RESEARCH LETTER

Comparative and evolutionary analysis of plasmid pREN isolated from *Lactobacillus rennini*, a novel member of the theta-replicating pUCL287 family

Ioanna-Areti Asteri[1,2], Konstantinos Papadimitriou[1,2], Effrossyni Boutou[2], Bruno Pot[3], Constantinos E. Vorgias[2] & Effie Tsakalidou[1]

[1]Laboratory of Dairy Research, Department of Food Science and Technology, Agricultural University of Athens, Athens, Greece; [2]Department of Biochemistry and Molecular Biology, Faculty of Biology, National and Kapodistrian University of Athens, Athens, Greece; and [3]Applied Maths N.V., Sint-Martens-Latem, Belgium

Correspondence: Konstantinos Papadimitriou, Laboratory of Dairy Research, Department of Food Science and Technology, Agricultural University of Athens, Iera Odos 75, 118 55 Athens, Greece. Tel.: +30 210 529 4661; fax: +30 210 529 4672; e-mail: kpapadimitriou@aua.gr

Received 29 November 2010; revised 14 January 2011; accepted 26 January 2011. Final version published online 22 February 2011.

DOI:10.1111/j.1574-6968.2011.02232.x

Editor: Diethard Mattanovich

Keywords
Lactobacillus; plasmid; ori; operon; mob; rep.

Abstract

Here, we describe plasmid pREN of *Lactobacillus rennini* ACA-DC 1534, isolated from traditional Kopanisti cheese. pREN is a circular molecule of 4371 bp. *Orf* calling revealed a novel *rep*A-*orf*2 operon with the deduced product of *orf*2 showing no similarity to other known proteins. Downstream of this operon, a gene cluster encoding different mobilization proteins, namely *mob*C, *mob*A1, *mob*A2 and *mob*B, was detected. Based on the sequence of the origin of replication (*ori*) and the similarity pattern of RepA, pREN was placed in the pUCL287 family of theta-replicating plasmids. Multiple sequence alignment demonstrated for the first time the degree of conservation in the pUCL287 *ori*s. Our analysis supported that the identified conserved repeats could drive similar secondary structures in the *ori*s of all plasmids. Furthermore, comparative mapping of pREN with its related plasmids (i.e. pLB925A03 and pLJ42) showed that they retain a unique combination in the architecture of their replication and mobilization elements within the pUCL287 family. Phylogenetic analysis also established that these plasmids have undergone a modular evolutionary process in order to acquire their *mob* genes. Research on plasmids from uncommon lactic acid bacteria will expand our appreciation for their divergence and will aid their rational selection for biotechnological applications.

Introduction

The plasmid content of more than a few lactic acid bacteria (LAB) has been shown to be vital for their technological traits. This is due to the fact that proteins involved in important functions, such as substrate utilization, bacteriocin or exopolysacharides production, etc, have been found in several instances to be encoded by plasmid-carried genes (Schroeter & Klaenhammer, 2009). Alternatively, plasmids may be solely dedicated to their replication, without conferring any obvious traits to the phenotype of the bacterium. However, such cryptic plasmids have often been used for the construction of LAB shuttle or delivery vectors. Furthermore, the biology of plasmids has attracted increasing attention with respect to their modular evolution processes by being potential vehicles for horizontal gene transfer (Thomas & Nielsen, 2005; Toomey *et al.*, 2009). LAB's plasmid research has been up to now biased in favor of well-characterized and established starter strains (Asteri *et al.*, 2010). The majority of LAB, which remain largely unexplored, constitute a vast pool for plasmids discovery so as to improve our understanding of plasmid evolution and divergence in these economically important bacteria.

Here, we report the isolation, cloning and characterization of the novel cryptic plasmid pREN deriving from *Lactobacillus rennini* strain ACA-DC 1534, isolated from traditional Kopanisti cheese (Asteri *et al.*, 2009). *Lactobacillus rennini* is a recently described species in LAB (Chenoll *et al.*, 2006) and its plasmid content has never been explored before.

Materials and methods

Bacterial strains

Lactobacillus rennini ACA-DC 1534 was routinely grown in MRS broth, pH 5.5 (Oxoid Ltd, Basingstoke, Hampshire, UK), supplemented with 2.5% NaCl (w/v), at 30 °C. *Escherichia coli* Mach1^{TM} (Invitrogen Corporation, Carlsbad, CA) was used as the transformation host and was cultivated in Luria–Bertani (LB) medium at 37 °C in a shaking incubator (250 r.p.m.). Ampicillin (Sigma, St. Louis, MO) was added to the LB medium at a concentration of 100 μg mL^{-1}.

Plasmid isolation, cloning and sequencing

Plasmid content was isolated from *L. rennini* and *E. coli* strains using the NucleoSpin Plasmid kit (Macherey-Nagel GmbH and Co. KG, Düren, Germany) according to the manufacturer's instructions. For *L. rennini* some modifications were incorporated into the original protocol so as to ensure proper cell lysis. In brief, lysozyme (20 mg mL^{-1}) and mutanolysin (50 U mL^{-1}) were added to the lysis buffer of the kit, followed by incubation at 37 °C for 1 h. Plasmid minipreps were subjected to agarose gel electrophoresis (0.8% w/v) and the plasmid under investigation (pREN) was excised from the gel and extracted using the QIAEX II Gel Extraction kit (Qiagen Inc., Valencia, CA). Plasmid DNA was then digested with XbaI restriction endonuclease or double digested with XbaI and Eco88I (both purchased from New England BioLabs Inc., Beverly, MA). The acquired fragments were ligated into the pUC18 vector, which was transformed in *E. coli* Mach1 competent cells. General cloning procedures, including the dephosphorylation of the digested pUC18 vector with antartic phosphatase (NEB), were performed according to established protocols (Sambrook *et al.*, 1989). The clones of interest were sequenced with the M13F(-20), M13R-pUC(-40) universal primers, as well as specific primers designed from the sequences, by Macrogen Inc. (Seoul, Korea). Primer-walking across the gaps facilitated sequencing of the complete pREN.

Plasmid sequence analysis and annotation

Plasmid assembly and annotation was performed as described previously (Asteri *et al.*, 2010). In brief, contigs were assembled using the CAP3 sequence assembly program (Huang & Madan, 1999). In order to identify potential protein encoding segments, three open reading frames (*orf*s) prediction programs were used: heuristic GENEMARK^{TM} (Besemer & Borodovsky, 1999), FGENESB (http://www.softberry.com) and METAGENEANNOTATOR (Noguchi *et al.*, 2006). BLASTN and BLASTP queries were performed at the NCBI server (Altschul *et al.*, 1997). Ribosome binding sites (RBS), putative promoter and terminator sequences were predicted by METAGENEANNOTATOR, BPROM and FINDTERM (http://www.softberry.com), respectively. KODON software (Applied Maths N.V., Sint-Martens-Latem, Belgium) was used for the construction of the genetic map of pREN plasmid, for the prediction of the DNA secondary structures and for the comparative mapping of pREN with its closely related plasmids.

Multiple sequence alignment and phylogenetic analysis methods

After BLASTP searches, protein sequences receiving top scores were retrieved from the GeneBank database. Multiple alignments of protein or nucleotide sequences were constructed using the MUSCLE program (Edgar, 2004). JALVIEW allowed the visualization and editing of the alignments (Waterhouse *et al.*, 2009). For phylogenetic analysis, the alignments were further curated with GBLOCKS (Castresana, 2000). Phylogenetic trees were constructed based on the maximum likelihood method using the PHYML program (Guindon & Gascuel, 2003) and TREEDYN for tree rendering (Chevenet *et al.*, 2006) with the WAG substitution matrix. Statistical validation for branch support (%) was conducted via a χ^2-based parametric approximate likelihood-ratio test (Anisimova & Gascuel, 2006). The MobB protein sequence was analyzed using INTERPROSCAN to determine functional protein domains (Mulder & Apweiler, 2007).

Nucleotide accession number

The full-length nucleotide sequence of the annotated pREN plasmid was deposited in the EMBL database under Accession No.: FR714836.

Results and discussion

Sequence analysis of pREN

The plasmid content of *L. rennini* ACA-DC 1534 was investigated. The strain harbors more than one plasmid and plasmid assigned as pREN was further analyzed. pREN was found to be a circular molecule of 4371 bp with a 43.3% GC content. *Ab initio orf* calling revealed that pREN carries six putative genes located on the same DNA strand (Fig. 1). The coding sequences (3513 nucleotides in total) cover ~80% of the plasmid. FGENESB indicated that *orf*1 (921 bp) and *orf*2 (330 bp) formed a single operon. Further analysis of this region supported this prediction. The two *orf*s shared a common promoter (-35 and -10 sequences) found upstream of *orf*1. Right after *orf*2, a terminator could be determined, while both *orf*s were preceded by typical RBS sequences. *orf*1 was identified as a replication initiation protein-coding gene. The deduced amino acid product (306 residues) showed the highest identity to RepA of

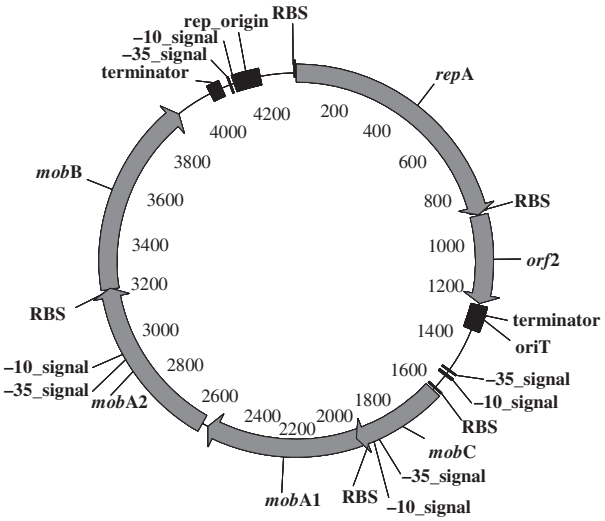

Fig. 1. Genetic map of the pREN plasmid with its annotated features.

plasmid pLJ42 from *Lactobacillus plantarum* (100% query coverage, 90% identity, *e*-value $7e^{-161}$) (Accession No.: DQ099911, direct submission). BLASTP of the Orf2 protein (109 amino acids) detected only a distant relationship with a hypothetical protein carried by plasmid pMCCL1 of *Macrococcus caseolyticus* (100% query coverage, 50% similarity, *e*-value $5e^{-3}$) (Baba *et al.*, 2009). However, TBLASTN returned two putative regions within plasmids pLJ42 and pLB925A03, the latter isolated from *Lactobacillus brevis* (Wada *et al.*, 2009), that could code for proteins with high identity to Orf2. Because *orf*2 was not included in the original annotation of either of the latter plasmids, we re-annotated them using the same bioinformatics tools as with pREN. *orf*2 was indeed predicted in the afore-mentioned plasmids (positions 5170–5499 nt for pLB925A03 and 2415–2744 nt for pLJ42). The deduced *orf*2 products exhibited a high degree of conservation (Fig. 2a). It should be mentioned that a terminator sequence within the *orf*2 locus (position 2515–2579 nt) was initially deposited for pLJ42; however, our analysis with FINDTERM did not support the existence of this terminator. In fact, *orf*2 was located downstream of *rep*A, followed by a terminator in both pLJ42 and pLB925A03, resulting in a conserved operon structure as shown for pREN.

The four remaining *orf*s were all found to encode different types of mobilization proteins. The *orf*3 product (112 amino acids) displayed the highest identity to MobC of pLJ42 (100% query coverage, 100% identity, *e*-value $9e^{-58}$). Orf4 (195 amino acids) and Orf5 (208 amino acids) proteins were identified as MobA (MobA1 and MobA2, respectively), both receiving top scores for the MobA of pLB925A03 (68% query coverage, 100% identity with *e*-value $2e^{-76}$ and 100% query coverage, 98% identity with *e*-value $5e^{-114}$, respectively). Initial analysis clearly excluded the possibility of a gene duplication event. INTERPROSCAN indicated that while MobA1 carried a significant proportion of the N-terminal pfam03432 signal of the family of relaxases, MobA2 carried the remaining distal sequence of the signal's C-terminus. The alignment of these proteins with the MobA of pLB925A03, carrying a full pfam03432 signal, demonstrated that MobA1 and MobA2 were originally a single full-length peptide (Fig. 2b). Inspection of the *mob*A1 gene revealed that it was disrupted by a frameshift mutation at position 2339 nt, causing premature termination at position 2526 nt. Furthermore, *orf*6 was predicted as a *mob*B gene. Interestingly, in contrast to the other pREN Mob proteins, this MobB molecule was detected only in a very limited number of bacteria, all of which were LAB. Sequence comparison among the MobB proteins showed a considerable degree of conservation that was more pronounced at the C-terminus (Fig. 2c). This annotation transfer through sequence identity was based on a previous observation that the protein product of an *orf* in plasmid pNZ4000 of *Lactococcus lactis* (van Kranenburg *et al.*, 2000) shared a moderate homology to MobB of *Staphylococcus aureus* plasmid pC223 (Smith & Thomas, 2004). However, when we performed BLASTP queries for pREN Orf6, as well as for the putative MobB of plasmid pNZ4000, we could not verify any relationship with the respective pC223 protein. In fact, no conserved motifs or domains were detected in any of the LAB MobB proteins with INTERPROSCAN and only two transmembrane regions were found that were similarly positioned among them (Fig. 2c). In our opinion, the actual annotation of Orf6 and its related proteins deserves further investigation.

It should be mentioned that FGENESB assigned the four *mob* genes to a single operon. However, this was not supported because all genes were preceded by individual promoters and RBS sequences, apart from *mob*A2, which lacked both *cis*-acting elements, further supporting our hypothesis for the disruption of an ancestral *mob*A gene into *mob*A1 and *mob*A2. In addition, a typical *ori*T detected upstream of *mob*C (position 1261–1355 nt) was almost identical to that of the pLJ42 plasmid (100% query coverage with 97% identity).

Features of the origin of replication (*ori*) of the pUCL287 theta-replicating family of plasmids

Among the top BLASTP hits for RepA of pREN was the homologous protein of plasmid pUCL287 isolated from *Tetragenococcus halophilus* (formerly *Pediococcus halophilus*) (99% query coverage, 70% identity with *e*-value $8e^{-121}$). pUCL287 is the prototype for a family of theta-type replicons (Benachour *et al.*, 1995, 1997). One of the main features of the pUCL287 family – apart from the actual sequence of its Rep protein – is the distinct structure of the *ori* (Benachour *et al.*, 1997). In pUCL287, *ori* spans 187 bp

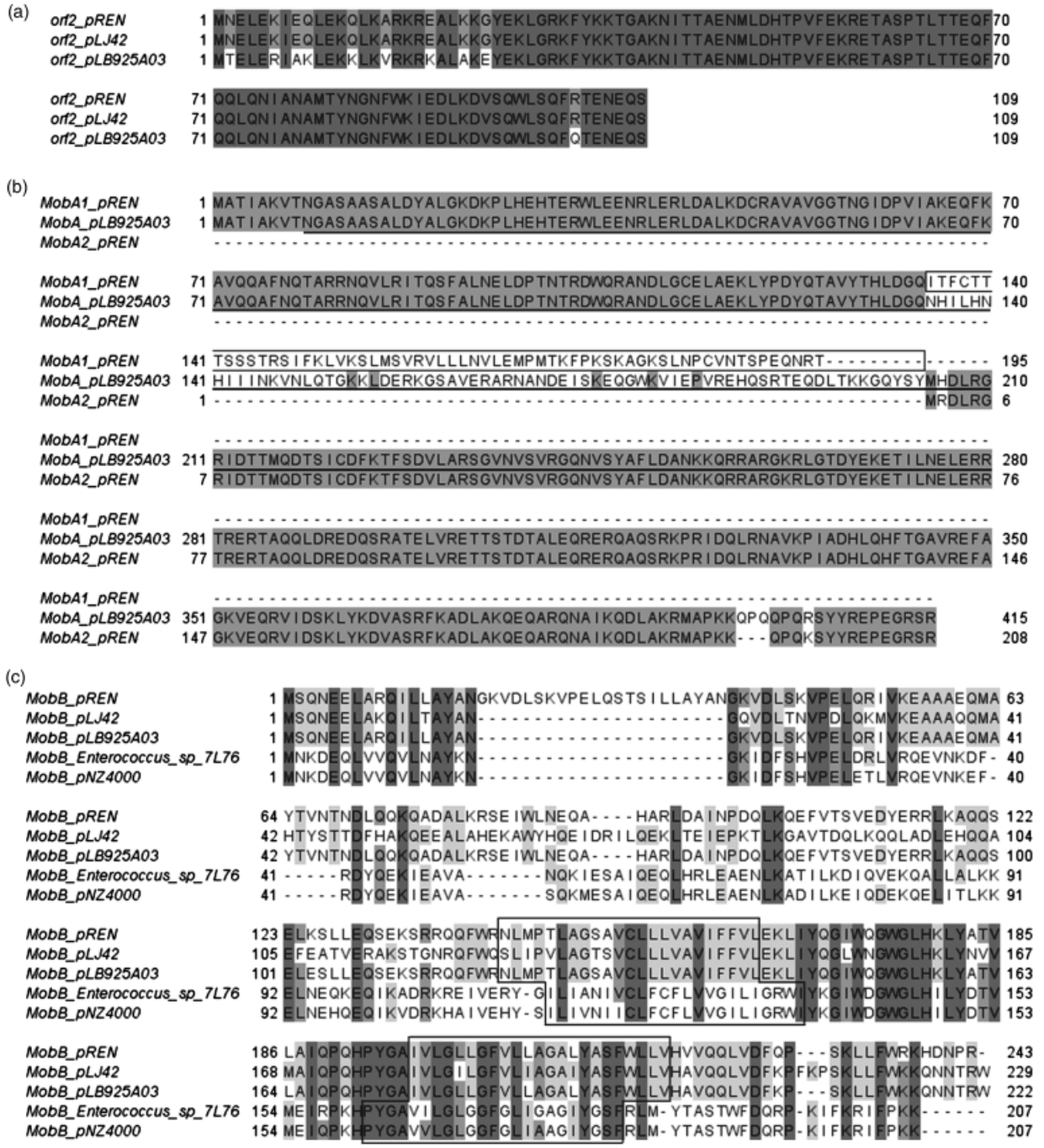

Fig. 2. Multiple sequence alignment using MUSCLE of Orf2 (a), MobA (b) and MobB (c) of plasmid pREN with their related proteins. The grades of shading in the alignment reflect the degree of amino acid conservation. In (b), pREN MobA1 and MobA2 were aligned against the pLB925A03 MobA full-length sequence. The underlined amino acids correspond to the full pfam03432 conserved signal of relaxases. The amino acids within the box represent the altered sequence of MobA1 due to the frameshift mutation at position 2339 nt causing premature termination at position 2526 nt. In (c), boxes include the amino acids of the predicted transmembrane regions for each MobB protein.

and is located upstream of the *rep*A gene. The sequence is arranged in three different regions (Fig. 3a). The first is an AT-rich region, necessary for the melting of the two strands during plasmid replication, containing four 11-mer direct repeats, while the third region, which is the binding site of the Rep protein, consists of a 22-mer iteron tandemly repeated 4.5 times. These two regions are separated by a 37-bp variable sequence. Indeed, an *ori* sequence, carrying

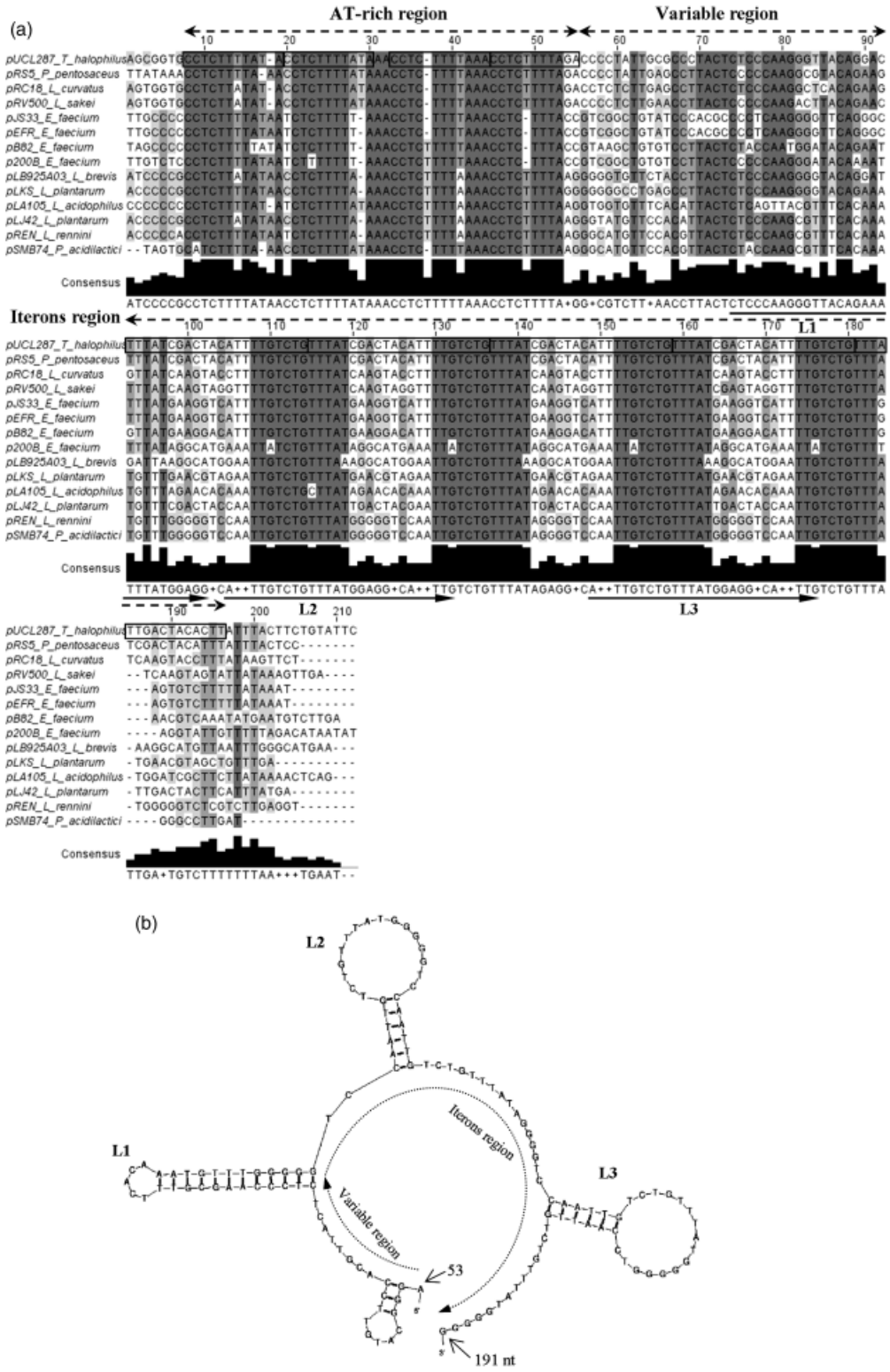

Fig. 3. Multiple sequence alignment using MUSCLE of the origins of replication (*oris*) of the pUCL287 family of replicons (a). The grades of shading in the alignment reflect the degree of nucleotide conservation. The dashed double-sided arrows mark the three typical regions of the pUCL287 *ori* as described in the text. The first four boxes correspond to the 11-mer direct repeats in the AT-rich region, while the five remaining correspond to the 22-mer iterons, as described previously. Genera are abbreviated as follows: L: *Lactobacillus*, E: *Enterococcus*, P: *Pediococcus* and T: *Tetragenococcus*. DNA secondary prediction of the pREN *ori* (variable and iterons regions) (b). Positions were numbered according to the consensus. Stem loops indicated as L1, L2 and L3 in a clock-wise fashion and the corresponding nucleotide sequences are also highlighted in (a) as one-sided arrows.

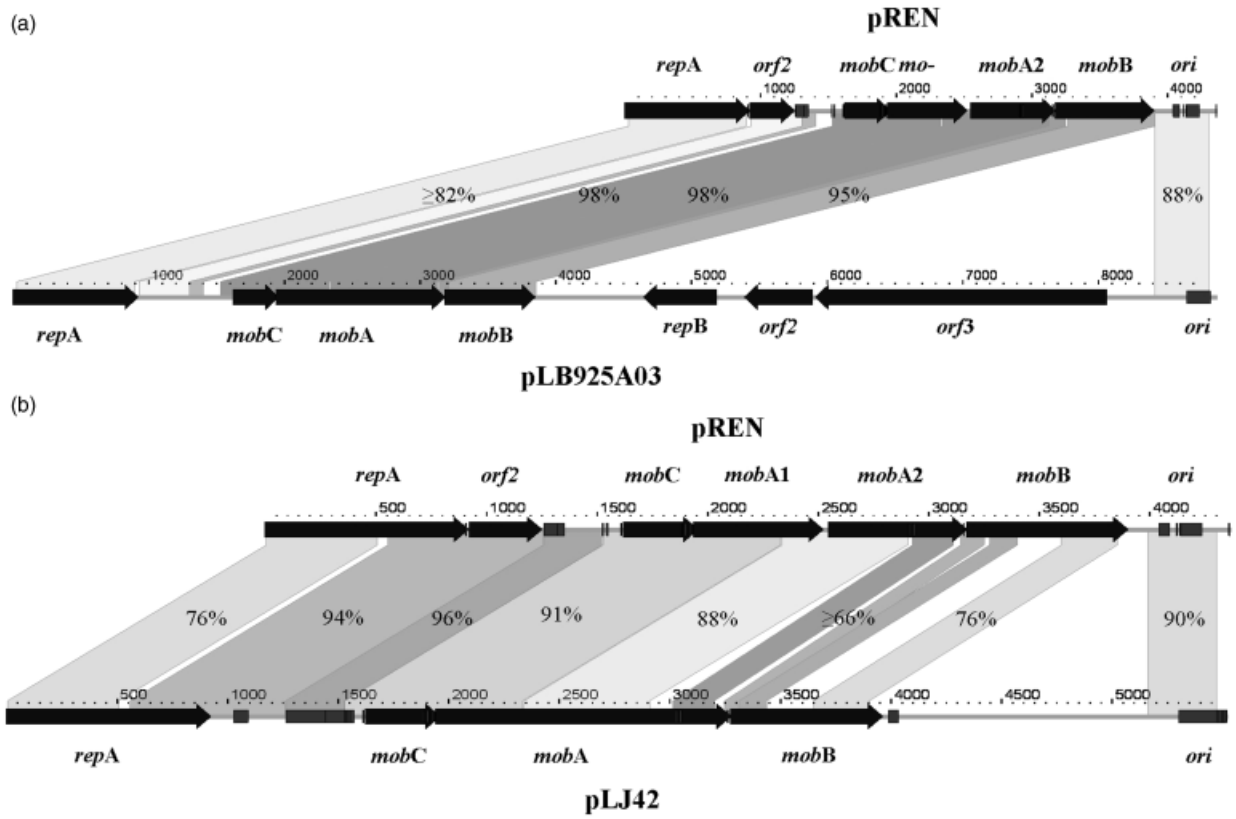

Fig. 4. Full length comparative mapping between plasmids pREN and pLB925A03 (a) and pREN and pLJ42 (b). Before this analysis, plasmid sequences and their annotated features were manually synchronized so that all plasmids have the first nucleotide of the *rep*A as position 1. Shaded regions between the plasmids indicate stretches of homology that are also presented as percent identity.

essentially all of the afore-mentioned features, was detected in pREN, although a certain degree of deviation from the expected sequence repeats (i.e. perfect 11-mer or 22-mer repeats) was observed. The same situation was evident for the *ori* sequences of other plasmids of the family. For this reason, we investigated whether a consensus could be determined for this region. Multiple sequence alignment of the *ori*s of the pUCL287 family, found either predeposited in the GenBank files or manually determined by us (data not shown), was performed (Fig. 3a). According to our findings, the AT-rich stretch was highly conserved over its full length, and showed the presence of the consensus 9-mer sequence CCTCTTTT(A/T), tandemly repeated four times. In the 37-bp variable region, no repeats could be identified, although a significant number of conserved positions were observed, indicating that the region may not be entirely variable after all. The iteron region located immediately downstream was characterized by four consecutive, almost identical, 12-mer blocks (TTGTCTGTTTAT). These four sequence blocks were separated by a variable to a certain degree among the plasmids 10-mer sequence that was identical for each plasmid. Of note, the same 10-mer sequence could also be found preceding the first 12-mer block. DNA folding simulations for pREN *ori* revealed a putative hairpin in the variable region and two identical stem loops in the iteron region (Fig. 3b). Similar secondary structure organizations could also be detected in the *ori*s of all other plasmids (data not shown). While the significance of these structures remains to be investigated, it is important to state that the similarity in secondary structures among the plasmids is clearly driven by sequence conservation (Fig. 3a).

Comparative analysis of pREN with plasmids of the pUCL287 theta-replicating family and evolutionary implications

The overall architecture of pREN was assessed in comparison with that of other members of the pUCL287 family of plasmids. Interestingly, while the replication backbone of pREN (*ori* and *rep*A) was highly conserved (data not shown), BLASTN queries returned only two hits showing identity over the entire plasmid sequence, i.e. pLB925A03 and pLJ42. pLB925A03 carries seven *orf*s on its 8881 bp sequence, consisting of two genes (*rep*A and *rep*B) involved

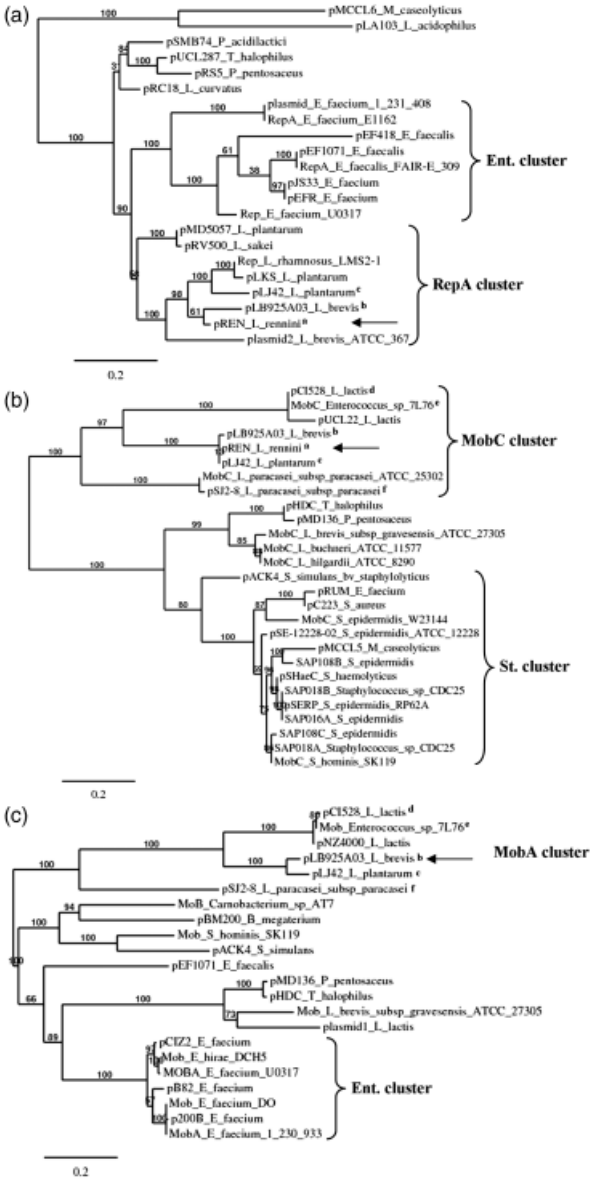

Fig. 5. Phylogenetic trees constructed using the PHYML of pREN RepA (a), pREN MobC (b) and pLB925A03 MobA (c). Top hits for each protein after BLASTP similarity searches were aligned with MUSCLE and the alignments were curated with GBLOCKS. Phylogenetic relationships among proteins were determined using the maximum likelihood method, based on the WAG substitution matrix via a χ^2-based parametric approximate likelihood-ratio test. Branch support is presented as %. Arrows point to the proteins used for the BLASTP queries. Parentheses include clusters that are further discussed in the text. Proteins of the same plasmid origin between the RepA, MobC and MobA clusters are indicated by the same superscript lettering. Genera are abbreviated as follows: L, *Lactobacillus* (with the exception of *L. lactis*, which stands for *Lactococcus lactis*); E, *Enterococcus*; S, *Staphylococcus*; P, *Pediococcus*; M, *Macrococcus*; B, *Bacillus* and T, *Tetragenococcus*.

in the replication process, three genes for mobilization and two unknown genes. pLJ42 (5529 bp in length) encodes a replication (RepA) and three mobilization (MobA, MobB and MobC) proteins. We synchronized all three plasmid annotations so as to start from the first nucleotide of the *rep*A gene in order to perform full-length plasmid sequence alignments (Fig. 4). This comparative mapping of plasmids demonstrated that they share a common organization not only in their replication backbone (*rep*A-*orf*2 operon and the *ori* regions) but also in the mobilization backbone. The three consecutive *mob* genes showed a high degree of identity among the plasmids, with the exception of pREN, which, due to the frameshift mutation mentioned earlier, had its *mob*A gene disrupted in two truncated pseudogenes. This organization of the replication and mobilization elements seems to be unique within the pUCL287 family. According to our analysis, only pREN and pLJ42 possess the basal backbone for this type of plasmids, because an insertion of approximately 4500 bp was evident downstream of the *mob* genes for plasmid pLB925A03.

Furthermore, the phylogeny of RepA, MobC and MobA was surveyed. MobB was excluded from this analysis because it could be detected in only five other bacteria, as mentioned earlier. In the case of MobA, the two truncated proteins of pREN were also omitted from the phylogenetic trees and therefore all conclusions presented below were based on the MobA sequence of plasmid pLB925A03. RepA of pREN clustered with the respective proteins of other *Lactobacillus* plasmids (Fig. 5a) and a clear relation of this cluster with several enterococci replication initiation proteins was observed. This fact, along with the high sequence identity of the *Lactobacillus* and the *Enterococcus ori*s of the pUCL287 family, may indicate that this type of replication backbones for these two genera may have originated from a common ancestral plasmid. MobC and MobA displayed an evolution pattern significantly different from RepA. LAB proteins clustering close to MobC derived from the same plasmids as those clustering with MobA (Fig. 5b and c). However, RepA clustered with LAB proteins of completely different origin, with the exception of pLB925A03 and pLJ42. These findings clearly indicate that the pREN, pLB925A03 and pLJ42 group of plasmids have acquired MobC and MobA as a single unit through a modular evolution process. This hypothesis was confirmed by TBLASTX searches, which identified the conserved *mob*CA region in all LAB plasmids common for the MobC and MobA clusters (Fig. 5b and c, data not shown). From the topology of the phylogenetic trees, it can also be inferred that the generation of the MobC and MobA modular unit took place in an ancestral plasmid because the former is related to proteins of staphylococci while the latter to proteins of enterococci.

Conclusion

In this report, we present the sequencing and characterization of plasmid pREN, a novel member of the pUCL287 family of theta-replicating plasmids. Throughout our study, we shed light on the plasmid's gene content, architecture and evolution. The typical features of the family's origin of replication were, for the first time, presented in a comparative manner. Additionally, plasmids pREN, pLB925A03 and pLJ42 were found to be unique within this family with respect to their actual combination of the replication and mobilization backbones. Finally, the three plasmids were shown to be products of a modular evolution process and an attempt was made to unveil the complex phylogenetic relationships underpinning this phenomenon. The current focus on characterizing plasmids mainly from industrial or widespread LAB strains obscures our view of their overall divergence. In our opinion, the development of an extended catalogue of plasmids in this group of bacteria, including those deriving from uncommon species, accompanied by appropriate comparative analysis, is necessary for the rational selection of plasmids for further functional applications.

Acknowledgements

Ioanna-Areti Asteri wishes to thank the State Scholarships Foundation of Greece (IKY-Idryma Kratikon Ypotrofion) for financial support.

Authors' contribution

I.-A.A. and K.P. contributed equally to this work.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Anisimova M & Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* **55**: 539–552.

Asteri I-A, Robertson N, Kagkli D-M *et al.* (2009) Technological and flavour potential of cultures isolated from traditional Greek cheeses - a pool of novel species and starters. *Int Dairy J* **19**: 595–604.

Asteri I-A, Papadimitriou K, Boutou E, Anastasiou R, Pot B, Vorgias CE & Tsakalidou E (2010) Characterization of pLAC1, a cryptic plasmid isolated from *Lactobacillus acidipiscis* and comparative analysis with its related plasmids. *Int J Food Microbiol* **141**: 222–228.

Baba T, Kuwahara-Arai K, Uchiyama I, Takeuchi F, Ito T & Hiramatsu K (2009) Complete genome sequence of *Macrococcus caseolyticus* strain JCSCS5402, [corrected] reflecting the ancestral genome of the human-pathogenic staphylococci. *J Bacteriol* **191**: 1180–1190.

Benachour A, Frere J & Novel G (1995) pUCL287 plasmid from *Tetragenococcus halophila* (*Pediococcus halophilus*) ATCC 33315 represents a new theta-type replicon family of lactic acid bacteria. *FEMS Microbiol Lett* **128**: 167–175.

Benachour A, Frere J, Flahaut S, Novel G & Auffray Y (1997) Molecular analysis of the replication region of the theta-replicating plasmid pUCL287 from *Tetragenococcus* (*Pediococcus*) *halophilus* ATCC33315. *Mol Gen Genet* **255**: 504–513.

Besemer J & Borodovsky M (1999) Heuristic approach to deriving models for gene finding. *Nucleic Acids Res* **27**: 3911–3920.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540–552.

Chenoll E, Macian MC & Aznar R (2006) *Lactobacillus rennini* sp. *nov.*, isolated from rennin and associated with cheese spoilage. *Int J Syst Evol Micr* **56**: 449–452.

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.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.

Huang X & Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Res* **9**: 868–877.

Mulder N & Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. *Method Mol Biol* **396**: 59–70.

Noguchi H, Park J & Takagi T (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34**: 5623–5630.

Sambrook J, Fritisch EF & Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schroeter J & Klaenhammer T (2009) Genomics of lactic acid bacteria. *FEMS Microbiol Lett* **292**: 1–6.

Smith MC & Thomas CD (2004) An accessory protein is required for relaxosome formation by small staphylococcal plasmids. *J Bacteriol* **186**: 3363–3373.

Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**: 711–721.

Toomey N, Monaghan A, Fanning S & Bolton DJ (2009) Assessment of horizontal gene transfer in Lactic acid bacteria–a comparison of mating techniques with a view to optimising conjugation conditions. *J Microbiol Meth* **77**: 23–28.

van Kranenburg R, Kleerebezem M & de Vos WM (2000) Nucleotide sequence analysis of the lactococcal EPS plasmid pNZ4000. *Plasmid* **43**: 130–136.

Wada T, Noda M, Kashiwabara F *et al.* (2009) Characterization of four plasmids harboured in a *Lactobacillus brevis* strain encoding a novel bacteriocin, brevicin 925A, and construction of a shuttle vector for lactic acid bacteria and *Escherichia coli*. *Microbiology* **155**: 1726–1737.

Waterhouse AM, Procter JB, Martin DM, Clamp M & Barton GJ (2009) Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.