Omics based comparative analysis of putative Mobile Genetic Elements in Lactococcus lactis

Andersen, Joakim Mark; Pedersen, Christine Møller; Bang-Berthelsen, Claus Heiner

Published in: FEMS Microbiology Letters

Link to article, DOI: 10.1093/femsle/fnz102

Publication date: 2019

Document Version
Peer reviewed version

Link back to DTU Orbit

Omics based comparative analysis of putative Mobile Genetic Elements in *Lactococcus lactis*

Joakim Mark Andersen*, Christine Møller Pedersen, Claus Heiner Bang-Berthelsen

National Food Institute, Technical University of Denmark, Kemitorvet, building 201, 2800 Kgs. Lyngby, Denmark

*Corresponding author: joakm@food.dtu.dk, +45 35 88 70 00

Abstract

*Lactococcus lactis* is globally used in food fermentation. Genomics is useful to investigate speciation and differential occurrence of (un)desired gene functions, often related to mobile DNA. This study investigates *L. lactis* for putative chromosomal mobile genetic elements through comparative genomics, and analyse how they contribute to chromosomal variation at strain level. Our work identified 95 loci may range over 10% of the chromosome size when including prophages and the loci display a marked differential occurrence in the analysed strains. Analysis of differential transcriptomics data revealed how mobile genetic elements may impact the host physiology in response to conditional changes.

The insight in the genetic variation of mobile genetic elements in *L. lactis* holds potential to further identify important functions related to food and biotechnology applications within this important species.

Introduction

*Lactococcus lactis* is one of the main species used in dairy starter cultures globally to ferment various dairy products such as buttermilk and cheese. The *L. lactis* strains used in dairy fermentation are thought to originate from a plant niche (Price 2012), but the historical repeated culturing and back-slopping of milk fermentations and modern strain selection have enforced a rapid adaptation into the milk environment of the used strains. The two subspecies lactis and cremoris contain strains that are used because they contribute differently to the manufacturing of cheese through protein breakdown, acidification and flavour production (Ward, Brown and Davey 1998).

*L. lactis* is part of gram-positive, acid-tolerant group of lactic acid bacteria (LAB). LAB produce lactic acid and other metabolites during, traditionally, milk fermentation. This affect the flavour, and texture of the dairy products (Cavanagh 2015). Thereby the strains have been specialised for growth in a nutrition rich environment, that provides all the necessary carbohydrates, proteins, fats, vitamins and minerals for cultivation (Kelly 2010, Leroy 2004, Kelleher 2017). This has also impacted the cell physiology, genetics and strain diversity - mainly by loss of gene functions through reductive genome evolution (Goh 2011, Kafsi 2014).

The food, feed and biotech industries need new and improved products based on LAB fermentations (Kandasamy et al. 2016, Chen et al. 2017, Poulsen, Derkx and Oregaard 2019) gives a continuous request for strains with alternative or better traits without the use for GMO technologies or outside of the dairy niche of food fermentation. Of particular note, LAB genomes are well-
known to harbour genes and loci involved with genetic mobility such as insertion sequences or integrative conjugative transposons. These actively may confer uptake of advantageous functions, loss of expendable loci (Selle 2015), or possible spread of undesired genes such as antibiotic resistance from related genera (Gawryszewska 2017) and thus play an active role in speciation and niche adaptation in food settings. To date the main work has been focused on the contributions of plasmids (Ainsworth et al. 2014) for horizontal gene transfer (HGT). HGT events in the chromosome on the other hand are unfortunately sparse and date back to pre-routine-genome sequence analysis but contain important findings such as nisin production related to the Tn5276 conjugative transposon (Rauch and De Vos 1992).

Comparative genomics in LAB is well established (Makarova 2006) and has greatly advanced our knowledge of how LAB species have diversified through genetic alterations (Smokvina 2013). Previous work has investigated mobile genetic elements (MGE) in LAB (Dahmane 2017, Eng 2011). However without known phenotypes associated to the MGE, prediction of both new MGE and their associated encoded functions are difficult (Johnson 2015), and more systematic approaches are needed to understand both the genetic diversity and impact on physiology on the species level. Furthermore, despite the routine use of genome sequencing, LAB genomes are rich in highly similar and redundant IS loci, which hamper identification of MGE when flanked by transposase genes (Darmon 2014). *L. lactis* specifically is known to host numerous transposase genes and other genes associated with transposable elements (Kelleher 2017), and some characterised MGE were found to have functional accessory genes like the bacteriocin nisin (Golomb 2017).

The aim of this study is to investigate *L. lactis* subsp. *lactis* and *cremoris* genomes with focus on the chromosomal for occurrence and diversity of putative chromosomal MGEs and to gain insight into physiological diversity and adaptation and speciation potential of highly variable DNA within the species through transcriptome analysis along with investigation for small regulatory RNA.

Materials and methods

**Sequence collection and analysis**

A total of 11 complete genomes and 100 draft genomes in scaffolds or contigs of *L. lactis* were obtained from NCBI (2. February 2018, Table 1 and table S1 respectively).

All DNA sequence analysis was performed in Geneious version 11.0.5 (Biomatters). Whole chromosome alignments were performed using progressive MAUVE (Darling 2004), standard settings. Prediction of bacteriocin encoding genes was done using BAGEL4 (April 2018, van Heel 2013). Antimicrobial resistance genes were predicted with ResFinder 3.0 (April 2018, Zankari 2012) using a sequence homolog cut off of 80% sequence identity and minimum length 60%. In silico MLST typing of available *L. lactis* genomes was performed using the method by Passerini et al. (Passerini 2010). The typing scheme utilized six conserved housekeeping genes (*pgk, glyA, recN, bcaT, deoA, pepX*), where the full length gene was identified within each genome. The
collection of loci were aligned using MAFFT (PAM1 K = 1 and default settings otherwise). Phylogenetic tree was constructed using Geneious.

**Identification of mobile genetic elements from genome sequences**

For identification of putative chromosomal MGE, we defined MGE as a genetic locus that contained 1) a gene putatively associated with DNA mobility (e.g., transposase, transposon, or integrase), 2) encoded accessory genes (i.e., genes not associated with genetic mobility), and 3) were differentially occurring in one or more pairwise MAUVE based alignment accessed by local conserved genetic synteny in the flanking region of the MGE.

Putative genes involved in genetic mobility were identified by text mining CDS annotations within each chromosome with the search inputs "phage", "integrase", "transposon" and "transposase".

Pairwise whole chromosome alignments (N=55) were performed for the 11 sequenced strains, (Table 1) with progressive MAUVE. Notably however strains MG1363 and NZ9000 were found to be of such high similarity that no differential MGE were identified in the pairwise comparison as NZ9000 being a strain derivative of MG1363 thus NZ9000 is omitted from the further analysis. Sequence alignment of all MGE was done to remove redundant MGE found to differentially occur among several strains. Identification of MGE in draft genomes was based on sequence homology of at least 60% length and 90% nucleotide identity (Table S1).

Subsequently, we investigated 5000 bases down and upstream each MGE from the parent chromosome to predict any repeats flanking the MGE (minimum length 30 bases, max 10% mismatches) using Geneious Prime (version 2019.0.4) to pinpoint the MGE boundaries. Additionally, we used the above flanking sequences adjacent to each MGE to compare genetic synteny across the 10 used genomes. In detail, the flanking regions for each MGE was identified in all 10 chromosomes, if possible, with or without MGE. These combined loci were extracted per genome and aligned using MAFFT (standard settings). The MGE endpoint and thus MGE length could be confirmed from the alignment gaps from genomes not hosting the MGE. From this combined analysis we noted the genetic context for each MGE (e.g., if it could be deduce if the MGE was a likely deletion or insertion).

**Transcriptional analysis of MGE**

Available transcriptome data were obtained from NCBI and re-analysed in Geneious as per the respective authors in the original publications.

**Analysis of short RNAs in MGE**

Available analysis of sRNA and asRNA in *L. lactis* subsp. *cremoris* MG1363, differentially expressed during stress conditions, were compared with presence in our predicted MGE. In short RNAs identified in MGE, a search for sequence homolog using BLAST with “somewhat similar sequences” as BLAST option (April 2018) on nucleotide-nucleotide level only in MG1363 was done to predict sequence homologs in MG1363 for putative regulation. Genes downstream to the sequences were noted.
Results

Identification of mobile genetic elements in Lactococcus lactis

*L. lactis* is well known to host numerous insertion sequences, transposase genes and other genes associated with conjugative and transposable elements (Kelleher 2017). The presence and transfer of plasmids have been the main determinant for the knowledge we currently have on associated mobile functions in *L. lactis* (van Mastriigt et al. 2018). Therefore we have directed our focus in this work on chromosomally integrated IS genes and other loci putatively involved with HGT while also excluding prophages as their impact is beyond the scope of this work and covered elsewhere (Kelleher et al. 2018). We identified a total of 687 IS genes in the chromosomes of 10 analysed strains.

The previous identification and physiological impact of IS genes is hampered by the variable nature of small genome structural changes and often also the lack of functional knowledge associated with the genes encoded within MGE. A phylogenetic tree (Figure 1) shows the natural genetic diversity of the *L. lactis* subsp. *lactis* and *cremoris* and how the 10 fully sequenced strains represent most of the main linages within the two subspecies. We investigated the occurrence and diversity of putative MGE in five *L. lactis* subsp. *lactis* and five subsp. *cremoris* (Table 1 and Table S1) to provide insight into how MGE is distributed and shape this important species, and how they may contribute to potential phenotypes of industrial importance in modern food fermentations. From the *in silico* identification of MGE we found 95 MGEs (Table S2) of which 23 were prophages. From the prediction of MGE by pairwise genome comparison we confirmed the end points of each MGE by comparing the flanking regions of each MGE to the synteny to chromosomes not encoding the MGE. Interestingly we found repeat sequences directly flanking or within 11 transposase associated MGE, which further supports the genetic end points or for the internal repeats suggests how MGE may be composite originating from multiple insertion or deletions. Based on inspection of the genetic synteny of IS associated MGE (N=14) it is in some cases not possible to deduce if the loci diversity is due to an insertion or IS mediated chromosomal deletion as listed in Table S2.

Notably upon comparison of predicted MGE to strain KW2, reported not to encode any IS genes, we initially predicted additional 14 putative MGE, originating from other chromosomes. It is likely that these were false positive from the analysis that are not part of the 95 predicted MGE. These loci are however still listed in Table S2 as negative examples of the systematic and rigorous approach which is needed in *in silico* analysis of genetic diversity.

Interestingly, the total size distribution of putative MGE per strain differed markedly ranging from one putative integrase associated MGE in KW2 to 8.7 % of the total chromosome, excluding prophages, in strain A76 (Figure 2). *L. lactis* subsp. *cremoris* SK11 was found to encode the highest number of putative MGEs (20 MGE) whereas IO-1 encoded the largest size fraction (198 kbp). The mean summarisation of MGE size in each chromosome was 153 kbp while the size of the individual MGE varied greatly (Figure S1 and Table S2), with the majority between 2-35 kbp. Some larger MGEs were found at up to 60 kbp whereas phage sizes were distributed around 30-50 kbp.

Interestingly, the accessory genes within MGE were found to be enriched with unknown functions (hypothetical proteins and domains with unknown functions) as compared to the remaining chromosomal genes, including prophage genes but excluding predicted MGE. The MGE was found to contain 40.0% genes with unknown functions compared to 21.4% in the chromosomal
genes, indicating that less is known about gene functions on MGE while they amount to a marked proportion of the genome.

**Distribution and diversity of MGE in Lactococcus lactis**

To assess the overall distribution of MGEs compared to phylogeny we identified sequence homologs of all identified MGEs within the tested 10 genomes (Figure 3, Table S3). From this analysis, we observed how some MGEs were seemingly exclusive to each subspecies (N=27 for cremoris and N=18 for lactis and N=9 shared for both subspecies) as they tended to be found in either subspecies suggestive of phylogenetic spread of the MGE indicative that transposase genes may become inactive through mutation post chromosomal integration. A few MGEs were widely spread, whereas other MGEs were uniquely found in one or only a few strains, indicating HGT.

Upon analysis of co-occurring MGEs identified across multiple strains we observed that two clusters of seemingly conserved MGEs based on full length and high nucleotide identity (above 90% identity) displayed variation within a subset of accessory genes (Figure 4). One strain (KLDS 4.0325) was found to host three highly similar MGEs at different genomic positions. This indicates how MGEs may originate and duplicate from an early insertion, and thereafter diverge through their independent evolution by differential acquisition of addition accessory genes over time and exemplify the high degree of variation caused by and within MGE.

**Transcriptional analysis of MGE under relevant physiological conditions**

To investigate the involvement of MGE during cellular processes, we used previously published transcriptional data. Notably, transcriptomics analysis may indicate if the MGE are indeed transcribed at all, and reveal regulatory patterns within the MGE, which could aid to further deduce the genetic structure and functions encoded within MGE. The analysis was limited to RNA sequencing data, excluding microarray data, because of the platform variation and DNA probe dependency of microarray platforms, which may not include genes in all MGE.

Upon analysis of the differential transcriptome for KF147 grown in complex M17 media, supplemented with glucose, compared to liquid Arabidopsis juice (Golomb 2015) we integrated our prediction of MGE within the KF147 chromosome to test any transcriptional response of MGE. We observed how all MGEs increased the overall gene expression when comparing the Arabidopsis grown culture to the M17 media (Figure 5, Table S4). The statistical analysis of single genes, differential regulation was used to identify genes within all KF147 MGE that were differential regulated (Table S5). Here it is obvious that the MGE encoded genes compose a marked proportion of the differential transcriptome as a response to diverse growth condition compared to rich optimized media. This is further supported by the observation that only three genes were identified to be downregulated. Statistical analysis was made to test whether there were a significantly higher percentage of MGE genes upregulated than chromosomal, with upregulation defined with a p-value < 0.05 and log2 fold change ≥ 1. Chi-square statistic found a p-value at 0.8×10⁻⁴ suggesting that MGE are significantly more enriched in upregulated genes.

To further analyse the gene contribution from MGE to the host functions, we analysed the transcriptome that displayed the highest fold change (Table S5 and S6). From inspection of the top 10% highest up-regulated genes (N=258) we observed that 55 genes were encoded in MGE and 24
were encoded in prophage regions. Based on putative gene annotations we identified a glycerol utilization system (locus tag llkF_1309-1311), previous identified to be an precursor in synthesis of lipids in Gram-positive bacteria (Bizzini 2010). Additionally, we identified several carbohydrate utilization systems, including an α-galactosidase utilization system, similar to that of Lactobacillus acidophilus (Andersen 2012), and three PTS transporters co-encoded with putative phospho-sucrase and phospho-β-glucosidases. Notably also the nisin gene cluster (Dodd HM, Horn N 1990) was found to be up-regulated.

Transcriptional data from RNA sequencing for MG1363 was analysed from van der Meulen et al. (van der Meulen 2017) (Table S7). Here we observed regulation of only few MGE encoded genes. These were mostly upregulated during cold stress and downregulated in heat, acid and starvation. Notably, genes in MG1363-41 were upregulated due to cold exposure, and were annotated as cold-shock proteins, while the gene, ABC transporter ATP-binding protein in MG1363-11, was upregulated during starvation likely in attempt to couple nutrient uptake through carbohydrate ABC transporters.

**sRNA and stress induced transcriptional regulation in MG1363**

Small RNA, such as non-coding regulatory RNA (sRNA) and antisense RNA (asRNA), are found to have a regulatory impact on the genomes through RNA-regulation (van der Meulen 2016). We investigated if sRNA and asRNA was found or even enriched in MGEs, using previously published data (van der Meulen 2016), with additional transcriptome data from RNA sequencing (van der Meulen 2017). Within our identified MGE we identified 41 sRNA out of 186 and 13 asRNA out of 60, which is respectively 22% sRNA and 22% asRNA encoded on MGE (Table S8 and Table S9 respectively).

For further analysis, we examined MG1363-41 encoding nine sRNA and two asRNA. In search for sequence homologs of the sRNA in the chromosome, we found three sRNA and one asRNA that displayed sequence homolog in the MGE. This might putatively regulate within the MGE, while the rest of the sequence alignments were located outside the MGE in the chromosome (Figure 6 and Table S10). These small RNAs were found to be regulated during different stress conditions, with two noticeable genes being upregulated during cold and having sequence alignment close to cold shock proteins.

**Putative function related to cell survival found in MGE**

To validate our analysis on how the identified MGEs may confer beneficial functions to *L. lactis*, we searched each MGE for genetic loci for genes putatively related to niche adaptation and survival, and metabolic diversity. Notably, we found genetic systems related to bacteriocin production, carbohydrate utilization, restriction modification systems and heavy metal resistance.
As confirmation of previous work we found a nisin gene cluster in CV56-8 and co-occurring in strains KF147, although some genes appear as pseudo genes, and IO-1 within the Tn5276 conjugative transposon (Rauch and De Vos 1992) which was found to differentially occurring within L. lactis (Rauch, Beerthuyzen and de Vos 1994). Interestingly we found glycosyl transferase family 2 and 8 (www.cazy.org) encoding genes in MG1363-27 that could potentially be used for glycosyl-modifications for exopolysaccharide production (Zeidan et al. 2017). For carbohydrate catabolism CV56-2 encoded an ABC transporter and a putative GH31 α-xylosidase. Both KF147-5 and KF147-9 encoded putative phosphotransferase systems associated with GH1 enzymes and a sucrose-6-phosphate hydrolase, and lastly KF147-19 encoding an α-galactoside utilization system (GH36 and an ABC transporter) as previously described (Golomb and Marco 2015).

Notably, restriction-modifications systems involved with phage resistance in L. lactis (Nyengaard, Vogensen and Josephsen 1995) were found in SK11-9 and UC509.9-1 (Type-I), and SK11-8, MG1363-11 and IO-1-8 (Type-II) supplementing previous findings of restriction-modification systems encoded on L. lactis plasmids (Twomey, McKay and O’Sullivan 1998). We found a magnesium-transporting ATPase in SK11-24 putatively involved with Mg2+ uptake and osmotic control. For MG1363 we found a putative tellurium resistance gene cluster in MG1363-4 that was found to have only limited sequence homology within L. lactis but displayed sequence similarity to a plasmid encoded tellurium resistance gene from Serratia marcescens (Whelan, Colleran and Taylor 1995) and an arsenical resistance gene cluster in MG1363-41. Additionally for IL1403 we found a putative cadmium exporter in IL1403-11 unrelated to previous work of analysis of cadmium resistance in L. lactis (Trotter et al. 2001; Sheng et al. 2016).

**Discussion**

Comparative genome, both chromosomal and extrachromosomal, analysis and the occurrence of MGE within the industrially important L. lactis species has potential for analysis of natural diversity and their related genetically encoded functions to be exploited in food, feed and biotechnological applications. In this study, we performed comparative analysis of 11 L. lactis chromosomes to systematically identify the occurrence and diversity of chromosomal MGE and proposing how MGE may impact strain specific phenotypes.

Based on our findings, we show how for most strains MGE accounts approximately 6% on average of the total chromosome, not including prophages, and how the distribution of MGEs can follow both phylogeny and strain to strain specific behaviour which can be reproduced also within draft genomes. The *in silico* approach applied here was used to identify MGEs with
great variation in size and number of accessory genes. This further shows how a systematic identification approach must be taken to encompass such a size diversity. However careful iterations of validation must be taken to avoid false positive prediction. We found how single IS genes may result in disruption of genome alignments and how comparison of genetic synteny flanking each MGE aids in endpoint and length estimations after pairwise genome comparisons. Future work could include MGE differentiation centred on the groups or families of IS genes found based on related species (Eraclio 2015). Albeit we cannot conclude if each MGE is truly mobile and able to copy or excise, previous work have identified specific L. lactis MGEs to be transferable from strain to strain (Machielsen 2011) or lost through environmental adaptation (Bachmann 2012). Alternatively, the abundance of chromosomal homologous IS elements, or shorter repeats as described above, can also drive gene activity (Solopova 2017) or genetic loss through homologous recombination (Selle 2015) as an indication of how differential occurrence of MGE might be due to gene loss rather than gain of mobile DNA.

The distribution of MGEs compared to the species phylogeny suggests some conservation of the MGEs within the two subspecies. It might indicate that some MGEs have lost their mobility, and rather followed a convergent evolution. The phylogeny also suggests that these MGEs could have been present in ancestral genomes, but lost in strains during adaption to new habitats. This is plausible as per the previous findings that MGE encoding pathways for utilization of plant carbohydrates, were lost during cultivated in milk (Bachmann 2012).

Interestingly, from the identification of MGEs we observed how they are enriched in genes with unknown functions (40.0% compared to 21.4% within the remaining chromosomal genes). The lack of putative functional assignment both limits the current prediction of physiological role but also highlights the need for further characterization. Interestingly, based on homology to previously characterized functions, we identified gene clusters related to diversity of carbohydrate utilization systems, restriction modifications systems, nisin production and heavy metal resistance. These putative functions confirm the metabolic diversity related to the differential occurrence of MGE within L. lactis.

To address the lack of knowledge of the single genes we applied a top-down approach by integrating previous transcriptional work to evaluate differential gene expression under various conditions (growth on plant material for KF147 (Golomb 2015) and stress in MG1363 (van der Meulen 2017) in combination with prediction tools for bacteriocins and antibiotic resistance genes. Interestingly, we found no antibiotic resistance gene clusters within our predicted MGE. We
observed a marked increase in gene expression within MGEs for strain KF147 when grown in a liquid plant media compared to complex laboratory medium. This indicates a potential for differential response to condition change for the MGE, and appear to be enriched in the *lactis* subspecies, which is often associated with plant isolates.

This observation also supports how horizontal transfer of MGE gives an initial low metabolic burden on the acquiring host chromosome, as the genes have low basal expression. Thus, acquired MGEs have a low risk of evolutionary selection for loss of MGE but may impact niche adaptation rapidly. Additionally, of the predicted functions within the MGE, several gene clusters were associated with carbohydrate metabolism associated with plants such as α-galactosides from the raffinose family oligosaccharides. To further support the niche adaptation, recent work (Golomb 2017) found a mobile element similar to KF147-20 identified in this study, which encoding an active hybrid non ribosomal peptide synthetase/polyketide synthase (NRPS/PKS) system. The NRPS/PKS system was involved in improving the hydrogen peroxide and oxygen stress tolerance to make the cells tolerate the aerobic conditions on leaf or as counter response to pathogenic bacteria.

Analysis of the presence of putative sRNA and asRNA in MG1363 suggested that MGEs may be enriched with these sequences and potentially transfer the associated regulatory functions horizontally. Previous work also found more than one third of sRNA to be in areas with prophages (van der Meulen 2016). It was suggested that the asRNA either had origin in these phages, or had evolved after integration of phages, and may gene silence prophage genes. Though functions of sRNA are still unknown, they suggested one sRNA to be involved in catabolite repression of carbon uptake and metabolism through stabilisation of the llmg_0957-llmg_0963 operon (van der Meulen 2016).

Our study found sRNA and asRNA homologs both within the MGE and the chromosomal genes. This indicates possible contribution to regulation of genes with a putative role in bacterial host transcriptome of the affected loci. It was seen that some sRNAs were upregulated during exposure to low temperatures. The sRNA sequence homologs were approximate to cold-shock proteins, indicating MGE to have putative differential regulation potential of relevant genes during stress.

Since recombinant DNA techniques cannot be used in the food starter cultures yet, we may instead look at the diversity of different genomes to discover new traits to apply in culture and food product development. The variable part of the genomes and a characterisation of these might
be useful in the search for new strains or for a wider genetic diversity. Focus on MGE and a plausible way to make selection for these elements might be useful in the search for strains with the desired genetic diversity.

In summary, we identified 95 putative MGE within 10 *L. lactis* chromosomes through a systematic comparative genomics approach. The MGE occurrence showed variation among strains with some conservation within subspecies, suggesting how MGE compose a marked, yet highly variable, part of the variable genome and may actively be part of the on-going speciation within *L. lactis* through horizontal gene transfer, or gene loss. This may confer additional functions applicable in food cultures.

Transcriptional analysis of MGE and prediction of sRNA within MGE further substantiate how mobile DNA contribute to niche adaptation and host physiology through transcriptionally targeting both chromosomal genes by sRNA and diverse carbohydrate utilization. This work advances our genomics insight into mobile DNA within a highly important species and its application in food fermentation and biotechnology applications.

**Acknowledgements**

We thank Peter Ruhdal Jensen for providing the opportunity to conduct the study. We also will like to thank Jun Chen and Miriam Meister for valuable feedback on the writing and fruitful discussions.

**Funding**

This work was supported by the Innovation Fund Denmark (Grant number 4106-00037B) for JMA. CHBB has been supported by the Beckett foundation (http://www.beckett-fonden.dk/) as well as AP Moller Foundation (https://www.apmollerfonde.dk/). Additionally, we have received funding for the study from a DTU discovery grant (Grant number 13451). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests**

The authors declare no competing interest: CHBB and JMA are employed at the Technical University of Denmark during the preparation of the manuscript. This does not alter our adherence to FEBS Microbiology Letters policies on sharing data and materials. CMP was a student at Technical University of Denmark during the preparation of the manuscript.

**Author contributions**

All Authors were involved in writing the original manuscript, as well as review & editing the final version. JMA and CHBB were involved in conceptualization of the project, JMA and CMP did the
sequence analysis, CHBB conducted the project administration and supervision together with JMA. All authors contributed to visualization and figure making.

References


Cavanagh D, Fitzgerald GF, McAuliffe O. From field to fermentation: The origins of *Lactococcus lactis* and its domestication to the dairy environment. *Food Microbiol* 2015, DOI: 10.1016/j.fm.2014.11.001.


Figure 1: The phylogeny of 113 *L. lactis* subsp. *lactis* and *cremoris* genomes. Strain names are highlighted for the 10 genomes used in comparative genome analysis for putative mobile genetic elements.
Figure 2. Distribution of MGE within the analysed strains. A: The size distribution of identified MGE (green), prophages (orange) and chromosomal genes (grey) per analysed genomes. B: Listed number of MGE, sum of MGE sizes, percentage of MGE in the chromosome, number of phage clusters, sum of kbp phage sizes, percentage of phages in the chromosome.

<table>
<thead>
<tr>
<th></th>
<th># MGE</th>
<th>kbp MGE</th>
<th>% MGE</th>
<th># phage</th>
<th>kbp phage</th>
<th>% phage</th>
<th>%MGE+phage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A76</td>
<td>5</td>
<td>214</td>
<td>8.7</td>
<td>2</td>
<td>57</td>
<td>2</td>
<td>11.1</td>
</tr>
<tr>
<td>KW/2</td>
<td>1</td>
<td>5</td>
<td>0.2</td>
<td>1</td>
<td>48</td>
<td>1</td>
<td>2.1</td>
</tr>
<tr>
<td>MG1363</td>
<td>13</td>
<td>194</td>
<td>7.7</td>
<td>2</td>
<td>86</td>
<td>3</td>
<td>11.1</td>
</tr>
<tr>
<td>SK11</td>
<td>20</td>
<td>168</td>
<td>6.9</td>
<td>3</td>
<td>105</td>
<td>3</td>
<td>11.2</td>
</tr>
<tr>
<td>UCS09.9</td>
<td>13</td>
<td>84</td>
<td>3.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.7</td>
</tr>
<tr>
<td>CV56</td>
<td>5</td>
<td>113</td>
<td>4.7</td>
<td>3</td>
<td>107</td>
<td>7</td>
<td>11.4</td>
</tr>
<tr>
<td>IL1483</td>
<td>9</td>
<td>150</td>
<td>6.3</td>
<td>5</td>
<td>166</td>
<td>4</td>
<td>9.1</td>
</tr>
<tr>
<td>IO-1</td>
<td>7</td>
<td>198</td>
<td>8.2</td>
<td>1</td>
<td>43</td>
<td>2</td>
<td>7.6</td>
</tr>
<tr>
<td>KS147</td>
<td>12</td>
<td>214</td>
<td>8.2</td>
<td>2</td>
<td>76</td>
<td>3</td>
<td>11.2</td>
</tr>
<tr>
<td>ELS15</td>
<td>14</td>
<td>186</td>
<td>7.2</td>
<td>4</td>
<td>153</td>
<td>6</td>
<td>13.1</td>
</tr>
</tbody>
</table>
Figure 3. Occurrence of all MGE excluding prophages in the 10 *L. lactis* genomes in comparison to their phylogeny. The presence of MGE is shown in grey whereas the absence is shown in white.
Figure 4. Five apparently conserved MGE based on full length and high nucleotide identity (90% identity) in 3 different strains. Gene color indicates: genes that are highly similar (90% identity) (green), structural varying (yellow), less than 90% similar (grey) or genes associated with mobility (purple).

Figure 5. Gene expression of *L. lactis* subsp. *lactis* KF147 in rich M17 media (blue) and in liquefied Arabidopsis (red) for chromosomal genes and genes within each identified MGE.
Figure 6. sRNA and asRNA in MG1363-41. A) Annotations for small RNA, as referrers to asRNA and LLMG referrers to sRNA. B) Sequence homologs of sRNA/asRNA in MGE is shown with arrows, while lack of arrows indicate sequence homologs in the chromosome.

Table 1: Complete *L. lactis* genomes used in this study with the number of TE/IN genes identified per genome

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Assembly</th>
<th>Accession number</th>
<th>number of TE/IN</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Lactococcus lactis</em> subsp. lactis IL1403</td>
<td>GCA_000006865.1</td>
<td>NC_002662.1/AE005176.1</td>
<td>74</td>
<td>Dairy</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. cremoris SK11</td>
<td>GCA_000014545.1</td>
<td>NC_008527.1/CP000425.1</td>
<td>119</td>
<td>Dairy cheese</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. cremoris MG1363</td>
<td>GCA_000009425.1</td>
<td>NC_009004.1/AM406671.1</td>
<td>78</td>
<td>Dairy</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. lactis KF147</td>
<td>GCA_000025045.1</td>
<td>NC_013656.1/CP001834.1</td>
<td>13</td>
<td>Mung Bean sprouts</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. lactis CV56</td>
<td>GCA_000192705.1</td>
<td>NC_017486.1/CP002365.1</td>
<td>22</td>
<td>Vaginal microbiota</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. cremoris A76</td>
<td>GCA_000236475.1</td>
<td>NC_017492.1/CP003132.1</td>
<td>155</td>
<td>Dairy cheese</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. cremoris UC509.9</td>
<td>GCA_000312685.1</td>
<td>NC_019435.1/CP003157.1</td>
<td>116</td>
<td>Dairy</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. lactis IO-1</td>
<td>GCA_000344575.1</td>
<td>NC_020450.1/AE005176.1</td>
<td>5</td>
<td>Drain water</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. cremoris KW2</td>
<td>GCA_000468955.1</td>
<td>NC_022369.1/CP004884.1</td>
<td>2</td>
<td>Fermented corn</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. lactis KLDS 4.0325</td>
<td>GCA_000479375.1</td>
<td>NC_022593.1/CP006766.1</td>
<td>25</td>
<td>fermented horse milk</td>
</tr>
</tbody>
</table>