

RESEARCH

Open Access



The lactococcal ICE-ome encodes a repertoire of exchangeable traits with potential industrial relevance

Simon van der Els^{1,2,3}, Jos Boekhorst¹, Peter A. Bron^{2,3} and Michiel Kleerebezem^{1,3*}

Abstract

Dairy industries apply selected lactococcal strains and mixed cultures to produce diverse fermented products with distinctive flavor and texture properties. Innovation of the starter culture functionality in cheese applications embraces natural biodiversity of the *Lactococcus* species to identify novel strains with alternative flavor or texture forming capacities and/or increased processing robustness and phage resistance. Mobile genetic elements (MGE), like integrative conjugative elements (ICEs) play an important role in shaping the biodiversity of bacteria. Besides the genes involved in the conjugation of ICEs from donor to recipient strains, these elements also harbor cargo genes that encode a wide range of functions. The definition of such cargo genes can only be achieved by accurate identification of the ICE boundaries (delimiting). Here, we delimited 25 ICEs in lactococcal genome sequences with low contig numbers using insertion-sites flanking single-copy core-genome genes as markers for each of the distinct ICE-integrases we identified previously within the conserved ICE-core genes. For ICEs in strains for which genome information with large numbers of contigs is available, we exemplify that CRISPR-Cas9 driven ICE-curing, followed by resequencing, allows accurate delimitation and cargo definition of ICEs. Finally, we compare and contrast the cargo gene repertoire of the 26 delimited lactococcal ICEs, identifying high plasticity among the cargo of lactococcal ICEs and a range of encoded functions that is of apparent industrial interest, including restriction modification, abortive infection, and stress adaptation genes.

Keywords *Lactococcus lactis*, Mobile genetic elements, Integrative conjugative element (ICE), Comparative genomics, ICE-cargo

Introduction

Lactococcus lactis serves as a paradigm representative of the lactic acid bacteria, and is used in the dairy industry for the production of cheese and buttermilk [1]. The

main function of *L. lactis* in the dairy fermentation process is the rapid acidification of milk by the conversion of the milk sugar lactose to lactic acid, which extends the shelf life of the product. In addition, fermentation by *L. lactis* adds important texture and flavor characteristics to the fermentation end-product. The applicability of individual strains of *L. lactis* depends on their performance during the entire process from starter culture production to their final application in dairy products. Robustness during starter culture production and processing conditions is critical for the applicability of a particular strain [2]. Moreover, reproducible performance of the starter culture in cheese applications conditions, including rapid

*Correspondence:

Michiel Kleerebezem
michiel.kleerebezem@wur.nl

¹ Host-Microbe Interactomics Group, Department of Animal Sciences, Wageningen University & Research, De Elst 1, Wageningen 6708 WD, The Netherlands

² NIZO B.V, Kernhemseweg 2, Ede 6718 ZB, The Netherlands

³ BE-Basic Foundation, Mijnbouwstraat 120, Delft 2628 RX, The Netherlands



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

milk acidification and (long-term) ripening, is required to obtain products with consistent flavor and texture properties [3]. In addition, application of dairy starter cultures is continuously challenged by bacteriophage predation, which may cause failure of production batches [4].

There is a strong impetus to harness the natural biodiversity of the *Lactococcus* genus to develop innovations in robustness, flavor, and shelf-life of the fermented products. Functional differences of strains that are based on core-genome functions present in all strains of the species are predominantly determined by variations in strain-specific gene expression levels, but do not encompass functions that are specific to a particular strain, which is often referred to as the 'variome'. Mobile genetic elements (MGE) play an important role in shaping the variome of bacterial strains. MGEs that can readily be transferred from one strain to another by mating and conjugation include plasmids and integrative conjugative elements (ICEs). These MGEs can harbor functions that are of importance for industrial application and may be critical for strain adaptation to, and fitness in the dairy environment [5–7]. For example, functions that have been reported to be encoded by plasmids include those that accelerate growth in the protein-rich dairy environments, such as extracellular proteases and oligopeptide transport systems (Opp), as well as the lactose utilization pathway involved in the efficient fermentation of lactose [8–10]. Similarly, several industrially important traits of *L. lactis* are encoded by ICEs, including the production of the bacteriocin nisin that can protect against food spoilage by other bacteria [11, 12], which is co-localized with the genes involved in sucrose utilization on the ICE designated Tn5276 [12]. Another example is Tn6098 which encodes the capacity to utilize raffinose [13], allowing *L. lactis* strains harboring this ICE to ferment substrates containing this carbon source, e.g., soy. The ICE-encoded genes that provide these phenotypes belong to the variable region of ICEs that are not predicted to play a role in the mobilization and transfer of ICEs, and are commonly referred to as ICE- 'cargo'.

ICEs are commonly present in a subset of strains and are therefore not part of the core-genome of the species. They are typically integrated in, and replicate passively within the replicative cycle of their host's chromosome, ensuring their maintenance in their host's progeny through vertical transfer. However, ICEs encode the capacity for excision from the host chromosome, leading to a circularized, extra-chromosomal form of the ICE. Following excision, ICEs can express a dedicated conjugation machinery that allows their transfer to a nearby and compatible recipient cell, a process termed horizontal transfer [14]. Notably, in their integrated state, ICEs do not express the functions involved in conjugal transfer,

indicating that excision is a prerequisite for the initiation of the conjugation process [15]. Transcription of the ICE-encoded genes that are involved in excision, including the integrase, has been proposed to be triggered by specific environmental or bacterial physiology conditions. For example, the *recA* dependent SOS response in ICE hosting cells [16], cell–cell signaling via quorum sensing to detect the presence of potential acceptor cells [17], and stationary phase of growth specific sigma factors [18], have been associated with the activation of ICE excision. However, ICE excision could also result from stochastic gene expression bistability, as has been described for bet-hedging strategies [19].

The insertion in, and excision from, the bacterial host's chromosome is facilitated by the ICE-encoded integrase, which commonly belongs to the tyrosine recombinase protein family [20, 21]. The insertion-site specificity is determined by the active site of the integrase, which recognizes specific attachment locations on the bacterial chromosome (*attB*) [22]. These *attB* sites are often located inside or in close proximity to tRNA encoding genes [23], which ensures successful integration in a variety of bacterial host strains due to the high degree of conservation of these integration loci. Recognition of the *attB* site is driven by the presence of an identical attachment site (*attP*) on the excised ICE, which facilitates their recombination that leads to the chromosomally integrated form of the ICE that is flanked by two identical copies of the attachment site (*attL* and *attR*). Recombination of the *attL* and *attR* site by the integrase during ICE excision generates the extrachromosomal and circular ICE form, leaving a single *attB* site behind in the chromosome, which has been referred to as excision 'scar'. Excision triggering as well as the capacity for conjugal transfer and the acceptor host range may vary considerably between ICEs, which will not be further detailed here, as the overall biology of ICEs has previously been reviewed [15].

The presence of ICEs in bacterial chromosomes can be investigated by the detection and recognition of conserved ICE functions (ICE-core) that play critical roles in the ICE lifecycle, including excision, conjugation and integration [24]. However, recognition of the ICE associated cargo genes is far from trivial, as delimitation of the ICE cannot be based on conserved genetic features that are associated with the 5'- and 3'-end of the ICE. The dual presence of identical *att* sites marking the boundaries of an integrated ICE should in theory allow for correct delimitation, however the variety in attachment site length, base pair composition and exact location is challenging in silico detection. Therefore, the precise detection of the boundaries of ICEs to determine their cargo remains a challenge.

In this study, we employed different strategies to achieve delimitation of ICEs in lactococcal genomes, allowing the definition of the cargo of 26 lactococcal ICEs. Using comparative genomics, 7 chromosomal ICE-insertion sites could be established that are largely conserved among strains of *L. lactis* and appear to reflect the target sites for the 7 integrase families recognized to be encoded by these ICEs [24]. Furthermore, we illustrate ICE delimitation strategies using comparative genomics or previously established CRISPR-Cas9 based ICE-curing [25] in strains for which genome information is available with low and high numbers of contigs, respectively. Taken together, our findings enable the effective delimitation of 7 distinctive ICE groups in *L. lactis* and cargo analysis of these ICEs identified a range functions of potential industrial interest. Importantly, the cargo of these distinctive ICE groups differed considerably, underpinning the high genetic plasticity and dynamic nature of these MGEs, which is in agreement with the high number of transposases (or their remnants) that are encoded in the ICE-cargo region.

Materials and methods

Bacterial strains and culturing conditions

The strains and their genome sequence-identifiers used in this study are listed in Table 1. *L. lactis* strains were grown at 30 °C without agitation in M17 medium (Tritium, Eindhoven, The Netherlands) routinely supplemented with 1% (wt/vol) glucose (Tritium, Eindhoven, The Netherlands). Erythromycin was added when appropriate at a final concentration of 10 µg/ml.

DNA extraction methods

PCR-grade chromosomal DNA was isolated by using InstaGene™ Matrix (Bio-rad, Veenendaal, The Netherlands). DNA for whole genome sequencing was isolated with a Promega Maxwell™ system, using standard protocols provided with the Promega Maxwell™ DNA Purification Kit (Promega, Leiden, The Netherlands).

ICE delimitation via comparative genomics

The ICE delimitation procedure was initiated by identifying the first *int*-neighboring single-copy gene belonging to the species' core-genome (*int* flanking core-gene) [26]. To identify the single-copy core-gene that flanks the

Table 1 Lactococcal strains used in this study. These were shown to containing a putative ICE [24]

| Genus | (Sub-)species | Strain | Isolation source | Nr contigs | Assembly | Reference |
|--------------------|-----------------|-------------|----------------------------|------------|-----------------|-------------|
| <i>Lactococcus</i> | <i>lactis</i> | KF147 | Mung bean sprouts | 2 | GCA_000025045.1 | [55] |
| <i>Lactococcus</i> | <i>lactis</i> | 184 | Dairy isolate | 7 | GCA_002078395.2 | [7] |
| <i>Lactococcus</i> | <i>lactis</i> | G423 | Protoplast shuffling | 1 | GCA_002804285.1 | [55] |
| <i>Lactococcus</i> | <i>lactis</i> | KF67 | Grapefruit juice | 110 | GCA_001456545.1 | [56] |
| <i>Lactococcus</i> | <i>lactis</i> | CV56 | Human isolate | 6 | GCA_000192705.1 | [57] |
| <i>Lactococcus</i> | <i>lactis</i> | IO-1 | Drain water | 1 | GCA_000344575.1 | [58] |
| <i>Lactococcus</i> | <i>lactis</i> | V4 | Raw sheep milk | 2 | GCA_032586255.1 | [24] |
| <i>Lactococcus</i> | <i>lactis</i> | A12 | Sourdough | 5 | GCA_900088425.1 | [59] |
| <i>Lactococcus</i> | <i>lactis</i> | N42 | Soil and grass | 6 | GCA_032464835.1 | [24] |
| <i>Lactococcus</i> | <i>lactis</i> | KF146 | Alfalfa and radish sprouts | 2 | GCA_032464735.1 | [24] |
| <i>Lactococcus</i> | <i>lactis</i> | AI06 | Açaí palm | 1 | GCA_000761115.1 | [60] |
| <i>Lactococcus</i> | <i>cremoris</i> | JM1 | Dairy isolate | 8 | GCA_002078895.1 | [7] |
| <i>Lactococcus</i> | <i>lactis</i> | KLDS 4.0325 | Koumiss | 7 | GCA_000479375.3 | [61] |
| <i>Lactococcus</i> | <i>lactis</i> | 229 | Dairy isolate | 6 | GCA_002078415.1 | [7] |
| <i>Lactococcus</i> | <i>lactis</i> | 275 | Dairy isolate | 5 | GCA_002078435.1 | [7] |
| <i>Lactococcus</i> | <i>lactis</i> | FM03 | Dairy isolate | 8 | GCA_002148215.1 | [62] |
| <i>Lactococcus</i> | <i>lactis</i> | UC06 | Dairy isolate | 4 | GCA_002078975.1 | [7] |
| <i>Lactococcus</i> | <i>lactis</i> | UC77 | Dairy isolate | 3 | GCA_002078615.1 | [7] |
| <i>Lactococcus</i> | <i>lactis</i> | NCDO2118 | Frozen Peas | 2 | GCA_000478255.1 | [63] |
| <i>Lactococcus</i> | <i>lactis</i> | 14B4 | Almond drupes | 2 | GCA_003176835.1 | [64] |
| <i>Lactococcus</i> | <i>lactis</i> | ML8 | Dairy starter | 4 | GCA_032464785.1 | [24] |
| <i>Lactococcus</i> | <i>lactis</i> | UL8 | Dairy isolate | 4 | GCA_002078855.1 | [7] |
| <i>Lactococcus</i> | <i>cremoris</i> | MG1363 | Dairy isolate | 1 | GCA_009661975.1 | [65] |
| <i>Lactococcus</i> | <i>lactis</i> | IL1403 | Dairy isolate | 1 | GCA_003722275.1 | [66] |
| <i>Lactococcus</i> | | M1734.1 | Artisanal cheese | 127 | GCA_007109175.1 | unpublished |

other boundary of the ICE, the *int* flanking core genes were mapped in the ICE-deficient *L. lactis* MG1363 genome and used for identification of its first neighboring single-copy core-gene in the direction where the ICE-encoded *int* was located in the ICE containing strain. These predicted ICE flanking single-copy core-genes was subsequently mapped in the chromosome of the ICE-containing strain to determine the genetic distance between the ICE flanking core genes in that strain to estimate the length and genetic content of the ICE. This synteny-driven procedure successfully identified the two core-genome functions that flank each of the 7 previously identified [24] *int*-discriminated ICE groups (Supplemental table ST1), and provides the genetic markers for the approximate delimitation of these ICEs, as well as the definition of their cargo in different strains. Since the approach used depends on conserved synteny of the ICE insertion loci, there can be specific strains where this approach fails to delimit their ICEs even if they belong to one of the distinct *int*-associated ICE groups. To obtain an inventory of all genes encoded by the delimited lactococcal ICEs, parsed genbank files were generated. Mapping of the two core-genes that were predicted to flank the ICE present in each of these genomes was used to assess the length of the ICE and extract the delimited ICE-locus, and ICE encoded genes were subsequently predicted and annotated with Prokka using default parameters [27], and ICE-cargo gene content was compared by BlastP (Supplemental file ICE-gb.zip). Further investigation of the selected cargo genes was performed using the conserved domain database hosted on the NCBI server [28]. Genome synteny dot plots were generated using the Gepard software tool with standard settings, and the images exported [29].

ICE_{KF67} curing and delimitation in *L. lactis* KF67

To obtain an ICE_{KF67} deficient derivative of strain *L. lactis* KF67, a previously described CRISPR-Cas9 based

curing approach was employed [25]. A short guide RNA (sgRNA) was designed to target a locus within the ICE_{KF67} and the corresponding annealed-oligonucleotides SG1 and SG2 were hybridized and (Table 2) ligated into Eco31I-digested pLABTarget [25]. The ligation mixture was transformed by electroporation [30] to *L. lactis* MG1363 as an intermediary cloning host that lacks the target sequence of the sgRNA. Upon recovery after electrotransformation, *L. lactis* cells were grown in recovery medium (M17, supplemented with 1% glucose, 200 mM MgCl₂, and 20 mM CaCl₂) [30]. The resulting colonies were used as templates in a colony-PCR using primers SG1 and SG3. One colony representing the anticipated amplicon profile was used for isolation of pLABTarget-KF67 using GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific, Breda, Netherlands), followed by transformation of the plasmid to electrocompetent KF67 cells [30]. The colonies obtained were anticipated to be ICE_{KF67} deficient derivatives of *L. lactis* KF67, which was verified by detection of the excision scar left behind by the excised ICE_{KF67}, as well as by the absence of the integrase encoded by the ICE_{KF67}, by PCR using primer pairs P1 + P2 and P1 + P3, respectively (Table 2). Two ICE_{KF67}-deficient isolates of *L. lactis* KF67 were subjected to whole genome sequencing using the ILMN Nextera XT library prep kit and sequenced on an Illumina NovaSeq 6000 platform (Baseclear BV, Leiden, The Netherlands). To determine ICE_{KF67} associated sequences, assembled contigs obtained for the ICE_{KF67}-deficient strain were mapped to the original *L. lactis* KF67 contigs, to identify contigs that were absent from the ICE_{KF67} cured strain. This procedure identified two *L. lactis* KF67 contigs (contig_8 and contig_14) that are partially absent in the ICE_{KF67}-deficient derivatives, thereby exactly defining the ICE_{KF67} sequence (39,264 bp) by their absence from the ICE_{KF67}-deficient derivative. Moreover, the contig comparison also confirmed the scar region left

Table 2 Plasmid and oligonucleotides used in this study for the curing and verification of ICE_{KF67}

| Plasmid | Relevant feature(s) or sequence | Source or reference |
|------------|--|-----------------------------|
| pLABTarget | EmR: pIL253 derivative harboring constitutively expressed Cas9 and a synthetic sgRNA expression handle | [25] |
| Primers | | |
| SG1 | TGATGACTATCTCGTTGTCATATA | KF67 sgRNA FW |
| SG2 | AAACTATATGACAACGAGATAGTC | KF67 sgRNA REV |
| SG3 | TTGAAGAACCCGATTACATGG | pLABTarget insert check REV |
| P1 | TGCCGGCATCATTCTACTCA | KF67 left core FW |
| P2 | CAACATCTTTGACCATCGTC | KF67 right core REV |
| P3 | ATTACAAGGAGAGCTGAAGCGG | KF67 <i>int</i> REV |

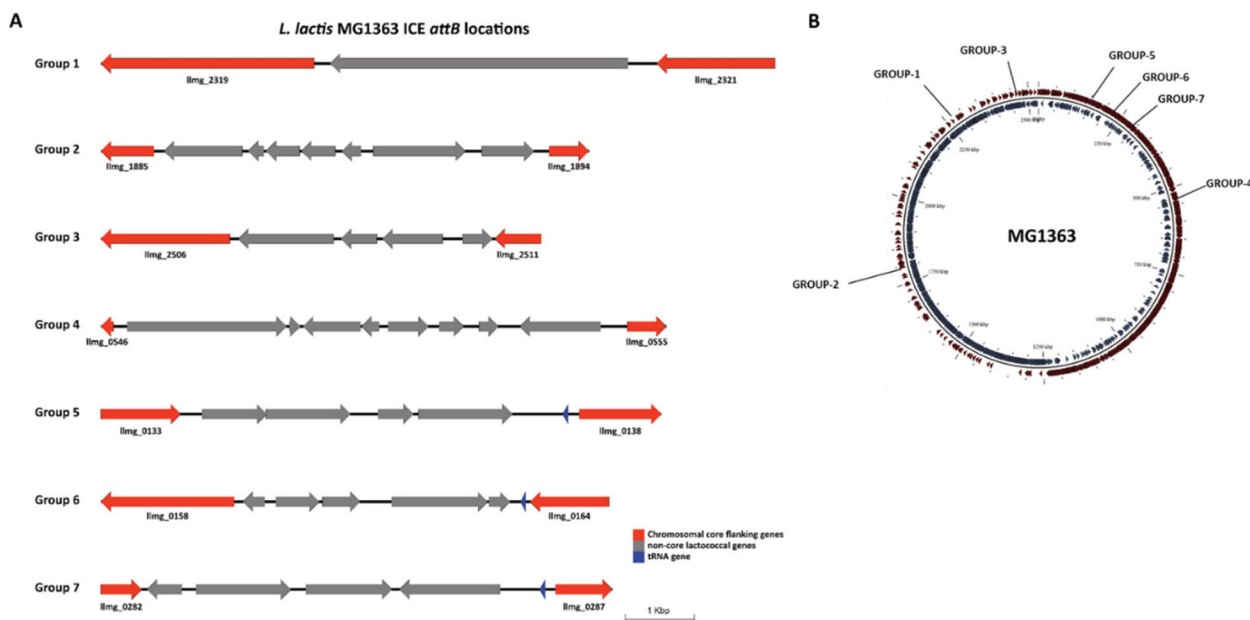


Fig. 1 Conserved ICE integration locations in MG1363. **A** Single-copy core-genome genes flanking the discriminated integrase (*int*) groups ICE insertion sites. **B** Distribution of the *int*-determined insertion locations across the ICE-deficient MG1363 chromosome

behind in the ICE_{KF67}-deficient derivatives following the excision of ICE_{KF67}.

Gene map image generation

All gene map images were generated using EasyFig [31], and whole genbank comparison pBLAST were performed using the automated feature from Easyfig.

Results

Lactococcal ICEs can be delimited in silico by identifying and using conserved integration sites

Previously we identified and classified a set of conserved ICE genes in the lactococcal genomes, identifying 7 distinct integrases in candidate lactococcal ICEs that were detected in the genome of more than one strain. Moreover, the same study confirmed that the integrase encoding *int* genes were consistently localized at the end of the ICE region [24]. Since integrases play a key-role in ICE excision and integration, these distinct integrases are predicted to correspond to 7 distinct chromosomal integration sites. However, complete ICE delimitation necessitates the localization of the 5'- and 3'-boundaries represented by *att* sites. For the delimitation of the 7 ICE groups, we selected 20 *L. lactis* strains in which 27 candidate ICEs were previously identified [24] and for which genome data were available that assembled into low contig numbers (Table 1). This set of strains encompasses several representatives of each of the integrase-classified ICEs [24]. Since ICEs are commonly only present in a

subset of strains, the ICE-encoded genes do not belong to the core-genome of the species. Therefore, exploiting the notion that the integrase encoding gene (*int*) is located close to, or directly flanking the end of the ICE-region [32] was used to detect the first *int*-neighboring lactococcal gene belonging to the species' single-copy core-genome (*int* flanking core-gene) [26]. This step identified distinct single-copy core-genes that consistently flanked each of the 7 *int* genes that were previously recognized in the ICE-containing strains [24]. This finding supports the role of the integrase function in site specific integration and corroborates the accuracy of the previously reported classification of the lactococcal integrases into 7 groups. One exception was observed for strain *L. lactis* KF146 that contains two candidate ICEs, of which ICE_{KF146_2} encodes a group 3 integrase (Supplemental table ST1) that is flanked by a different single-copy core-gene (limg_0350, KF146_01199) as compared to the other members of this group, indicating that this group 3 lactococcal ICE inserted at a different location in the *L. lactis* KF146 chromosome. Interestingly, 3 of the 7 identified conserved integration sites are adjacent to a tRNA gene (Fig. 1A). The *int*-flanking core gene was used to identify the first neighboring core-gene in the genome of ICE-deficient strain *L. lactis* MG1363 (in the direction across the *int*-gene of the ICE), to obtain an approximate delimitation of the chromosomal position of the integrase-specific ICE groups on basis of the flanking core-genes. This analysis yielded 7 integrase-specific ICE *attB*

locations (Fig. 1A) which we mapped on the MG1363 genome (Fig. 1B).

The genomic distance between the 7-pairs of ICE-flanking core-genome genes was subsequently determined in each of the ICE-containing strains and compared to their distance in the ICE-deficient MG1363. This analysis revealed that the difference in genomic distance between the 7-pairs of ICE-flanking core genes in 25 (out of 27) ICE-containing genomic loci was between 27 and 90 kb enlarged relative to the distance observed in strain MG1363, which is in agreement with the commonly observed length of ICEs in various bacteria [20], suggesting that this approach is able to accurately delimit the large majority of the lactococcal ICEs in genome assemblies that have low numbers of contigs. Following this delimitation analysis, ICEs can readily be visualized by genome comparison between close relative strains, as is exemplified by the alignment of the ICE insertion locus in ICE-containing *L. lactis* KLDS 4.0325 (ICE_{KLDS}) with its close relative *L. lactis* M1734.1 that has no ICE inserted in this location (Fig. 2).

The two ICEs where strongly deviating estimated ICE lengths were predicted using the approach above were ICE_{KF146_2} and ICE_{V4} (Supplemental Table ST1). For ICE_{KF146_2} this was anticipated based on the observation that this ICE integrated at another position in the chromosome compared to the other members of the lactococcal group 3 ICEs (see above). The predicted distance of the flanking core-genes of group 2 lactococcal ICE in strain *L. lactis* V4 was larger than 1.3 Mb, which is hugely exceeding the expected ICE-size. This observation could be explained by comparative analysis of the V4 genome with that of *L. lactis* IL1403, revealing a large genomic inversion and loss of synteny precisely at the location where the ICEs of this family were predicted to be located (Supplemental figure SF1). This finding indicates that the assumed local genomic synteny that underlies this ICE delimitation strategy is confirmed for the vast majority of the *L. lactis* strains, but at the same time highlights that this local genomic synteny is a prerequisite for the successful application of the delimitation-analysis.

The ICE-flanking single-copy core-gene approach only succeeds to delimit ICEs in strains for which a genome assembly with relatively low contig numbers is available.

Previously we also identified 36 candidate ICEs in 29 *L. lactis* strains for which only draft genome sequences (i.e., characterized by high numbers of contigs; >100 contigs) were available [24]. Analysis of the contigs that encode the ICE-associated *int* confirmed that the chromosomal genes flanking these *int* genes consistently were the identified single-copy core-genome genes associated with the *int*-grouping, corroborating the *int*-determined insertion site-preference described above. However, delimiting the ICEs in these draft genomes failed since the ICE associated genes consistently were spread over more than one contig, which is most likely due to the prevalence of repetitive sequence elements in ICEs (e.g., transposase encoding genes; see below) that are known to lead to contig-breaks in genome sequence assemblies. To illustrate this, we employed the ICE-flanking core gene approach to identify the two contigs that encode these genes in the draft genome of *L. lactis* KF67 that contains the group-1 ICE_{KF67} (Contig_8 and Contig_14). This indicates that draft genome sequence information can still be used to confirm the insertion site of an ICE, based on the group-classification of the *int* gene it encompasses, but does not allow ICE delimitation, especially in low-coverage genomes encompassing many contigs.

CRISPR-Cas9 ICE-curing and re-sequencing enables ICE delimitation

The analysis above showed that ascertaining the genetic content of ICEs in draft genomes is not successful. Therefore, we proceeded to delimit the group-1 ICE in *L. lactis* KF67 (i.e., ICE_{KF67}) using an experimental curing approach. Based on the confirmation that ICE_{KF67} is inserted at the typical group-1 ICE location in the chromosome, we defined primers for the PCR-based detection of its inserted, chromosomal localization (establishing the chromosome-*int* genetic linkage) as well as its excised, circularized state (emergence of the chromosomal ICE- 'scar') (Fig. 3A). These two stages of the ICE lifecycle are anticipated to co-exist in a culture, although the relative abundance of the subpopulation in which the ICE in its excised state is commonly present in low abundance (data not shown). Chromosomal DNA obtained from an overnight culture of *L. lactis* KF67 allowed the PCR detection of both life-cycle stages of the

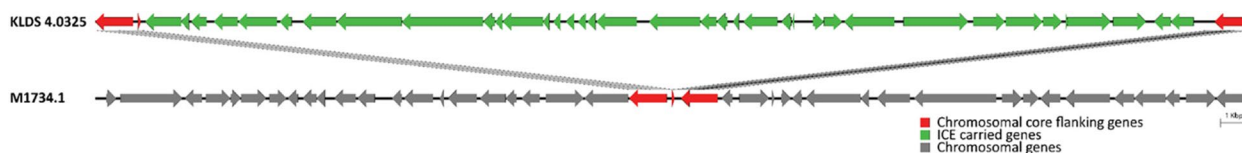


Fig. 2 Example alignment of the group 5 insertion region. Comparing an ICE carrying *L. lactis* strain (KLDS 4.0325) and a non-carrying close relative *L. lactis* strain (M1734.1), illustrating the precise detection of the ICE insertion sequence (*att*)

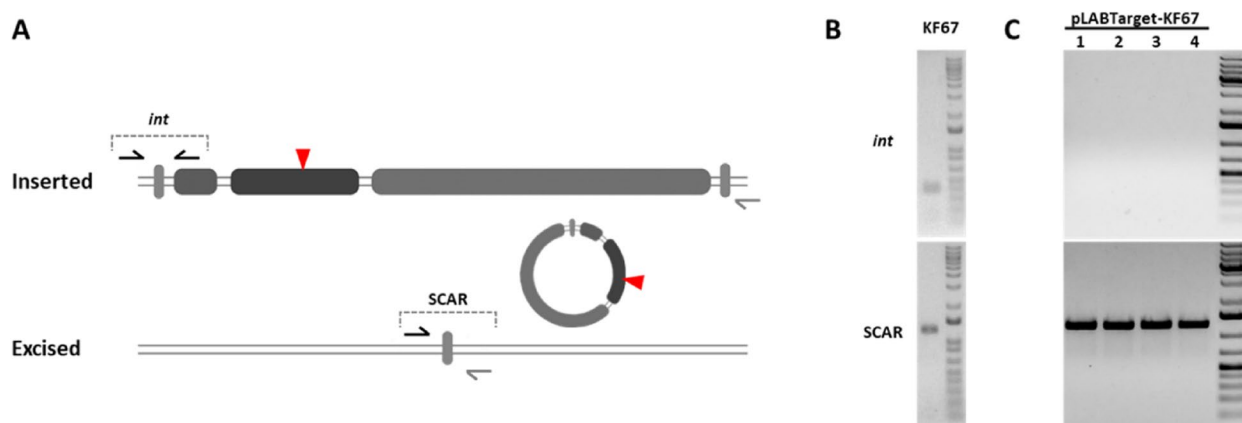


Fig. 3 Visual representation of experimental design to verify ICE activity. **A** The ICE can exist in two configurations in a culture: inserted and excised. The arrows represent oligonucleotides used in PCR reactions to verify the presence of both states. The red triangle indicates the location of Cas9 targeted curing. **B** Gel electrophoresis image of the PCR-products used for the detection of the inserted and excised coexisting forms of the ICE_{KF67} in an exponentially growing KF67 culture. **C** Gel electrophoresis image of the PCR-products used for the detection of the inserted (not detected, upper panel) and excised (consistently detected, lower panel) in colonies obtained after Cas9 targeting of the KF67 ICE. In all four colonies the loss of the ICE_{KF67} in these colonies is confirmed

ICE_{KF67} (Fig. 3B), confirming that it can be mobilized in this strain at a detectable level. Subsequently, we transformed *L. lactis* KF67 with a plasmid that encodes the Cas9 endonuclease and a sgRNA that targets the ICE_{KF67} sequence (LKF67_0087) (pLABTarget [25] derivative). Previous work established that sgRNA guided Cas9 targeting of a chromosomally located sequence induces a dsDNA break, which is lethal for most bacteria [33, 34]. Targeting of an extrachromosomal sequence leads to loss of the element containing this target while leaving the chromosome unaffected [25]. This principle was used to select for a subpopulation in which the ICE_{KF67} was in its excised extrachromosomal state (Fig. 3A). Transformation of *L. lactis* KF67 with the ICE_{KF67} targeting pLABTarget-KF67 vector yielded a strongly reduced (>100-fold) number of transformants as compared to the transformation with the corresponding pLABTarget empty vector, which is in agreement with the anticipated elimination of the cells in which ICE_{KF67} was chromosomally integrated [35, 36]. The selective survival of *L. lactis* KF67 derivatives from which the ICE_{KF67} had been cured was confirmed by PCR-detection of the ICE-‘scar’ in combination with the failure to detect the ICE encoded *int* gene. Two colonies displaying the ICE_{KF67} cured amplicon profile were subjected to whole genome sequencing (Fig. 3C).

The sequence information obtained was mapped to the *L. lactis* KF67 (containing ICE_{KF67}) genome, establishing that part of the previously assembled contigs 14, and 8 were missing, and indicating that these contigs contain parts of the ICE_{KF67}. Importantly, no further contigs from the *L. lactis* KF67 draft genome assembly appeared

to be absent in the newly obtained sequence information, nor were genetic regions detected in the cured variant that were not represented in the contigs of the KF67 draft genome. Moreover, the sequence data obtained for the cured derivatives of *L. lactis* KF67 contained the predicted scar sequence region that would result from ICE_{KF67} excision, confirming the accuracy of the excision prediction based on the ICE-flanking single-copy core-gene approach. These findings indicate that the two identified contigs (8 and 14) contain the complete ICE_{KF67}, demonstrating that the in silico detection of ICE-associated contigs using the flanking single-copy core-genes combined with resequencing of an experimentally cured derivative enables the delimitation and subsequent cargo definition of the ICE.

Comparative analysis of lactococcal ICEs reveal high degree of cargo region plasticity

Comparative analysis of the 27 delimited ICE sequences confirmed the substantial variation in ICE-size (Table 3) that we predicted on basis of the flanking core-genome marker genes. Moreover, the analysis also revealed that several ICEs are very closely related or even identical. For instance, *L. lactis* ML8 and UL8 harbor identical group 7 ICEs, possibly reflecting a recent conjugal transfer event between these strains or these strains have recently derived from a common ancestor harboring this ICE. The latter scenario appears to be supported by genome comparison that confirms the close relatedness of the ML8 and UL8 strains (NCBI genome phylogeny, data not shown).

Table 3 Delimited ICEs of each ‘integrase’-group and their predicted size in base pairs

| Group | ICE | predicted ICE size |
|-------|----------------------|--------------------|
| 1 | KF147 (Tn6098) | 58,117 |
| | 184_2 | 50,528 |
| | G423 | 44,398 |
| | KF67 ^a | 39,411 |
| 2 | CV56_1 | 65,722 |
| | IO-1 | 64,356 |
| 3 | CV56_2 | 40,968 |
| | A12 | 42,225 |
| | N42_1 | 42,214 |
| | KF146_2 ^b | 31,067 |
| 4 | AI06 | 57,146 |
| | JM1 | 36,633 |
| 5 | KLDS 4.0325 | 36,472 |
| | KF146_1 | 42,680 |
| | N42_2 | 89,199 |
| 6 | 184_1 | 79,683 |
| | 229 | 88,817 |
| | 275 | 53,323 |
| | FM03 | 55,661 |
| | UC06_1 | 97,385 |
| | UC77 | 88,818 |
| 7 | NCDO2118 | 47,923 |
| | 14B4 | 36,822 |
| | KF147_1 | 47,925 |
| | ML8 | 36,440 |
| | UC06_2 | 34,461 |
| | UL8 | 36,495 |

^a delimited via Cas9 curation

^b inserted in an alternative insertion site compared to other group 3 members

The conserved modular genetic make-up of ICEs allowed us to differentiate the canonical ICE functions that are involved in excision, transfer and integration [24] from those that are considered as ICE- ‘cargo’. Cargo analysis revealed that the ICEs belonging to group 7 consistently encoded the smallest cargo region (4–16 kb), whereas group 6 ICEs consistently encoded the largest cargo (19–62 kb). The genes detected in the cargo regions of all ICEs were annotated and subjected to comparative analysis, which initially focused on the ICEs of group 6 because their cargo regions were the largest and may display the largest degree of variation. Full-length alignment of the group 6 ICEs immediately revealed the high-level conservation and terminal-localization of the conserved ICE-lifecycle machinery functions (assigned as ICE-core functions in Fig. 4), with a deletion of 4 genes in ICE_{184_1}. It remains unclear whether this deviation in ICE_{184_1} relative to its group members has any functional consequences and may render this ICE non-mobile or conjugation incompetent. Moreover, this analysis also revealed that ICE₂₂₉ and ICE_{UC77} in strains *L. lactis* 229 and UC77 are identical, reiterating either a recent transfer event or a recent shared ancestor. Finally, the cargo regions of the ICEs belonging to group 6 are very variable, both in size and gene content, although regional similarities are frequently observed as is exemplified by several genes and functions that appear to be encoded by each of the members of this ICE-group (Fig. 4). This genetic plasticity within the ICE cargo region may be facilitated by the scattered presence in these regions of (truncated-) transposase and resolvase encoding genes that play important roles in gene exchange reactions and genetic recombination [37]. Notably, the length of the cargo region appears to correlate with the number of transposases encountered

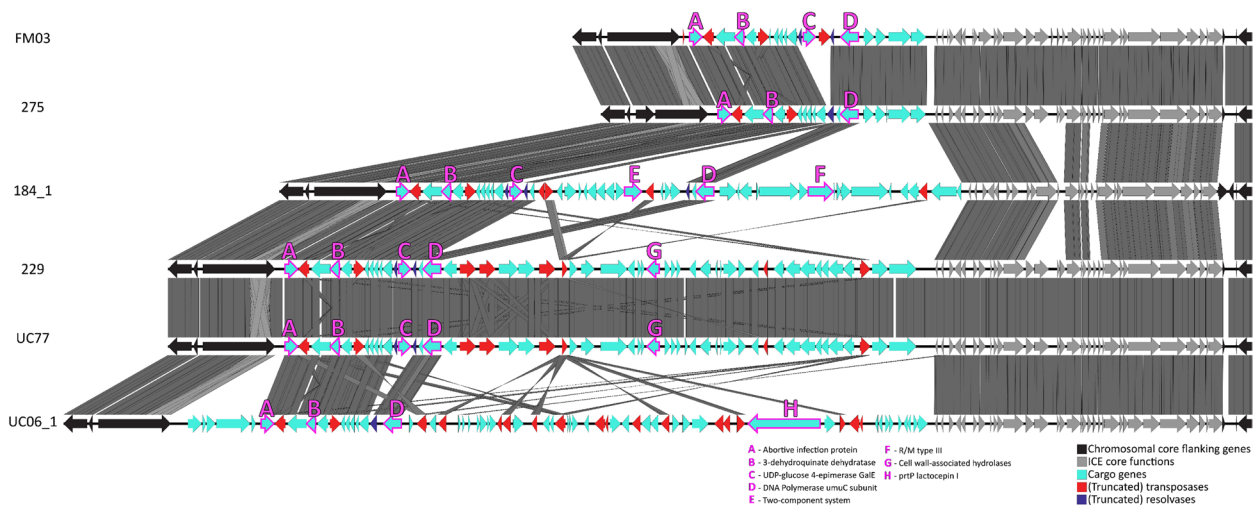


Fig. 4 Alignment of the complete sequences of the delimited group 6 ICEs. Gray lines indicate BLASTp comparison and conservation between entries. The explanatory legend is included in the image

in this region, which accumulated to no less than 6 intact, and 12 truncated transposase genes present in the longest representative of this ICE-group (ICE_{UC06_1}). Importantly, the transposase (and resolvase) encoding genes appear to be localized at the junctions between functionally related gene clusters within the ICE cargo region, suggesting that transposase- (or resolvase-) facilitated cargo recombination events can readily drive functional diversification of the ICEs, thereby dynamically contributing relevant and fitness improving genetic traits to the host organism. For example, the UDP-galactose epimerase encoding *galE* gene, which is relevant for sugar nucleotide biosynthesis and galactose metabolism in *L. lactis* [38], is present in 4 of the 6 group 6 ICEs (Fig. 4) where it is consistently flanked by two non-identical resolvase genes, that is replaced by again another resolvase in the group 6 ICEs that lack the *galE* gene.

Although, we exemplified these findings only focusing on group 6 lactococcal ICEs, similar observations were also made when analyzing the other ICE groups (data not shown). Taken together, this clearly demonstrates the highly dynamic nature of ICE-cargo regions and pinpoints the possible role of the canonical transposases and resolvases in functional diversification of these MGEs.

Lactococcal ICE cargos encode a wide range of adaptive gene functions

The intrinsic mobility of ICEs enables their transfer among strains using natural-mating procedures, allowing the expansion of specific ICE-encoded traits in a host strain of interest. The cargo annotations for the 26 ICEs revealed a wide range of encoded functions, including those with potential industrial relevance such as stress resistance, phage abortive infection, the utilization of different carbon sources, antimicrobial production and restriction modification. We focused on the functional traits encoded by the ICEs of group 6 and their variability among the members of these ICEs. In addition, we analyzed whether the functional modules recognized in the group 6 ICEs are possibly exchanged and/or shared with ICEs of the other groups. This analysis was performed at high stringency (at least 90% amino acid sequence identity and at least 95% sequence overlap) to allow the detection of exchanges of genetic modules among the ICE groups. Each ICE in group 6 carries a tandem of two cold shock protein (CSP) encoding genes, a feature that appears to be shared with the majority of the ICE-cargo regions of the other groups (19/26), where in some instances 3 CSP encoding genes are present. Although CSPs are predicted to contribute to (cold) stress tolerance and may thereby contribute to the environmental fitness of the ICE containing host strain, their mechanism of action is highly diverse [39, 40]. Intriguingly, the

CSPs contain a highly conserved nucleic acid binding domain [41], which may suggest a role of these proteins in a particular stage of the ICE lifecycle. Another function that is consistently encoded by the group 6 ICEs is a protein containing an bacteriophage abortive infection (AbiH) domain that has been described to be involved in blocking phage multiplication and inducing premature bacterial cell death upon phage infection, thereby reducing phage progeny size and phage spread in the population [42]. The *abi* genes found in group 6 ICEs were not found in the other ICE groups (Fig. 4). Another genetic module that is found uniquely in all representatives of the group 6 ICEs is a cluster of genes that encompasses a 3-dehydroquinate dehydratase encoding gene, which is involved in the shikimate pathway [43] that is part of the aromatic amino acid biosynthesis pathway (labelled “B” in Fig. 4) and a protein containing a 2A78 domain, which is part of a family of transporters involved in metabolite (amino acid) or drug efflux [44]. The combination of an aromatic amino acid metabolism function combined with an efflux function could play a role in the flavor volatile formation and its extrusion [45]. Similarly, a protein containing an UmuC subunit of DNA Polymerase V, involved in error prone replication belonging to the bacterial SOS response, is present in all group 6 ICEs (Fig. 4). Under severe stress conditions, cells can activate Pol V dependent error-prone replication to drive environmental adaptation of the host cell [46]. Notably, the group 6 ICE encoded Pol V is shared with the group 4 ICE_{AI06}. As mentioned above, 4 of the group-6 ICEs encode a UDP-Glucose 4-epimerase (*galE*) (labelled “C” in Fig. 4), that is important in galactose metabolism [38] and is involved in the synthesis of sugar nucleotides for cell-wall and exopolysaccharides (EPS) production, where the latter molecules contribute to texture properties of fermented dairy products [47]. ICE_{184_1} encodes a unique cluster of genes flanked by an ABC-type multidrug transport ATPase component and a two-component regulator system (labelled “E” in Fig. 4), which are known to regulate cellular responses to environmental stimuli, which could be of interest in the dairy environment. Also only found on ICE_{184_1} we identify a type III restriction-modification system as well as a type I hsdR type restriction-modification R protein (labelled “F” in Fig. 4). Restriction-modification systems are important in defense against foreign DNA such as bacteriophages and other MGEs and could thereby be industrially relevant. Unique for the ICE shared by strains 229 and UC77 is a cell-wall associated hydrolase that contains a SPP1 phage holin motif (labelled “G” in Fig. 4) that are known to induce membrane permeability during late stage phage infection [48], but could also affect autolysis of strains harboring this functionality, which has been proposed to contribute

to cheese ripening [49]. Notably, this holin function is genetically linked to a function that contains a BacA like motif, suggesting a bacteriocin-like function [50] and a cell-wall associated hydrolase and N-acetylmuramoyl-L-alanine amidase. The clustering of these functions implies a possible role in autolysis or the lysis of neighboring cells, which could be relevant for flavor formation. ICE_{UC06_1} encodes a *prtP* lactocepain I, which is a cell wall associated proteinase that degrades casein into shorter peptides to support growth on dairy substrates (labelled “H” in Fig. 4). Specificity of lactocepains can differ wildly, and can contribute to different flavor profiles [51].

Taken together, ICE cargo encompasses a large variety of functionalities that are potentially involved in industrially relevant traits, whereby the ICEs cargo investigation presented here specifies a reservoir of potentially mobilizable and transferable functions for industrial optimization of *L. lactis* strains.

Discussion

ICEs contribute a wide range of different functions to the *L. lactis* variome. In this study we show that there are seven preferred insertion locations based on seven conserved integrases. These integration sites can be used for future determination of ICE presence in lactococcal genomes that were not included in this study. The comparative genomics approach illustrates the ability to delimit ICEs in the case of genome sequences with low numbers of contigs. With the increasing availability and reduction in costs of long read sequencing methods such as MinION, deciphering the genetic context needed for ICE delimiting in-silico becomes readily available. Especially when combined in hybrid assemblies with high sequence-fidelity short reads (e.g., Illumina based sequencing) ICE delimitation by genomic context is quite straightforward to achieve. In contrast, genomes assemblies with a large number of contigs are likely to disallow the definition of the ICE cargo region, especially since high abundance of (remnants of) transposons in ICE-cargo regions is bound to result in contig breaks. Nevertheless, the recognition of preferred insertion sites based on the MG1363 genome can support the recognition of ICE associated contigs, provided that the strain's genome shares synteny with MG1363 or another reference strain used. Notably, delimiting the ICE in the V4 strain did fail due to a lack of synteny between this strain and MG1363 or IL1403. Therefore, delimitation of ICEs in genomes assembled into many contigs or for strains that lack synteny can be resolved using the Cas9 curing strategy employed in this study, where delimitation can be achieved by comparative genome sequencing of the original and ICE-cured variant of the strain. This approach depends on the genetic accessibility of the

strain and the excision activity of the ICE itself. In particular when the frequency of excision events is low, in combination with a low transformation efficiency of the targeted strain, it can be challenging to retrieve colonies of the ICE-cured (transformation-surviving) variants. Nevertheless, the curing method also verifies the biological activity of the ICE, providing additional functional information of the ICE that is predicted in silico. Many ICEs are shown to integrate close to the *oriC* of their host and it has been proposed that this bias might be due to the conservation of highly expressed housekeeping genes in those regions [52]. The lactococcal ICEs are no exception to this bias; most identified lactococcal ICE insertion sites can be found relatively close to the chromosomal origin of replication (*oriC*). One outlier is the insertion location for group 2 in MG1363. A major genomic inversion occurred in this strain [9], with the group 2 insertion location on the axis of inversion (Supplemental figure SF2). The conserved nature of the housekeeping, highly expressed genes could facilitate the ICE's host-range expansion options by increased chance of conservation of ICE-insertion location in a new host. The same rationale is likely applicable for tRNA encoding regions, which is frequently found to be the chromosomal insertion region for lactococcal ICEs [15].

The amount of modularity and variation in the cargo regions we found in tandem with the high presence of (disrupted) transposons hints towards a highly adaptive nature of ICEs. This finding underpins the importance of investigation of novel ICEs in individual strains, as their cargo content can differ between isolates. *L. lactis* strains are commonly isolated from dairy environments and plant environments. The dairy isolates are proposed to have been evolved from plant isolates and harbor a wide range of adaptations specific for the dairy niche [53, 54]. Furthermore, ICE investigation may be more fruitful in plant-derived lactococcal isolates where the ICE-cargo gene repertoire might be more variable as a reflection of the more variable and dynamic plant associated niches. The cargo gene functions could provide a selective advantage to the lactococcal host strain in these plant-associated environments. Intriguingly, typical plant isolate associated ICEs (exemplified for the *L. lactis* KF147 Tn6098) might prove to be a burden to host fitness when they reside in the dairy environment, as was exemplified by the loss of Tn6098 in *L. lactis* KF147 when grown for approximately 1000 generations in milk [53]. As shown in this study, cargo analysis of ICEs in *L. lactis* is worthwhile, since the investigation of the potential functions based on automated in silico annotations already provides ample leads for further investigation. For example, some of the lactococcal ICEs listed in this study harbor partial or potentially complete restriction

modification systems which have established roles in the protection against foreign DNA agents such as infective bacteriophages, plasmids or other ICEs. It is also possible that the presence of such a restriction-modification system further modifies the host's genome methylation state, leading to downstream changes in gene transcription potentially changing its performance under different growth conditions. Abortive infection proteins might offer more phage resistance for industrial strains and genes potentially involved in stress resistance could contribute to robustness of the strains when exposed to industrial processing conditions. For the further investigation or elucidation of the ICE functions, the generation of isogenic 'cured' variants of the strains and performing comparative performance evaluations in a wide range of phenotype tests could support deciphering of the ICE-derived functional potential. Such a knowledge base could be capitalized by rationally designed mating (conjugation) experiments to obtain strains with an optimized ICE-derived gene repertoire (cargo-based selection of the best possible ICE-donor strain) for specific industrial applications.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10646-y>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.

Acknowledgements

Not applicable

Authors' contributions

The study design was conceived by my MK, PAB, and SvdE. The Lactococcal genomic data was collected and compiled by SvdE under supervision of JB, PAB and MK. Data analyses were performed by SvdE and JB. SvdE wrote the first draft of the manuscript to which JB, PAB and MK provided their feedback. All authors agree to the submission and consent to the publication of the final manuscript.

Funding

This work was carried out within the BE-Basic R&D program (grant F10.002.01), which was granted an FES subsidy from the Dutch Ministry of Economic Affairs.

Availability of data and materials

All data used in this study is available through public repositories (NCBI genomes), and data-identifiers are specified in Table 1 of the manuscript.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 29 May 2024 Accepted: 22 July 2024

Published online: 30 July 2024

References

1. Leroy F, De Vuyst L. Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends Food Sci Technol.* 2004;15:67–78.
2. Ghandi A, Powell IB, Howes T, Chen XD, Adhikari B. Effect of shear rate and oxygen stresses on the survival of *Lactococcus lactis* during the atomization and drying stages of spray drying: A laboratory and pilot scale study. *J Food Eng.* 2012;13:194–200.
3. Johnson ME. A 100-Year Review: Cheese production and quality. *J Dairy Sci.* 2017;100:9952–65.
4. Coffey A, Ross RP. Bacteriophage-resistance systems in dairy starter strains: Molecular analysis to application. *Antonie van Leeuwenhoek.* 2002;82:303–21.
5. Kelleher P, Mahony J, Bottacini F, Lugli GA, Ventura M, Van Sinderen D. The *Lactococcus lactis* pan-Plasmidome. *Front Microbiol.* 2019;10:707.
6. Siezen RJ, Renckens B, Van Swam I, Peters S, Van Kranenburg R, Kleerebezem M, De Vos WM. Complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment. *Appl Environ Microbiol.* 2005;71:8371–82.
7. Kelleher P, Bottacini F, Mahony J, Kilcawley KN, van Sinderen D, Sinderen DV. Comparative and functional genomics of the *Lactococcus lactis* taxon; insights into evolution and niche adaptation. *BMC Genomics.* 2017;18:267.
8. Christensson C, Pillidge CJ, Ward LJH, O'Toole PW. Nucleotide sequence and characterization of the cell envelope proteinase plasmid in *Lactococcus lactis* subsp. *cremoris* HP. *J Appl Microbiol.* 2001;91:334–43.
9. Wegmann U, Overweg K, Jeanson S, Gasson M, Shearman C. Molecular characterization and structural instability of the industrially important composite metabolic plasmid pLP71. *Microbiology (United Kingdom).* 2012;158:2936–45.
10. Yu W, Gillies K, Kondo JK, Broadbent JR, McKay LL. Plasmid-mediated oligopeptide transport system in lactococci. *Dev Biol Stand.* 1995;85:509–21.
11. Mills S, Griffin C, O'Connor PM, Serrano LM, Meijer WC, Hill C, Ross RP. A multibacteriocin cheese starter system, comprising nisin and lactacin 3147 in *Lactococcus lactis*, in combination with plantaricin from *Lactobacillus plantarum*. *Appl Environ Microbiol.* 2017;83:e00799–00717.
12. Rauch PJG, De Vos WM. Characterization of the novel nisin-sucrose conjugative transposon Tn5276 and its insertion in *Lactococcus lactis*. *J Bacteriol.* 1992;174:1280–7.
13. Machiels R, Siezen RJ, Van Hijum SAFT, Van Hylckama Vlieg JET. Molecular description and industrial potential of Tn6098 conjugative transfer conferring alpha-galactoside metabolism in *Lactococcus lactis*. *Appl Environ Microbiol.* 2011;77:555–63.
14. Burrus V, Pavlovic G, Decaris B, Guédon G. Conjugative transposons: The tip of the iceberg. *Mol Microbiol.* 2002;46:601–10.
15. Johnson CM, Grossman AD. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Ann Rev Genet.* 2015;49:577–601.
16. Beaber JW, Hochhut B, Waldor MK. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature.* 2004;427:72–4.
17. Auchtung JM, Lee CA, Garrison KL, Grossman AD. Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICEBs1 of *Bacillus subtilis*. *Mol Microbiol.* 2007;64:1515–28.
18. Miyazaki R, Minoia M, Pradervand N, Sulser S, Reinhard F, van der Meer JR. Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. *PLoS Genet.* 2012;8:e1002818.
19. Veening J-W, Smits WK, Kuipers OP. Bistability, Epigenetics, and Bet-Hedging in Bacteria. *Ann Revi Microbiol.* 2008;62:193–210.
20. Cury J, Touchon M, Rocha EPC. Integrative and conjugative elements and their hosts: Composition, distribution and organization. *Nucleic Acids Res.* 2017;45:8943–56.

21. Van Houdt R, Leplae R, Lima-Mendez G, Mergeay M, Toussaint A. Towards a more accurate annotation of tyrosine-based site-specific recombinases in bacterial genomes. *Mob DNA*. 2012;3:6.
22. Grainge I, Jayaram M. The integrase family of recombinases: Organization and function of the active site. *Mol Microbiol*. 1999;33:449–56.
23. Williams KP. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: Sublocation preference of integrase subfamilies. *Nucleic Acids Res*. 2002;30:866–75.
24. van der Els S, Sheombarsing R, van Kempen T, Wels M, Boekhorst J, Bron PA, Kleerebezem M. Detection and classification of the integrative conjugative elements of *Lactococcus lactis* BMC Genomics. 2024;25(1):324.
25. van der Els S, James JK, Kleerebezem M, Bron PA. Versatile Cas9-driven subpopulation selection toolbox for *Lactococcus lactis*. *Appl Environ Microbiol*. 2018;84:e02752–e02717.
26. Wels M, Siezen R, van Hijum S, Kelly WJ, Bachmann H, Hijum Sv, Kelly WJ, Bachmann H. Comparative genome analysis of *Lactococcus lactis* indicates niche adaptation and resolves genotype/phenotype disparity. *Front Microbiol*. 2019;10:4.
27. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
28. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020;48:D265–8.
29. Krumsiek J, Arnold R, Rattei T. Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*. 2007;23:1026–8.
30. Wells JM, Wilson PW, Le Page RWF. Improved cloning vectors and transformation procedure for *Lactococcus lactis*. *J Appl Bacteriol*. 1993;74:629–36.
31. Sullivan MJ, Petty NK, Beatson SA. Easyfig: A genome comparison visualizer. *Bioinformatics*. 2011;27(7):1009–10.
32. Ambroset C, Coluzzi C, Guédon G, Devignes MD, Loux V, Lacroix T, Payot S, Leblond-Bourget N. New insights into the classification and integration specificity of streptococcus integrative conjugative elements through extensive genome exploration. *Front Microbiol*. 2016;6:1483.
33. Cui L, Bikard D. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Res*. 2016;44(9):4243–51.
34. Gomaa AA, Klumpe HE, Luo ML, Selle K, Barrangou R, Beisel CL. Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*. 2014;5(1):e00928–00913.
35. Selle K, Klaenhammer TR, Barrangou R. CRISPR-based screening of genomic island excision events in bacteria. *Proc Natl Acad Sci USA*. 2015;112:8076–81.
36. Vercoe RB, Chang JT, Dy RL, Taylor C, Gristwood T, Clulow JS, Richter C, Przybilski R, Pitman AR, Fineran PC. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet*. 2013;9(4):e1003454.
37. Hallet B, Sherratt DJ. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev*. 2006;21:157–78.
38. Boels IC, Ramos A, Kleerebezem M, De Vos WM. Functional Analysis of the *Lactococcus lactis* galU and galE Genes and Their Impact on Sugar Nucleotide and Exopolysaccharide Biosynthesis. *Appl Environ Microbiol*. 2001;67:3033–40.
39. Graumann PL, Marahiel MA. A superfamily of proteins that contain the cold-shock domain. *Trends Biochem Sci*. 1998;23:286–90.
40. Yamanaka K, Fang L, Inouye M. The CspA family in *Escherichia coli*: Multiple gene duplication for stress adaptation. *Mol Microbiol*. 1998;27:247–55.
41. Graumann P, Marahiel MA. Some like it cold: Response of microorganisms to cold shock. *Arch Microbiol*. 1996;166:293–300.
42. Chopin MC, Chopin A, Bidnenko E. Phage abortive infection in lactococci: Variations on a theme. *Curr Opin Microbiol*. 2005;8:473–9.
43. Herrmann KM, Weaver LM. The shikimate pathway. *Ann Rev Plant Biol*. 1999;50:473–503.
44. Västermark Å, Almén MS, Simmen MW, Fredriksson R, Schiöth HB. Functional specialization in nucleotide sugar transporters occurred through differentiation of the gene cluster *EamA* (DUF6) before the radiation of Viridiplantae. *BMC Evol Biol*. 2011;11:123.
45. Ardö Y. Flavour formation by amino acid catabolism. *Biotechnol Adv*. 2006;24:238–42.
46. Robinson A, McDonald JP, Caldas VEA, Patel M, Wood EA, Punter CM, Ghodke H, Cox MM, Woodgate R, Goodman MF, et al. Regulation of Mutagenic DNA Polymerase V Activation in Space and Time. *PLoS Genet*. 2015;11:e1005482.
47. Looijesteijn PJ, Boels IC, Kleerebezem M, Hugenholtz J. Regulation of Exopolysaccharide Production by *Lactococcus lactis* subsp. *cremoris* by the Sugar Source. *Appl Environ Microbiol*. 2001;65:5003–8.
48. Wang IN, Smith DL, Young R. Holins: The protein clocks of bacteriophage infections. *Ann Rev Microbiol*. 2000;54:799–825.
49. Smid EJ, Kleerebezem M. Production of aroma compounds in lactic fermentations. *Ann Rev Food Sci Technol*. 2014;5:313–26.
50. Kurushima J, Hayashi I, Sugai M, Tomita H. Bacteriocin Protein BacL1 of *Enterococcus faecalis* Is a Peptidoglycan d-Isoglutamyl-l-lysine Endopeptidase. *J Biol Chem*. 2013;288:36915–25.
51. Broadbent JR, Barnes M, Brennand C, Strickland M, Houck K, Johnson ME, Steele JL. Contribution of *Lactococcus lactis* cell envelope proteinase specificity to peptide accumulation and bitterness in reduced-fat cheddar cheese. *Appl Environ Microbiol*. 2002;68:1778–85.
52. Carraro N, Burrus V. The dualistic nature of integrative and conjugative elements. *Mob Genet Elements*. 2015;5:98–102.
53. Bachmann H, Starrenburg MJC, Molenaar D, Kleerebezem M, van Hylckama Vlieg JET. Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. *Genome Res*. 2012;22:115–24.
54. Kelly WJ, Ward LJM, Leahy SC. Chromosomal Diversity in *Lactococcus lactis* and the Origin of Dairy Starter Cultures. *Genome Biol Evol*. 2010;2:729.
55. Tian K, Li Y, Wang B, Wu H, Caiyin Q, Zhang Z, Qiao JJ. The genome and transcriptome of *Lactococcus lactis* ssp. *lactis* F44 and G423: Insights into adaptation to the acidic environment. *Dairy Sci*. 2019;102:1044–58.
56. Backus L, Wels M, Boekhorst J, Dijkstra AR, Beerthuyzen M, Kelly WJ, Siezen RJ, van Hijum SAFT, Bachmann SAFT. Draft Genome Sequences of 24 *Lactococcus lactis* Strains. *Genome Announc*. 2017;5(13):e01737–16.
57. Gao Y, Lu, Y, Teng KL, Chen ML, Zheng ML, Zhu YQ, Zhong J. Complete genome sequence of *Lactococcus lactis* subsp. *lactis* CV56, a probiotic strain isolated from the vaginas of healthy women. *J Bacteriol*. 2011;193:2886–7.
58. Kato H, Shiwa Y, Oshima K, Machii M, Araya-Kojima T, Zendo T, Shimizu-Kadota M, Hattori M, Sonomoto K, Yoshikawa H. Complete genome sequence of *Lactococcus lactis* IO-1, a lactic acid bacterium that utilizes xylose and produces high levels of l-lactic acid. *J Bacteriol*. 2012;194:2102–3.
59. Guellerin M, Passerini D, Fontagné-Faucher C, Robert H, Gabriel V, Loux V, Klopp C, Le Loir Y, Coddeville M, Daveran-Mingot ML, Ritzenthaler P, Le Bourgeois P. Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* A12, a Strain Isolated from Wheat Sourdough. *Genome Announc*. 2016;4:e00692–16.
60. McCulloch JA, de Oliveira VM, de Almeida Pina AV, Perez-Chaparro PJ, de Almeida LM, de Vasconcelos JM, de Oliveira LF, da Silva DEA, Rogez HLG, Cretenet M, et al. Complete Genome Sequence of *Lactococcus lactis* Strain AI06, an Endophyte of the Amazonian Acai Palm. *Genome Announc*. 2014;2(6):e01225–14.
61. Yang X, Wang Y, Huo G. Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* KLD54.0325. *Genome Announc*. 2013;1(6):e00962–13.
62. van Mastriigt O, Abee T, Smid EJ. Complete Genome Sequences of *Lactococcus lactis* subsp. *lactis* bv. *diacytylactis* FM03 and *Leuconostoc mesenteroides* FM06 Isolated from Cheese. *Genome Announc*. 2017;5(28):e00633–17.
63. Oliveira LC, Saraiva TDL, Soares SC, Ramos RTJ, Sá PHCG, Carneiro AR, Miranda F, Freire M, Renan W, Júnior AFO, et al. Genome sequence of *Lactococcus lactis* subsp. *lactis* NCDO 2118, a GABA-producing strain. *Genome Announc*. 2014;2(5):e00980–14.
64. Tran TD, Huynh S, Parker CT, Han R, Hnasko R, Gorski L, McGarvey JA. Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* Strain 14B4, Which Inhibits the Growth of *Salmonella enterica* Serotype Poona In Vitro. *Microbiol Resour Announc*. 2018;7(19):e01364–18.

65. Linares DM, Kok J, Poolman B. Genome sequences of *Lactococcus lactis* MG1363 (revised) and NZ9000 and comparative physiological studies. *J Bacteriol.* 2010;192:5806–12.
66. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarne K, Weissenbach J, Ehrlich SD, Sorokin A. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res.* 2001;11:731–53.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.