198 in S. macedonicus and of pORI198 in L. lactis was determined based on the qPCR method described by Skulj et al. (2008) [33] and as performed previously [34]. In brief, total DNA was extracted as described by Leenhouts et al. [35]. Internal fragments of the single copy genes mcdM and tufA were used as chromosomal references for S. macedonicus and L. lactis, respectively. For PCN determination an internal segment of the single copy plasmid gene rep
of both plasmids was amplified. qPCR was performed on a MX3005P (Stratagene, La Jolla, CA) using the KAPA SYBR FAST qPCR Kit (Kapa Biosystems, Inc. Woburn, MA) according to the manufacturer’s instructions. The PCN of pSMA198 was determined after three independent experiments.

For the assessment of the presence of pSMA198 in different strains of S. macedonicus, cultures were grown to stationary phase before total DNA isolation with the GenElute Bacterial Genomic DNA Kit. The presence/absence of pSMA198 was verified by PCR using primers reb1F/reb1R with the Taq PCR 2X ReadyMix kit (Kapa Biosystems, Inc.). To avoid any false positive results the presence of the plasmid was also verified with primers mobb1F/mobb1R.

Results and Discussion

pSMA198 belongs to the narrow host range pC1305/pWV02 family of lactococcal plasmids

Streptococcus macedonicus ACA-DC 198 carries a novel plasmid of 12,728 bp assigned as pSMA198 [15]. The plasmid has a 35.0% G+C content, which is lower than that of the S. macedonicus chromosome (37.6%), suggesting that it may have been acquired from another organism. Overall, 17 CDSs were identified on pSMA198 (Fig. 1 and Table 2). The first gene codes for a replication initiation protein (Rep). The rep gene showed 87% of DNA sequence identity (e-value 2.0e⁻¹⁹⁶) with the respective gene found on plasmid 1 (pl1) of L. lactis subsp. cremoris SK11 [36]. Among the best WU-Blast hits of Rep we identified the RepB proteins of the pC1305 and the pWV02 plasmids (78% identity, e-value 1.7e⁻¹⁶¹ and 75% identity, e-value 5.9e⁻¹⁵², respectively) that are the prototypes of the pC1305/pWV02 family of the lactococcal theta-replicating plasmids [37–39]. Multiple sequence alignment of the top Rep WU-Blast hits, including the pC1305 and pWV02 RepB proteins, revealed a high degree of conservation among them (Fig. 2). Analysis of these proteins with InterProScan determined four Pfam sequence signatures corresponding to the initiator Rep protein (Rep_3, PF01051), the L. lactis RepB C-terminal protein (L_lactis_RepB_C, PF06430) and two consecutive winged helix-turn-helix transcription DNA-binding domains (Wing_hlx_DNA_bd, G3DSA:1.10.10.10). The first signature is typically shared by the theta replicon-type Rep proteins [40], while according to
Table 2. 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%/ 2.0e-106 | Initiator RepB protein                    |
| SMA_p0002  | orfX       | 585     | SPJ1_2227/ Streptococcus parauberis KRS-02083/ 91%/ 2.1e-101          | Replication associated protein           |
| SMA_p0003  | ydeE       | 858     | IS71_01319/ Enterococcus durans ATCC 6056/ 99%/ 1.0e-184              | AraC family transcriptional regulator    |
| SMA_p0004  | -          | 582     | OGW_05102/ Enterococcus faecium EnGen0004/ 99%/ 8.2e-122              | Integral membrane protein                |
| SMA_p0005  | yoeC       | 591     | OIU_05588/ Enterococcus faecium EnGen0039/ 98%/ 5.3e-121              | Integrate/recombinase plasmid associated  |
| SMA_p0006  | -          | 459     | CAC42047/ Listeria innocua Clp1126/ 99%/ 4.2e-34                    | Putative pseudo                          |
| SMA_p0007  | -          | 438     | LACR_D31/ Lactococcus lactis subsp. cremoris SK11 plasmid 4/ 99%/ 5.6e-90 | Universal stress protein family           |
| SMA_p0008  | mntH       | 1578    | SPJ2_0535/ Streptococcus parauberis KRS-02109/ 99%/ 0.0              | Manganese transport protein MntH         |
| SMA_p0009  | -          | 480     | Img_pseudo_13/ Lactococcus lactis subsp. cremoris MG1363 pseudogene/ 97%/ 3.7e-214 | Putative pseudo                          |
| SMA_p0010  | -          | 195     | Img_pseudo_13/ Lactococcus lactis subsp. cremoris MG1363 pseudogene/ 97%/ 3.7e-214 | Putative pseudo                          |
| SMA_p0011  | -          | 276     | Img_pseudo_13/ Lactococcus lactis subsp. cremoris MG1363 pseudogene/ 97%/ 3.7e-214 | Putative pseudo                          |
| SMA_p0012  | yoeC       | 465     | OGQ_02346/ Enterococcus faecium EnGen0017/ 98%/ 2.5e-93              | Integrate/recombinase plasmid associated  |
| SMA_p0013  | -          | 132     | pIL7_28/ Lactococcus lactis subsp. lactis IL584 plasmid pIL7/ 84%/ 1.4e-13 | Putative pseudo                          |
| SMA_p0014  | mobC       | 366     | HMPREF9519_01999/ Enterococcus faecalis TX1348/ 89%/ 6.7e-51           | Mobilization protein                     |
| SMA_p0015  | rlx        | 1233    | C16MOBP0R/ Lactococcus lactis subsp. cremoris UC503 pCIS28/ 99%/ 3.8e-268 | Mobilization protein                     |
| SMA_p0016  | -          | 627     | ENT_30400/ Enterococcus sp. 7L76/ 99%/ 7.1e-124                     | Conserved hypothetical protein           |
| SMA_p0017  | -          | 603     | BN193_11500/ Lactococcus raffinosiactis 4877/ 99%/ 4.0e-125           | Fic family protein                       |

doi:10.1371/journal.pone.0116337.t002

Analysis of Plasmid pSMA198 of S. macedonicus

In our analysis its combination with the second signature seems to be specific and characteristic of the Rep proteins of the pCI305/pWV02 family. The product of the gene orfX, found immediately downstream of rep, has been suggested to participate in the control of the PCN [39]. OrfX is significantly less conserved than Rep (data not shown). This is not completely unexpected, as in some plasmids orfX is absent, indicating that it may not be necessary for the replication of the plasmids carrying it [39].

Other potentially interesting proteins encoded by pSMA198 are a transcriptional regulator of the AraC family (YdeE), a universal stress protein (SMA_p0007) and a manganese transport protein (MntH). The fourth gene codes for an integral membrane protein of unknown function, while the fifth and the twelfth genes are both coding for an integrase/recombinase YoeC protein. Five ORFs are putative pseudogenes. Four (SMA_p0006, SMA_p0009, SMA_p0010 and SMA_p0011) consist of fragments of genes coding for transposase elements. The remaining pseudogene SMA_p0013 is devoid of an RBS sequence. It is also noteworthy that, according to current annotations, analogs of all these pseudogenes could be identified in a potentially functional form in different lactococcal plasmids related to pSMA198 (e.g. SMA_p0006 in pILS, SMA_p0009, SMA_p0010 and SMA_p0011 in pGdh442 and SMA_p0013 in pCI605, data not shown). These findings indicate that pSMA198 has been subjected to gene decay processes. The last four genes of the plasmid are probably dedicated to the mobilization of the plasmid. Apart from mobC and rlx coding for mobilization proteins, two additional genes (SMA_p0016 and SMA_p0017) encode a conserved hypothetical protein and a Fic family protein, respectively. Genes like SMA_p0017 are often found in the same context with mobilization-associated genes and are probably implicated in cell cycle regulation [41].

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). 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 pCI305/pWV02 family of replicons, which are normally found in *Lactococcus* species [37–39].

In addition, 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. Eight bases after the end of IR3 we determined an identical nick site to those previously proposed for plasmids pS7a and pS7b found in *L. lactis* subsp. *lactis* biovar. diacetylactis S50 isolated from butter [42]. Once more, these structures are highly conserved among several lactococcal plasmids including pCI305 [42,43].

pSMA198 was acquired by *S. macedonicus* from *L. lactis* probably in the dairy environment

The relation of pSMA198 to other plasmids was further investigated (Fig. 5). WU-Blastn searches supported that the closest plasmids were of lactococcal origin. Top hits were aligned against pSMA198 in a circular manner using Circoletto. We initiated our search with the pSMA198 ori and oriT that showed highest identity to the corresponding features of plasmids
pSK11b (92% identity, e-value 8.4e$^{-44}$) and pCIS8 (98% identity, e-value 5.1e$^{-95}$), respectively. Both pSK11b and pCIS8 have been isolated from *L. lactis* subsp. *cremoris* strains used as dairy starters [44,45]. We also looked 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 the cheese starter *L. lactis* subsp. *lactis* IL594 [43]. pIL5 showed 97% identity (e-value 0.0) with a query coverage of approximately 30% of a central part between ori and oriT of plasmid pSMA198 (Fig. 5). It should be emphasized that apart from the close similarity hits mentioned above, the overriding majority of the top hits for the different features annotated on pSMA198 at the protein and nucleotide level originated from *L. lactis* dairy strains. For example, nine out of ten top hits for the replication backbone originated from strains isolated from milk or its products. Therefore, it is most likely that the original donor of pSMA198 was a dairy *L. lactis* strain.

To provide further evidence for this hypothesis we performed a phylogenetic analysis of the RepB protein (S1 Fig.). The pSMA198 Rep sequence was placed within a branch in which 28 proteins out of 32 originated from dairy lactococcal strains. Only four proteins within the branch could be traced back to non-dairy species. Furthermore, the fact that non-lactococcal proteins were limited to 47 out of the 200 RepB included in the phylogenetic tree is in line with a relatively restricted dispersion of pCI305/pWV02 plasmids to species other than lactococci. This is in accordance with the experimental data that suggested a narrow host range nature of these plasmids [37,46]. Another potential explanation of the high prevalence of lactococcal sequences in the phylogenetic tree could be that the sequence database is biased due to the high number of sequenced *L. lactis* strains especially of dairy origin. Even though we acknowledge this possibility, there is currently no direct way to validate or refute it. Nevertheless, the absence of pCI305/pWV02 plasmids from the high number of sequenced LAB strains other than *L. lactis* is also an indirect indication of the lactococcal origin of this family of replicons.
In order to experimentally assess the possibility of an ancestral genetic exchange event between S. macedonicus and L. lactis concerning pSMA198, we constructed the pORI198 E. coli—S. macedonicus shuttle vector based on the replication backbone of pSMA198 (Fig. 6A) and we tested the ability of this plasmid to replicate in L. lactis MG1363. Indeed, plasmid pORI198 was successfully cloned in the latter host (Fig. 6B). Interestingly, the plasmid was 100% stable in L. lactis for up to 100 generations, since no statistically significant difference was observed between cells that could grow in GM17 in the presence and absence of erythromycin (p > 0.05, Fig. 6C). Furthermore, the PCN of pSMA198 in S. macedonicus and of pORI198 in L. lactis was low (2–3 copies per cell in both cases), suggesting that the pSMA198 backbone has equal replication efficiencies in the two hosts. The low copy number of the pSMA198 replicon is in agreement with the low copy number of other constructs based on the pCI305/pWV02 replicon [47]. The efficient replication of the pSMA198 replicon in L. lactis shows the full compatibility of this plasmid with this species, as well. This property coincides with the close phylogenetic relationship of the pSMA198 Rep protein to the respective proteins in dairy lactococcal pCI305/pWV02 plasmids mentioned above. Taken as a
whole, our *in silico* analysis and our experimental findings provide strong evidence for the acquisition of pSMA198 by *S. macedonicus* from *L. lactis*. They also suggest that this acquisition took place in the dairy environment.

### The acquisition of pSMA198 by *S. macedonicus* seems not to be a recent event

As mentioned above, pSMA198 exhibits a high percentage of pseudogenes similarly to the chromosome of *S. macedonicus* ACA-DC 198 [16], suggesting that the two molecules must have co-evolved under the same gene decay processes.

Given this inference on such a long co-existence, we examined possible genetic exchanges between pSMA198 and the chromosome of *S. macedonicus* ACA-DC 198. This hypothesis was also supported by the reduced size of pSMA198 as compared to some of its related plasmids, e.g. pCIS8 and pIL5 (12 kb vs. 80 kb and 23 kb, respectively) indicative of loss of genetic material, some of which could have moved to the chromosome. To identify such regions, we initially searched for the presence of chromosomal genes showing high identity to the genes found on pSMA198 or its related plasmids (Fig. 7). Using this strategy, a number of transposase-encoding genes were found on the chromosome that exhibited high sequence identity to the pSMA198 SMA_p0006 transposase pseudogene (e.g. SMA_0311 and SMA_0486, data not shown). Genes coding for similar transposases are also present in different lactococcal plasmids related to pSMA198 (e.g. pIL5, pCIS8, etc., data not shown). We then examined the genome of *S. macedonicus* for the existence of genes that could probably have derived from plasmids of the pCI305/pWV02 family in general. We found a 2.5 kb region within a large genomic island of *S. macedonicus* ACA-DC 198 containing genes SMA_0309 and SMA_0310 that showed ≥ 98% identity with *cadC* and *cadA* of pAH82. Plasmid pAH82 has been isolated from the dairy *L. lactis* subsp. *lactis* biovar diacetylactis DPC220 [48]. Noteworthy, the cassette of the cadmium resistance regulator (*cadC*) and the cadmium efflux ATPase (*cadA*) was previously suggested to have been also transferred horizontally from *L. lactis* to *S. thermophilus* [49]. An additional chromosomal region was found where five out of six genes (namely SMA_0488 to SMA_0493, involved in nucleotide biosynthesis and its regulation) showed ≥ 93% identity to the respective genes of plasmid pGdh442, which belongs to a plant isolate of *L. lactis* [50]. Strikingly, the two chromosomal regions (*cadC-cadA* and SMA_0488 to SMA_0493) are
flanked at least on one side by one of the transposases showing high identity to pSMA198 SMA_p0006 mentioned above (i.e. SMA_0311 and SMA_0486). All these observations suggest a long co-existence with obvious genetic exchanges between the chromosome and the plasmid of *S. macedonicus*. It should be mentioned that a similar integration of plasmid sequences in the chromosome of *Streptococcus salivarius* has also been reported [51].

When looking at the distribution of pSMA198 in more than 10 randomly selected *S. macedonicus* strains of the ACA-DC collection, the plasmid was always present (data not shown). It should be emphasized that the strains belonged to different biotypes according to PFGE analysis and were all isolated from traditional Greek dairy products (our unpublished results). The widespread distribution of pSMA198 among *S. macedonicus* strains was also

![Figure 6. Cloning plasmid pORI198 in *L. lactis* MG1363 and assessing its stability in this host.](image-url)
supported by its detection in strains from the Nestlé culture collection isolated in Switzerland (data not shown). Again, these findings suggest that the acquisition of pSMA198 from \textit{S. macedonicus} is not a recent event. This seemingly universal presence of pSMA198 in strains of \textit{S. macedonicus} and its potential acquisition in the milk environment from \textit{L. lactis} led us to investigate if the plasmid plays a role in the growth of \textit{S. macedonicus} in milk. Unfortunately, stress- or antibiotic-induced curing of the plasmid was not possible, presumably due to its high replication stability.

**Conclusion**

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 \textit{Lactococcus} species [37,46]. pSMA198 is the first streptococcal plasmid to be described within this family. The levels of sequence identity of the pSMA198 replication/mobilization backbone reflect an evolutionary history linked with different lactococcal plasmids supporting that \textit{S. macedonicus} acquired pSMA198 from \textit{L. lactis} and that this
exchange took place most probably in milk or milk products. Indeed, the closest related plasmids to pSMA198 (i.e. pSK11b, pCIS8 and pIL5) have been isolated from L. lactis of dairy origin, while the pSMA198 Rep showed high phylogenetic relatedness to RepB proteins of L. lactis dairy isolates. The possibility that L. lactis may have acted as a donor of the pSMA198 plasmid is not surprising, since the species has been characterized as promiscuous to genetic exchange due to owning conjugative plasmids [44]. Finally, the acquisition of pSMA198 by S. macedonicus ACA-DC 198 seems not to be a recent event. This conclusion is based on: i) the high proportion of pseudogenes found in both pSMA198 and the chromosome of S. macedonicus ACA-DC 198, suggesting co-evolution of the two molecules, ii) the detection of different putative genetic exchange events between the two replicons and iii) the high prevalence of pSMA198 in divergent strains of S. macedonicus.

Our analyses thus point towards the long presence of S. macedonicus in the milk environment. This is in full agreement with the fact that most S. macedonicus strains (including S. macedonicus ACA-DC 198) have been isolated from dairy products that according to the current literature seem to be their primary ecological niche [6,16]. It should be noted that a plasmid related to lactococcal plasmids has been also found in S. infantarius CJ18, the only other member of the SBSEC that was isolated from a dairy product [12]. Similar plasmids were absent from the S. gallolyticus and S. pasteurianus genomes sequenced to date that were all clinical isolates. Whether the presence of S. macedonicus ACA-DC 198 in the dairy environment has resulted in a diminished pathogenic potential compared to other streptococci, as suggested in some relevant studies [15,16,52–54], remains to be elucidated.

Supporting Information

S1 Fig. Maximum likelihood tree of the pSMA198 RepB generated using the Phylogeny.fr pipeline [27] and as described in the Materials and Methods section. The arrow indicates the position of the pSMA198 Rep and the bracket denotes the branch with its closest related proteins. Branch support values are presented in the tree, while branches showing <80% support were collapsed.

(PDF)

Acknowledgments

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

Author Contributions

Conceived and designed the experiments: KP BP ET. Performed the experiments: KP PS SF TP EM RA. Analyzed the data: KP TP NCP SJH PR. Contributed reagents/materials/analysis tools: PS SF BP. Wrote the paper: KP RA.

References

1. Ljungh A, Wadstrom T (2006) Lactic acid bacteria as probiotics. Curr Issues Intest Microbiol 7: 73–89. PMID: 16875422
2. Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, et al. (2004) Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. Nat Biotechnol 22: 1554–1558. doi:10.1038/nbt1034 PMID: 15543133
3. Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, et al. (2005) New insights in the molecular biology and physiology of Streptococcus thermophilus revealed by comparative genomics. FEMS Microbiol Rev 29: 435–463. doi:10.1016/j.femsre.2005.04.008 PMID: 16125007
4. Liu M, Siezen RJ, Nauta A (2009) In silico prediction of horizontal gene transfer events in Lactobacillus bulgaricus and Streptococcus thermophilus reveals protocooperation in yogurt manufacturing. Appl Environ Microbiol 75: 4120–4129. doi: 10.1128/AEM.02898-08 PMID: 19395564

5. Delorme C (2008) Safety assessment of dairy microorganisms: Streptococcus thermophilus. Int J Food Microbiol 126: 274–277. doi: 10.1016/j.ijfoodmicro.2007.08.014 PMID: 17822794

6. De Vuyst L, Tsakalidou E (2008) Streptococcus macedonicus, a multi-functional and promising species for dairy fermentations. International Dairy Journal 18: 476–485.

7. Schlegel L, Grimont F, Collins MD, Regnault B, Grimont PA, et al. (2000) Identification of streptococci from raw and fermented dairy products from East and West Africa. Int J Syst Evol Microbiol 50 Pt 4: 1425–1434. doi: 10.1099/00207713-50-4-1425 PMID: 10939646

8. Abdelgadir W, Nielsen DS, Hamad S, Jakobsen M (2008) A traditional Sudanese fermented camel’s milk product, Gariss, as a habitat of Streptococcus infantarius subsp. infantarius. Int J Food Microbiol 127: 215–219. doi: 10.1016/j.ijfoodmicro.2008.07.008 PMID: 18774196

9. Jans C, Kaindi DW, Bock D, Njage PM, Kouame-Sina SM, et al. (2013) Complete genome sequence of the African fermented camel milk isolate Streptococcus infantarius subsp. infantarius strain CJ18. J Bacteriol 194: 2105–2106. doi: 10.1128/JB.00160-12 PMID: 22461547

10. Malfij J, Kimmitt PT, Ou HY, Bhasker PS, Khare M, et al. (2008) Identification of Streptococcus galolyticus subsp. macedonicus as the etiological agent in a case of culture-negative multivalve infective endocarditis by 16S rDNA PCR analysis of resected valvular tissue. J Heart Valve Dis 17: 589–592. PMID: 18980096

11. Herrero IA, Rouse MS, Piper KE, Alyaseen SA, Steckelberg JM, et al. (2002) Reevaluation of Streptococcus bovis endocarditis cases from 1975 to 1985 by 16S ribosomal DNA sequence analysis. J Clin Microbiol 40: 3848–3850. doi: 10.1128/JCM.40.10.3848-3850.2002 PMID: 12354897

12. Papadimitriou K, Ferreira S, Papandreou NC, Mavrogonatou E, Supply P, et al. (2012) Complete genome sequence of the dairy isolate Streptococcus macedonicus ACA-DC 198. J Bacteriol 194: 1838–1839. doi: 10.1128/JB.06804-11 PMID: 22408241

13. Papadimitriou K, Anastasiou R, Mavrogonatou E, Blom J, Papandreou NC, et al. (2014) Comparative genomics of the dairy isolate Streptococcus macedonicus ACA-DC 198 against related members of the Streptococcus bovis/Streptococcus equinus complex. BMC Genomics 15: 272. doi: 10.1186/1471-2164-15-272 PMID: 24710345

14. Jans C, Follador R, Hochstrasser M, Lacroix C, Meile L, et al. (2013) Comparative genome analysis of Streptococcus infantarius subsp. infantarius CJ18, an African fermented camel milk isolate with adaptations to dairy environment. BMC Genomics 14: 200. doi: 10.1186/1471-2164-14-200 PMID: 23521820

15. Popa O, Dagan T (2011) Trends and barriers to lateral gene transfer in prokaryotes. Curr Opin Microbiol 14: 615–623. doi: 10.1016/j.mib.2011.07.027 PMID: 21856213

16. Papadimitriou K, Ferreira S, Papandreou NC, Mavrogonatou E, Supply P, et al. (2012) Complete genome sequence of the dairy isolate Streptococcus macedonicus ACA-DC 198. J Bacteriol 194: 1838–1839. doi: 10.1128/JB.06804-11 PMID: 22408241

17. Papadimitriou K, Anastasiou R, Mavrogonatou E, Blom J, Papandreou NC, et al. (2014) Comparative genomics of the dairy isolate Streptococcus macedonicus ACA-DC 198 against related members of the Streptococcus bovis/Streptococcus equinus complex. BMC Genomics 15: 272. doi: 10.1186/1471-2164-15-272 PMID: 24710345

18. Jans C, Follador R, Hochstrasser M, Lacroix C, Meile L, et al. (2013) Comparative genome analysis of Streptococcus infantarius subsp. infantarius CJ18, an African fermented camel milk isolate with adaptations to dairy environment. BMC Genomics 14: 200. doi: 10.1186/1471-2164-14-200 PMID: 23521820

19. Popa O, Dagan T (2011) Trends and barriers to lateral gene transfer in prokaryotes. Curr Opin Microbiol 14: 615–623. doi: 10.1016/j.mib.2011.07.027 PMID: 21856213

20. Aziz RK, Bartels D, Best AA, DeJongh M, Diz T, et al. (2006) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9: 75. doi: 10.1186/1471-2164-9-75 PMID: 18261238

21. Van Domselaar GH, Stothard P, Srivastava S, Cruz JA, Guo A, et al. (2005) BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res 33: W455–459. doi: 10.1093/nar/gki593 PMID: 15980511

22. Lopez R, Silventoinen V, Robinson S, Kibria A, Gish W (2003) WU-Blast2 server at the European Bioinformatics Institute. Nucleic Acids Res 31: 3795–3798. doi: 10.1093/nar/gkg573 PMID: 12824421

23. Papadimitriou K, Ferreira S, Papandreou NC, Mavrogonatou E, Supply P, et al. (2012) Complete genome sequence of the dairy isolate Streptococcus macedonicus ACA-DC 198. J Bacteriol 194: 1838–1839. doi: 10.1128/JB.06804-11 PMID: 22408241
24. Mulder N, Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol 396: 59–70. PMID: 18025686
25. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552. doi: 10.1093/oxfordjournals.molbev.a026334 PMID: 10742046
26. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704. doi: 10.1080/10635150390235520 PMID: 14530136
27. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36: W465–469. doi: 10.1093/nar/gkn180 PMID: 18424797
28. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 55: 539–552. doi: 10.1086/50600755453 PMID: 16785212
29. Chevenet F, Brun C, Banuls AL, Jacq B, Christen R (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7: 439. doi: 10.1186/1471-2105-7-439 PMID: 17032440
30. Darzentas N (2010) Circoletto: visualizing sequence similarity with Circos. Bioinformatics 26: 2620–2621. doi: 10.1093/bioinformatics/btq484 PMID: 20736393
31. Maguin E, Prevost H, Ehrlich SD, Gruss A (1996) Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. J Bacteriol 178: 931–935. PMID: 8550537
32. Hol H, Nes IF (1989) High-Frequency Transformation, by Electroporation, of Lactococcus lactis subsp. cremoris Grown with Glycine in Osmotically Stabilized Media. Appl Environ Microbiol 55: 3119–3123. PMID: 16348073
33. Skulj M, Okršlar V, Jalien S, Jevsevar S, Slanc P, et al. (2008) Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. Microb Cell Fact 7: 6. doi: 10.1186/1475-2859-7-6 PMID: 18326904
34. Ainsworth S, Zomer A, de Jager V, Bottacini F, van Hijum SAFT, et al. (2013) Complete Genome of Lactococcus lactis subsp. cremoris UC509.9, Host for a Model Lactococcal P335 Bacteriophage. Genome Announcements 1. doi: 10.1128/genomeA.00119-12 PMID: 23405300
46. Hayes F, Vos P, Fitzgerald GF, de Vos WM, Daly C (1991) Molecular organization of the minimal replicon of novel, narrow-host-range, lactococcal plasmid pCI305. Plasmid 25: 16–26. doi: 10.1016/0147-619X(91)90003-F PMID: 1852014

47. Shareck J, Choi Y, Lee B, Miguez CB (2004) Cloning vectors based on cryptic plasmids isolated from lactic acid bacteria: their characteristics and potential applications in biotechnology. Crit Rev Biotechnol 24: 155–208. doi: 10.1080/07388550490904288 PMID: 15707158

48. O’Sullivan D, Twomey DP, Coffey A, Hill C, Fitzgerald GF, et al. (2000) Novel type I restriction specificities through domain shuffling of HsdS subunits in Lactococcus lactis. Mol Microbiol 36: 866–875. doi: 10.1046/j.1365-2958.2000.01901.x PMID: 10844674

49. Schirawski J, Hagens W, Fitzgerald GF, Van Sinderen D (2002) Molecular characterization of cadmium resistance in Streptococcus thermophilus strain 4134: an example of lateral gene transfer. Appl Environ Microbiol 68: 5508–5516. doi: 10.1128/AEM.68.11.5508-5516.2002 PMID: 12406744

50. Tanous C, Chambellon E, Yvon M (2007) Sequence analysis of the mobilizable lactococcal plasmid pGdh442 encoding glutamate dehydrogenase activity. Microbiology 153: 1664–1675. doi: 10.1099/mic.0.2006/002246-0 PMID: 17464081

51. Heng NC, Haji-Ishak NS, Kalyan A, Wong AY, Lovric M, et al. (2011) Genome sequence of the bacteriocin-producing oral probiotic Streptococcus salivarius strain M18. J Bacteriol 193: 6402–6403. doi: 10.1128/JB.06001-11 PMID: 22038965

52. Boleij A, Muytjens CM, Bukhari SI, Cayet N, Glaser P, et al. (2011) Novel clues on the specific association of Streptococcus galalcticus subsp galalyticus with colorectal cancer. J Infect Dis 203: 1101–1109. doi: 10.1093/infdis/jiq169 PMID: 21451000

53. Maragkoudakis PA, Papadelli M, Georgalaki M, Panayotopoulou EG, Martinez-Gonzalez B, et al. (2009) In vitro and in vivo safety evaluation of the bacteriocin producer Streptococcus macedonicus ACA-DC 198. Int J Food Microbiol 133: 141–147. doi: 10.1016/j.ijfoodmicro.2009.05.012 PMID: 19515446

54. Danne C, Entenza JM, Mallet A, Briandet R, Debarbouille M, et al. (2011) Molecular characterization of a Streptococcus galalyticus genomic island encoding a pilus involved in endocarditis. J Infect Dis 204: 1960–1970. doi: 10.1093/infdis/jir666 PMID: 22043018

55. Smith WM, Pham TH, Lei L, Dou J, Soomro AH, et al. (2012) Heat resistance and salt hypersensitivity in Lactococcus lactis due to spontaneous mutation of ilmg_1816 (gdpP) induced by high-temperature growth. Appl Environ Microbiol 78: 7753–7759. doi: 10.1128/AEM.02316-12 PMID: 22923415