NAR Genomics and Bioinformatics, 2025, **7**, lqaf085
https://doi.org/10.1093/nargab/lqaf085
Standard Article

OXFORD

A 69.9-kb long inverted repeat increases genome instability in a strain of *Lactobacillus crispatus*

Lorenzo Colombini, Francesco Santoro*, Mariana Tirziu, Anna Maria Cuppone, Gianni Pozzi, Francesco Iannelli**

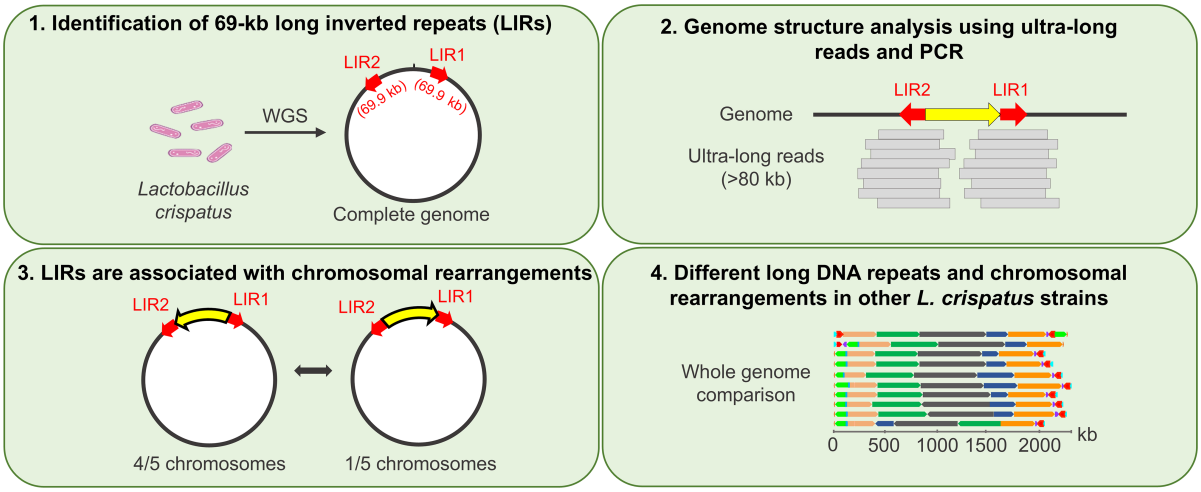
Laboratory of Molecular Microbiology and Biotechnology, Department of Medical Biotechnologies, University of Siena, 53100 Siena, Italy

*To whom correspondence should be addressed. Email: francesco.iannelli@unisi.it
Correspondence may also be addressed to Francesco Santoro. Email: santorof@unisi.it

Abstract

Long inverted repeats (LIRs) of DNA sequences longer than 30 kb are rare in prokaryotes. Here, we identified two 69.9-kb LIRs in the genome of *Lactobacillus crispatus* M247_Siena, a derivative of strain M247. Complete genome sequence of M247_Siena was determined using Nanopore and Illumina technologies, while genome structure was analyzed using ultra-long Nanopore read mapping and polymerase chain reaction (PCR). In the parental M247 genome, there was only one copy of the 69.9-kb segment, while a 15.4-kb DNA segment was present instead of the second 69.9-kb segment copy. Both segments were delimited by the same insertion sequences (IS*1201* and IS*Lcr2*), and PCR analysis of the M247 population revealed low rates (~1.28 per 10^4 chromosomes) of chromosomal rearrangements involving these regions. In contrast, the 69.9-kb LIRs in M247_Siena increased genomic instability, as evidenced by two alternative chromosomal structures detected at frequencies of 23.3% and 76.7% (~1 out of 5 chromosomes). Comparative analysis of *L. crispatus* genomes revealed no LIRs similar to those of M247_Siena. However, long repeats of other DNA segments and chromosomal rearrangements, mostly associated with insertion sequences, were detected in 8 and 9 out of 25 *L. crispatus* genomes, respectively, highlighting genomic instability as a trait of the species.

Graphical abstract



Introduction

Long inverted repeats (LIRs) are palindromic DNA sequences longer than 100 bp, either immediately adjacent or separated by a unique sequence, found in both eukaryotic and prokaryotic genomes [1, 2]. LIRs can form intra-strand DNA secondary structures (e.g. hairpins) that act as replication stall sites and promote intrachromosomal homologous recombination, resulting in genomic rearrangements such as translocations, duplications, deletions, and inversions, ultimately leading to genome instability [3–7]. An elevated number of LIRs in a genome does not necessarily correlate with increased genome instability; rather, instability is directly proportional to the size of the LIRs and inversely proportional to the length of the unique sequence separating the LIRs [2, 8]. LIRs are found in high numbers in eukaryotic genomes [9–11], while are less common in prokaryotes [12] mainly due to the

Received: February 3, 2025. **Revised:** May 13, 2025. **Editorial Decision:** May 30, 2025. **Accepted:** June 3, 2025

activity of removal systems such as the *Escherichia coli* SbcC–SbcD nuclease complex [1, 6, 13]. In prokaryotic genomes, the following LIRs longer than 30 kb have been described: (i) 553-kb LIRs in a *Bacillus thuringiensis* serovar *israelensis* laboratory strain [14]; (ii) 70-kb LIRs in *Pseudomonas koreensis* P19E3 [12]; and (iii) LIRs located at the replication terminus, ranging from 38 to 76 kb, in five *Lactobacillus delbrueckii* ssp. *bulgaricus* strains [15]. Bacterial strains, which are sub-cultured in specific laboratory media and conditions, often for decades, adapt their genomes to these growth conditions leading to the loss or amplification of genes that enhance *in vitro* fitness [16, 17]. The *Lactobacillus crispatus* M247 strain is a newborn fecal isolate studied for its probiotic activity and reported to exert beneficial effects on intestinal inflammatory disorders, to help counteracting vaginal dysbiosis, and likely contribute to papilloma virus clearance [18–22]. We previously characterized the M247 genome [23] revealing a single 2 336 109-bp chromosome (37% GC content) enriched in insertion sequences (ISs), which account for >10% of the genome length, and containing two novel mobile genetic elements, including an integrative and mobilizable element carrying a bacteriocin biosynthesis gene cluster. In this work, genome sequencing was performed on a derivative of M247 that has been repeatedly sub-cultured and maintained in our strain collection at the University of Siena since early 1996, hence named M247_Siena. The genome of this derivative strain was characterized by an unusual duplication of a 69.9-kb chromosomal segment producing LIRs located 224.4 kb apart. We investigated the impact of LIRs on genome stability and discussed possible molecular mechanisms underlying the generation of the M247_Siena LIRs.

Materials and methods

Bacterial strains and growth conditions

The *L. crispatus* M247_Siena strain is a derivative of the M247 strain [18, 23], which was repeatedly sub-cultured and maintained in the University of Siena laboratory strain collection since early 1996. M247_Siena was grown in de Man–Rogosa–Sharpe (MRS) broth (Oxoid) or in MRS supplemented with 1.5% agar (BD Difco) at 37°C in the presence of a 5% CO_2-enriched atmosphere.

DNA purification and quantification

Bacterial cells were grown at 37°C in 500 ml of MRS broth until reaching middle exponential phase (OD_{590} of 1.9), and then harvested by centrifugation (5000 × *g* for 30 min at 4°C). High-molecular-weight genomic DNA was purified using a raffinose-based method, as reported for strain M247 [23–25]. Briefly, the cell pellet was vortex-mixed dry and incubated in 15 ml of protoplasting buffer [20% raffinose, 50 mM Tris–HCl, pH 8.0, 5 mM ethylenediaminetetraacetic acid (EDTA), 4 mg/ml lysozyme] for 1 h at 37°C. Protoplasts were centrifuged (5000 × *g* for 5 min), and osmotic lysis was achieved by resuspending the pellet in 15 ml of deionized H_2O containing 100 μg/ml proteinase K (Merck KGaA) and incubating for 30 min at 37°C, with 0.5% Sodium Dodecyl Sulfate (SDS) added after 15 min. Then, 0.55 M NaCl was added and the mixture was incubated for 10 min at room temperature. High-molecular-weight DNA was purified three times with 1 volume of chloroform–isoamyl alcohol (24:1, v/v) and precipitated with 0.6 volumes of ice-cold isopropanol and spooled on a glass rod. The DNA was air-dried and resuspended in 10-fold diluted saline–sodium citrate (SSC) 1× buffer, and then adjusted to 1× SSC. The DNA solution was homogenized using a rotator mixer and stored at +4°C. DNA was quantified with a Qubit 4.0 Fluorometer (Invitrogen, Life Technologies) using the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific) and with a spectrophotometer (Implen). DNA integrity and size were assessed by horizontal gel electrophoresis using 0.6% Seakem LE (Lonza) agarose in 0.5× Tris–borate–EDTA running buffer.

Illumina whole genome sequencing

Illumina sequencing was performed at MicrobesNG (University of Birmingham, UK) using the Nextera XT library preparation kit (Illumina), followed by sequencing on a NovaSeq 6000 device (Illumina) (2 × 250 bp paired-end sequencing). Illumina reads were trimmed with Trimmomatic v0.30 (https://github.com/usadellab/Trimmomatic) [26] and analyzed with FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Illumina reads properties are reported in Supplementary Table S1.

Nanopore whole genome sequencing

Nanopore whole genome sequencing was carried out as already described [23, 27, 28]. Briefly, sequencing libraries were prepared in 1.5-ml LoBind tubes (Sarstedt) using wide bore (Ø 1.2 mm) tips for DNA manipulation to minimize physical shearing. Genomic DNA size selection was obtained with 0.5 volumes of AMPure XP beads (Beckman Coulter) according to manufacturer's instructions. Two micrograms of size-selected DNA were used for library construction with the SQK-LSK 108 kit (Oxford Nanopore Technologies). Library preparation was performed following the manufacturer's protocol with the following modifications: (i) incubation on a rotator mixer for 15 min and (ii) library loading beads (LLBs) were omitted. Finally, 1 μg of DNA library was loaded onto an R9.4 flow cell (Oxford Nanopore Technologies). A 48-h sequencing run was performed on a GridION device (Oxford Nanopore Technologies) using MinKNOW v.19.12.5. Real-time base calling was performed with Guppy v3.2.6 (Oxford Nanopore Technologies), filtering out reads with a quality cutoff <Q7. Base called reads were analyzed with NanoPlot v1.18.2 (https://github.com/wdecoster/NanoPlot) [29]. Nanopore reads properties are reported in Supplementary Table S1.

Genome assembly and annotation

M247_Siena Nanopore reads were filtered twice with Filtlong v0.2.0 (https://github.com/rrwick/Filtlong), first to retain reads longer than 80 kb ("*--min_length 80000*"), and then to obtain a 110× coverage taking 2.3 Mb as the genome size estimate ("*--target_bases 253000000*"). Filtered Nanopore reads were assembled with Unicycler v0.4.7 (https://github.com/rrwick/Unicycler) [30]. The resulting circular contig was polished with Medaka v0.7.1 (https://github.com/Nanoporetech/medaka) using Nanopore reads longer than 80 kb, followed by two additional rounds of polishing with Pilon v1.22 using Illumina reads (https://github.com/broadinstitute/pilon) [31]. Assembly completeness was assessed with Bandage v.0.8.1 (https://github.com/rrwick/Bandage) [32], whereas assembly quality was evaluated with both Ideel

(https://github.com/mw55309/ideel) [33] and CheckM v1.1.3 (https://github.com/Ecogenomics/CheckM) [34]. BWA v0.7.17 (https://github.com/lh3/bwa) [35] and minimap2 v2.13 (https://github.com/lh3/minimap2) [36] were used to align Illumina and Nanopore reads to the assembled genome, respectively. Genome read alignments were visually inspected with Tablet v1.17.08.17 (https://github.com/cropgeeks/tablet) [37, 38], and the coverage graph obtained from genome mapping with all Nanopore reads longer than 80 kb was used to manually investigate and validate the genome structure. The whole genome was automatically annotated with the NCBI Prokaryotic Genome Annotation Pipeline v4.10 [39]. In addition, manual annotation of DNA sequences of interest was carried out by BLAST homology searches against the National Center for Biotechnology Information protein database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins, accessed in October 2024) and Pfam (available under the InterPro consortium at https://www.ebi.ac.uk/interpro/, accessed in October 2024), whereas KEGG orthology assignments were obtained using BlastKoala (https://www.kegg.jp/blastkoala/, accessed in October 2024) [40]. Default parameters were used for all software, unless otherwise specified.

Genome analysis

Chromosomal structural variants were investigated using the Sniffles v1.0.12 structural variation caller (https://github.com/fritzsedlazeck/Sniffles) [41] and the npInv v.1.24 tool (https://github.com/haojingshao/npInv) [42] for non-allelic homologous recombination (NAHR)-mediated genomic inversion, and then visualized using the Integrative Genomics Viewer v.2.19.3. (https://software.broadinstitute.org/software/igv/, accessed in September 2024) [43]. The alternative chromosomal structures of M247_Siena, detected by npInv, were designed *in silico* and aligned to ultra-long Nanopore reads (>80 kb in length). Aligned reads were processed with samclip v.0.4.0 (https://github.com/tseemann/samclip) with parameter "*--max 100*" and then filtered with SAMtools (https://github.com/samtools/) [44] using the command "*view*" with the parameter "*-b*" specifying the chromosomal region of interest to retain only reads spanning the repeated regions relevant for supporting the different chromosomal structures. The previously described *L. crispatus* M247 genome (GenBank accession no. CP088015) was used for genome comparison analysis [23]. In addition, *L. crispatus* complete genomes were downloaded from the NCBI Microbial Genome Database (https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/1815/, accessed in December 2024), with GenBank accession numbers provided in Supplementary Table S2. Prior to analysis, the complete genome sequences were synchronized using Circlator v.1.5.5 (https://github.com/sanger-pathogens/circlator) [45] with option "*fixstart*" to reorient and rotate each sequence based on the coordinates of the *dnaA* gene to ensure the same starting position and polarity. Genome comparisons were conducted with the following tools: (i) Mauve v2.4.0 (https://github.com/PATRIC3/mauve-viewer) [46], (ii) D-genies (https://dgenies.toulouse.inrae.fr/run, accessed in December 2024) [47], (iii) Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi), (iv) Artemis and Artemis Comparison Tool (ACT) v17.0.1 (https://github.com/sanger-pathogens/Artemis/blob/master/act) [48], (v) MUMmer v3.23 (https://github.com/mummer4/mummer) [49], and (vi) PopPUNK tool v2.6.5 (https://github.com/johnlees/PopPUNK) [50] using the "*--fit-model lineage*" parameter for data fitting as described [51]. Fisher's exact test was used to assess the odds ratio between the presence of repeats and genomic inversions in the genomes. Default parameters were used for all software, unless otherwise specified.

PCR and Sanger sequencing

PCR and direct PCR sequencing were carried out following an already described protocol [52–55] to investigate the genome structure of M247 and M247_Siena. Oligonucleotide primers and their coordinates are reported in Supplementary Table S3. Specifically, primer pairs IF1110/IF1111 and IF1118/IF1119 were used to detect the right and left junctions of the first 69.9-kb DNA sequence with the chromosome in M247_Siena, while primer pairs IF1118/IF1120 and IF1111/IF1121 were used to detect the junctions of the second 69.9-kb DNA sequence with the M247_Siena chromosome. For M247, primer pairs IF1110/IF1111 and IF1118/IF1120 were used to amplify the junctions of the 69.9-kb DNA sequence with the chromosome, while IF1119/IF1122 and IF1117/1121 were used to amplify the junctions of the 15.4-kb DNA sequence. Rearranged chromosomal structures in M247 were detected with primer pairs IF1118/IF1119 and IF1111/IF1121 for the 69.9-kb DNA sequence, while IF1120/IF1122 and IF1110/1117 were used for the rearranged junctions of the 15.4-kb DNA sequence. Sanger sequencing was performed to verify the PCR results for all rearranged junctions.

Long-range PCR and direct PCR Nanopore sequencing

In addition, long-range PCR reactions were performed on the parental strain M247: primer pairs IF1502/IF1503 and IF1504/IF1505 were used to amplify the junctions of the 69.9-kb DNA sequence with the chromosome producing amplicons of ∼15 kb in length, while IF1506/IF1508 and IF1505/1509 were used to amplify the junctions of the 15.4-kb DNA sequence producing amplicons of ∼9 kb in length. Analogously, rearranged chromosomal structures in M247 were detected with primer pairs IF1504/IF1506 and IF1503/IF1121 for the 69.9-kb DNA sequence producing amplicons of ∼15 kb in length, while IF1508/IF1505 and IF1509/IF1502 were used for the rearranged junctions of the 15.4-kb DNA sequence producing amplicons of ∼9 kb in length. The long-range PCR products were purified using 1 volume of AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions. Direct PCR sequencing was performed on purified PCR products using the SQK-LSK 108 kit along with the EXP-NBD104 kit (Oxford Nanopore Technologies) for library construction according to manufacturer's instructions for amplicon sequencing, followed by a 2-h sequencing run on a GridION device (MinKNOW v.19.12.5) using an R9.4 flow cell (Oxford Nanopore Technologies). As for genome sequencing, real-time base calling was performed with Guppy v3.2.6 (>Q7). Base called reads were aligned against the M247 genome using nucleotide BLAST to verify the PCR results.

qPCR

Quantitative PCR (qPCR) experiments were carried out with the KAPA SYBR FAST qPCR Kit Master Mix Universal (2×) (Merck KGaA) on a LightCycler 1.5 apparatus (Roche Di-

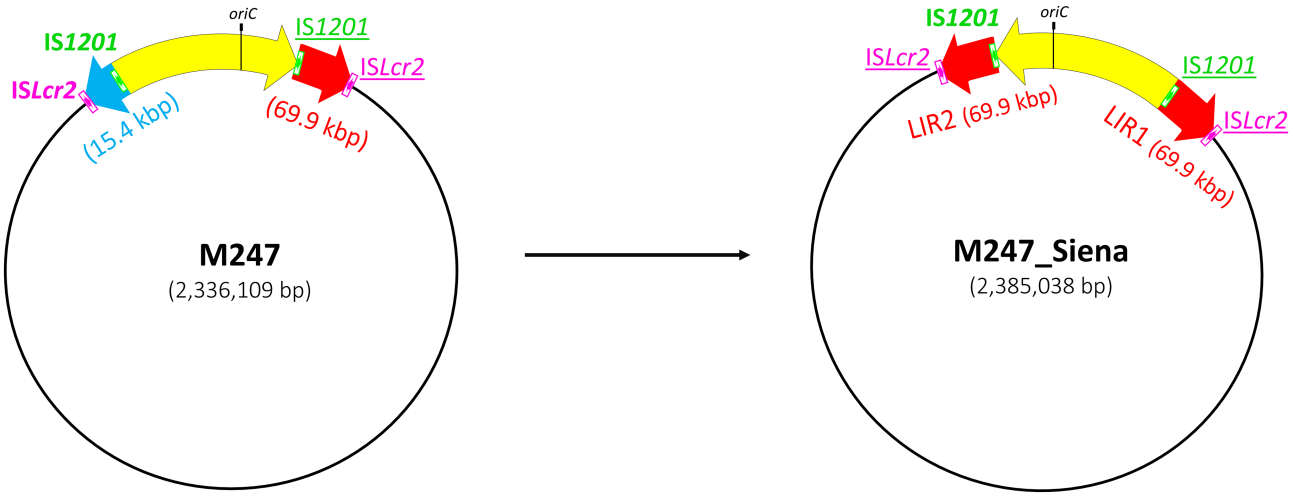


Figure 1. Schematic representation of *L. crispatus* M247 and its derivative M247_Siena genomes. In M247_Siena genome, an additional copy of a 69.9-kb DNA segment (red arrow) replaced a 15.4-kb DNA segment (blue arrow). The two 69.9-kb copies were arranged as long inverted repeats, LIR1 and LIR2, separated by a 224.4-kb DNA segment (yellow arrow) containing the chromosomal origin of replication *oriC*. The 224.4-kb LIR-separating region in M247_Siena genome is inverted compared to the 225.8-kb homologous region of M247. The LIRs and the 15.4-kb segment contained a copy of IS*1201* at 5′ end and a copy of IS*Lcr2* at 3′ end. IS copies can contain nucleotides changes; 100% identical ISs are indicated with the same font style (bold or underlined). IS*Lcr2* copies differ in two nucleotide changes, while IS*1201* copies differ in four nucleotide changes and six nucleotide insertions. LIRs positions on genome sequence (GenBank no. CP046589) are nucleotides 38 618–108 536 (LIR1) and nucleotides 2 129 303–2 199 215 (LIR2). To highlight inversion of the *oriC* containing DNA segment, LIR1 and LIR2 are represented as being 185 823 bp and 38 618 bp downstream of the chromosomal origin of replication on the right and left replichores, respectively. Figure is not scaled.

agnostics). The qPCR reaction mixture contained, in a final volume of 20 µl, 1× KAPA SYBR FAST qPCR reaction mix, 5 pmol of each primer, and 20 ng of bacterial gDNA. The thermal cycling profile consisted of an initial 4-min denaturation step at 95°C, followed by 40 cycles of repeated denaturation (10 s at 95°C), annealing (15 s at 60°C), and polymerization (3 min and 30 s at 72°C). The temperature transition rate was 20°C/s during the denaturation and annealing steps and 5°C/s during the polymerization step. Serial dilutions of *L. crispatus* M247 chromosomal DNA with known concentration were used to build a *gyrB* gene standard curve by plotting the threshold cycle against the number of chromosome copies. This external standard curve was used to quantify the number of rearranged M247 chromosomal structures in each sample. Primer pairs were the same used in endpoint short-range PCR (see above and Supplementary Table S3). The frequency of rearranged M247 chromosomal structures was calculated as the average of the rearranged DNA junctions frequencies. Melting curve analysis was performed to discriminate the desired amplification products from primer–dimer products.

Results

Lactobacillus crispatus M247_Siena genome contains 69.9-kb long inverted repeats

Genome sequencing was performed on an *L. crispatus* M247 derivative strain, maintained in our laboratory for over 20 years, which we named M247_Siena. The complete genome sequence was obtained combining Illumina sequencing reads with ultra-long Nanopore sequencing reads (>80 kb), which resolved the genome complexity of the strain. Sequence analysis showed that the M247_Siena genome is organized in one circular chromosome of 2 385 038 bp in length, containing 2336 open reading frames (ORFs) with an average GC content of 37.07%. Compared to the parental M247 strain [23], M247_Siena genome was 48 929 bp longer due to the presence of an additional copy of a 69.9-kb long chromosomal DNA segment, which replaced a 15.4-kb DNA segment. The two 69.9-kb copies were arranged as long inverted repeats (LIRs) located 224 440 bp apart. LIRs positions on genome sequence (GenBank no. CP046589) are nucleotides 38 618–108 536 (LIR1) and nucleotides 2 129 303–2 199 215 (LIR2). This duplication was associated with the inversion of the 224 440-bp DNA region separating the LIRs, compared to the parental strain (Fig. 1). According to the genome schematic representation, LIR1 and LIR2 are located 185 823 bp and 38 618 bp downstream of the chromosomal origin of replication on the right and left replichores, respectively. Nanopore read mapping integrated by PCR analysis confirmed that the genomic differences between M247 and M247_Siena are genuine (Supplementary Fig. S1). Each 69.9-kb LIR harbored 72 ORFs and was delimited by a copy of IS*1201* at the 5′ end and by a copy of IS*Lcr2* at the 3′ end (Fig. 1 and Supplementary Table S4). The LIRs differed by four nucleotide changes and a 6-bp insertion located in the IS*1201* copy. Manually curated annotation assigned a putative function to 59 out of 72 ORFs, including 18 genes encoding for metabolic enzymes, 18 for cellular signaling and cellular processes, 8 for transcriptional regulation, 3 rRNA, and 5 tRNA (Supplementary Table S4). In the parental M247 genome, the 15.4-kb DNA segment, harboring 17 ORFs, also contained an IS*1201* and an IS*Lcr2* copy at 5′ and 3′ ends, respectively (Fig. 1 and Supplementary Table S5). A putative function was assigned to 14 out of 17 ORFs including a peptidase, a protease, a peroxide stress protein YaaA, and an ABC transporter system (Supplementary Table S5).

Chromosomal rearrangements in M247_Siena and parental M247 strain

Since M247_Siena contained a large inversion compared to the parental strain, we investigated the presence of alternative chromosomal structures with two different bioinformatic

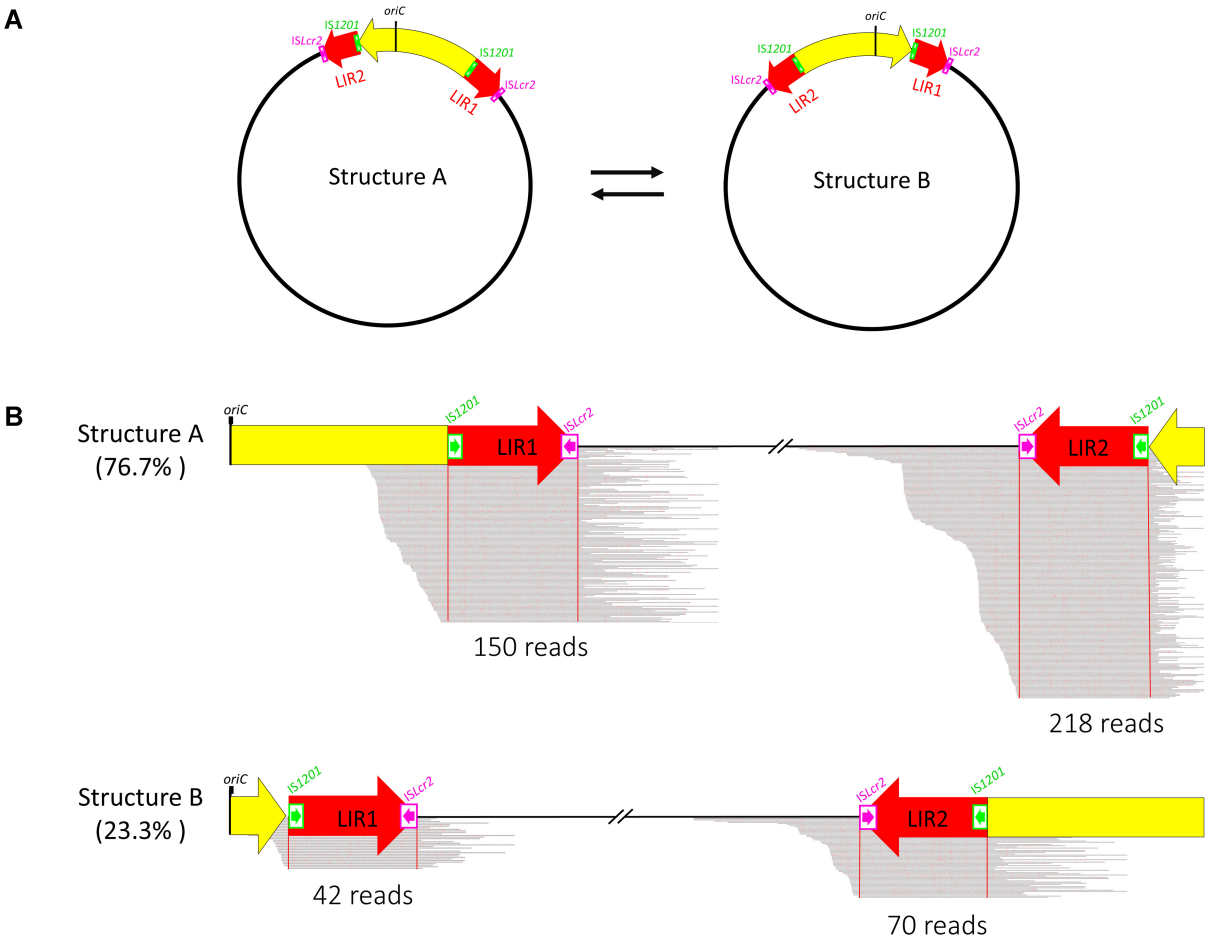


Figure 2. (**A**) Detection and (**B**) quantification of chromosomal rearrangements in the M247_Siena genome. (A) Genome mapping with ultra-long Nanopore reads indicated the presence of chromosomal structures A and B. In addition to the assembled structure A, the inversion of the 224.4-kb sequence containing the *oriC* (yellow arrow) and flanked by the two LIRs (red arrows) produced structure B. (B) Chromosomal rearrangements occurred at a frequency of ~1 per 5 chromosomes. Nanopore reads longer than 80 kb spanning LIRs and the respective flanking regions are indicated by gray lines and were generated with the Tablet v1.17.08.17 software. Figure is not scaled.

tools (Sniffles and npInv) and by mapping Nanopore reads longer than 80 kb, which accounted for a 363× genome coverage. These approaches revealed the presence of two different chromosomal structures, which we named structures A and B (Fig. 2 and Supplementary Fig. S2). Structure A resulted from the genome assembly, while structure B was characterized by the inversion of the 224 440-bp DNA region containing the *oriC* and flanked by the LIRs. Semi-quantitative analysis of 480 sequencing reads spanning both 69.9-kb LIRs revealed that 368 reads (76.7%) contained structure A, while 112 reads (23.3%) contained structure B (Fig. 2). Analysis of alternative chromosomal structures in the parental M247 genome failed to identify the presence of the 224 440-bp inversion, which was therefore investigated by long-range PCR analysis using PCR primers in different pair combinations. In addition to structure I, which corresponds to the assembled genome, analysis of PCR amplicons suggested the presence of putative alternative rearranged structures (II, III, IV, V, VI, VII) (Fig. 3). It is likely that the inversion of the 225 800-bp segment containing *oriC* produced structure II, whereas a larger inversion (311 201 bp), including also the flanking 69.9- and 15.4-kb segments, produced structure III. Additional rearranged structures could derive from the duplication of the 69.9-kb segment and the 15.4-kb segment deletion (structure IV) or from the duplication of the 15.4-kb segment and the 69.9-kb segment deletion (structure VI). In structures IV and VI, the inversion of 225.8-kb DNA sequence could produce structures V and VII, respectively. Real-time quantitative PCR analysis indicated that rearranged structures occurred at a frequency of ~1.28×10^{-4} chromosomes (Table 1).

Structural analysis of *L. crispatus* complete genomes suggests genome instability

Analysis of 25 *L. crispatus* complete genomes available in the NCBI database (accessed in December 2024), revealed that these genomes contain DNA segments homologous to the 69.9- and 15.4-kb segments found in M247 genome (Fig. 4), whereas the 69.9-kb segment duplication (LIRs), as reported in M247_Siena, was not detected. The CO3MRSI1 and Lc1226 genomes contained a 15.4-kb segment delimited by IS*1201* and IS*Lcr2* as in M247, with an additional IS*982* inserted in strain Lc1226 (Supplementary Table S6). In contrast, 13 genomes contained an IS*30* copy at the 3′ end, and in 9 of these, the 5′ end lacked an insertion sequence, while, in the remaining 10 genomes, the 15.4-kb segment was not delimited by insertion sequences. Insertions and deletions of ISs and/or ORFs (up to 12.1 kb) were present at different posi-

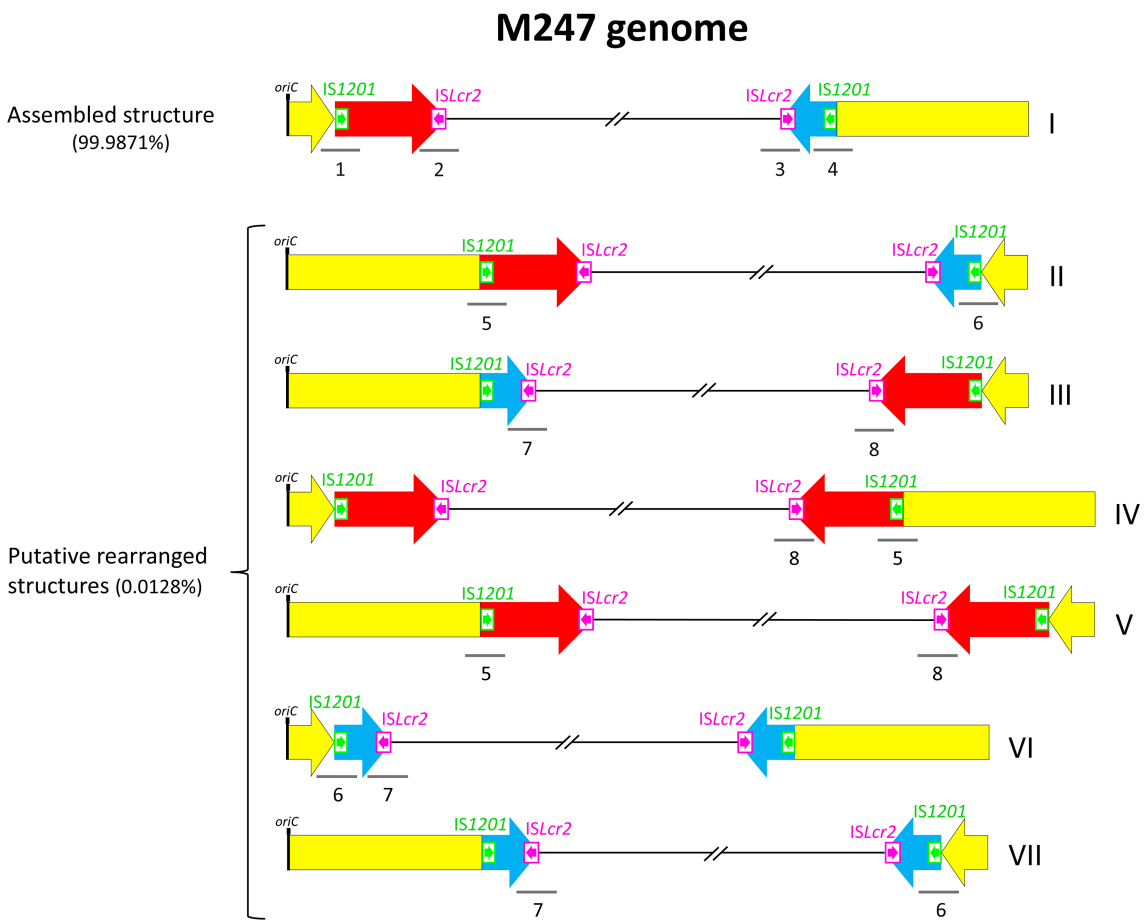


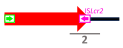
Figure 3. Schematic representation of chromosomal rearrangements in M247 genome. Genome mapping with long-range PCR suggested the presence of alternative chromosomal structures in addition to the assembled genome structure I. The inversion of the 225.8-kb DNA sequence containing the *oriC* (yellow arrow) produced structure II, whereas a larger inversion including the flanking 69.9-kb (red arrow) and 15.4-kb (blue arrow) segments produced structure III. Structure IV contains a duplication of the 69.9-kb segment and 15.4-kb segment deletion, while structure VI contains a duplication of the 15.4-kb segment and 69.9-kb segment deletion. Inversion of 225.8-kb DNA sequence in structures IV and VI produces structures V and VII, respectively. Frequencies of PCR junction fragments are reported in Table 1. Chromosomal rearrangements occurred at a frequency of ~1.28×10^4 chromosomes. Gray lines indicate PCR amplicons (numbered 1–8). Figure is not scaled.

tions within the DNA segment (Supplementary Table S6). The 69.9-kb segment showed higher variability among genomes: (i) 22 segments contained a duplication of ribosomal RNA operon genes; (ii) 7 segments exhibited large deletions at the 5′ end, ranging from 8.7 to 22 kb; and (iii) various insertions or deletions of ISs and/or ORFs were present (Supplementary Table S7). Long direct and inverted repeats of other DNA segments ranging from 6 985 to 39 820 bp were detected in eight genomes. Out of these, two repeats were associated with mobile genetic element DNA, while the remaining six cases mostly involved duplication of metabolic genes, similarly to what was observed in the M247_Siena LIRs. Synteny analysis, performed by pairwise genomic comparisons using the M247 genome as reference, identified large inversions in other nine *L. crispatus* genomes, namely PMC201, VSI04, AB70, lc31, VSI24, FDAARGOS_743, 2029, CO3MRSI1, and DC21.1 (Fig. 4 and Supplementary Fig. S3). These inversions spanned the origin–terminus axis, involving both replichores, with recombination sites characterized by the presence of ISs in five genomes, a group II intron in three genomes, and the 16S rRNA gene in one genome. Although a statistically significant correlation between the presence of repeats and genomic inversions was not identified (Fisher's exact test, $P = 0.1077$), four genomes (FDAARGOS_743, VSI04, 2029, CO3MRSI1) contained both a long repeat and a large inversion. Phylogenetic analysis, conducted with PopPUNK, identified six partitioning clusters based on core genomic sequences (Fig. 4). The genomes containing duplications were distributed across different clusters, while six out of the nine genomes containing large inversions were grouped into two clusters (Fig. 4).

Discussion

In this work, we reported the unusual genomic structure of M247_Siena, an *L. crispatus* M247 derivative maintained in our laboratory for over 20 years. The peculiarity of the M247_Siena genome is the presence of a 69.9-kb chromosomal DNA segment duplication, resulting in inverted repeats located 224.4 kb apart. Large structural variations have mostly remained undetected when sequencing genomes with short read technologies [56–60]. In addition, the formation of very long inverted repeats in bacterial genomes is considered an extremely rare event; it has been reported that among over 9600 prokaryotic genomes, only a small subset (3%) contains long nearly identical repeats above 30 kb in length [12]. The mechanisms underlying the formation of

Table 1. Real-time PCR quantification of chromosomal junctions containing the ends of the 69-kb and the 15-kb DNA segments in the M247 genome

Junction fragment	Primer pair	Genome position	Fragment copies[a]	Genomic structure (copies)[b]	Frequency of rearranged structures[c]
1	IF1110 - IF1111		6.59×10^{-2}		
2	IF1118 - IF1120		2.80×10^{-1}	Assembled structure (1.71×10^{-1})	
3	IF1119 - IF1122		9.21×10^{-2}		
4	IF1117 - IF1121		2.45×10^{-1}		
5	IF1111 - IF1121		4.46×10^{-5}		1.28×10^{-4}
6	IF1110 - IF1117		6.47×10^{-6}	Putative rearranged structures (2.20×10^{-5})	
7	IF1120 - IF1122		3.46×10^{-5}		
8	IF1118 - IF1119		2.14×10^{-6}		

[a] Concentrations of junction fragments are expressed as copies per chromosomes.
[b] Copies of genomic structures are calculated as the average number of the junction fragments copies.
[c] Frequency of putative rearranged genomic structures is expressed relative to the canonical chromosomal structure. Schematic representations of rearranged structures are shown in Fig. 3.

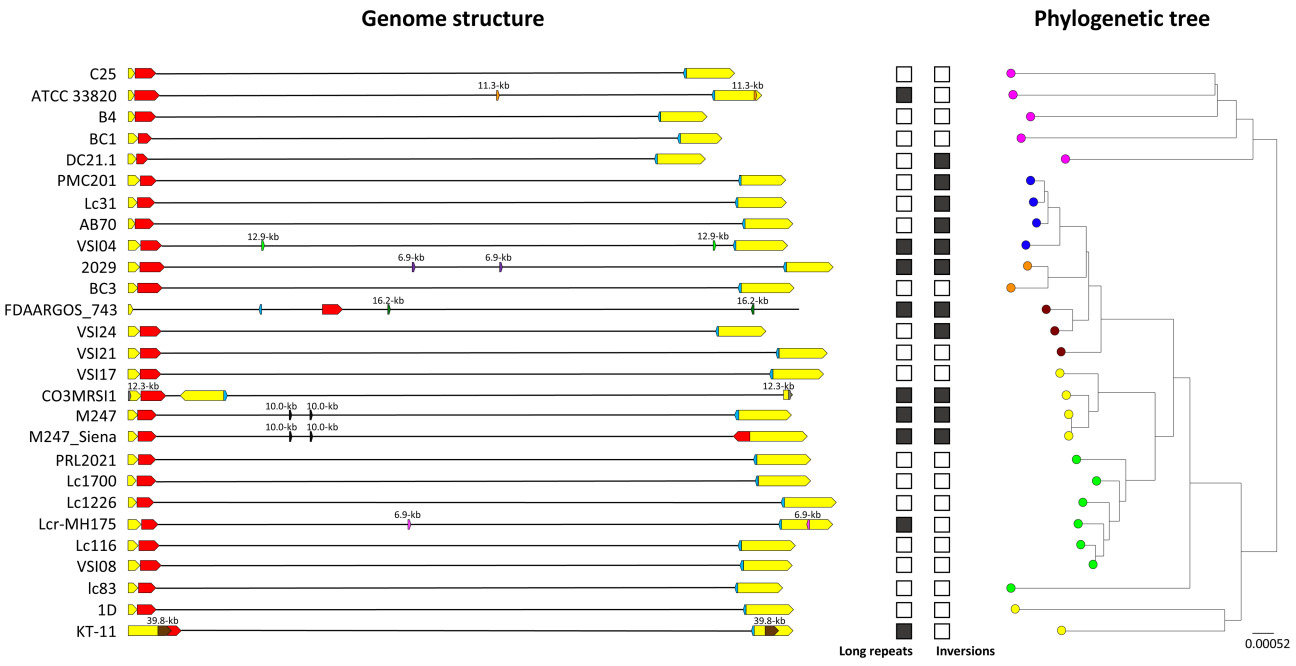


Figure 4. *Lactobacillus crispatus* complete genome structures and their genetic relatedness. *Lactobacillus crispatus* genomes contain DNA segments homologous to the 69.9- and 15.4-kb segments of M247, indicated with red and blue arrows, respectively. A duplication of the 69.9 kb as reported in M247_Siena was not found, while eight genomes contain a duplication of other DNA segments ranging in size from 6.9 to 39.8 kb, arranged as direct (DR) or inverted (IR) repeats. The size of duplicated segments is reported. Additionally, synteny analysis using M247 as reference indicated the presence of large inversions in other nine genomes (Supplementary Fig. S3). The presence of long DNA repeats and inversion in a genome is reported as a black box in the heatmap. The phylogenetic tree was generated based on whole genome sequences using PopPUNK (branch lengths indicate the number of nucleotide substitutions per site as indicated by the scale bar), whereas six population clusters, indicated with dots of different color, were obtained from core genome sequences. Names of the genome strains are reported on the left.

LIRs are not yet fully understood. Hypothetically, LIRs arise during DNA replication due to double-strand DNA breaks near short inverted repeats, which fold back (intra-strand self-annealing) and serve to prime DNA synthesis of the complementary strand, thereby facilitating chromosomal rescue by DNA damage repair mechanisms [15]. Genome comparison of M247_Siena with its parental M247 strain identified a 15.4-kb chromosomal DNA segment, which was lost and replaced by the 69.9-kb LIR2. DNA sequence analysis of both 15.4- and 69.9-kb segments excluded the hypothesis of a mobile DNA origin, as neither *int*/*xis* nor transfer genes were found [61, 62], and the GC content was comparable to the rest of the genome [63, 64]. The 69.9-kb DNA segment contains genes encoding surface proteins likely involved in *L. crispatus* adhesion to epithelial cells [65, 66]. Thus, the presence of segment duplication likely resulting in an increased gene dosage could improve adhesion, which is a critical trait for probiotic persistence. Additionally, duplication of the D-lactate dehydrogenase gene may influence lactic acid production, providing a competitive advantage to the M247_Siena strain [67, 68]. Both the 15.4- and 69.9-kb segments were delimited by the same ISs elements, namely IS*Lcr2* and IS*1201*. ISs have been described as source of genomic instability within bacterial genomes and can induce large chromosomal inversions [69–72]. In line with this, our PCR analysis revealed the presence of chromosomal rearrangements involving the 69.9- and the 15.4-kb DNA segments in the M247 bacterial population, likely associated with the presence of IS*Lcr2* and IS*1201* copies. Homologous recombination can lead to deletions, duplications, translocations (for direct repeats), and inversions (for inverted repeats) [73–76]. Indeed, genome mapping with Nanopore reads showed the presence of two alternative chromosomal structures in the M247_Siena population characterized by a 224 440-bp inversion of the region flanked by the LIRs and containing the origin of replication. Of note, the M247_Siena inversion does not significantly alter the length of replichores, as reported for other bacterial large chromosomal inversions [77, 78]. The 69.9-kb LIRs of M247_Siena represent large regions of homology with 99.99% of sequence identity, which are likely subject to nonallelic homologous recombination (NAHR), a mechanism associated with recurrent DNA rearrangements [79, 80]. In principle, the LIRs could favor the formation of a very large non-B DNA cruciform structure consisting of two 70-kb stems and two 224-kb loops, where the single-stranded loops are complementary and capable of reciprocally annealing. This large cruciform would be possibly subject to random double-strand breakage and could cause DNA polymerase stalling with consequent large deletions [81]. Since we consistently detected only two main chromosomal structures without deletions, we speculate that NAHR is the most likely mechanism involved in the generation of the chromosomal inversion observed in M247_Siena. A semi-quantitative estimation by sequencing read counting showed that 112 out of 480 (23.3%) ultra-long nanopore reads spanning the repeats contained the inverted structure (Fig. 2), indicating that the chromosomal structure is rearranged in ∼1 out of 5 bacterial cells. Quantitative PCR performed on M247 DNA showed that the frequencies of chromosomal rearrangements were significantly lower (∼1.28 out of 10^4 cells), suggesting that the newly generated LIRs of M247_Siena increased the intrinsic genomic instability of the strain. Whole genome alignment of M247 with 25 other *L. crispatus* complete genomes revealed that no other genome contains a repeated DNA sequence like M247_Siena and that in each genome a 69.9-kb and a 15.4-kb DNA homologous segment are present. These findings support the hypothesis of a laboratory-based evolution of the M247 strain, which led to its derivative M247_Siena. Furthermore, the genome alignment revealed that (i) eight genomes contain long DNA repeats (6.9–39.8 kb), in six cases the duplication involved metabolic genes, as observed in the M247_Siena LIRs (Supplementary Table S4), and (ii) nine genomes contain large inversions, mostly associated with IS elements. Thus, genomic instability is likely a trait of the *L. crispatus* species and a source of genomic diversity within the bacterial population, as observed in other bacterial species [71, 72].

Acknowledgements

Illumina genome sequencing was provided by MicrobesNG (http://www.microbesng.com).

Author contributions: Lorenzo Colombini (Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing—original draft), Francesco Santoro (Conceptualization, Formal analysis, Funding acquisition, Visualization, Writing—review & editing), Mariana Tirziu (Data curation, Formal analysis), Anna Maria Cuppone (Data curation, Formal analysis), Gianni Pozzi (Conceptualization, Funding acquisition, Methodology, Resources, Writing—review & editing), and Francesco Iannelli (Conceptualization, Methodology, Funding acquisition, Resources, Supervision, Visualization, Writing—review & editing).

Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

Conflict of interest

None declared.

Funding

This work was supported in part by the Italian Ministry University and Research (MUR-Italy) under grant numbers 202089LLEH "CoDiCo project" (call "Progetti di Ricerca di Rilevante Interesse Nazionale—Bando 2020"), 2022ZJE2FH "MyMo project" (call "Progetti di Ricerca di Rilevante Interesse Nazionale—Bando 2022"), P202228W7E "CrispOmics" (call "Progetti di Ricerca di Rilevante Interesse Nazionale—Bando 2022 PNRR"), and in part by EU funding within the MUR PNRR Extended Partnership Initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT, "One Health Basic and Translational Research Actions addressing Unmet Needs on Emerging Infectious Diseases", Spoke 3, WP1). L.C. holds a junior lecturer position funded by the PNRR Extended Partnership initiative on Emerging Infectious Diseases.

Data availability

Lactobacillus crispatus M247 and M247_Siena complete genomes are available under GenBank accession nos. CP088015 and CP046589, respectively. All Nanopore and Illumina sequencing data used in this study are available in the

Sequence Read Archive (SRA), accession nos. SRR17479173, SRR17479172, SRR10902282, and SRR10902283. Properties of sequencing reads generated in this study are reported in Supplementary Table S1, whereas a list of complete genomes used for genome comparison can be found in Supplementary Table S2.

References

1. Leach DRF. Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioessays* 1994;**16**:893–900. https://doi.org/10.1002/bies.950161207
2. Wang Y, Leung FCC. Long inverted repeats in eukaryotic genomes: recombinogenic motifs determine genomic plasticity. *FEBS Lett* 2006;**580**:1277–84. https://doi.org/10.1016/j.febslet.2006.01.045
3. Gordenin DA, Resnick MA. Yeast ARMs (DNA at-risk motifs) can reveal sources of genome instability. *Mutat Res* 1998;**400**:45–58. https://doi.org/10.1016/S0027-5107(98)00047-5
4. Waldman AS, Tran H, Goldsmith EC *et al.* Long inverted repeats are an at-risk motif for recombination in mammalian cells. *Genetics* 1999;**153**:1873–83. https://doi.org/10.1093/genetics/153.4.1873
5. Kurahashi H, Inagaki H, Ohye T *et al.* Palindrome-mediated chromosomal translocations in humans. *DNA Repair (Amst)* 2006;**5**:1136–45. https://doi.org/10.1016/j.dnarep.2006.05.035
6. Darmon E, Eykelenboom JK, Lincker F *et al. E. coli sbc*CD and RecA control chromosomal rearrangement induced by an interrupted palindrome. *Mol Cell* 2010;**39**:59–70. https://doi.org/10.1016/j.molcel.2010.06.011
7. Asaf S, Khan AL, Lubna K *et al.* Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Sci Rep* 2020;**10**:3881. https://doi.org/10.1038/s41598-020-60803-y
8. Lobachev KS, Shor BM, Tran HT *et al.* Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics* 1998;**148**:1507–24. https://doi.org/10.1093/genetics/148.4.1507
9. Houck CM, Rinehart FP, Schmid CW. A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol* 1979;**132**:289–306. https://doi.org/10.1016/0022-2836(79)90261-4
10. Lu L, Jia H, Dröge P *et al.* The human genome-wide distribution of DNA palindromes. *Funct Integr Genomics* 2007;**7**:221–7. https://doi.org/10.1007/s10142-007-0047-6
11. Wang Y, Leung FCC. A study on genomic distribution and sequence features of human long inverted repeats reveals species-specific intronic inverted repeats. *FEBS J* 2009;**276**:1986–98. https://doi.org/10.1111/j.1742-4658.2009.06930.x
12. Schmid M, Frei D, Patrignani A *et al.* Pushing the limits of *de novo* genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. *Nucleic Acids Res* 2018;**46**:8953–65. https://doi.org/10.1093/nar/gky726
13. Bowater RP, Bohálová N, Brázda V. Interaction of proteins with inverted repeats and cruciform structures in nucleic acids. *Int J Mol Sci* 2022;**23**:6171. https://doi.org/10.3390/ijms23116171
14. Bolotin A, Quinquis B, Roume H *et al.* Long inverted repeats around the chromosome replication terminus in the model strain *Bacillus thuringiensis* serovar *israelensis* BGSC 4Q7. *Microb Genomics* 2020;**6**:12. https://doi.org/10.1099/mgen.0.000468
15. El Kafsi H, Loux V, Mariadassou M *et al.* Unprecedented large inverted repeats at the replication terminus of circular bacterial chromosomes suggest a novel mode of chromosome rescue. *Sci Rep* 2017;**7**:44331. https://doi.org/10.1038/srep44331
16. Fux CA, Shirtliff M, Stoodley P *et al.* Can laboratory reference strains mirror 'real-world' pathogenesis? *Trends Microbiol* 2005;**13**:58–63. https://doi.org/10.1016/j.tim.2004.11.001
17. Barrick JE, Yu DS, Yoon SH *et al.* Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 2009;**461**:1243–7. https://doi.org/10.1038/nature08480
18. Cesena C, Morelli L, Alander M *et al. Lactobacillus crispatus* and its nonaggregating mutant in human colonization trials. *J Dairy Sci* 2001;**84**:1001–10. https://doi.org/10.3168/jds.S0022-0302(01)74559-6
19. Marcotte H, Ferrari S, Cesena C *et al.* The aggregation-promoting factor of *Lactobacillus crispatus* M247 and its genetic locus. *J Appl Microbiol* 2004;**97**:749–56. https://doi.org/10.1111/j.1365-2672.2004.02364.x
20. Castagliuolo I, Galeazzi F, Ferrari S *et al.* Beneficial effect of auto-aggregating *Lactobacillus crispatus* on experimentally induced colitis in mice. *Immunol Med Microbiol* 2005;**43**:197–204.
21. Pierro FD, Bertuccioli A, Cattivelli D *et al. Lactobacillus crispatus* M247: a possible tool to counteract CST IV. *Nutrafoods* 2018;**17**:169–72.
22. Pierro FD, Criscuolo AA, Giudici AD *et al.* Oral administration of *Lactobacillus crispatus* M247 to papillomavirus-infected women: results of a preliminary, uncontrolled, open trial. *Minerva Obstet Gynecol* 2021;**73**:621–31.
23. Colombini L, Santoro F, Tirziu M *et al.* The mobilome of *Lactobacillus crispatus* M247 includes two novel genetic elements: Tn*7088* coding for a putative bacteriocin and the siphovirus prophage ΦM247. *Microb Genomics* 2023;**9**:001150. https://doi.org/10.1099/mgen.0.001150
24. Teodori L, Colombini L, Cuppone AM *et al.* Complete genome sequence of *Lactobacillus crispatus* type strain ATCC 33820. *Microbiol Resour Announc* 2021;**10**:e0063421. https://doi.org/10.1128/MRA.00634-21
25. Pinzauti D, Iannelli F, Pozzi G *et al.* DNA isolation methods for Nanopore sequencing of the *Streptococcus mitis* genome. *Microb Genomics* 2022;**8**:000764. https://doi.org/10.1099/mgen.0.000764
26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. https://doi.org/10.1093/bioinformatics/btu170
27. Cuppone AM, Colombini L, Fox V *et al.* Complete genome sequence of *Streptococcus pneumoniae* strain Rx1, a Hex mismatch repair-deficient standard transformation recipient. *Microbiol Resour Announc* 2021;**10**:e00799–21. https://doi.org/10.1128/MRA.00799-21
28. Colombini L, Cuppone AM, Tirziu M *et al.* The mobilome-enriched genome of the competence-deficient *Streptococcus pneumoniae* BM6001, the original host of integrative conjugative element Tn*5253*, is phylogenetically distinct from historical pneumococcal genomes. *Microorganisms* 2023;**11**:1646. https://doi.org/10.3390/microorganisms11071646
29. De Coster W, D'Hert S, Schultz DT *et al.* NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018;**34**:2666–9. https://doi.org/10.1093/bioinformatics/bty149
30. Wick RR, Judd LM, Gorrie CL *et al.* Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;**13**:e1005595. https://doi.org/10.1371/journal.pcbi.1005595
31. Walker BJ, Abeel T, Shea T *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;**9**:e112963. https://doi.org/10.1371/journal.pone.0112963
32. Wick RR, Schultz MB, Zobel J *et al.* Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 2015;**31**:3350–2. https://doi.org/10.1093/bioinformatics/btv383
33. Watson M, Warr A. Errors in long-read assemblies can critically affect protein prediction. *Nat Biotechnol* 2019;**37**:124–6. https://doi.org/10.1038/s41587-018-0004-z
34. Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single

cells, and metagenomes. *Genome Res* 2015;**25**:1043–55. https://doi.org/10.1101/gr.186072.114
35. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;**25**:1754–60. https://doi.org/10.1093/bioinformatics/btp324
36. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100. https://doi.org/10.1093/bioinformatics/bty191
37. Milne I, Stephen G, Bayer M *et al.* Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* 2013;**14**:193–202. https://doi.org/10.1093/bib/bbs012
38. Milne I, Bayer M, Cardle L *et al.* Tablet—next generation sequence assembly visualization. *Bioinformatics* 2010;**26**:401–2. https://doi.org/10.1093/bioinformatics/btp666
39. Tatusova T, DiCuccio M, Badretdin A *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;**44**:6614–24. https://doi.org/10.1093/nar/gkw569
40. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;**428**:726–31. https://doi.org/10.1016/j.jmb.2015.11.006
41. Smolka M, Paulin LF, Grochowski CM *et al.* Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol* 2024;**42**:1571–80. https://doi.org/10.1038/s41587-023-02024-y
42. Shao H, Ganesamoorthy D, Duarte T *et al.* npInv: accurate detection and genotyping of inversions using long read sub-alignment. *BMC Bioinf* 2018;**19**:261. https://doi.org/10.1186/s12859-018-2252-9
43. Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6. https://doi.org/10.1038/nbt.1754
44. Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9. https://doi.org/10.1093/bioinformatics/btp352
45. Hunt M, Silva ND, Otto TD *et al.* Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015;**16**:294. https://doi.org/10.1186/s13059-015-0849-0
46. Darling ACE, Mau B, Blattner FR *et al.* Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;**14**:1394–403. https://doi.org/10.1101/gr.2289704
47. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 2018;**6**:e4958. https://doi.org/10.7717/peerj.4958
48. Carver TJ, Rutherford KM, Berriman M *et al.* ACT: the Artemis comparison tool. *Bioinformatics* 2005;**21**:3422–3. https://doi.org/10.1093/bioinformatics/bti553
49. Marçais G, Delcher AL, Phillippy AM *et al.* MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol* 2018;**14**:e1005944. https://doi.org/10.1371/journal.pcbi.1005944
50. Lees JA, Harris SR, Tonkin-Hill G *et al.* Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 2019;**29**:304–16. https://doi.org/10.1101/gr.241455.118
51. De Giorgi S, Ricci S, Colombini L *et al.* Genome sequence typing and antimicrobial susceptibility testing of infertility-associated *Enterococcus faecalis* reveals clonality of aminoglycoside-resistant strains. *J Glob Antimicrob Resist* 2022;**29**:194–6. https://doi.org/10.1016/j.jgar.2022.03.017
52. Iannelli F, Giunti L, Pozzi G. Direct sequencing of long polymerase chain reaction fragments. *Mol Biotechnol* 1998;**10**:183–5. https://doi.org/10.1007/BF02760864
53. Iannelli F, Santoro F, Oggioni MR *et al.* Nucleotide sequence analysis of integrative conjugative element Tn*5253* of *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* 2014;**58**:1235–9. https://doi.org/10.1128/AAC.01764-13
54. Santoro F, Oggioni MR, Pozzi G *et al.* Nucleotide sequence and functional analysis of the *tet*(M)-carrying conjugative transposon Tn*5251* of *Streptococcus pneumoniae*: tn*5251* of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* 2010;**308**:150–8.
55. Lazzeri E, Santoro F, Oggioni MR *et al.* Novel primer-probe sets for detection and identification of mycobacteria by PCR-microarray assay. *J Clin Microbiol* 2012;**50**:3777–9. https://doi.org/10.1128/JCM.02300-12
56. Chaisson MJP, Huddleston J, Dennis MY *et al.* Resolving the complexity of the human genome using single-molecule sequencing. *Nature* 2015;**517**:608–11. https://doi.org/10.1038/nature13907
57. Chaisson MJP, Sanders AD, Zhao X *et al.* Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun* 2019;**10**:1784. https://doi.org/10.1038/s41467-018-08148-z
58. De Coster W, De Rijk P, De Roeck A *et al.* Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res* 2019;**29**:1178–87. https://doi.org/10.1101/gr.244939.118
59. Huddleston J, Chaisson MJP, Steinberg KM *et al.* Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res* 2017;**27**:677–85. https://doi.org/10.1101/gr.214007.116
60. Pendleton M, Sebra R, Pang AWC *et al.* Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat Methods* 2015;**12**:780–6. https://doi.org/10.1038/nmeth.3454
61. Wozniak RAF, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 2010;**8**:552–63. https://doi.org/10.1038/nrmicro2382
62. Getino M, de la Cruz F. Natural and artificial strategies to control the conjugative transmission of plasmids. *Microbiol Spectr* 2018;**6**:mtbp-0015-2016. https://doi.org/10.1128/microbiolspec.mtbp-0015-2016
63. Bohlin J, Eldholm V, Pettersson JHO *et al.* The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC Genomics* 2017;**18**:151. https://doi.org/10.1186/s12864-017-3543-7
64. Sueoka N. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc Natl Acad Sci USA* 1962;**48**:582–92. https://doi.org/10.1073/pnas.48.4.582
65. Von Ossowski I, Satokari R, Reunanen J *et al.* Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic *Lactobacillus rhamnosus* GG. *Appl Environ Microb* 2011;**77**:4465–72. https://doi.org/10.1128/AEM.02497-10
66. Ye Q, Lao L, Zhang A *et al.* Multifunctional properties of the transmembrane LPxTG-motif protein derived from *Limosilactobacillus reuteri* SH-23. *J Dairy Sci* 2023;**106**:8207–20. https://doi.org/10.3168/jds.2023-23440
67. Abdelmaksoud AA, Koparde VN, Sheth NU *et al.* Comparison of *Lactobacillus crispatus* isolates from *Lactobacillus*-dominated vaginal microbiomes with isolates from microbiomes containing bacterial vaginosis-associated bacteria. *Microbiology* 2016;**162**:466–75. https://doi.org/10.1099/mic.0.000238
68. O'Hanlon DE, Moench TR, Cone RA. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PLoS One* 2013;**8**:e80074. https://doi.org/10.1371/journal.pone.0080074
69. Daveran-Mingot M-L, Campo N, Ritzenthaler P *et al.* A natural large chromosomal inversion in *Lactococcus lactis* is mediated by homologous recombination between two insertion sequences. *J Bacteriol* 1998;**180**:4834–42. https://doi.org/10.1128/JB.180.18.4834-4842.1998
70. Lee H, Doak TG, Popodi E *et al.* Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*. *Nucleic Acids Res* 2016;**44**:7109–19.
71. Weigand MR, Peng Y, Batra D *et al.* Conserved patterns of symmetric inversion in the genome evolution of *Bordetella*

respiratory pathogens. *mSystems* 2019;**4**:e00702–19 https://doi.org/10.1128/msystems.00702-19
72. Jespersen MG, Hayes AJ, Tong SYC *et al.* Insertion sequence elements and unique symmetrical genomic regions mediate chromosomal inversions in *Streptococcus pyogenes*. *Nucleic Acids Res* 2024;**52**:13128–37. https://doi.org/10.1093/nar/gkae948
73. Achaz G, Coissac E, Netter P *et al.* Associations between inverted repeats and the structural evolution of bacterial genomes. *Genetics* 2003;**164**:1279–89. https://doi.org/10.1093/genetics/164.4.1279
74. Romero D, Palacios R. Gene amplification and genomic plasticity in prokaryotes. *Annu Rev Genet* 1997;**31**:91–111. https://doi.org/10.1146/annurev.genet.31.1.91
75. Roth JR, Benson N, Galitski T *et al.* Rearrangements of the bacterial chromosome: formation and applications. In: Neidhardt FC, Curtiss R (eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Vol. **2**. Washington, D.C.: ASM Press, 1996, 2256–76.
76. Smith GR. Homologous recombination in prokaryotes. *Microbiol Rev* 1988;**52**:1–28. https://doi.org/10.1128/mr.52.1.1-28.1988
77. Eisen JA, Heidelberg JF, White O *et al.* Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol* 2000;**1**:research0011. https://doi.org/10.1186/gb-2000-1-6-research0011
78. Tillier ERM, Collins RA. Genome rearrangement by replication-directed translocation. *Nat Genet* 2000;**26**:195–7. https://doi.org/10.1038/79918
79. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* 2016;**17**:224–38. https://doi.org/10.1038/nrg.2015.25
80. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet* 2002;**18**:74–82. https://doi.org/10.1016/S0168-9525(02)02592-1
81. Voineagu I, Narayanan V, Lobachev KS *et al.* Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci USA* 2008;**105**:9936–41. https://doi.org/10.1073/pnas.0804510105

Received: February 3, 2025. **Revised:** May 13, 2025. **Editorial Decision:** May 30, 2025. **Accepted:** June 3, 2025