CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Molecular characterisation of a conjugative Bifidobacterium megaplasmid
Authors	Dineen, Rebecca L.
Publication date	2023
Original Citation	Dineen, R. L. 2023. Molecular characterisation of a conjugative Bifidobacterium megaplasmid. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2023, Rebecca L. Dineen. - https://creativecommons.org/publicdomain/zero/1.0/
Download date	2026-11-03 00:33:46
Item downloaded from	https://hdl.handle.net/10468/15491

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork



Molecular Characterisation of a Conjugative *Bifidobacterium* Megaplasmid

A thesis presented to the National University of Ireland for the Degree of Doctor of Philosophy by

Rebecca L. Dineen, B.Sc. (Hons) Genetics

https://orcid.org/0000-0003-4752-3629

School of Microbiology

University College Cork

Head of Department: Dr. David Clarke
Supervisor: Prof. Douwe van Sinderen
Co-supervisor: Dr. Mary O'Connell-Motherway

2023

TABLE OF CONTENTS

TABLE OF CONTENTS i
DECLARATION iv
LIST OF FIGURES v
LIST OF TABLES viii
LIST OF PUBLICATIONS x
ABBREVIATIONS xi
GENERAL ABSTRACT xiv

CHAPTER I 1

GENERAL INTRODUCTION 1

The Genus *Bifidobacterium* and its Relevance to Human Health 2

Functionalities and Genetic Organization of Plasmids 5

Replication of Circular Plasmids: An Overview 7

Mechanisms of Vertical Plasmid Propagation: Plasmid Stability and Maintenance 14

Phylogeny and Genetic Diversity of Plasmids from *Bifidobacterium* 21

Current Plasmid-based Genetic Tools for use in Bifidobacteria 34

Engineering *Bifidobacterium* for Food and Biomedical Applications: Limitations and Future Perspectives 39

Horizontal Gene Transfer by Conjugation 40

Type IV Secretion Systems: Mechanism and Structure 42

The Ever-Expanding Diversity of Conjugative Relaxases 46

Mating Pair Formation: Identification and Classification 51

Coupling Proteins of T4SSs 55

Mating Pair Stabilization in Gram Positive Conjugative Systems: Cell-Cell Attachment 59

Conclusion 66

REFERENCES 68

CHAPTER II 89

Molecular Analysis of the Replication Functions of the Bifidobacterial Conjugative Megaplasmid pMP7017 89

ABSTRACT 90

INTRODUCTION 91

EXPERIMENTAL PROCEDURES 93

RESULTS 101

DISCUSSION 112

AUTHOR CONTRIBUTIONS 115

FIGURES AND TABLES 116

SUPPLEMENTAL MATERIAL 123

REFERENCES 139
CHAPTER III 145
Molecular Aspects of the Conjugative Transfer System of Megaplasmid pMP7017 from *Bifidobacterium breve* JCM7017 145
ABSTRACT 146
INTRODUCTION 147
MATERIALS AND METHODS 150
RESULTS 155
DISCUSSION 163
FIGURES AND TABLES 167
REFERENCES 187
CHAPTER IV 194
Conjugal Transfer of Bifidobacterial Megaplasmid pMP7017 Depends on Cell Wall Associated Structures 194
ABSTRACT 195
INTRODUCTION 196
EXPERIMENTAL PROCEDURES 199
RESULTS 204
DISCUSSION 210
FIGURES AND TABLES 214
SUPPLEMENTAL MATERIAL 233
REFERENCES 235
CHAPTER V 241
Transcriptional Landscape of the pMP7017 Megaplasmid and its Impact on the *Bifidobacterium breve* UCC2003 Transcriptome 241
ABSTRACT 242
INTRODUCTION 243
MATERIALS AND METHODS 246
RESULTS 250
DISCUSSION 261
FIGURES AND TABLES 265
SUPPLEMENTAL MATERIAL 296
REFERENCES 300
CHAPTER VI 306
GENERAL DISCUSSION AND FUTURE PERSPECTIVES 306
REFERENCES 314

DEDICATION **319**
ACKNOWLEDGEMENTS **320**

DECLARATION

I hereby declare that the research presented in this thesis is my own work and effort, and that it has not been submitted for any other degree, either at University College Cork or elsewhere. Whatever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

This work was completed under the guidance of Prof. Douwe van Sinderen at the APC Microbiome Institute & School of Microbiology, Biosciences Institute, University College Cork.

Signature:…….

Date:…….

LIST OF FIGURES

CHAPTER I

Figure 1. Mechanism of rolling-circle replication 10
Figure 2. Mechanism of strand displacement replication initiation (plasmid RSF1010) 11
Figure 3. Model of replication initiation in class A theta plasmids 12
Figure 4. The contribution of stability mechanisms to plasmid maintenance..... 17
Figure 5. Graphical representation of the genetic organisation of partitioning modules from various plasmids of gram positive bacteria..... 19
Figure 6. Species distribution and general characteristics of plasmids from *Bifidobacterium*..... 27
Figure 7. Phylogenetic relatedness among plasmid replicases identified in *Bifidobacterium*..... 28
Figure 8. Sequence identity among natural plasmids from *Bifidobacterium*..... 30
Figure 9. General translocation pathway and structural comparison of canonical T4SSs of G- and G+ origin 44
Figure 10. A schematic representation of the relationship between relaxase MOB groups..... 49
Figure 11. Mating pair classification of conjugative ssDNA secretion systems 53
Figure 12. Molecular architecture of the different T4CPs 58
Figure 13. The PrgABC system of the enterococcal sex-pheromone responsive plasmid pCF10 63
Figure 14. Schematic representation of the pCW3 *cnaC* region and $CnaC_{pCW3}$ domain organisation..... 64

CHAPTER II

Figure 1. pRD derivative plasmids; stability and copy number..... 116
Figure 2. The binding behavior of RepA to its corresponding promoter region and of RepB to multi-site probes in EMSA 118
Figure 3. Identification of the critical nucleotide residues for the binding of RepB to iterons..... 119
Figure 4. Development of a novel bifidobacterial shuttle vector, pRD1.3 120
Figure 5. Working model of pMP7017 replication initiation 121
Figure S1. Promoter analysis of the *repA* and *repB* genes and HPLC chromatograms 123
Figure S2. EMSAs representing the lack of affinity of RepB for its promoter and coding regions 125
Figure S3. EMSA data illustrating the affinity of RepB for single-site probes, E and H, and probe G containing 3 putative iterons 126
Figure S4. Heuristic modelling of RepB to replicases of IRPs, HPLC chromatogram of RepB_ΔC and EMSA displaying the lack of affinity of RepB_ΔC for probe, P17 127
Figure S5. Comparative analysis of the pMP7017 basic replicon across bifidobacterial megaplasmids..... 128
Figure S6. Alignment of the repeat contain region between 13 Bifidobacterial megaplasmids..... 130

CHAPTER III

Figure 1. Effect of methylation on pMP7017 conjugation efficiency 167
Figure 2. Genome map of pMP7017 indicating genes targeted for mutagenesis 168
Figure 3. Mutational analysis of the pMP7017 candidate conjugation-associated genes.....169
Figure 4. Genetic complementation of ATPase-specifying gene mutations in pMP7017.....171
Figure 5. Presence/absence of conjugation-associated gene clusters of pMP7017-related megaplasmids.....172
Figure 6. Proposed functional model of the pMP7017 conjugal system 173

CHAPTER IV

Figure 1. Domain architecture and structural homology of Spd1.....214
Figure 2. Domain architecture and structural homology of PilA.....216
Figure 3. Domain architecture and structural homology of PilB.....218
Figure 4. Domain architecture and structural homology of Spd2.....219
Figure 5. Mutation and complementation analysis of pMP7017 genes encoding surface proteins 220
Figure 6. Transcriptional profile of the pMP7017-adhesion locus 221
Figure 7. Schematic comparative representation of the spd1-srtC-pilABC putative adhesion loci among pMP7017-homologs from various strains of *B. longum* subsp. *longum*.....222
Figure 8. Mechanism of polymerisation and cell-wall anchoring of the PilABC structure and proposed phenotypes resulting from site-directed mutagenesis of individual pilin subunits.....223
Figure S1. Structural homology of the PilA N-terminal subdomains, CBM1 and CBM3, with the carbohydrate binding modules of various carbohydrate-active enzymes.....233

CHAPTER V

Figure 1. Intrinsic features of pMP7017 promoters, transcriptional unit elements and termination signals, as a function of transcriptional rate 265
Figure 2. Conserved putative regulatory motif present within the 5' untranslated region of fourteen pMP7017 genes.....267
Figure 3. Pie chart representation of transcriptionally active predicted pMP7017-genes, grouped by functional categories 268
Figure 4. Codon usage bias of individual genes within the pMP7017 genome based on RSCU values.....269
Figure 5. Correlation analysis of transcription level with overall nucleotide composition.....270
Figure 6. Codon usage frequencies of pMP7017 genes vs JCM7017: tRNAs and codon usage.....271
Figure 7. Differential expression of *B breve* UCC2003 chromosomal genes in the presence of pMP7017.....272
Figure 8. Codon usage comparison of chromosomal DEGs in the presence of pMP7017 and its relationship with pMP7017 encoded tRNA anti-codons 273
Figure 9. Mapping of genes from 43 *B. longum* pMP7017-replicase positive strains to the pMP7017 genome 274
Figure 10. Synteny and variability of Bifidobacterial megaplasmid genomes 276
Figure 11. Genome synteny and co-linearity of pMP7017 family plasmids 277

Figure S1. Codon usage comparison of pMP7017-encoded transcribed genes vs non-transcribed genes at mid-log growth of *B. breve* UCC2003 297

LIST OF TABLES

CHAPTER I
Table 1. Comparison of the three main modes of plasmid replication initiation described for circular plasmids 13
Table 2. Features of the eight classes of TA systems 20
Table 3. *Bifidobacterium* plasmids 32
Table 4. Current genome editing tools for use in *Bifidobacterium* 36
Table 5. Features of conjugative relaxases based on their current MOB family classification....50
Table 6. Features of the eight MPF classes....54
Table 7. Defining features of the six different classes of sortases 65

CHAPTER II
Supplementary Table S1. Strains and plasmids used in this study 131
Supplementary Table S2. List of primers used in this study....134
Supplementary Table S3. Probe information for EMSAs....137

CHAPTER III
Table 1. Functional analysis of pMP7017-encoded candidate conjugal genes....174
Table 2. Co-mobilisation assay 177
Supplementary Table S1. Strains and plasmids used in this study 178
Supplementary Table S2. Oligonucleotide primers used in the study 184

CHAPTER IV
Table 1. Strains and plasmids used in this study....225
Table 2. List of primers used in this study 229
Table 3. Pairwise comparison of the predicted structures of candidate pMP7017-encoded surface proteins and experimentally determined target structure 231
Table S1. Functional analysis of the alternative pilin proteins encoded within the adhesion *loci* of megaplasmids from *B. longum* subsp. *longum* 234

CHAPTER V
Table 1. Updated annotation of transcriptionally active genes 278
Table 2. pMP7017-expressed Type II TA systems: activity and process inhibition281
Table 3. Analysis of CRISPR spacer targets of the pMP7017 Type IE CRIPSR-cas system....282
Table 4. pMP7017-expressed RNAs....283
Table 5. Correlation analysis of transcription level with overall nucleotide composition....284
Table 6. Down-regulated *B. breve* UCC2003 chromosomal genes in the presence of pMP7017 285
Table 7. Gene Ontology enrichment analyses of down-regulated *B. breve* UCC2003 chromosomal genes in the presence of pMP7017....287
Table 8. Up-regulated *B. breve* UCC2003 chromosomal genes in the presence of pMP7017 288
Table 9. Gene Ontology enrichment analyses of up-regulated *B. breve* UCC2003 chromosomal genes in the presence of pMP7017....293

Table 10. Prevalence of pMP7017 among metagenomic samples.....295
Supplementary Table S1. Comparison of codon usage frequencies.....298

LIST OF PUBLICATIONS

Chapter II: **R. L. Dineen, C. Penno, P. Kelleher, M. J. B. Bourin, M. O'Connell-Motherway, and D. van Sinderen**, 'Molecular analysis of the replication functions of the bifidobacterial conjugative megaplasmid pMP7017', *Microb Biotechnol*, May 2021, doi: 10.1111/1751-7915.13810.

ABBREVIATIONS

Abbreviation	Definition
3'OH	3' hydroxyl group
AAD	all-alpha domain
ANI	average nucleotide identity
Ka	binding constant
CBM	carbohydrate binding module
CWA	cell wall associated
CWSS	cell wall sorting signal
CRISPR-Cas	clustered regularly interspaced short palindromic repeats-CRISPR-associated
COG	clusters of orthologous groups
CDS	coding sequence
CAI	Codon Adaptation Index
CBI	Codon Bias Index
CP	coupling protein
CWSS	carboxy-terminal cell wall-sorting signal
DEGs	differentially expressed genes
Kd	dissociation constant
Pol III	DNA polymerase III holoenzyme
dso	double stranded origin
EMSAs	Electrophoretic Mobility Shift Assays
EPS	exopolysaccharide
eDNA	extracellular DNA
FDR	false discovery rate
FOS	fructo-oligosaccharides
GOS	galacto-oligosaccharides
GIT	gastrointestinal tract
GW	Glycine-Tryptophan
G-	Gram-negative
G+	Gram-positive
G+C	guanine + cytosine
nH	Hill coefficient
CHAP	histidine-dependent amidohydrolases/peptidases
HUH	histidine-hydrophobe-histidine
HGT	horizontal gene transfer
HRB	human-residential bifidobacteria
IMCC	inner membrane core complex
IS elements	insertion elements
ICEs	integrative and conjugative elements
IRs	inverted repeats
IRP	iteron-regulated plasmid

KO	KEGG Orthology
LTA	lipoteichoic acids
LB	Luria-Bertani (broth)
MCL	Markov clustering
MPF	mating pair formation
MTase	methytransferase
MIC	minimal inhibitory concentration
MGEs	mobile genetic elements
mob	mobilisation
mMRS	modified de Man, Rogosa, and Sharpe medium
MMRS	multimer resolution systems
ncRNA	non-coding RNA
NBD	N-terminal binding domain
NTPase	nucleoside triphosphate hydrolase
NBD	nucleotide binding domain
ORF	open reading frame
oriT	origin of transfer
OMCC	outer membrane core complex
par	partition
PAM	plasmid artificial modification
PCN	Plasmid copy number
PSK	post-segregational killing
PE	primer extension
RPKM	reads per kilobase per million mapped
RCM	reinforced clostridial medium
RSCU	relative synonymous codon usage
REase	restriction endonuclease
RM	restriction-modification
RV	retention volume
RNAseq	RNA sequencing
RNAP	RNA-polymerase
RCR	rolling circle replication
RMSD	Root Mean Square Deviation
ssDNA	single-stranded DNA
sso	single-stranded origin
SDM	site-directed mutagenesis
SDPs	sortase-dependent proteins
ter	terminator sites
TA	toxin–antitoxin
TSS	transcription start site
TSS	transcriptional start site
TUs	transcriptional units
tra	transfer (gene)

tmRNA	transfer messenger RNA
t-strand	Transfer-strand
TM	transmembrane
TMD	transmembrane domain
TNF-α	tumour necrosis factor α
T4CP	Type IV coupling protein
T4SS	Type IV Secretion System
UTR	untranslated region
WT	wild type

GENERAL ABSTRACT

Bifidobacterium species are highly abundant autochthonous bacteria of the human gut microbiota, particularly during host infancy. Various members of the *Bifidobacterium* genus have been associated with a plethora of health-promoting attributes, among which maintaining gut homeostasis, limiting pathogenic bacterial invasion as well as their purported ability to modulate host immune responses are notable examples. Due to their positive association with human health, this genus has received substantial scientific attention and commercial interest. Plasmids were once considered an atypical feature of *Bifidobacterium* and those identified within this genus were primarily small cryptic plasmids which were presumed to replicate by a so-called rolling circle mechanism. However, the recent availability of long-read single-molecule sequencing technologies precipitated the resolution of the first reported bifidobacterial megaplasmid isolated from the common and abundant inhabitant of the human gut, *Bifidobacterium breve* JCM7017. The discovery of this >190 kb conjugative megaplasmid, denoted pMP7017, and the subsequent identification of pMP7017 homologs in several *B. longum* subsp. *longum* strains, highlights the prevalence of this megaplasmid family within this genus, representing an unexplored feature. Conjugative plasmids such as pMP7017 play a central role in bacterial evolution and have the potential to significantly influence the activity of the microbiome community and, by extension, impact human health and physiology.

The research described in this thesis covers the replication functions of megaplasmid pMP7017 and exploits these functions for the development of important molecular tools that facilitate the genetic engineering of these genetically recalcitrant bacteria. As pMP7017 represents the first and, thus far, only experimentally validated conjugative plasmid of bifidobacterial origin, research performed within the context of this thesis also concerned examination of the conjugative functions of this megaplasmid. While the molecular characterisation of pMP7017 represents an opportunity for the development of much needed molecular tools and provides a starting point in the understanding of bifidobacterial DNA transfer systems, an integrated approach of comparative analyses and metagenomic data mining has generated highly relevant and insightful information concerning the biology and distribution of pMP7017 and related megaplasmids.

Chapter I summarizes the current knowledge pertaining to the mechanisms and control of plasmid replication and includes a comprehensive overview of the general features of known resident plasmids of bifidobacterial origin and their phylogenetic relatedness based on their encoded replicase proteins. The review further presents a general overview of scientific literature concerning conjugative systems of gram-positive bacteria.

Chapter II of this thesis describes the molecular dissection of the replication system of plasmid pMP7017 encoding two putative Rep proteins, RepA and RepB. Comprehensive functional analyses identify RepB as the replicative initiator of pMP7017 and the minimal replicon was determined to be a 2.4 kb region encompassing the *repB* gene and seventeen 20-mer repeat sequences located downstream, termed iterons. The essentiality of individual iterons was investigated and the affinity of RepB for these iterons was determined and based on these findings the pMP7017 origin of replication and a putative stability locus were identified. Utilizing this information, we subsequently constructed a series of vectors with varying stabilities and produced a novel low-copy shuttle vector, pRD1.3, which can be employed for cloning and expression of large genes/gene clusters in bifidobacteria.

Chapter III outlines the investigation of conjugative functions encoded by pMP7017. While intraspecies conjugal transfer of pMP7017 had previously been demonstrated, the efficiencies reported were relatively low prompting the development of an improved protocol using a native conjugation system. The cloning and expression of a recipient-derived methylase in the donor allowed for a substantial increase in conjugation efficiency, thus revealing the negative impact of residential restriction-modification systems in the donor, while also enabling molecular investigation of pMP7017-encoded genes. Following their genetic delineation, conjugation-associated genes were demonstrated to be organized within two genetically unconnected gene clusters, designated *tra1* and *tra2*, encoding proteins which collectively constitute a ssDNA type IV secretion system as described for other characterised conjugation systems of various Gram-positive bacteria.

Chapter IV explores pMP7017-encoded cell-surface adhesions and sortase-dependent pili as potential mediators of cell-cell contact facilitating the horizontal acquisition of pMP7017. This work enabled the identification of a putative cell adhesion locus, *srtC-pilABC,* encoding a class C sortase and three pilin substrates. Due to their essentiality for the conjugal transfer of pMP7017, we propose that PilABC

may form an adhesion complex polymerised by a corresponding sortase, SrtC, that promotes adhesion to recipient cells, thus facilitating formation of the mating pair. Comparative analysis of the adhesion locus among pMP7017 homologs indicates a high degree of variability in structural organisation and pilin substrates, suggesting varying host specificity determinants.

Chapter V describes the in-depth investigation of the pMP7017 temporal transcriptome during exponential growth and assesses the impact of pMP7017-carriage on the host strain. Data generated from these analyses exposed the basic features of promoter elements which correlate with a high level of transcription, in addition to the identification of terminator and gene regulatory elements of transcriptionally active genes. Furthermore, this work permitted the identification of novel features including a highly transcribed transfer messenger RNA, which likely provides enhanced translational capacity to the host strain. Differential expression analysis revealed an upregulation of chromosomal genes involved in genetic information processing in the presence of pMP7017, while altered expression of genes involved in biofilm production and polysaccharide biosynthesis was also observed. Moreover, analyses of the synteny and variability among pMP7017 homologs indicated substantial diversity in gene organisation and accessory gene cargo which may denote diverse (co-)evolution and potential host-specific rearrangements and adaptations. In addition, mining of metagenomic data suggests the presence of pMP7017-homologs in ~10% of human faecal samples from various geographical locations reinforcing the relatively high prevalence of this horizontal transfer system and the necessity for a deeper understanding of its biology and host impact(s).

Collectively the scientific research described in these experimental chapters represents novel information on the replication and conjugative mechanisms of megaplasmid pMP7017, while providing insights into the prevalence, synteny and variability of pMP7017 and related plasmids. As it is likely that there are yet to be identified members of this megaplasmid family, which seems to be unique to the *Bifidobacterium* genus, it is important to continue to investigate the functions encoded by these megaplasmids particularly if they are involved in interactions between the human host and these common and prevalent human gut commensals.

CHAPTER I

GENERAL INTRODUCTION

The Genus *Bifidobacterium* and its Relevance to Human Health

Bifidobacteria were first isolated by Henry Tissier in 1899 from the faeces of breast-fed infants [1]. Because of their peculiar morphology these bacteria were initially assigned the species name *Bacillus bifidus* and subsequently reclassified into various other taxonomic groups throughout the mid-20th century until their designation as a separate taxon, the genus *Bifidobacterium,* in 1973 [2]. Bifidobacteria represent anaerobic, Gram-positive (G+), non-spore-forming, saccharolytic bacteria, which contain a genome with a relatively high guanine + cytosine (GC) content (ranging from 55-67%). Bifidobacterial morphology is variable, with some species typically exhibiting a bifurcated or bifid shape, while other species are pleomorphic, displaying both bifid and rod-shaped forms [3]. *Bifidobacterium* species are commonly cultivated on rich media, which would make them nutritionally fastidious, while they typically encode a plethora of glycosyl hydrolases enabling them to metabolise various glycans, such as non-digestible and complex plant and milk-derived oligo- and polysaccharides. The latter also include a number of commercial prebiotics, such as fructo-oligosaccharides (FOS) and galacto-oligosaccharides (GOS), in this context also referred to as 'bifidogenic factors' [4]–[6]. Furthermore, bifidobacteria possess a unique metabolic pathway for the degradation of both pentose and hexose sugars, referred to as the fructose six phosphate phosphoketolase pathway or bifid shunt [7].

Currently, the *Bifidobacterium* genus consists of 105 recognized species (NCBI taxonomy, accessed August 2023), which have been isolated from several environmental niches, including the mammalian and insect gastrointestinal tract (GIT), sewage, human blood, human dental caries and the vaginal mucosa. Considering that most bifidobacterial species have been isolated from faecal samples it is assumed that their natural niche is the GIT, particularly the lower gut compartments (colon and cecum) of social animals [8]. Bifidobacteria can be loosely divided into human-residential bifidobacteria (HRB) and non-HRB. While HRB species have been further subdivided into infant-type or adult-type, where *Bifidobacterium breve, Bifidobacterium longum* subsp. *infantis* and *Bifidobacterium bifidum*, or *Bifidobacterium adolescentis* and *Bifidobacterium catenulatum* are the prevailing species, respectively, members of the most prevalent bifidobacteria

(sub)species, the *B. longum* subsp. *longum* taxon, are found across all ages of the host lifespan [9], [10]. HRB constitute a dominant microbial group of gut commensals, colonizing the GIT of the neonate within the first few days following birth [11].

In addition to the establishment of a healthy microbiota [12]–[14], members of this genus have been reported to exert various health-promoting activities upon their host, such as easing of lactose intolerance [15], [16] and reduction of blood serum cholesterol [17]–[19]. Furthermore, bifidobacteria have been shown to protect against lethal enteropathogenic infection in mouse models [20] and have been linked to the alleviation of type 2 diabetes in humans, due to the production of the short chain fatty acid acetate, a by-product of their unique fermentative metabolic pathway, commonly referred to as the bifid shunt [21]. In addition, bifidobacterial abundance and prevalence have been found to be positively associated with anti-tumour immunity and cancer immunotherapy [22]–[26]. Furthermore, a growing body of evidence supports various immunomodulatory effects elicited by bifidobacteria, such as the promotion of colonic epithelial cell proliferation [27] and evocation of a tumour necrosis factor α (TNF-α) response through the production of pili [28]. More recently, exopolysaccharide (EPS) produced by *B. breve* was shown to prevent maturation of dendritic cells and activate antigen specific $CD4^+$ T cell responses in a species-specific manner [29], consistent with previous reports [30]. In addition, the EPS of *B. breve* UCC2003 has been suggested to act as a potential dietary substrate modulating early life microbiota and facilitate the reduction of apoptotic epithelial cell shedding under highly inflammatory conditions [31], [32]. Moreover, the consumption of certain bifidobacterial strains has been shown to elicit positive effects on stress-related behaviours, brain physiology and cognitive performance [33], [34]. Furthermore, administration of *B. breve* UCC2003 in pregnant, germ-free C57BL/6J mice has been demonstrated to modulate placental structure, maternal body adaptations and nutrient transporter capacity [35].

While there is a growing body of evidence in support of the presumed probiotic effect(s) of certain strains of *Bifidobacterium*, it is important to note that the majority of bifidobacterial research has been limited to a few representative bifidobacterial species/strains, hence the probiotic potential of non-representative and newly identified species/strains requires exploration as strain/species-specific probiotic activities may yet to be discovered.

In the subsequent sections of this introductory chapter, we will provide a review of the literature surrounding plasmid replication and horizontal gene transfer by conjugation in model systems of monoderms and further present an in-depth characterisation of known native plasmids of bifidobacterial origin while highlighting their importance for future biotechnological applications.

Functionalities and Genetic Organization of Plasmids

The term ‘plasmid’ was first coined in 1952 by Joshua Lederberg to describe extrachromosomal (at least in part during replication) genetic elements in bacteria [36]. However, this description was later refined to describe extrachromosomal elements which replicate autonomously, therefore excluding bacterial viruses which were encompassed within the original definition.

Plasmids are highly dynamic, modular in structure and extremely diverse in terms of their general characteristics, physical properties, composition, and gene-cargo. Naturally occurring plasmids generally range in size from 5–500 kb, although plasmids with a size of <2 kb which just encode replication functions and those with a size of >1 Mb have been reported [37]–[40]. Plasmids measuring >100 kb (an arbitrarily assigned threshold) have been assigned the descriptor ‘megaplasmid’, a term first coined in the 1980s by Rosenberg and colleagues [41], [42]. Although plasmids most commonly possess a circular geometry, linear plasmids have also been identified [43]. In addition to size and geometry, plasmids differ substantially in their copy number which refers to the number of molecules of a plasmid per chromosome contained within a cell. Plasmid copy number (PCN) is commonly maintained by negative regulatory systems that control the frequency of plasmid replication initiation [44]. Several additional factors may influence PCN, including the host strain, temperature, cellular growth rate and extracellular stresses [44]–[47]. PCN can be described as low (1–10 copies), medium (11–20 copies) or high (>20 copies) [48], [49]. While low copy number plasmids encode active partitioning mechanisms to ensure the faithful transmission of plasmids to each daughter cell during division, high copy number plasmids commonly rely on random diffusion to ensure that each daughter cell inherits at least one plasmid molecule [50].

Plasmids may encode traits which provide a competitive advantage and facilitate the persistence of plasmid-containing cells in natural populations, thus promoting rapid evolution and adaptation of the host bacterium to various environments, with plasmid-encoded antimicrobial resistance genes being a prominent example [51], [52]. Other plasmid-encoded traits which have been demonstrated to contribute to the persistence of the host cell include, among others, the production of virulence factors, resistance to metal ions, metabolic functions for nutrient utilization

such as the biodegradation of pesticides and hydrocarbons, production of bacteriocins, and the presence of restriction-modification (RM) systems or other bacteriophage resistance systems [53]–[57]. While plasmids may encode proteins with assigned functions, some plasmids encode no discernible phenotype outside of their replication (and/or mobilisation) functions and are referred to as ‘cryptic’ [58].

Replication of Circular Plasmids: An Overview

A defining feature of plasmids is their ability to replicate as autonomous genetic elements within a host cell. To enable replication, plasmids contain an essential region comprised of the origin of replication and all necessary replication functions and associated control elements, termed the minimal replicon.

Replication of plasmid DNA follows three distinct stages: initiation, elongation, and termination. The three main modes of replication that have been described for circular plasmids are (i) rolling circle replication (RCR), (ii) strand displacement and (iii) theta-type replication, which differ primarily in their strategy of initiation, utilizing plasmid-encoded replication initiation protein/s, or alternatively, RNA species to facilitate initiation and recruitment of the replicative helicase to the replication fork (discussed in detail below, and outlined in (**Table 1**) [59]. Elongation of plasmid replication is subsequently accomplished by the DNA polymerase III holoenzyme (Pol III) with participation of additional host proteins, collectively forming the replisome. Termination of plasmid replication involves the interaction of *cis*-acting terminator (*ter*) sites and *trans*-acting Ter-binding proteins, and the formation of this nucleoprotein complex antagonises the action of the replicative helicase in a polar fashion, thus arresting replication forks [60].

RCR is a common replication strategy for plasmids less than 10 kb, particularly those of G+ Eubacteria, it is a unidirectional and asymmetrical process due to its uncoupled leading and lagging strand synthesis [61], [62]. Mechanistically this mode of replication follows several distinct stages (**Figure 1**). Replication commences upon binding of the Rep protein to the double stranded origin (*dso*) of the leading strand, followed by cleavage at the flanking *nic* site by the catalytic tyrosine residue, resulting in the covalent attachment of the Rep protein to the 5' end of the DNA and release of the 3' hydroxyl group (3'OH) which acts as a primer for leading strand synthesis by DNA polymerase III. The elongation process of the leading strand causes displacement of the nicked parent strand. The replication fork continues until it reaches the *dso*, upon which Rep catalyzes a strand transfer reaction resulting in a dsDNA molecule and a ssDNA parental intermediate, a hallmark of this mode of replication [63], [64]. The single-stranded origin (*sso*) within the displaced parental

strand is then available for binding by the host initiation factor and converted into dsDNA by host DNA polymerases.

In contrast to RCR, strand displacement is a bidirectional process whereby leading and lagging strand synthesis are coupled, replication via this strategy involves the activity of three plasmid encoded proteins: a 3'->5' helicase, a primase, and an initiator, denoted RepA, RepB and RepC, respectively [65]. In this model, RepC binds its cognate iteron sequences within the origin, promoting DNA bending which facilitates melting of the flanking AT-rich region [66], [67] (**Figure 2**). Unwinding of the DNA is catalyzed by RepA assembly around the exposed ssDNA allowing the adjacent palindromic sequences within the unwound region to form stem-loop structures. Exposure of these hairpins induce RepB assembly and priming at the origin [68], [69]. As replication via this method is continuous the complementary strand becomes displaced, replication of this ssDNA is initiated at the origin independent of host factors. The best described example of this model is the promiscuous IncQ plasmid family, and a thoroughly studied member of this family is the broad-host-range plasmid RSF1010 [67].

Replication via theta-mode has been extensively studied in G- bacteria, yet is less well characterized in G+ bacteria, with the majority of characterized replicons of G+ bacteria being of streptococcal and lactococcal origin [70]. DNA synthesis through this mechanism resembles that of chromosomal DNA replication as both strands are replicated in concert, with continuous leading and discontinuous lagging strand synthesis. Two distinct modalities of initiation occur in theta replicons as duplex melting can be mediated by plasmid-encoded Rep proteins or rely on transcription through the origin. Based on this theta replicating plasmids can be sub-divided into four classes A-D; all classes, excluding class B, encode Rep proteins (**Table 1**) [71]. Class B plasmids instead depend on the successful formation of an R-loop structure, a DNA-RNA stem-loop attained through the hybridization of the 3' terminus of a constitutively transcribed ~600 bp pre-primer (RNA II) in complex with the 5' end of the lagging strand template DNA [72]. Recognition of the poly-A motif in RNA II by endonuclease RNase H results in cleavage at the DNA-RNA junction, releasing the 3' hydroxyl group for extension by PolC on the leading strand. Class A theta replicating plasmids depend on binding of *trans*-acting Rep proteins to repeat or reiterated sequences, termed iterons, within the plasmid origin (**Figure 3**). The replicases of class C differ from that of class A due to their additional primase activity, plasmids

encoding this type of Rep protein include that of ColE2 and ColE3 from *E. coli* [73]. Plasmids categorized as class D encompass features of both modes of initiation, these plasmids require binding of self-encoded replicative initiators in addition to transcription through the origin and are represented by large, low-copy, broad-range plasmids from *Streptococcus* species.

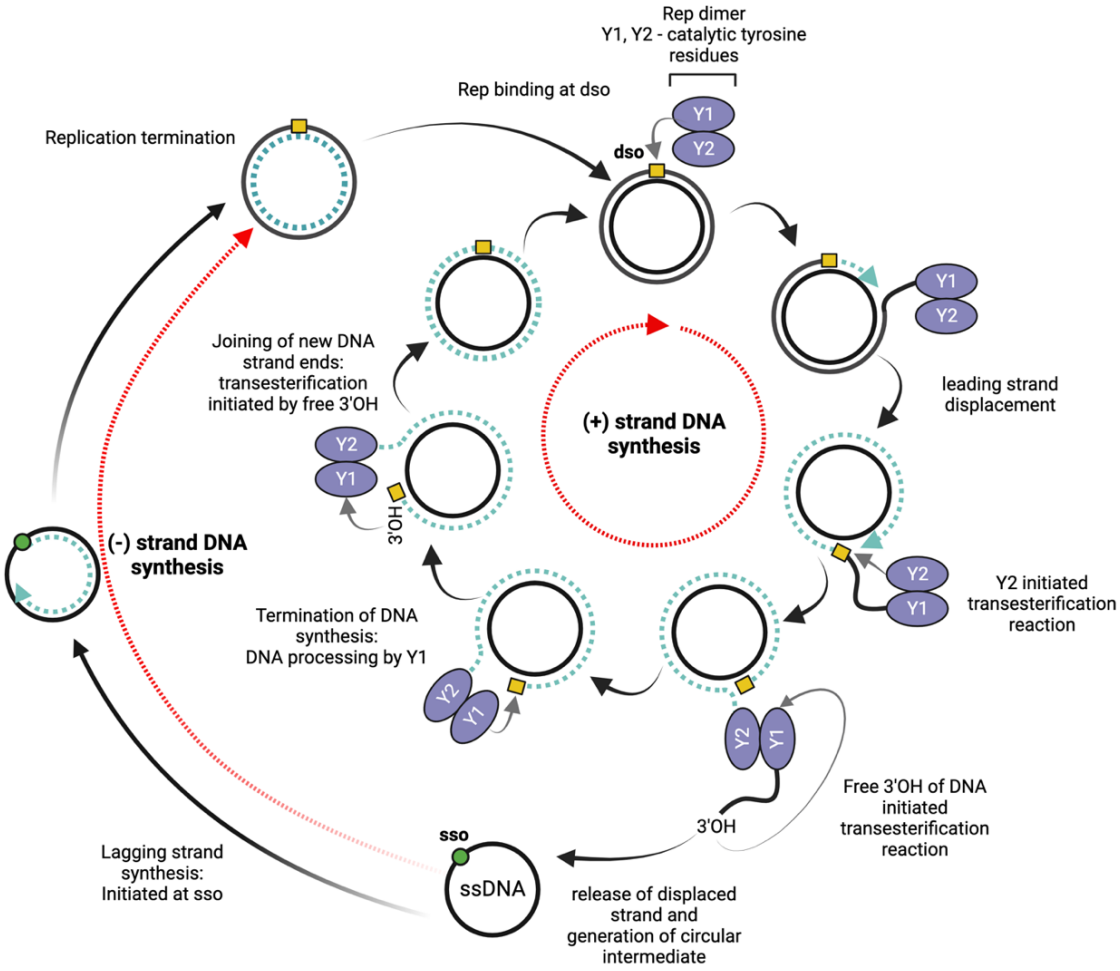


Figure 1. Mechanism of rolling-circle replication

Separated cycles of leading (+) and lagging (-) DNA strand synthesis are indicated by red dotted arrows, initiated at the *dso* and *sso* origins, respectively. The RCR initiation protein, Rep (Purple), initiates (+) strand replication at the nick site (yellow square). Dashed pale blue lines indicate the nascent DNA strand with arrow heads representing the direction of synthesis. Grey arrows represent nucleophilic attacks (transesterification reactions), of either the catalytic tyrosine residue (Y1/Y2) of the Rep protein on the phosphodiester bond in the DNA or the free 3'OH end of the DNA on the phospho-tyrosine bond between the Rep protein and the DNA. The steps of synthesis are additionally described in text and outlined in the figure. Adapted from [74].

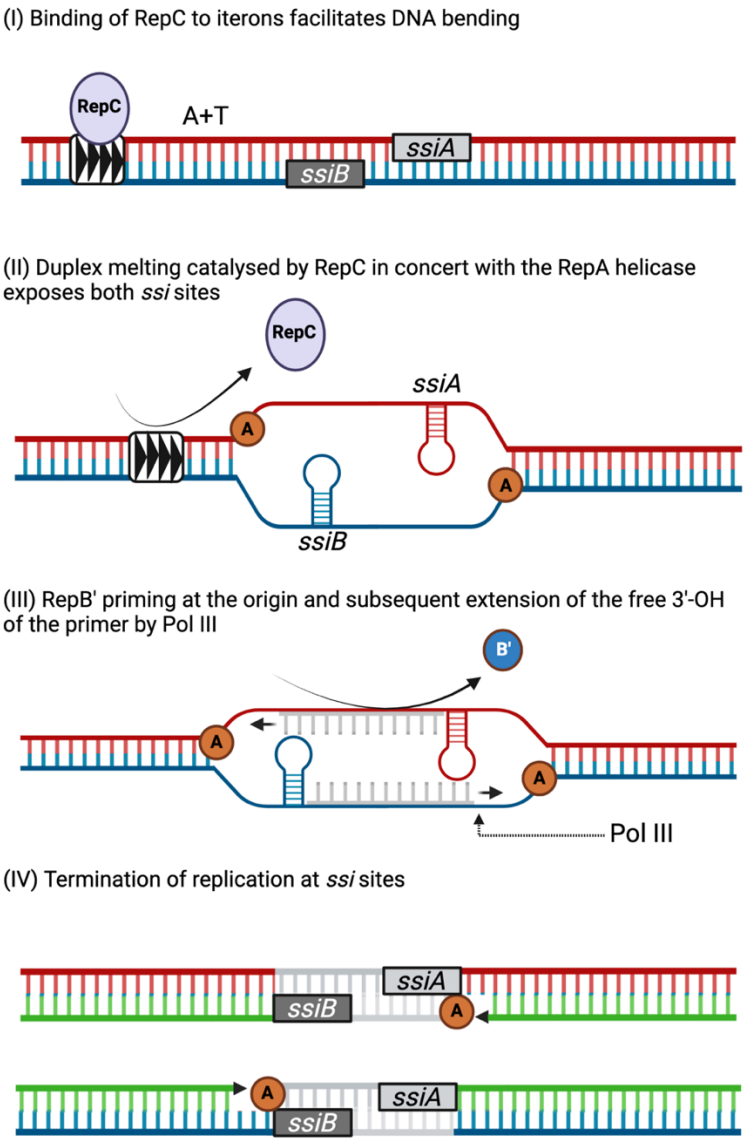


Figure 2. Mechanism of strand displacement replication initiation (plasmid RSF1010)

Replication occurs independently and with opposite polarities from origins *ssiA* and *ssiB*. (I) Interactions between the plasmid-encoded initiator of replication, RepC (purple) and it's cognate iterons induces DNA bending which facilitates melting of the AT-rich region. (II) RepA (orange) catalyses unwinding of the DNA and assembles around the exposed ssDNA. (III) Following D-loop formation RepB (blue circles) assemble and catalyse priming at the origins (grey strands). (IV) Green strands indicate newly synthesized DNA, and the direction of synthesis indicated by black arrows/arrowheads. The steps of synthesis are further described in text and outlined in the figure.

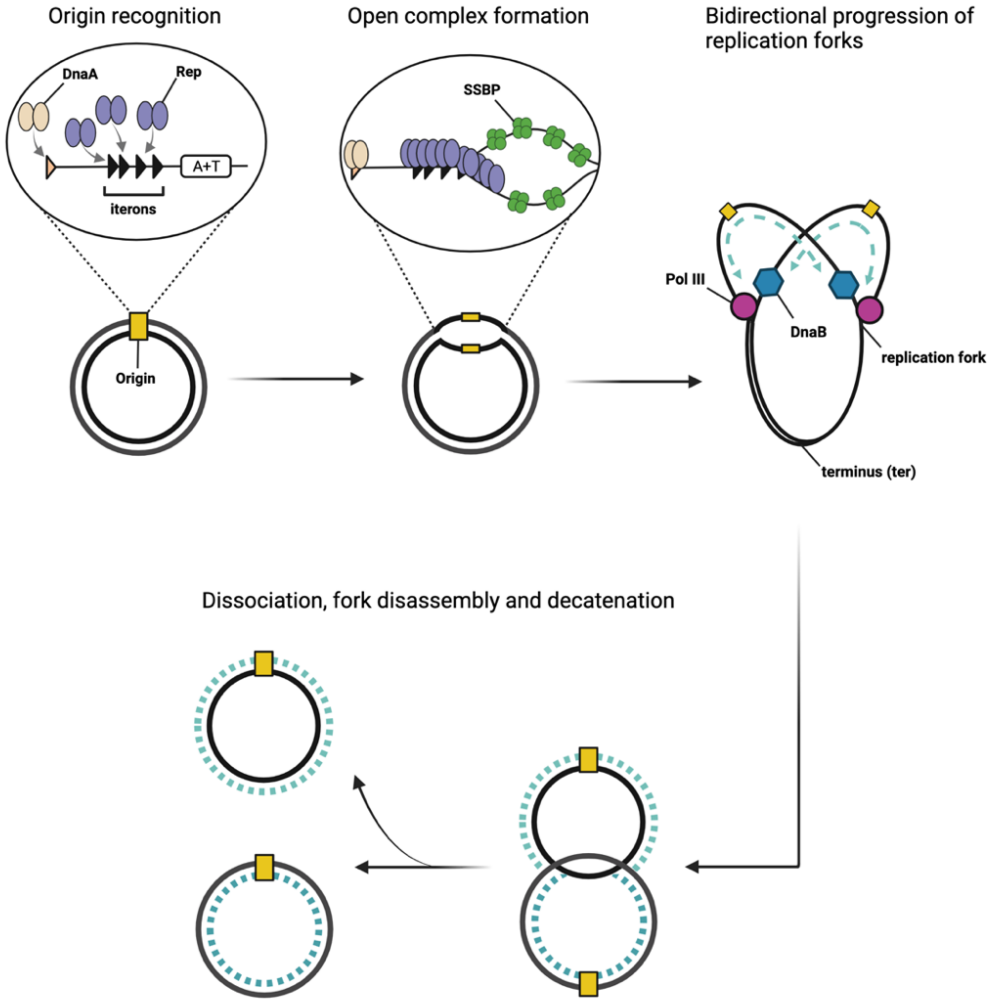


Figure 3. Model of replication initiation in class A theta plasmids

During plasmid origin recognition (stage 1), the replication initiation proteins (Rep) (purple) cooperatively bind iterons (black arrow heads) until saturation. The formation of an open complex (stage 2) involves unwinding of the DNA unwinding element (DUE) (represented as a white box containing 'A+T') followed by continued binding of Rep to the bottom strand of the ssDNA and binding of tetramers of the single-stranded binding protein (SSB) (green circles) to the ssDNA. Host-encoded DnaA (pale orange) enhances/stabilizes the formation of the Rep-mediated open complex through binding of the adjacent DnaA box (orange arrowhead) and assists with the recruitment of the DnaB helicase (blue hexagons). The bi-directional replication forks continue to progress as Pol III duplicates plasmid DNA until reaching the termination site (*ter*).

Table 1. Comparison of the three main modes of plasmid replication initiation described for circular plasmids

	Continuous	Initiation factor	Initiation factor binding site	Primer	Helicase
	Leading strand synthesis				
RC	Y	Rep (binding, nicking)	*dso* (hairpin cruciform)	Rep-generated nick (3'-OH)	DnaB (PcrA)
SD	Y	RepC (DnaA-like)	*ssi* sites (hairpins)	RepB' (at hairpin)	RepA
Theta A	Y	Rep	Multiple iterons (induced bending)	DnaG primase (host)	DnaB (PcrA)
Theta B	Y	N/A	N/A	RNAse H processed transcript	PriA + DnaB (PcrA)
Theta C	Y	Rep (also primase activity)	Iterons	ppApGpA (produced by Rep)	PriA + DnaB (PcrA)
Theta D	Y	Rep (duplex melting, primer generation)	No iterons (No bending)	Rep transcript	PriA + DnaB (PcrA)
	Lagging strand synthesis				
RC	N	host factors	*sso* (host-recognized)	DnaG	N/A
SD	Y	N/A	N/A	RepB'	RepA (+ Pol III)
Theta A	N	N/A	N/A	DnaG	PriA+ DnaB (PcrA)
Theta B	N	N/A	N/A	DnaG	PriA+ DnaB (PcrA)
Theta C	N	N/A	N/A	DnaG	PriA+ DnaB (PcrA)
Theta D	N	N/A	N/A	DnaG	PriA+ DnaB (PcrA)

Including the differentiating factors of each of the four classes, A-D, of theta mode. RC = rolling circle replication, SD = strand displacement replication.

Mechanisms of Vertical Plasmid Propagation: Plasmid Stability and Maintenance

As expression of plasmid-specific genes may solely confer a selective advantage in certain environments, plasmids must ensure the faithful transmission of plasmid copies to daughter cells during cell division to prevent plasmid loss when conditions do not necessarily favour retention of the plasmid by the host bacterium. Strategies adopted by plasmids for their vertical propagation can be divided into three categories; random segregation, active partitioning systems and post-segregational killing (PSK) (**Figure 4**).

As mentioned above, high copy plasmids tend to be small in size, and genes encoding active partition systems or PSK mechanisms are commonly absent. Therefore, these plasmids rely on random segregation for their vertical propagation. This strategy assumes the free diffusion of plasmid copies throughout the cytoplasm leading to random segregation during cell division. Random segregation of plasmid copies is facilitated by multimer resolution systems (MMRS) which use site-specific DNA recombination to resolve plasmid multimers. Plasmid multimers have been demonstrated to significantly hinder the equipartitioning of high copy plasmids resulting in a high rate of plasmid loss, thus the conversion of multimers into monomeric substrates by MMRS maximizes the number of plasmid molecules for segregation, ensuring proper partitioning of high copy plasmids [75], [76].

While plasmids with a low copy number exert less metabolic burden on their host, they subsequently incur a higher probability of loss resulting from mis-segregation during cell division, and thus have evolved specific mechanisms to ensure transfer to both daughter cells at the end of a cell division event.

The most thoroughly studied partition (*par*) systems originate from plasmids of Gram-negative (G-) bacteria. *par* systems constitute a centromere-like *cis*-acting region of DNA (partitioning site), *parS* and two *trans*-acting proteins, a centromere-binding protein and a nucleoside triphosphate hydrolase (NTPase), represented by ParB and ParA, respectively. ParB proteins bind the *parS* site with high affinity with multiple ParB-ParB interaction interfaces resulting in the formation of higher-order complexes [77]. ParA then binds the ParB-DNA complex, forming a nucleoprotein

partitioning complex also called the segrosome. ParA supplies energy to the scaffold which moves the two sister molecules in opposite directions toward each cell pole. Partitioning systems can be classified into three types (four at the subgroup level), Ia, Ib, II and III. All currently characterised type Ia *par* loci are from plasmids of G- origin, while type I *par* loci from plasmids of G+ origin are exclusively of the Ib type [78]. Although the general mechanism of action (as outlined above) is similar for all *par* types, they are differentiated based on the size of the Par proteins, structure of the NTPase and the genetic organisation at the *par* locus [78], [79].

Type I *par* loci encode Walker-type ATPases with subtype Ia loci encoding large ParA (251–420 aa) and ParB (182–336 aa) proteins and a *parS* site located downstream of *parB*, while Ib loci typically encode small ParA (208–227 aa) and ParB (46–113 aa) proteins and a *parS* site upstream of the *parAB* operon [80]. The best characterised examples of type Ib Par systems of G+ bacteria include the Inc18 plasmid pSM19035 from *Streptococcus pyogenes* and pAD1 from *Enterococcus faecalis* (**Figure 5**) [81], [82].

Type II partitioning systems encode actin-like ATPases such as that of the 46.4 kb conjugative staphylococcal plasmid pSK41 (**Figure 5**). The pSK41 *par* gene cluster encodes a centromere-binding protein, ParR, that binds to a series of eight 10 bp sequences representing the *parS* site. The *parS* site is located directly upstream of the *parM* gene which encodes the actin-like NTPase ParM. In a coordinated action with ParR, ParM produces actin filaments in the presence of ATP which facilitates accurate plasmid segregation [83]. The *alfAB* partitioning operon of the 70 kb cryptic plasmid pLS32 from *Bacillus subtilis* is another well characterised example of a type II partitioning system [84].

Type III partitioning systems encode tubulin-like GTPases, an example of this system in plasmids of G+ origin includes the *tubRZ* partitioning system of plasmid pBtoxis from *Bacillus thuringiensis* subsp. *israeliensis* (**Figure 5**). The *tubRZ* partitioning cassette encodes the centromere-binding protein, TubR, and tubulin-like GTPase, TubZ. The TubZ GTPase binds the TubR-DNA nucleoprotein complex and assembles into polymers that migrate the plasmids to the two opposite cell poles [85]. Binding of TubR to the centromere-like region upstream of the *tubRZ* operon additionally serves in the autoregulation of its own expression.

PSK partitioning systems, which sometimes are also referred to as addiction systems or toxin–antitoxin (TA) systems, are almost exclusively encoded by large,

low copy conjugative plasmids. PSK systems are typically composed of a protein toxin and an antitoxin which differ significantly in their half-life; the stable protein toxin is long-lasting, while the labile antitoxin is degraded rapidly. This means that in plasmid-free cells which do not anymore produce TA proteins *de novo*, the labile antitoxin disappears from any pre-existing TA complex. Once the toxin is liberated from such a TA complex, it selectively kills the plasmid-free daughter-cell [86], [87]. Although the killing of plasmid-free daughter cells does not effectively increase the frequency of a plasmid's vertical inheritance but ensures that the majority of cells in a population remain plasmid-containing, hence the term "addiction". TA systems are classified based on the type (RNA/protein) and mode of inhibition of the antitoxin. Currently, eight TA types exist; type-I, -III and -VIII antitoxins represent small noncoding RNAs, while type-II, -IV, -V, -VI and -VII TA antitoxins are small proteins [88]. Most toxins are proteinaceous affecting cellular processes such as translation or DNA replication, however, other intracellular molecular targets such as the biosynthesis of cell membranes, biofilm and cell morphology have also been described (**Table 2**).

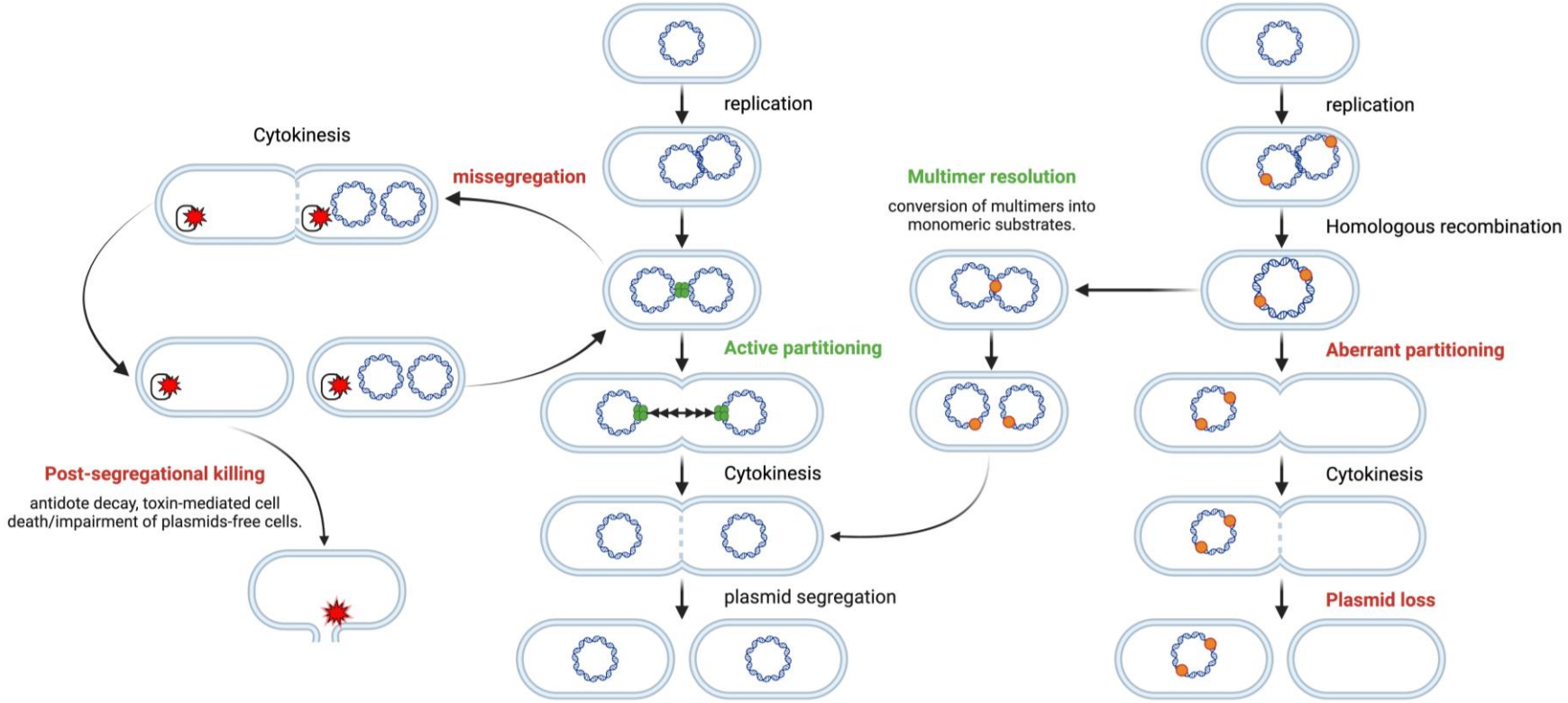


Figure 4. The contribution of stability mechanisms to plasmid maintenance

Active partitioning systems (centre) position replicated plasmids appropriately within the elongating bacterial cell such that each daughter cell receives a plasmid copy after cell division. If homologous recombination (right) dimerizes two plasmid copies (recombination sites, orange circles), the plasmids cannot be distributed equally at cell division which may lead to aberrant partitioning and subsequent plasmid loss. However, multimer resolution systems may resolve plasmid dimers to monomers which may be partitioned accurately by active partition systems or indeed by random diffusion (hcn plasmids). Plasmid-free cells arising from mis-segregation or defective replication (left) can be killed or growth impaired due to PSK mechanisms. While the plasmid-encoded toxin (red star) is sequestered by the antitoxin (white square) in plasmid-containing cells, in plasmid-free

derivatives, the labile antitoxin is more quickly degraded by host enzymes, liberating the toxin which exerts its effects on the intracellular molecular target. Green and red text indicate productive and non-productive steps respectively, in accurate plasmid segregation. The host chromosome is not depicted in the figure for the sake of clarity. Adapted from [89].

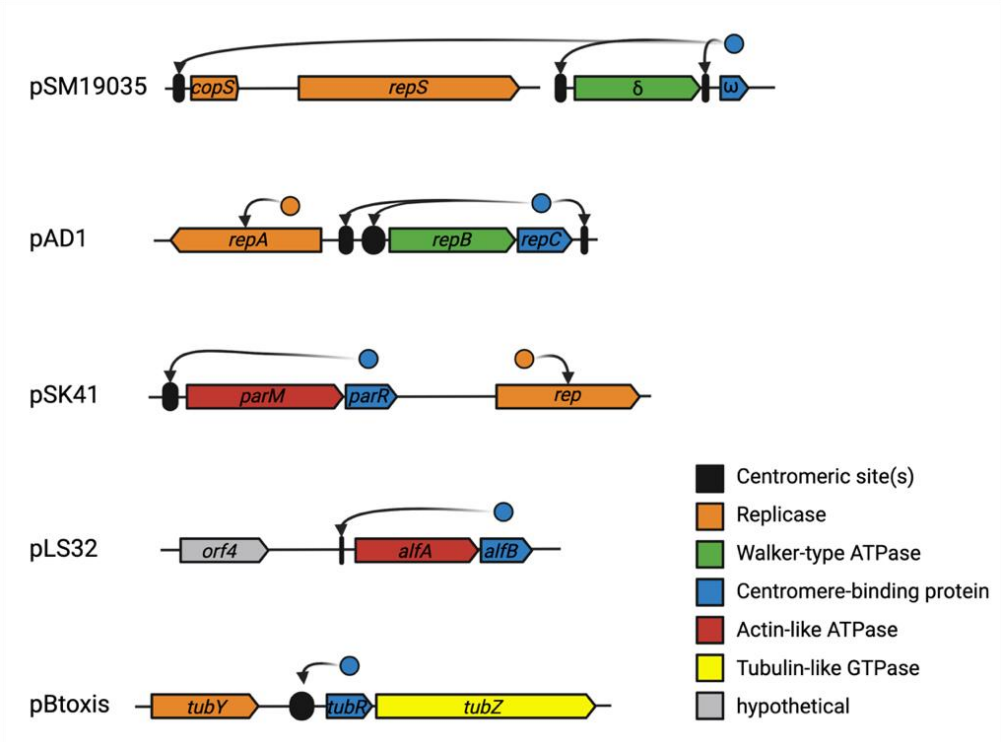


Figure 5. Graphical representation of the genetic organisation of partitioning modules from various plasmids of gram positive bacteria

Type Ib partitioning systems of plasmids pSM19035 and pAD1. Type II systems of plasmids, pKS41 and pLS32. Type III system of pBtoxis. Walker-type ATPases are shown in green, actin-like ATPase are shown in red, tubulin-like GTPase in yellow and centromere binding proteins in blue. Genes encoding replication proteins are also displayed, shown in orange. The location of the cognate centromeric sequences to which the centromere binding proteins bind are depicted as black curved squares on the DNA line and arrows are included for clarity.

Table 2. Features of the eight classes of TA systems

Adapted from [88]

TA type	Toxin	Antitoxin	Toxin targets	Affected cellular processes	Example	References
I	Protein	RNA	Bacterial membrane	Biosynthesis of cell membrane	*hoh-sok*	[90]
					tisB-istR	[91]
II	Protein	Protein	DNA gyrase	DNA replication	*ccdAB*	[92]
			Sequence-specific mRNA	Translation	*parDE*	[93]
			Initiator tRNA	Translation	*mazEF*	[94]
			EF-Tu elongation factor	Translation	*phd-doc*	[95]
			glu-tRNA synthetase	Translation	*hipBA*	[96]
III	Protein	RNA	mRNA	Translation	*toxIN*	[97]
IV	Protein	Protein	MreB and FtsZ	Cell morphology	*cbtA-cbeA*	[98]
V	RNA	Protein	Cell membrane	Biosynthesis of cell membrane	*ghoST*	[99]
VI	Protein	Protein	β-sliding clamp	DNA replication	*socAB*	[100]
VII	Protein	Protein	Biofilm	Biofilm	*hha-tomB*	[101]
VIII	RNA	RNA	Repression of YhcB	Cell morphology	*sdsR-ryeA*	[102]

Phylogeny and Genetic Diversity of Plasmids from *Bifidobacterium*

Comparative genomics and pan-genome oriented studies have been fundamental in the expansion of our understanding of a wide range of biological functions encoded by bifidobacteria. These analyses have also significantly contributed to the identification of synteny regions and species- or strain-specific genes [103], [104]. Furthermore, information was generated concerning the mobilome, which represents genes encompassing mobile genetic elements (MGEs), plasmids, insertion elements (IS elements) and prophages [105]. Plasmids were once considered an atypical feature of bifidobacteria, and those identified in members of this genus were shown to be predominantly small 'cryptic' plasmids (whose encoded functions outside of replication are unknown), until the discovery of a 190,178 bp conjugative megaplasmid in *B. breve* JCM7017, designated pMP7017 [106]. Since then pMP7017-homologs have been identified in several strains of *B. longum* subsp. *longum* [10].

Currently (NCBI, accessed August 2023) forty-nine plasmids have been identified in nine species of *Bifidobacterium*: *B. asteroides*, *B. bifidum*, *B. breve*, *B. catenulatum*, *B. choerinum*, *B. longum*, *B. pseudocatenulatum*, *B. pseudolongum* and *B. pullorum*. In most cases individual strains harbour a single plasmid, with the exception of six strains of (sub)species *B. longum*, one strain of *B. asteroides* and one strain of *B. catenulatum* subsp. *kashiwanohense*, in which two distinct plasmids have been identified (**Table 3**). The majority (65.3%) of bifidobacterial plasmids have been isolated from strains belonging to *B. longum* (**Figure 6A**), and as such it has been inferred that plasmids are more prevalent among this species than any other *Bifidobacterium* species. However, when comparing the number of plasmids to the number of fully sequenced genomes per *Bifidobacterium* species, a moderate degree of correlation (r_s = 0.522) was observed (**Figure 6B**). While the distribution of bifidobacterial plasmids may be species-dependent, it is likely that the distribution of plasmids is biased towards the (sub)species *B. longum* due to the overrepresentation of fully sequenced genomes for this species.

All currently known plasmids identified in the genus *Bifidobacterium* are circular in their topology, with a size range between 1,847 bp for pMB1 to 198,121 bp

for the recently sequenced megaplasmid, P1. The median sequence length of bifidobacterial plasmids is 4,488 bp (**Figure 6C**) and the G+C content ranges from 41.7% (pBIF10) to 66.3% (pRY68), with a mean of 60.73%, which is comparable to that calculated for *Bifidobacterium* genomes (61.02%) (**Figure 6D**).

To broaden our view of the diversity of plasmids present in *Bifidobacterium*, phylogenetic analysis of their replication (Rep) initiator proteins was undertaken, which revealed that *Bifidobacterium* plasmids can be grouped into five distinct clades (**Figure 7**), consistent with previous phylogenetic studies on bifidobacterial plasmids [107]. The tree highlights considerable variation in divergence between and within clades I, II, III and IV, while limited phylogenetic divergence is observed between members of the V^a, V^c and V^d subclades of predicted Rolling Circle (RC) replicases.

Clade I plasmids encode a single Rep-like protein containing a Viral_rep superfamily domain (cl03577). This family of Rep-like proteins may have originated through interspecies recombination and are represented in viral, bacterial, plasmid and parasitic protozoan genomes. While Rep_{p4M} does not display homology to other plasmid replicases, it is closely related to the Rep initiator protein of circovirus ($E = 7 \times 10^{-15}$) which replicates by a RC mechanism [108].

Clade II Reps can be collectively defined by their Rep_1 super family (pfam01446) conserved domain, and due to an absence of genes encoding Mob protein homologs their cognate plasmids are presumed to lack mobility. Clade II plasmids range in size from 4,815 - 5,750 bp and include pB21a, pCIBb1, BIF195_P1 and pNBb1 from *B. breve* and pASV479 from *B. pseudolongum* subsp. *globosum*. While no experimental validation of replication mode exists for plasmids pB21a, BIF195_P1 or pNBb1, pCIBb1 has been proposed to replicate via a RC mechanism due to the presence of single-stranded (ssDNA) intermediates in cell lysates which were sensitive to S1 nuclease degradation [109]. Furthermore, each identified replication protein within this clade possesses a pHuHuuu motif, believed to be involved in metal binding, and an N-terminal Cx_5CPxC motif, both of which are also present in all characterised Rep proteins of *Streptomyces* RCR plasmids [109]–[111].

Clade III Rep proteins can be found in plasmids originating from a variety of species, including pBC1 from *B. catenulatum*, pFI2576, pDOJH10S and pMB1 from *B. longum*, pBC from *B. choerinum* and pBBKW-2 from *B. catenulatum* subsp. *kashiwanohense*. All clade III replicases contain an N-terminal Replicase superfamily

(RepA) domain (pfam03090) and a C-terminal Primase domain (pfam08708). Plasmids pBC1, pFI2576, pDOJH10S and pBBKW-2 have been predicted to replicate via the theta-type replication mode due to the absence of ssDNA intermediates in S1 nuclease analyses in addition to the lack of features consistent with the RC mode of replication [112]–[115]. pMB1 was the first bifidobacterial plasmid to be analysed at a molecular level, and while the replication mode for this plasmid has not yet been established, it is considered to replicate via a theta-type mechanism due to the homology of its Rep protein to Reps of characterised theta-type replicating plasmids [116], [117]. While plasmids, pBC1, pMB1, pFI2576 and pBBKW-2 do not appear to be mobilizable, plasmids pDOJH10S and pBC encode MobA_MobL (pfam03389) and NicK super family (COG2946) Mob proteins, respectively.

Clade IV replicases share a common Rep_3 super family domain (pfam04796). This clade is divided into two sister monophyletic sub-clades, designated IV^{a} and IV^{b}. Rep proteins from the characterised conjugative megaplasmid pMP7017 from *B. breve* JCM7017 and megaplasmid P1 from *B. longum* I2-2-3 form the sub-clade IV^{a} [106]. Both plasmids are complex encoding two predicted replicases, denoted RepA and RepB. Molecular analysis of the replication functions of pMP7017 identified RepB as the replicative initiator of this plasmid as the *repA* gene was discovered to be non-essential [118]. Furthermore, the minimal replicon containing the *repB* gene and downstream iteron-containing region were determined to be essential for efficient replication, while an *E. coli-Bifidobacterium* shuttle vector based on the pMP7017 replicon was demonstrated to facilitate autonomous replication in two bifidobacterial species [118]. Due to (i) the large size of these plasmids, (ii) organisation of the essential region of replication, and (iii) homology of the replicases to characterised theta-type Rep proteins, a theta mechanism of replication is assumed for these plasmids.

Subclade IV^{b} replicases originate from plasmids pAP1 and pCIBAO89 of *B. asteroides* DSM20089, pBBKW-1 of *B. catenulatum* subsp. *kashiwanohense* JCM 15439, pNAC3 of *B. longum* RW041, BLNIAS_P1 of *B. longum* subsp. *longum* and p1CACC514 of *B. pullorum* subsp. *gallinarum*, and range in size from 2,111 bp (pCIBAO89) – 10,224 bp (pNAC3). While the mechanism of replication for these six plasmids has not yet been experimentally validated, it is speculated that they replicate via a theta mechanism based on the in-depth molecular characterization of plasmids pCIBAO89 and pNAC3 [119], [120].

Clade V is the largest monophyletic group containing 32 Rep proteins from 30 plasmids, that share a conserved Rep_3 domain (PF01051). 84.4% of plasmids identified in the *B. longum* (sub)species fall within this clade. Seven plasmids pSP02, pDOJH10L, pMG1, pKJ50, pKJ36, pNAC1 and pNAL8L have been demonstrated to accumulate ssDNA intermediates, an observation consistent with a RCR mode of replication [109], [112], [120]–[124]. Four subclades of higher similarity can be distinguished within this clade, designated as V^a-V^d.

V^a replicases derive almost exclusively from plasmids of the *B. longum* (sub)species, except for replicase 2 (label: pBBKW-1 (2), WP008783726) from complex plasmid pBBKW-1 of *B. catenulatum* subsp. *kashiwanohense.* V^a plasmids possess an average GC % of 65.1%, which is much higher than that of any other (sub)clade. Additionally, each of these plasmids may be considered mobilizable as they each encode a MobQ-type mobilization protein characterised by a conserved MobA_MobL (pfam03389) domain.

Two replicases are present in the V^b subclade, originating from plasmids pBIF10 and an unnamed complex plasmid from *B. pseudocatenulatum* YIT11953 (YIT11953_unnamed). Plasmid pBIF10 was previously indicated (**Figure 1D**) to present with a GC content which is 18.3% lower than its *Bifidobacterium* host. It was therefore speculated that this plasmid had recently been acquired from the G- obligate anaerobic bacteria *Bacteroides*, substantiated by the presences of large regions of homology within the plasmid to chromosomal regions of intestinal strains of this genus [107].

V^c replicases derive predominantly from small non-mobilizable cryptic plasmids of the *B. longum* (sub)species, with an average plasmid size of 3,648.7 bp ± 235.4 bp, in addition to one of the two identified replicases (rep (2)) from the unnamed complex plasmid from *B. pseudocatenulatum* YIT11953 (YIT11953_unnamed (2)).

V^d replicases are encoded by plasmids from *B. bifidum* B80 (pB80) and *B. breve* BR3 (pBR3) in addition to the *B. longum* (sub)species, these plasmids encode predicted mobilization functions and are larger than their V^c counterparts with a plasmid size range of 4,891 bp (pBR3) – 10,073 bp (pDOJH10L).

To evaluate the coherence of the phylogenetic grouping based on the encoded replicase(s) of native bifidobacterial plasmids, we performed pair-wise analyses of the identity between plasmid sequences (**Figure 8**). The clustering of plasmids is largely in agreement with the Rep-protein phylogeny of Figure 7 and highlights the limited

accrual of variation among RC replicating plasmids of the five (sub)clade(s), while illustrating a large degree of variation in genetic diversity among plasmids replicating by a theta mechanism. The current analysis shows inter-plasmid similarities in the range of 76.1–99.9% for members of subclade V^a, containing six near-identical plasmids originating from the (sub)species *B. longum* (i.e., plasmids pKJ36, p2-2B, p72B, pB44, p1-6B2 and p157F-NC2; >98.6% identity, 3,624.2 bp ± 0.41 plasmid length, GC 65.1% ± 0.1). Inter-plasmid identity values for subclade V^d plasmids range from 57.6 to 99.9% identity; eight plasmids within this sub-clade, i.e., pKJ50, pBIFA24, pNAL8M, p6043A, p157F-NC1 and pSP02, found in strains of the (sub)species *B. longum* and, pB80 and pBR3 found in strains of *B. breve,* are near identical (>99.4% identity, 4,904.75 bp ± 23.07 plasmid length, GC 61.9% ± 0.05).

Furthermore, based on the aligned genome fraction and genome to length ratio, clear evidence of cointegration is observed between plasmids within the V clade, consistent with previous reports [113], [125]. Plasmid pBBKW-1 (V^a) isolated from *B. catenulatum* subsp. *kashiwanohense* appears to be a fusion of two copies of the *B. longum* plasmid, pMG1 (V^a). Plasmid pDOJH10L (V^c) appears to be a cointegrate of two *B. longum* plasmids, pKJ50 (V^d) and pNAC2 (V^a), while plasmid p1-6B1 (V^a) appears to constitute a fusion of three units and is homologous to plasmids pNAL8M (V^d), pNAC1 (V^c) and pFI2576 (III). As expected, plasmid pBIF10 (V^b), presumed to originate from *Bacteroides* spp., shows no regions of homology with other *Bifidobacterium* plasmids. Like pBIF10, p4M (I) is unique among known plasmids of bifidobacterial origin.

Outside of their region of replication, no similarity is observed for clade II plasmids pB21a or pASV479 with any other plasmid, while plasmids, pCIBb1 and BIF195_p1 found in strains of *B. breve* are near identical (99.9% identity, 5,749.5 bp ± 0.71 plasmid length, GC 57.0% ± 0.07).

Clade IV plasmids, pAP1 and pCIBAO89 both originating from *B. asteroides* DSM20089 are highly related (98.8% identity, 2,125.5 bp ± 20.5, GC 52.2% ± 0.2). Although plasmid size and intergenomic similarity between the remaining clade IV plasmids differ substantially, the presence of homologous replicons suggests that they are probably derived from the same ancestral plasmid and have evolved divergently in different host backgrounds. Interestingly, clade IV^b megaplasmids pMP7017 and

P1 share 46.3% sequence identity across 60% of their genomes indicating their relatedness.

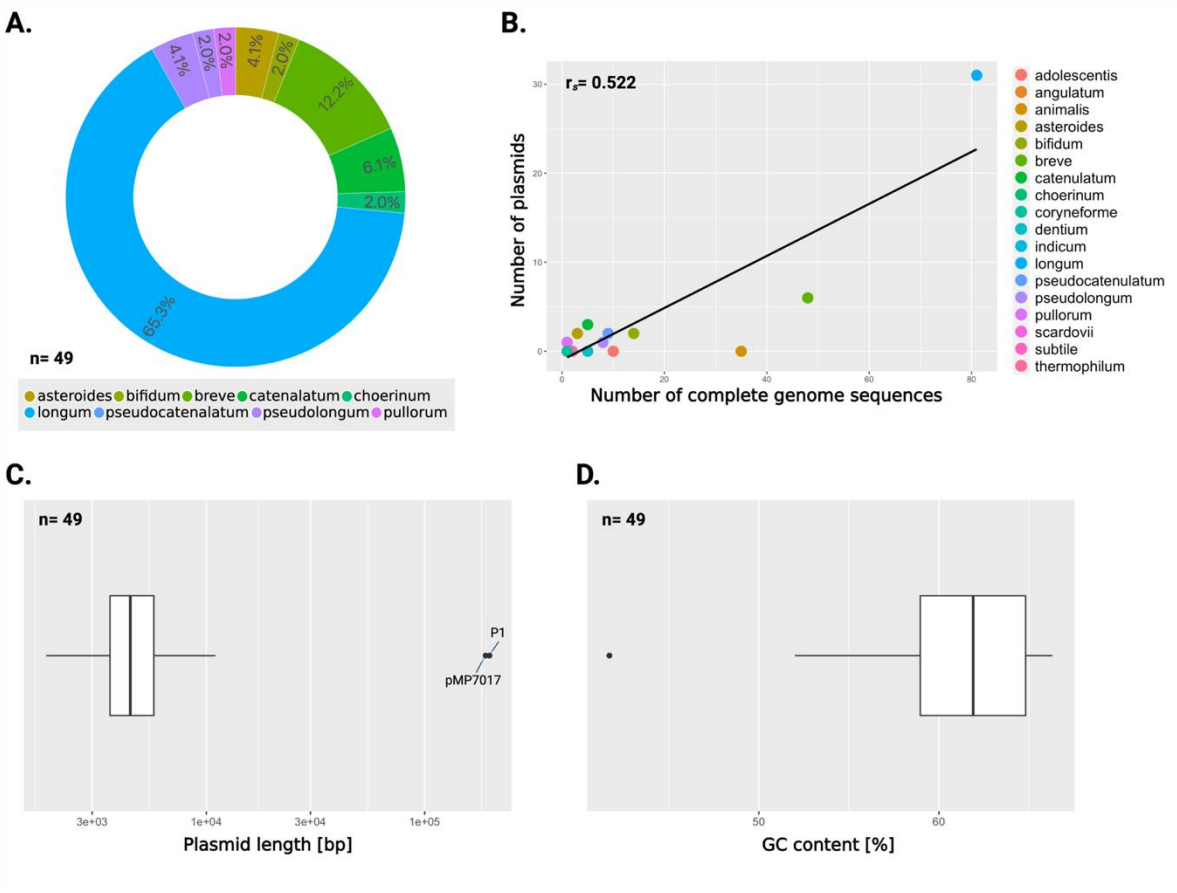


Figure 6. Species distribution and general characteristics of plasmids from *Bifidobacterium*

A. Pie chart representing the distribution of natural bifidobacterial plasmids by species. **B**. Scatterplot displaying the relationship between the number of complete genome sequences available on the NCBI database using a dataset pertaining to eighteen species (including subspecies) with the highest number of complete genome sequences (ranking 1-11) and the number of plasmids originating from each species. As the data within both variables was positively skewed a linear relationship could not be assumed, as such, the nonparametric Spearman Rank Correlation coefficient (r_s) was employed to evaluate the monotonic relationship between the ranked values, calculated at 0.522, suggesting a moderate degree of correlation (based on grading standards). **C**. Boxplot illustrating the variance of plasmid length in base pairs [bp], of the normal distribution or plasmid locality, spread and skewness groups of numerical data through their quartiles. **D**. Boxplot illustrating the G+C content [%] of bifidobacterial plasmids, the data is marginally negatively skewed but symmetrically distributed within the inter-quartile range.

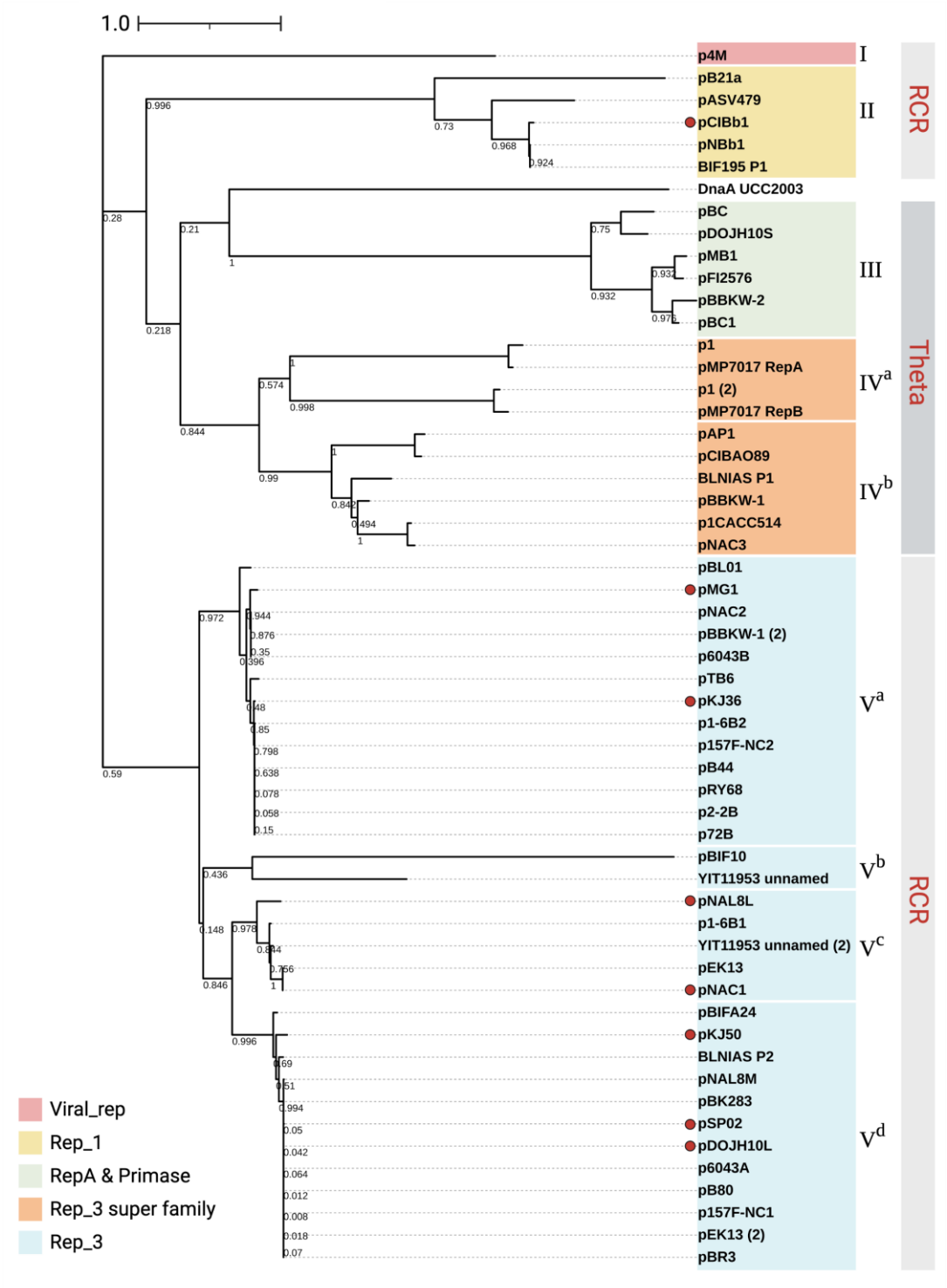


Figure 7. Phylogenetic relatedness among plasmid replicases identified in *Bifidobacterium*

A phylogenetic tree was generated from 54 replicase amino acid sequences originating from 49 naturally occurring plasmids of *Bifidobacterium*. Branch lengths represent evolutionary divergence (black lines) and bootstrap values are displayed at the base of each branch to show robustness of tree topology (500 replicates). Scale bar (top), refers to a phylogenetic distance

of 1.0 amino acid substitutions per site. Plasmid labels are lined up for ease of comparison (dashed grey lines). Colour underlay across labels represent conserved domains; viral_rep cl03577 (red), Rep_1 super family Pfam01446 (yellow), Replicase superfamily (RepA) domain Pfam03090 and C-terminal Primase C terminal 1 Pfam08708 (green), Rep_3 super family domain Pfam04796 (orange) and Rep_3 domain Pfam01051 (blue). Red circles to the left of labels indicates experimental evidence of RC replication, due to the detection of ssDNA intermediates in cell lysates.

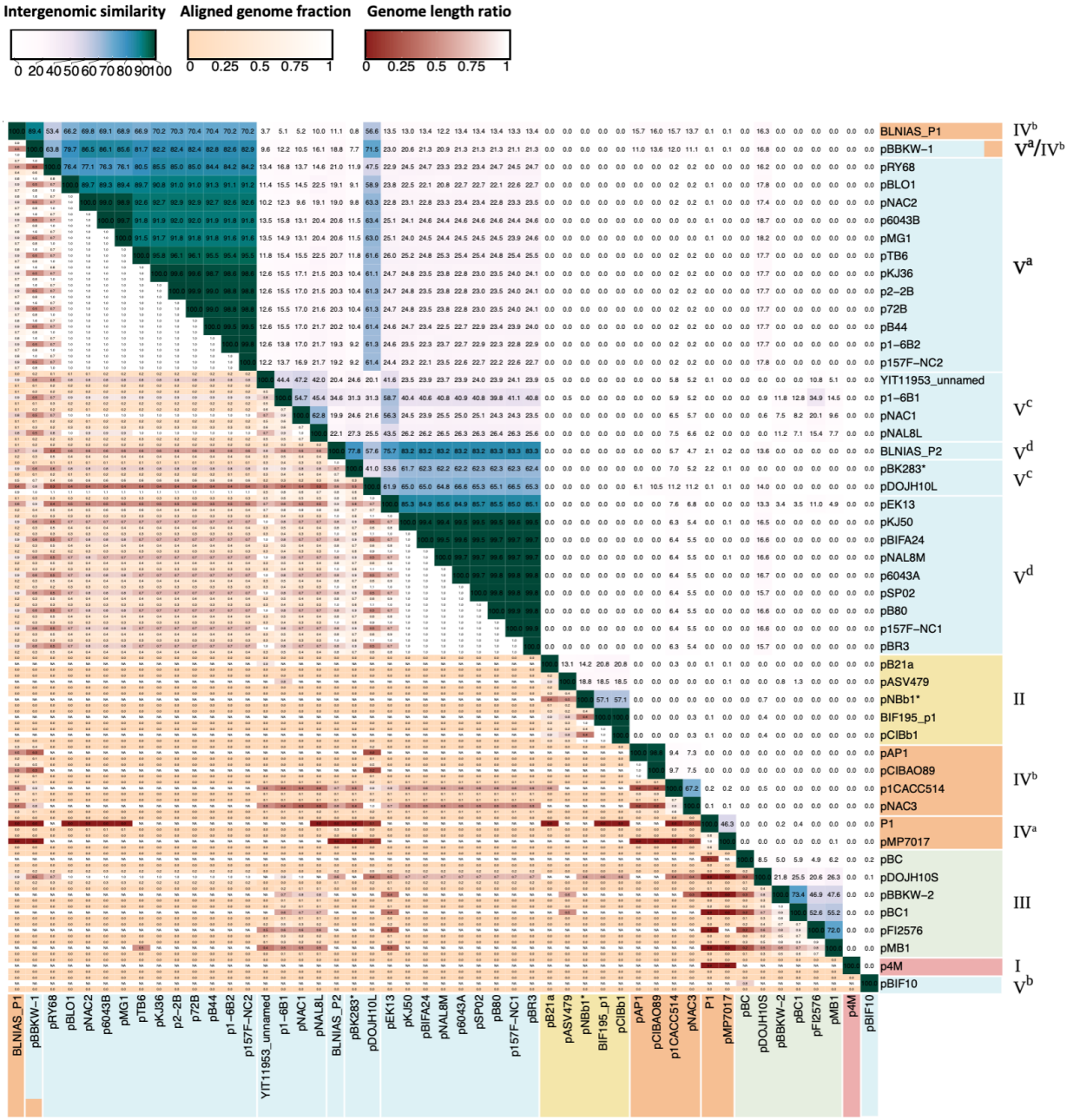


Figure 8. Sequence identity among natural plasmids from *Bifidobacterium*

Heatmap incorporating inter-plasmid sequence identity values (right) and alignment indicators (left). Identity values (rounded to the first decimal) are shown for each plasmid pair (right) based on the percentage identity obtained in whole plasmid sequence pairwise alignments using BLASTN. Clustering of pair-wise inter-plasmid sequence identity values agrees largely with the clustering of the plasmid-replicase protein tree in Figure 7, the colour scale (top-left) shows a gradation of colour from white to purple to green representing increasing plasmid-plasmid identity. In the left half, three indicator values are represented for each plasmid pair: aligned fraction of plasmid named in row (top), plasmid length ratio of the pair (middle) and aligned fraction of plasmid named in column (bottom). Darker colours in the colour scale

represent low values for these alignment indicators. Colour coding is consistent with the clade grouping outlined in the text, where clades -II and -IV replicate via the theta mechanism and all other clades utilise a RC mechanism.

Table 3. *Bifidobacterium* plasmids

(sub)species	Plasmid	GC [%]	Length [bp]	GBK accession	Identical plasmid(s)* GBK accession(name)
asteroides	pAP1	52.0	2140	Y11549	
	pCIBAO89	52.3	2111	NC010908	
bifidum	pB80	61.9	4898	NC011332	
breve	pMP7017	57.4	190178	KM406416	
	pB21a	56.3	5206	NC010930	
	pCIBb1	56.9	5750	NC002133	
	pBR3	61.9	4891	CP010414	
	BIF195 p1	57.0	5749	CP102537	
	pNBb1	59.0	2279	E17316	
catenulatum	pBC1	63.7	2540	NC007068	
catenulatum subsp. *kashiwanohense*	pBBKW-2	62.6	2920	NC021876	
	pBBKW-1	64.9	7716	NC021875	
choerinum	pBC	55.7	11012	CP018045	
longum	pBK283	59.8	6500	AB495342	
	p6043A	61.8	4896	NC010857	
	p6043B	65.1	3680	NC010861	
	pMG1	65.1	3682	NC006997	
	pNAC1	58.8	3538	NC004770	
	PNAC2	64.7	3684	NC004769	
	pNAC3	62.0	10224	NC004768	
	pB44	65.1	3624	NC004443	
	pBIF10	41.7	9275	DQ093580	
	pKJ36	65.1	3625	NC002635	
	pRY68	66.3	2638	CP010454	
	p1	57.1	198121	CP065398	CP065396 (p1)
	pFI2576	61.9	2197	NC011139	
	pDOJH10S	66.2	3661	NC004253	
	pDOJH10L	62.2	10073	NC004252	
	pSP02	61.9	4896	NC019200	
	pTB6	65.1	3624	NC006843	
	pKJ50	61.8	4960	NC004978	
	pBLO1	64.8	3626	NC004943	
	pMB1	62.0	1847	X84655	
	pNAL8M	61.9	4910	AM183144	
	pNAL8L	59.0	3489	NC025162	
	pBIFA24	61.8	4892	NC010164	
longum subsp. *infantis*	p157F-NC2	65.0	3624	NC015066	
	p157F-NC1	61.9	4895	NC015053	
longum subsp. *longum*	p1-6B1	61.1	3919	KP691636	KP691640 (p1-5B1) KP691641 (p17-1B)
	p2-2B	65.1	3624	KP691633	KP691634 (p35B)
	pEK13	60.7	7050	KP691637	
	p1-6B2	65.0	3624	KP691635	KP691639 (p1-5B2)
	p72B	65.2	3624	KP691638	
	BLNIAS_P2	61.1	6230	NC017222	
	BLNIAS_P1	64.4	4233	NC017220	

pseudocatenulatum	p4M	53.1	4488	NC003527
	YIT11953_ unnamed	58.0	4776	CP079234
pseudolongum subsp. *globosum*	pASV479	59.3	4815	NC010877
pullorum subsp. *gallinarum*	p1CACC514	64.2	8720	CP035465

*Bifidobacterial plasmids sharing 100% sequence identity with another resident plasmid of a different name/host strain origin.

Current Plasmid-based Genetic Tools for use in Bifidobacteria

Due to the ever-expanding portfolio of evidence in support of the numerous purported health promoting attributes associated with *Bifidobacterium*, there has been an increasing scientific interest in their application for functional food and biomedical purposes. Of particular interest is the bioengineering of these bacteria to enhance their probiotic characteristics and to utilise them as delivery vehicles for vaccines and anticancer agents in the GIT [126], [127]. However, the current capacity for advancing these innovative applications is limited by the lack of functional information regarding the causation and molecular mechanisms underpinning these purported probiotic activities. This limitation is mainly attributed to the recalcitrance of bifidobacteria to genetic modification [14], [128], combined with a paucity of available molecular tools to study them. Thus, given the increasing commercial and clinical relevance of HRB, it is necessary to improve these genetic engineering techniques particularly in the area of DNA manipulation in order to capitalize on their probiotic potential.

Traditional genome editing tools such as temperature-sensitive and non-replicative plasmids are still widely used in bifidobacteria for the generation of recombinant bifidobacterial strains (**Table 4**) [129], [130]. Gene-disruption for functional genomic studies has predominantly been achieved by site-directed insertional mutagenesis via single-crossover homologous recombination utilizing insertion vectors such as the non-replicative plasmid pORI19-tet [128], and suicide vector developed by Sakurama *et al*. [131]. Protocols for both marker-less [132], [133] and antibiotic replacement [133] double-cross over approaches that generate 'clean' deletions have also been established but have not been as widely applied due to laborious protocols and limited success [27]. Random transposon mutagenesis libraries such as the Tn5-based library for *B. breve* UCC2003 [134], [135] and the IS3-based library for *B. longum* subsp. *longum* [136], although limited in their application to the delineation of genes which contribute to a phenotype of the bacterium, the transposon mutagenesis libraries have been an indispensable tool for high-throughput screening and characterisation of genes involved in carbohydrate utilization, iron responses and biofilm production [137], [138]. Emerging genome

editing tools, such as clustered regularly interspaced short palindromic repeats-CRISPR-associated (CRISPR-Cas) proteins have been extensively applied in eukaryotes, however, their application for genome engineering in probiotic bacteria is lagging, likely hindered by a number of intrinsic factors including nuclease lethality. Nevertheless, the endogenous type I-G CRISPR-Cas system of *B. animalis* subsp. *lactis* was recently reconfigured and implemented in the generation of a 500 bp deletion within the *tet*(W) gene thereby abolishing tetracycline resistance [139]. This successful application of a bifidobacterial type I-G CRISPR-Cas system opens the opportunity for its application for further probiotic enhancement.

The primary limitation for successful mutagenesis utilizing the aforementioned methods is a dependency on sufficiently high transformation efficiencies, this in many cases being affected by the susceptibility of the vector to RM systems in the target strain. This barrier has been partially overcome by bypassing the RM systems present in the target strain through the use of the so-called plasmid artificial modification (PAM) strategy, which involves methylation of the plasmid prior to transformation [128], [140], [141]. While this approach has allowed for an increase in transformation efficiencies up to 5 orders of magnitude, the construction of recombinant strains expressing these bifidobacterial MTases is laborious, furthermore, due to the diversity of RM systems in bifidobacteria, few species are currently accessible [142]. A newly adapted method to bypass bifidobacterial RM systems is through the use of synthetically engineered plasmids where relevant restriction recognition sites have been removed [143].

In addition to their use as vectors for mutagenesis, plasmids are essential tools for cloning and complementation studies (**Table 4**). Reversal of the altered phenotype in mutant strains and hence reinstatement of the WT phenotype by complementation is considered ‘gold standard’, as it unequivocally verifies the role of a genetic determinant in a specific phenotype. Complementation studies in bifidobacteria are typically performed by introducing an intact copy of the mutated gene into the mutant strain using multi-copy autonomous plasmids (**Table 4**). Due to a lack of characterised resident plasmids in bifidobacteria, there are currently only a small number of suitable vectors for cloning and complementation studies, hence a limitation on the size of the insert to be cloned poses a major constraint, partly as a result of the transformation strategy but most significantly due to the natural size limitation of the vector.

Table 4. Current genome editing tools for use in *Bifidobacterium*

Technology	Tool	Reference(s)
Cloning and complementation vectors	pBRASTA100-1003 and pDSO44Sp, *E. coli* - *Bifidobacterium* shuttle vectors based on the pTB6 replicon	[144]
	pAM1, pAM2 and pAM3-6 *E. coli* - *Bifidobacterium* shuttle vectors based on the pBC1 replicon	[114], [145]
	pSP1, pSP2 and pSP3 *E. coli* - *Bifidobacterium* shuttle vectors based on the pSP02 replicon	[124]
	pTK2064, pBS423 with a pSC101 *ori* in replace of the ColE1 *ori*	[131]
	pBES2 *E. coli* - *Bifidobacterium* shuttle vector based on the pMG1 replicon	[146]
	pKKT427, 3.9 kb *E. coli-Bifidobacterium* shuttle vector, Sp^r	[140]
	pPKCm vector series and pSKEm, *E. coli* - *Bifidobacterium* shuttle vectors based on the pCIBAO89 replicon	[119]
	pRD1 series of *E. coli* - *Bifidobacterium* shuttle vectors based on the replicon of megaplasmid pMP7017	[118]
Controlled gene expression	Bile-inducible gene expression system based on the promoter of transporter gene *betA*	[147]
	Arabinose-inducible *araC*-P*BAD* gene expression system from *E. coli*	[148]
	The promoter of gene *BL1518* encoding α-galactosidase, induced by raffinose	[149]
	Temperature-inducible lambda phage promoter PRPL	[150]
	Carbohydrate-inducible promoters P*fruEKFG,* P*scrP* and P*cscBA*	[151]
	Bifidobacteria Expression SysTem (BEST) based on the stress-inducible promoter, P*dnaK*	[152]
	Iron-dependent inducible gene expression system based on the *bfeU* gene promoter	[153]
	The promoter of genes *phoRP*, inducible by phosphate starvation	[154]
Constitutive gene expression	Promoter of gene *gap* encoding glyceraldehyde 3-phosphate dehydrogenase	[149], [155], [156]
	Promoter of gene *hup* encoding histone-like protein	[155], [157]
	Promoter of gene *amy* encoding α-amylase	[158]
	Promoter of 16S rDNA gene	[159]

Technology	Tool	Reference(s)
	A luciferase reporter system based on the *repC* gene promoter of pBC1	[160]
	Optimized promoter P*help* from *Listeria monocytogenes*	[160]
	Phage T5 promoter	[161]
	The promoter of gene *BL1613* encoding a LacI-type transcriptional regulator	[149]
	Promoter P919	[162]
	Promoter P*xfp*	[151]
	Promoter of gene encoding the elongation factor P (EF-P)	[163]
	Promoter of gene encoding the elongation factor Tu	[164], [165]
	pBCMAT, chloramphenicol acetyltransferase gene (cat) reporter system for promoter and terminator evaluation	[166]
	pKO403-TPCTcon plasmid series, *hup* promoter connected with a chloramphenicol acetyltransferase as a reporter gene for the evaluation of bifidobacterial promoters	[167]
	Constitutive p44 promoter from a lactococcal chromosome	[168]
Secretion	Signal peptide Sec2	[169]
	Signal peptide SPBL1181 from *B. longum*	[152]
	Signal peptide of the β-galactosidase (BIF3)	[170]
	AmyB and ApuB signal peptides	[171]
	Exo-xylanase (XynF) signal peptide (Xs)	[148]
	BIF_1734 Sialidase signal peptide	[172]
	Cell surface protein encoded by gene, *gltA,* used for fusion of the target sequence	[173], [174]
Genome engineering	Single-crossover insertion based on non-replicative plasmid	[128], [131], [175]
	Double-crossover replacement by antimicrobial resistance marker, based on a non-replicative plasmid	[20]
	Inducible plasmid self-destruction (IPSD) for gene knock-out and knock-in	[176]
	Double-crossover marker-less gene deletion, based on a non-replicative plasmid	[132], [177]

Technology	Tool	Reference(s)
	Temperature-sensitive plasmid mediated homologous recombination	[129], [178]
	Transposon mutagenesis libraries	[134], [136]
	Synthetic broad purpose vector pFREM28 for single-crossover mutagenesis, RM insensitive, for use in *B. breve*	[143]
	Type I-G CRISPR-Cas system and exogenous CRISPR based editor for use in *B. animalis* subsp. *lactis*	[139]

Engineering *Bifidobacterium* for Food and Biomedical Applications: Limitations and Future Perspectives

While the beneficial effects of particular *Bifidobacterium* strains have been well documented, due to the paucity of genetic tools little information is available pertaining to the precise molecular mechanisms on which these purported health-promoting properties are based. With the ever-expanding availability of bifidobacterial genome sequences providing an ample array of genetic data, new and improved vectors for cloning, knockout, integration, and gene expression studies are required to discover and study the mode of action behind their probiotic properties and subsequently facilitate the application of synthetic biology in this genus.

Although establishing the genetic basis of relevant phenotypes can be achieved through mutagenesis and complementation studies, harnessing the potential of these discoveries in industrial strains without applying genetic modification methodologies is difficult. Therefore, there is increased interest in non-artificial methods of gene transfer systems such as conjugation, as it is considered an acceptable method of genome engineering by the European Food Safety Authority, once the use of recombinant DNA is avoided [179]. A major benefit for the use of conjugal transfer over electroporation protocols is a lack of size limitation. In addition, conjugation incurs a reduced RM effect as DNA is transferred by conjugation as a ssDNA and the complementary strand is synthesised in the host, thus acquiring host-specific epigenetic modifications. To date just one conjugation protocol has been developed for use in *Bifidobacterium*, being based on the transmissible *E. coli–Bifidobacterium* shuttle vector pDOJHR-WD2, containing the conjugative machinery of the promiscuous RP4 plasmid from *E. coli*. Utilizing this protocol, pDOJHR-WD2 has been successfully and consistently transferred from *E. coli* strain WM3064 (pBB109) to several strains of *Bifidobacterium* representing four species at efficiencies which correlated with *E. coli* to bifidobacteria ratios [180].

Horizontal Gene Transfer by Conjugation

Horizontal gene transfer (HGT), also referred to as lateral gene transfer, is the non-sexual transmission of genetic material between organisms [181]. Similar to the genetic exchange by sexual reproduction in eukaryotes, HGT allows the acquisition of genetic information between two organisms. Horizontally acquired genetic material serves as the main source of functional novelty in prokaryotes, underpinning their high adaptability and resilience to diverse and extreme ecological niches (in addition to the mechanisms which allow their rapid response to abiotic stresses). Newly acquired genes may be paralogous to an existing gene or xenologous, involving HGT-mediated displacement of a gene by an orthologous gene from another line of descent or genetic lineage [182]. Both categories of horizontally transferred genes possess the capacity to propel the expansion of protein families in the presence of selective pressure. While the acquisition of paralogous genes may lead to increased gene dosage or indeed adaptation through sub-functionalization or neo-functionalization, xenologs are estimated to persist longer in lineages (lines of descent) and have been linked to the rapid acquisition of novel functions [183], [184]. Comparative analyses of the *B. breve* pangenome has estimated that approximately one third of the variable/accessory genome (10% of the overall number of gene families) has been acquired by HGT, and a number of these genes are associated with environmental adaptations, including genes involved in the production of pili and EPS; defence mechanisms such as CRISPR-Cas systems and metabolic capabilities [185].

There are three principal mechanisms of HGT in prokaryotes: transformation, where DNA is taken up from the environment; transduction, concerning the movement of DNA by bacteriophage; and conjugation, defined as the direct transfer of DNA from a donor cell to a recipient cell, mediated by direct cell-to-cell contacts [186]. Conjugation allows the recipient genome to acquire non-homologous genes and has a much broader host range when compared to transformation or transduction [187], [188]. Additional HGT mechanisms, such as gene transfer agents [189], nanotubes [190] and extracellular vesicles [191], [192], have more recently been described, although their ecological impact and extent of HGT contribution is currently not clear.

The process of conjugation was first experimentally demonstrated in 1946 by Lederberg and Tatum, following the observation of an exchange of genetic markers

between two co-cultured *E. coli* strains, resulting in recombinant progeny [193], a discovery which earned them the 1958 Nobel Prize in Physiology or Medicine. This phenomenon was later attributed to the activity of a 100 kb extrachromosomal element referred to as the F-factor (though the F-factor is in some *E. coli* strains integrated in the chromosome). Based on this discovery, conjugation was used to map the first bacterial chromosome [194]. In the early 1960s the transfer of antimicrobial resistance between species of *Enterobacteriaceae* by means of conjugation was documented [195]. Since then, conjugation has been identified as the main causal agent in the spread of antimicrobial resistance genes among pathogenic bacteria, therefore, due to its significance for human health it remains the most extensively researched mechanism of HGT.

While plasmid propagation may be limited by phylogenetic barriers, bacterial conjugation has been observed to occur between members of the same or different genera, and, though less frequently, between kingdoms, as demonstrated by the activity of the G- pDPT51 and Ti plasmids, which can form productive junctions from bacteria to yeast or plants, respectively [196], [197].

Conjugative systems are found predominately on large plasmids as well as other MGEs which include integrative and conjugative elements (ICEs), conjugative transposons and phage-inducible chromosomal islands, although classes of MGEs are not strictly defined due to substantial overlap in both genetic structure and function [198]. Furthermore, small plasmids which lack their own conjugative machinery may be mobilized by hijacking the conjugative system of a self-transmissible plasmid residing in the same cell, thus are referred to as 'mobilizable' plasmids.

Note: for clarity we will refer to a protein/gene of a specific MGE as, (Protein/gene$_{MGE}$) in the following sections, for example, the TraB protein encoded on plasmid pSG5 will be denoted as TraB$_{pSG5}$.

Type IV Secretion Systems: Mechanism and Structure

There are two recognized mechanisms of conjugation in bacteria that differ significantly in complexity: i) the transfer of ssDNA via multiprotein membrane-spanning complexes called Type IV Secretion Systems (T4SSs) observed for G- and unicellular G+ bacteria, and ii) the transfer of dsDNA described for the multicellular mycelial Actinomycetota *Streptomyces*, requiring a single protein translocase, $TraB_{pSG5}$ [199]–[201].

The prototypical conjugative T4SS is that of the VirB/D4 system of the G- soil dwelling bacterium *Agrobacterium tumefaciens*, responsible for crown galls disease in plants [197]. This system is composed of twelve VirB/D4 proteins encoded by two operons located on the Ti plasmid [202]. The *virB* operon encodes proteins $VirB1_{Ti}$ - $VirB11_{Ti}$, responsible for mating pair formation (mpf), and the *virD* operon encoding the $VirD4_{Ti}$ coupling ATPase, $VirD1_{Ti}$ and $VirD2_{Ti}$ involved in the processing of the DNA substrate, in addition to non-essential components $VirD3_{Ti}$ and $VirD5_{Ti}$ [202]. These proteins can be categorized based on function and/or subassemblies (**Figure 9**); three ATPases, $VirD4_{Ti}$, $VirB4_{Ti}$ and $VirB11_{Ti}$ form the energy centre; polytopic $VirB3_{Ti}$ and $VirB6_{Ti}$ together with bitopic $VirB8_{Ti}$ form the inner membrane core complex (IMCC); $VirB7_{Ti}$, $VirB9_{Ti}$ and $VirB10_{Ti}$ form the outer membrane core complex (OMCC); and $VirB2_{Ti}$ together with $VirB5_{Ti}$ make up the conjugative pilus, while the transglycosylase activity of $VirB1_{Ti}$ promotes pilus formation [203, p. 1].

While this well-characterized system serves as the archetype for ssDNA secretion systems, allowing the unification of the genetic nomenclature among related systems, conjugative T4SSs display significant structural diversity and may contain additional components, or indeed, be composed of a minimal subset of proteins homologous to those of the VirB/D4 system. Considering this, alternative naming systems have been introduced such as *tra* (for transfer) or *mob* (for mobilisation), and a consensus regarding nomenclature of T4SS components has therefore not, as of yet, been universally accepted.

The T4SS conjugation systems are thought to have evolved in G- bacteria and later spread to G+ bacteria [204]. Due to compositional differences in cell wall structure, G+ T4SS conjugation systems lack homologs of VirB7, VirB9, VirB10 that make up the OMCC found in G- T4SSs (**Figure 9**). Furthermore, G+ systems lack

homologs of the pilus-associated components, VirB2 and VirB5, favouring alternative mechanisms for adhesion to recipient cells. The trafficking ATPase, VirB11, which plays an essential role in supplying energy for pilus biogenesis in G- systems is also absent in characterised T4SSs of G+ origin [205], [206]. The IMCC components VirB3, VirB6 and VirB8, in addition to the ATPase-containing components VirD4 and VirB4 are common to both G- and G+ systems. While present in G+ T4SSs, G+ VirB1 homologs differ from prototypical G-VirB1 proteins in both structure and significance. In addition to facilitating the assembly of the secretion channel across the peptidoglycan layer, G+ VirB1 homologs form part of the channel's core via its N-terminal transmembrane (TM) domain and although VirB1 is dispensable in G- systems, it is an essential component of the G+ T4SSs.

In addition to the T4SS apparatus (also referred to as the transferosome) which facilitates the active translocation of the nucleoprotein substrate across the cell wall, ssDNA conjugation systems require a second protein complex, the relaxosome, composed of a mobilization protein (relaxase/Mob) and accessory proteins responsible for the recruitment and processing of the DNA to be transferred (**Figure 9**). At the initial stages of conjugation, the relaxome assembles at the origin of transfer (*oriT*), a ~ 40 – 500 bp stretch of DNA within which the relaxase catalyses a strand-specific transesterification at its cognate recognition sequence, the *nic* site. The product of this isoenergetic reaction is a nucleoprotein complex comprising the relaxase which is covalently bound via a phosphotyrosyl linkage to the 5'-end of the cleaved strand. Once the cleaved strand is displaced, the relaxase shuttles the transfer strand (t-strand) to the T4SS, where the ATPase coupling protein (CP) associated with the T4SS links the nucleoprotein complex to the transferosome, thus is integral for the concerted action of these two protein complexes. The relaxase-bound t-strand is then pumped into the recipient cell, fuelled by energy supplied from ATP hydrolysis. Once the nucleoprotein complex enters the recipient cell, the relaxase/Mob protein catalyses a reverse transesterification, initiated by the free 3'-OH of the t-strand, resulting in circularisation of the transferred ssDNA. Subsequently, recircularization and complementary synthesis of the transfer strand occurs. Upon expression of the plasmid genes the recipient is phenotypically converted to a transconjugant. As the transferred DNA encodes the conjugation machinery, this in turn enables the erstwhile recipient to further spread the plasmid DNA as a donor.

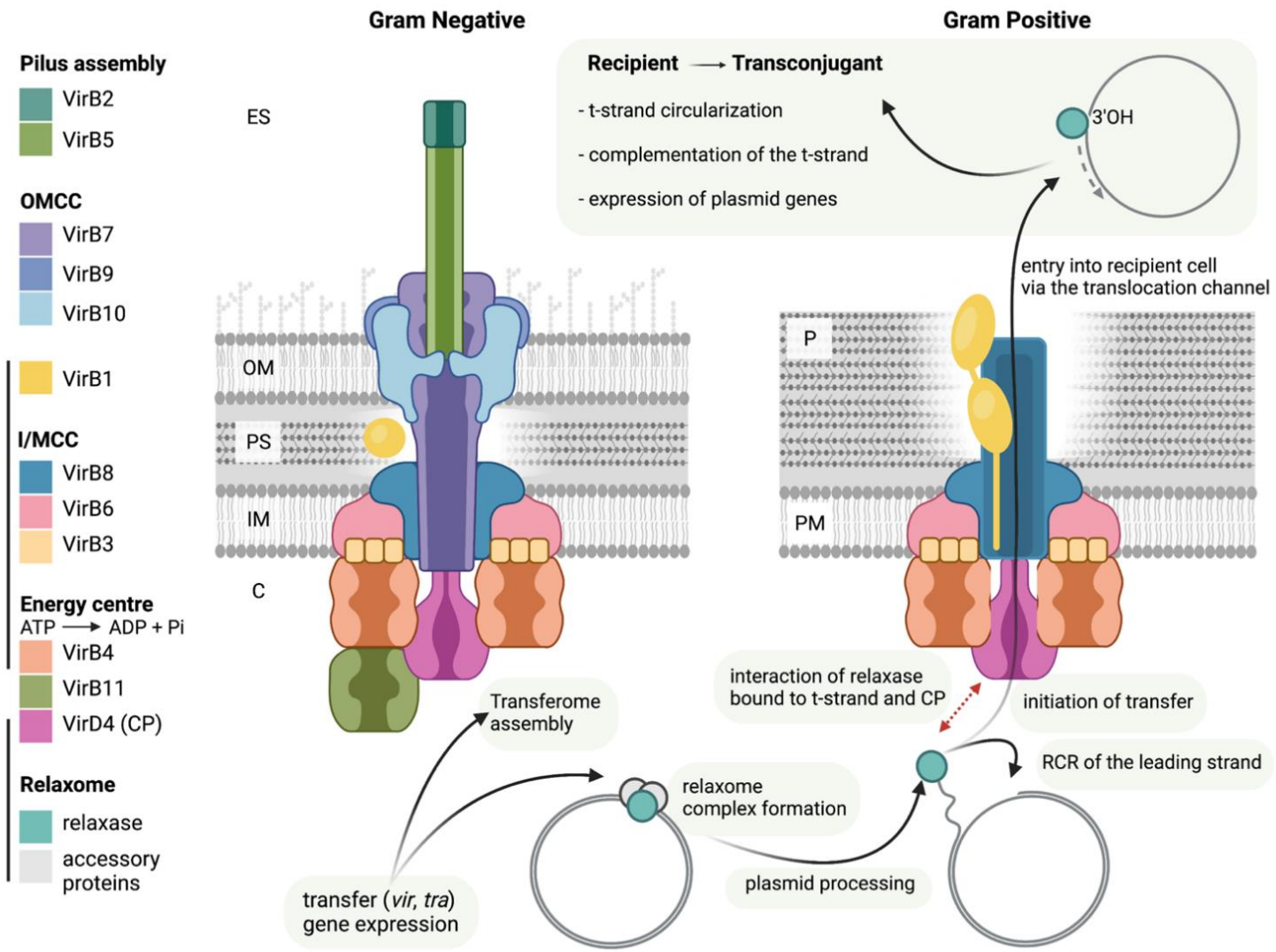


Figure 9. General translocation pathway and structural comparison of canonical T4SSs of G- and G+ origin

The initiation of conjugation first requires the expression of the *tra* genes. Some of the produced Tra proteins form the T4SS complex (transferome) which mediates mating pair stabilization. Tra proteins constituting the relaxosome bind to the *oriT* and prepare the plasmid for transfer by inducing the nicking reaction by the relaxase. Interaction between the relaxosome and the T4CP initiates the transfer of the T-strand through the translocation channel into the recipient. Transfer of the relaxase-bound T-strand in the recipient is concomitant with the conversion of the ssDNA into dsDNA by RCR in the donor. Once fully internalized the relaxase performs the circularization of the fully internalized T-strand.

The Ever-Expanding Diversity of Conjugative Relaxases

Conjugative relaxases are multi-domain proteins responsible for the site- and strand-specific endo-nucleolytic cleavage of plasmid DNA and strand transfer reactions which initiate and complete the conjugative transfer process, in the donor and recipient, respectively [207], [208]. The name relaxase was coined in honour of the pioneering work of Clewell and Helinski, who discovered the conversion of protein-bound *Col*E1 plasmid DNA, which they termed 'relaxation complexes', from supercoiled DNA to an open circular state upon treatment with denaturing agents. Furthermore, Clewell and Helinski revealed the strand-specificity event of the reaction in later works [209], [210].

While the vast majority of relaxases which have been biochemically and structurally characterized to date belong to the histidine-hydrophobe-histidine (HUH) endonuclease superfamily, additional relaxase families have been proposed, based on distinct protein folds, mechanistic differences and catalytic alternatives to accomplish conjugative DNA processing, thus indicating a much larger diversity than previously thought [211]. Considering this, relaxases can be clustered into one of nine established protein families based on sequence similarity, referred to as MOB groups, thus representing a key taxonomic trait of transmissible mobile elements [211]. The defining features of each family are detailed in **Table 5** and represented visually in **Figure 10**.

Six MOB groups (MOB_P, MOB_F, MOB_Q, MOB_H, MOB_V and MOB_B) fall collectively within the aforementioned HUH endonuclease superfamily. HUH endonucleases catalyse breakage and joining of ssDNA by utilizing a nucleophilic tyrosine residue to create a 5′-phosphotyrosine intermediate and a free 3′-OH at the cleavage site [212]. The 3′-OH group then serves as the nucleophile in the strand transfer reaction to resolve the phosphotyrosine intermediate in the termination step of conjugative transfer. These reactions are cofactor-dependent, requiring divalent metal ions to facilitate cleavage, which are coordinated by the histidine pair of the HUH motif together with a third ligand, an additional polar residue whose location varies across families [212].

MOB_P represents the largest relaxase group and displays marked diversity within the HUH family, represented by its distribution into a number of phylogenetic clades, each with a specific prototype. Due to its extensive characterization, $TraI_{RP4}$ is the overall prototype of the MOB_P group [213]. Furthermore, a previously described MOB group, MOB_{HEN}, is now embedded within the MOB_P family, this subclade solely contains relaxases of mobilizable plasmids which contain a so-called HEN signature, as opposed to the classical His-triad [211], [214]. Relaxase Rel_{pLS20} from G+ *B. subtilis*, although recently put forward as the prototype for a new family of relaxases (MOB_L), also appears to belong to a subfamily of the MOB_P group [215].

MOB_Q is phylogenetically related to the MOB_P family and contains relaxases of conjugative plasmids from G+ bacteria, including staphylococcal pGO1, lactococcal pMRC01 and TraA from the broad host range streptococcal plasmid pIP501 [214], [216]. The prototype of the MOB_Q family is represented by MobA of the G- broad-host range mobilizable plasmid pRSF1010, due to its extensive biochemical and structural characterisation [217].

MOB_B represents a small family of relaxases that are found exclusively in the class *Bacteroidia* (formerly *Bacteriodetes*) and contain a domain of unknown function (DUF5712) [218].

MOB_F relaxases are distinct from other HUH endonucleases, as they contain two active tyrosine residues within their N-terminal catalytic centre, while containing a C-terminal helicase domain. Furthermore, cognate *oriT*'s of this MOB family, whose prototypes are $TraI_F$ and $TrwC_{R388}$, are characteristically large and complex [208], [219].

MOB_V family relaxases are mostly found in cyanobacteria and are the dominant group of relaxases identified in *S. aureus*. MobM from the promiscuous streptococcal plasmid pMV158 is considered the prototype of this family [220]. MOB_V family relaxases employ a histidine nucleophile instead of tyrosine in the cleavage reaction, resulting in the formation of a covalent phosphoramidate adduct [221].

Members of the MOB_H relaxase family are predominantly found on large plasmids, ICEs and genomic islands of *Pseudomonadota* [211]. The prototype of this family is TraI encoded within the gonococcal genetic island (GGI) of *Neisseria gonorrhoeae* [222]. The defining genetic signatures of the MOB_H family are a triple

histidine motif (PASE-x-HHH-x3-GG-x3-H-x-L) and a conserved (HDxGK) of its HD-hydrolase domain [222].

The remaining families, i.e., MOB_C, MOB_M and MOB_T, differ significantly from the classical HUH-relaxases and are therefore termed non-canonical relaxases. Members of the MOB_C family contain a C-terminal DEE motif for the coordination of metal ions, resembling Type I restriction endonucleases. Mutagenesis of acidic residues D152, E170, and E172 within the active site of the MOB_C prototype, $TraX_{pAD1}$ from *E. faecalis*, verified the essentiality of these acidic residues for the functioning of the relaxase, and demonstrated that they are responsible for coordinating Mg^{+2} cations to allow the nucleophilic attack [223].

MOB_T relaxases display similarity to the Rep_trans family of rolling-circle replication (RCR) proteins and are widespread in ICEs of G+ origin, particularly within the Bacillota phylum. Biochemical characterisation of the RelSt3 relaxase encoded by ICESt3 from *S. thermophilus*, demonstrated that MOB_T relaxases employ a DDE motif for the coordination of metal ions [224].

MOB_M family relaxases are related to tyrosine recombinases and have been exclusively associated with conjugative plasmids from members of the genus *Clostridioides* (formerly *Clostridium*). As such, the MOB_M family prototype is TcpM encoded by pCW3 from *Clostridioides perfringens* [225]. Unlike HUH endonucleases, the strand-transfer reactions catalysed by Tyr recombinases are cofactor-independent and generate 3′-phosphotyrosine intermediates and free 5′-OH groups.

In addition to their requirement for conjugation and their importance in mobilisation of secondary targets, relaxases may perform other functions associated with their nicking and strand transfer ability, such as catalysing site-specific recombination or initiation of plasmid replication (**Table 5**).

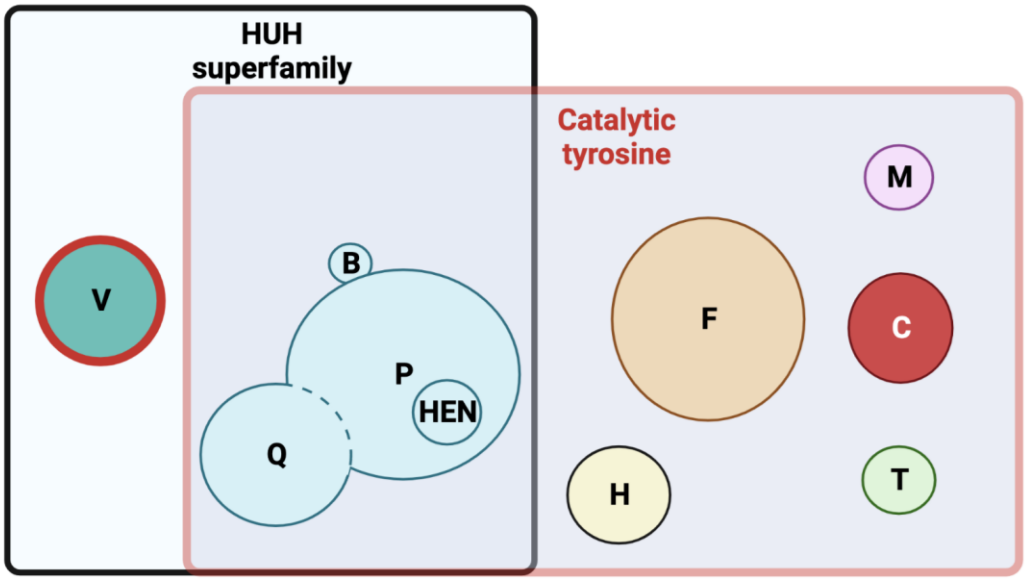


Figure 10. A schematic representation of the relationship between relaxase MOB groups

The first cluster (black box) represents the canonical 'HUH' endonuclease superfamily, containing MOB groups; V, B, Q, P and HEN. MOB groups falling outside of this cluster (F, H, M, C and T) represent non-canonical relaxases which bear no relatedness with each-other or any other MOB family. Relaxases of the MOB_V family (green circle with thick red outline) differ from all other groups due to their use of a catalytic histidine in place of a catalytic tyrosine in transesterification reactions. The areas of circles are proportional to the number of relaxases known for each MOB group. Some groups may overlap with others due to significant sequence homology at particular clades or family subgroups, as illustrated with the MOB_P group which contains both MOB_{HEN} and MOB_{Q1} (see [211] for detailed phylogenetic analysis).

Table 5. Features of conjugative relaxases based on their current MOB family classification

MOB family	Prototype	Signature motif/ 3D fold	Catalytic residue	Covalent complex	NS Endonuclease activity	Domains	Additional Function(s)*		
P	$TraI_{RP4}$	HUH, HEN	Tyr	Y	N		PRE	Mob	rTn
Q	$MobA_{RSF1010}$	HUH	Tyr	Y	N		PRE	Mob	
F	$TrwC_{R388}$	HUH	Tyr, Tyr	Y	N	Helicase	PRE	Mob	
V	$MobM_{pMV158}$	HUH	His	Y	N		PRE	Mob	Cop
B	$MobB_{p5482}$	HUH	His	Y	N	DUF5712			
H	$TraI_{GGI}$	3H, HD	Tyr	Y	N		PRE	Rep	
C	$TraX_{pAD1}$	RE	Tyr	N	Y		PRE		
T	$Orf20_{Tn916}$	Rep_Trans	Tyr	Y	Y	PF02486		Rep	
M	$TcpM_{pCW3}$	Tyr_rec	Tyr	Y	Y	cl00213			

* Secondary biological functions outside of its main relaxase activity: *in trans* activity on heterologous *oriT* sequences (Mob); Plasmid Recombination Enzyme (PRE); initiator of plasmid replication (Rep); plasmid copy number control (Cop); retro-transposition enhancer (rTn).

Mating Pair Formation: Identification and Classification

While MOB typing is very useful in the classification of relaxases from various transmissible elements, it does not distinguish mobilizable elements from those which encode a T4SS (**Figure 11A**). The mating pair formation (MPF) classification system is based on the phylogeny of the ubiquitous VirB4 ATPase, thus is specific to conjugative ssDNA secretion systems. MPF-typing was introduced to standardise gene nomenclature within related conjugal systems, in response to the extremely diverse gene naming currently observed among homologous conjugal systems that has complicated literature analysis and comparative analysis of conjugal systems [204].

Among the large class of bacterial ATPases involved in various cellular processes, VirB4 proteins form a distinct clade. VirB4 proteins can be further divided into eight subclades which underpins their MPF classification, each associated with specific bacterial lineage (**Figure 11B**, **Table 6**). Six MPF classes are specific to diderms, MPF_T, MPF_F, MPF_I, MPF_C, MPF_B, and MPF_G, four of which are predominantly found in *Pseudomonadota* (MPF_T, MPF_F, MPF_I and MPF_G), while classes MPF_C and MPF_B, are found exclusively within the Cyanobacteria and *Bacteroidia* phyla, respectively. The two remaining classes, MPF_{FATA} and MPF_{FA}, can be found in monoderm bacteria and are discussed below [204].

The MPF_{FATA} group is quite unique as it contains systems from monoderms with diverse cell envelopes. The group is composed of bacteria from the phyla *Bacillota* and *Actinomycetota*, *Mycoplasmatota* which are devoid of a cell wall, and members of the domain Archaea. MPF_{FATA} T4SSs are associated with classical Type Four Coupling Proteins (T4CPs) and with canonical relaxases belonging to the MOB families MOB_P, MOB_Q, and MOB_C [204], [226].

Contrastingly, the majority of conjugative plasmids and ICEs belonging to MPF_{FA} T4SSs encode atypical TcpA-like CPs, which are closely related to the dsDNA translocase, FtsK, involved in DNA segregation during cell division [204], [226]. Furthermore, MPF_{FA} T4SSs have been associated with non-canonical relaxases such as the MOB_T family relaxase, orf20 of the Tn916 conjugative transposon from *B. subtilis*, and the MOB_M family relaxase, $TcpM_{pCW3}$ from *C. perfringens* [225], [226]. In addition, it has been demonstrated that members of the MPF_{FA} group contain proteins with significant sequence similarity to the PG-hydrolase, VirB1, described

for G- systems. $TcpG_{pCW3}$ was experimentally determined to hydrolyse peptidoglycan in *C. perfringens*, consistent with VirB1-like activity [227].

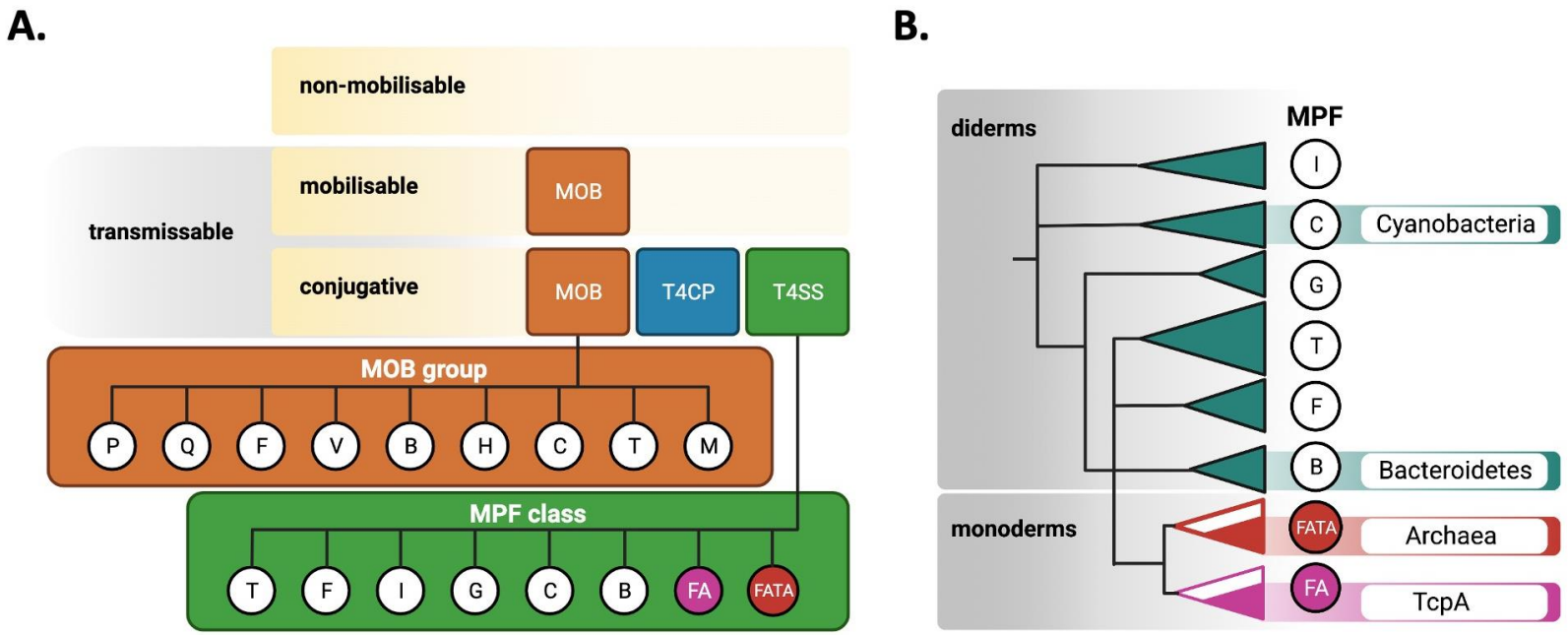


Figure 11. Mating pair classification of conjugative ssDNA secretion systems

A. Schematic view of the genetic constituents of transmissibility. Self-transmissible, i.e. conjugative plasmids, code for all components of a conjugative apparatus: a relaxase (MOB) (orange), a type IV coupling protein (T4CP) (blue), and a type IV secretion system (T4SS) (green). Mobilizable plasmids contain just a MOB module (they may additionally encode a T4CP, not shown) and require the MPF of a co-residing conjugative plasmid to become transmissible. While a certain MOB group may be more frequently associated with a particular MPF class, any combination of MOB and MPF can occur. **B.** Diversification of MPF groups within diderms (G-) and transfer to monoderms (G+), the series of clades (as depicted) is the basis of the MPF classification system. MPF classes currently associated with a specific G- phylum are indicated. A subset of MPF_{FATA} classes can also be found in the domain Archaea. Some elements among MPF_{FA} utilise (vastly different system) a TcpA-like coupling protein in lieu of a VirD4 homologue.

Table 6. Features of the eight MPF classes

MPF Class	T	F	I	G	C	B	FA	FATA
Prototype	Ti	F	R64	ICEHin1056	pCC7120a	CTnDOT	Tn916	pGO1
Cell envelope	diderm	diderm	diderm	diderm	diderm	diderm	monoderm	monoderm
VirB4	VirB4	VirB4	TraU	VirB4	TraU	VirB4	VirB4	VirB4
Coupling protein	T4CP	T4CP	T4CP	T4CP	T4CP	T4CP	TcpA	T4CP
Number of associated genes	4	9	10	5	ND	17	7	13
Frequently associated MOB groups	H, F, P, V, C	H, F, Q	P, Q, F	H	V	non-canonical	non-canonical	P, Q, C

ND = not determined.

Coupling Proteins of T4SSs

Type IV coupling proteins (T4CPs) are part of the mobilisation module of transmissible plasmids and are most often encoded by a gene that flanks the relaxase-encoding gene. They are present in all conjugative systems and are essential constituents of most. Due to their ubiquity, T4CPs are good indicators of the presence of conjugative systems. As their name suggests, T4CPs play a key role in connecting the substrate to the transfer apparatus through substrate recruitment, and consequently, T4CPs have co-evolved with their translocated substrates and cognate secretion systems.

In terms of their structure, most T4CPs contain a transmembrane domain (TMD), which anchors the protein to the inner-membrane, and a cytosolic domain essential for substrate recognition. The cytosolic domain of T4CPs contains two subdomains, the nucleotide binding domain (NBD) and the all-alpha domain (AAD), a third C-terminal variable domain is occasionally present. The NBD is present in all T4CPs and is essential for their *in vivo* activity. This domain contains two conserved ATP-binding motifs, i.e. the Walker A motif characterised by (G-x(4)-GK-[TS]) and the Walker B motif represented by [RK]xxxGxxx-LhhhDE [228]. The AAD is proposed to interact with DNA-processing factors functioning as a substrate specificity determinant thus constituting the basis for selective coupling [229].

T4CPs display various domain architectures and can be categorised into five subfamilies: VirD4-type, TMD-less, TraG-J pairs, FtsK-type and Archaeal T4CPs (**Figure 12**) [230].

The VirD4-type subfamily is named after the $VirD4_{At}$ T4CP from the T-plasmid of *A. tumefaciens*, although its prototype is $TrwB_{R388}$ of the *E. coli* IncW plasmid R388. $PcfC_{pCW3}$ from plasmid pCW3 of *C. perfringens* represents a VirD4-prototype T4CP of G+ origin [231]. Members of this subfamily display low sequence identity and are 500–750 aa residues in length with ≥2 N-terminal TMDs [232].

The TMD has been determined to be an essential structure within VirD4-like T4CPs for *in vivo* conjugation, as the TMD is both critical for the activation of the secretion apparatus and for oligomerisation [233]–[235]. The TMD of $TrwB_{R388}$ has been extensively studied, and remains the only T4CP which has been successfully

purified in its native form [236]. Comparative structural and biochemical analyses of $TrwB_{R388}$ and its TMD deletion derivative $TrwB1N70_{R388}$, have highlighted significant differences in *in vivo* conjugation, nucleotide binding ability and ATPase activity, in addition to subcellular localization, oligomerisation and stability. Stability analyses comparing the native protein and the soluble truncated form that lacks the 70 N-proximal residues comprising the TMD, TrwBΔN70, established that the native form is significantly more stable than its truncated counterpart under denaturing conditions. This suggests that the TMD of $TrwB_{R388}$ and by extension those of other VirD4-type CPs play a role in maintaining the functional conformation of the cytosolic domain [236]. It has therefore been concluded that the TMD of $TrwB_{R388}$ is an essential element for protein functionality [231], [234], [236]–[238]. While these conclusions have previously been extrapolated to other members of this subfamily, recent comparative studies of $MobB_{CloDF13}$ from mobilizable plasmid CloDF13 and its cognate TMD deletion variant, MobBTMD, observed that the TMD of $MobB_{CloDF13}$ is essential for plasmid mobilisation. However, it did not reveal a significant impact on the aforementioned functions as observed for TrwB1N70, indicating this is not the case for all VirD4-type T4CPs [239].

Unlike VirD4-type T4CPs whose TMD is critical for their activity, T4CPs that naturally lack TMDs have the capacity to facilitate the transfer of their cognate substrate without impediment. These TMD-less T4CPs form the second subfamily, and have been indicated to interact with small membrane proteins, which are thought to come together to form a complex resembling that of a VirD4-type T4CP. $TraJ_{pIP501}$ and $TraI_{pIP501}$ are the most studied pair within this group [239].

TraG-TraJ pairs form the third subfamily and are similar in size and structure to VirD4-type T4CPs but contain a shorter periplasmic domain. TraG-type CPs are integral membrane proteins which physically and functionally associate with an ~200 aa hydrophobic TraJ protein. The prototype of this subfamily is the $TraG_{R27}$ T4CP of the R27 plasmid origination from *Salmonella enterica,* which associates with the 224 aa $TraJ_{R27}$ protein containing four TMDs, being required for DNA translocation [240]. Interestingly, the conserved TraG-TraJ module has been identified in a variety of sources that do not contain R27-like *Mpf* genes, suggesting this conserved module can be inherited independently [240], [241].

FtsK-like CPs represent the largest of all T4CP subfamilies with a size of >750 aa residues, containing several TMDs and a nucleotide binding domain NBD

separated by a large linker. In addition, FtsK-like CPs contain a domain for ATP-binding and hydrolysis, in place of an AAD [242]. $TcpA_{pCW3}$ is an FtsK-type coupling protein encoded by the pCW3 plasmid from *C. perfringens*, functional genetic analysis of $TcpA_{pCW3}$ has established that its Walker A and Walker B ATP-binding motifs, in addition to its RAAG motif are essential for pCW3 conjugative transfer [242]. Interestingly, deletion of the TMD of the FtsK-type T4CP, $TcpA_{pCW3}$, was shown to result in a lower conjugation efficiency in comparison to the wild type (WT), but did not abolish the plasmid's transfer capacity [242]. Furthermore, it has been demonstrated by bacterial two-hybrid analyses, that $TcpA_{pCW3}$ functionally interacts with $Orf13_{Tn916}$-like $TcpC_{pCW3}$, PG hydrolase $TcpG_{pCW3}$, and VirB6-like $TcpH_{pCW3}$ [243]. Similar to the conjugative system of pCW3, it has been demonstrated through deletion and complementation analyses that conjugative transfer of the 350 kb pXO16 conjugative plasmid from *B. thuringiensis* serovar *israelensis* relies on an FtsK-type coupling ATPase, $TipL_{pXO16}$ [244]. The fifth subfamily is composed solely of archaeal T4CPs that have been identified through sequence prediction utilizing T4CP N-terminal binding domains (NBDs) [245].

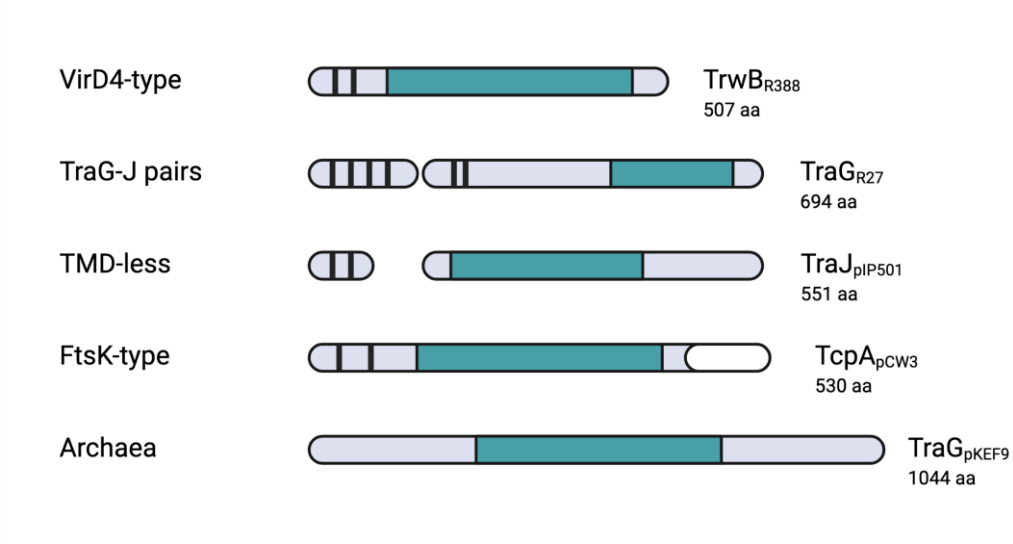


Figure 12. Molecular architecture of the different T4CPs

The molecular architecture of each prototype of the five subfamilies are represented. The family name appears to the left of the architecture and Protein name$_{MGE}$ and sequence length to the right. Transmembrane α-helices are represented as black vertical lines, the conserved central nucleotide-binding domain in teal green and the glutamate- and aspartate-enriched C-terminal domain in white.

Mating Pair Stabilization in Gram Positive Conjugative Systems: Cell-Cell Attachment

The G+ cell wall is a rigid but porous, three-dimensional structure, with low selective permeability which is composed of long covalently cross-linked polysaccharide chains of alternating *N*-acetylglucosamine and *N*-acetylmuramic acid connected by β-1,4-O-glycosidic bonds, known as peptidoglycan [246]. This structure primarily serves to maintain cellular integrity and provide protection against environmental stressors, while additionally serving as a scaffold for polar lipids, teichoic acids, lipoteichoic acids (LTA), polysaccharides, cell wall binding proteins and sortase-displayed cell-surface structures.

Sortases are membrane-anchored cysteine transpeptidases that utilise a His-Cys-Arg catalytic triad to catalyse transpeptidation reactions, joining proteins via a conserved C-terminal cell wall sorting signal (CWSS) motif to an amino group on their polypeptide substrate, i.e. the cell wall/ pilin subunits. The CWSS (which is recognised by and cleaved through the activity of a Sortase) of sortase-dependent proteins is a five-residue amino acid sequence (LPXTG, or variant thereof), followed by a hydrophobic stretch of amino acids and a positively charged tail. Sortases can be grouped into six different classes, A-F, based on their amino acid sequence and can be further differentiated based on cleavage site recognition sequence, main function and/or associated bacterial genus (**Table 7**) [247].

Briefly, G+ cell wall associated (CWA) proteins interact with key constituents of the extracellular matrix and can adhere to a wide variety of polymers, including, collagen, extracellular DNA (eDNA), LTA and glucans [248]–[252]. Several classes of G+ cell CWA adhesins have been proposed based on structural and functional analysis, including; microbial surface components recognizing adhesive matrix molecules, near iron transporter motif family adhesins, G5-E domain repeat family, cadherin-like family, aggregation substances, in addition to the well categorized long hair-like filamentous supramolecular protein structures referred to as pili [253]–[259].

G+ surface proteins have various functions and can be involved in surface attachment, biofilm formation, cell-to-cell or host-to-cell interactions, in addition to facilitating mating pair stabilization and efficient DNA transfer during conjugation due its contact-dependent nature. In contrast to the extensively characterized

conjugative systems of G- bacteria, G+ systems do not produce long retractile pili as part of their secretion systems in order to stabilise mating pairs, instead, it has been established that cell surface adhesins mediate cell-cell contact in G+ ssDNA conjugation mechanisms. While this area of research is still in its infancy, recent studies have shed light on the diverse strategies of cell-to-cell contact utilized by G+ bacteria to facilitate these interactions. To date, the best characterised plasmid-encoded conjugation-associated adhesins from G+ bacteria are those of the PrgABC system of the enterococcal sex-pheromone responsive plasmid pCF10. The attachment module of pCF10 is proximal to the 24 gene *prgQ* conjugation operon and encodes three sortase-dependent surface proteins, $PrgA_{pCF10}$, $PrgB_{pCF10}$ and $PrgC_{pCF10}$, in addition to a small cytoplasmic RNA-binding protein, $PrgU_{pCF10}$ (**Figure 13A**). In-frame deletion of $prgB_{pCF10}$ was shown to abolish cellular aggregation and reduce the plasmid transfer ability by ~ 2 orders of magnitude in liquid assays compared with WT transfer rates, however, individual deletion mutations of $prgA_{pCF10}$ or $prgC_{pCF10}$ did not diminish plasmid transfer ability [260]. It was subsequently demonstrated that only $PrgB_{pCF10}$ is required for cellular aggregation and pCF10 transfer at WT frequencies.

It has been established that the N-proximal adhesin domain of the 140 kDa $PrgB_{pCF10}$ adopts a lectin-like fold, resembling the carbohydrate binding-domain of the Streptococcal Antigen I/II (AgI/II) and the variable region of the *S. gordonii* (commensal bacterium of the oral cavity) SspB adhesion [252], [261], [262]. While this structural fold is common among G^+ adhesins, the adhesin domain of $PrgB_{pCF10}$ differs substantially in its polymer preference mediating aggregation through the binding and compaction of eDNA, the strength of which is positively correlated with eDNA length [252], [260]. Genes encoding predicted aggregation proteins of other enterococcal sex-pheromone responsive plasmids, pAD1 and pPD1, display >90% amino acid identity with $PrgB_{pCF10}$ thus are likely functionally comparable [263]–[265].

In 1985, Dunny *et al.* first reported that $PrgA_{pCF10}$ functions as a surface exclusion factor [266]. Surface exclusion functions to prevent donor–donor contacts, thereby averting plasmid transfer to a cell which already harbours the element. Although deletion of $prgA_{pCF10}$ did not affect cellular aggregation nor impact pCF10 conjugative transfer, interestingly, in pCF10Δ*prgA* cells, large amounts of the full length 140 kDa PrgB species were detected by Western immunoblotting, while two

PrgB species of 140 kDa and a putative processed form of 73 kDa (which accumulated in the extracellular milieu) were detected in WT cells, suggesting that PrgA $_{pCF10}$ may contribute to PrgB $_{pCF10}$ processing [260]. Structural analysis of PrgA $_{pCF10}$ (891 aa) later revealed a centrally located CAP protease domain that is presented 40 nm from the cell wall via its long coiled-coil stalk, successful docking of the predicted PrgB $_{pCF10}$ proteolysis region within the PrgA $_{pCF10}$ CAP region lead the authors to propose proteolytic cleavage of $PrgB_{pCF10}$ is likely mediated by $PrgA_{pCF10}$, in agreement with previous reports [267]. Therefore, in addition to providing surface exclusion, $PrgA_{pCF10}$ decreases cellular adhesion by antagonizing $PrgB_{pCF10}$-mediated aggregation and conjugative transfer of pCF10 (**Figure 13B**).

PrgC $_{pCF10}$ (285 aa) contains a large centrally located proline-rich repeat region characteristic of other G+ adhesins, while PrgC $_{pCF10}$ has been indicated as a virulence factor, its role in conjugation is unknown. It was recently established that *prgU* $_{pCF10}$ encodes a novel RNA-binding regulator which functions to suppress sex pheromone-mediated cellular toxicity by preventing the overproduction of $PrgABC_{pCF10}$ surface adhesins [268].

Upstream of the *tcp* conjugation locus of pCW3 resides the *cnaC* region (nucleotide coordinates 17,927–24,024) which encompasses eight genes, *cnaC*, *srtD*, *dam* and five hypothetical genes (**Figure 14A**). Efforts to functionally annotate the protein products of the *cnaC* region were previously unsuccessful due to a lack of identifiable conserved domains and little homology to proteins of other bacterial species [269]. However, recent reclassification of the pCW3-encoded SrtD sortase coupled with the identification of a C-terminal CWSS and two predicted central collagen binding domains within the 1,187 aa CnaC (**Figure 14B**), prompted an investigation into their potential for mediating cell-cell interactions and promoting mating pair stabilisation during conjugal transfer of pCW3 [270].

To investigate the role of the genes encoded within the *cnaC* region on pCW3 conjugation, mutational analyses and genetic complementation of each individual gene was undertaken. Intraspecies conjugal mating assays with *srtD* and *pcw324* mutants displayed a five order of magnitude reduction in conjugation frequency as compared to the WT, while no significant effect on conjugation efficiency was observed for the *cnaC* mutant. In interspecies mating assays, a ten-fold reduction in conjugation efficiency was observed for the *cnaC* mutant, while no transconjugants were recovered for either the *srtD* or *pcw324* mutants. Additionally, an approximate

20-fold reduction in conjugation frequency was observed due to the disruption of the *dam* gene, encoding a DNA adenine methyltransferase, in both intra- and inter-strain matings. While genetic complementation of the *dam* gene was unsuccessful, the authors speculate this protein may act as a positive regulator of pCW3 conjugation as *dam*-mediated regulation has been observed in other systems [271], [272]. These data demonstrate the importance of the *srtD*, *pcw324* (subsequently renamed *tcpN*) and *dam* gene products for efficient pCW3 conjugative transfer, while highlighting the requirement of CnaC for the optimal transfer of pCW3 to recipient cells with different strain backgrounds [270].

$CnaC_{pCW3}$ shares common features with the putative surface adhesions encoded on other conjugative plasmids from G+ bacteria, including $TraO_{pIP501}$ and $PrgB_{pCF10}$. Interesting however, while $CnaC_{pCW3}$ is required for the efficient transfer of pCW3 in interspecies mating, conjugal mating with the deletion variant of $PrgB_{pCF10}$ was shown to result in a minimal reduction in transfer frequencies [273]. Furthermore, deletion of *traO* encoding the adhesion of the conjugative plasmid pIP501 completely abolished its transfer ability (indicated as preliminary data in the publication, thus not discussed here) [274]. The variability of these data regarding the essentiality of these surface adhesions further highlights the need for more research into adhesions of other conjugative systems.

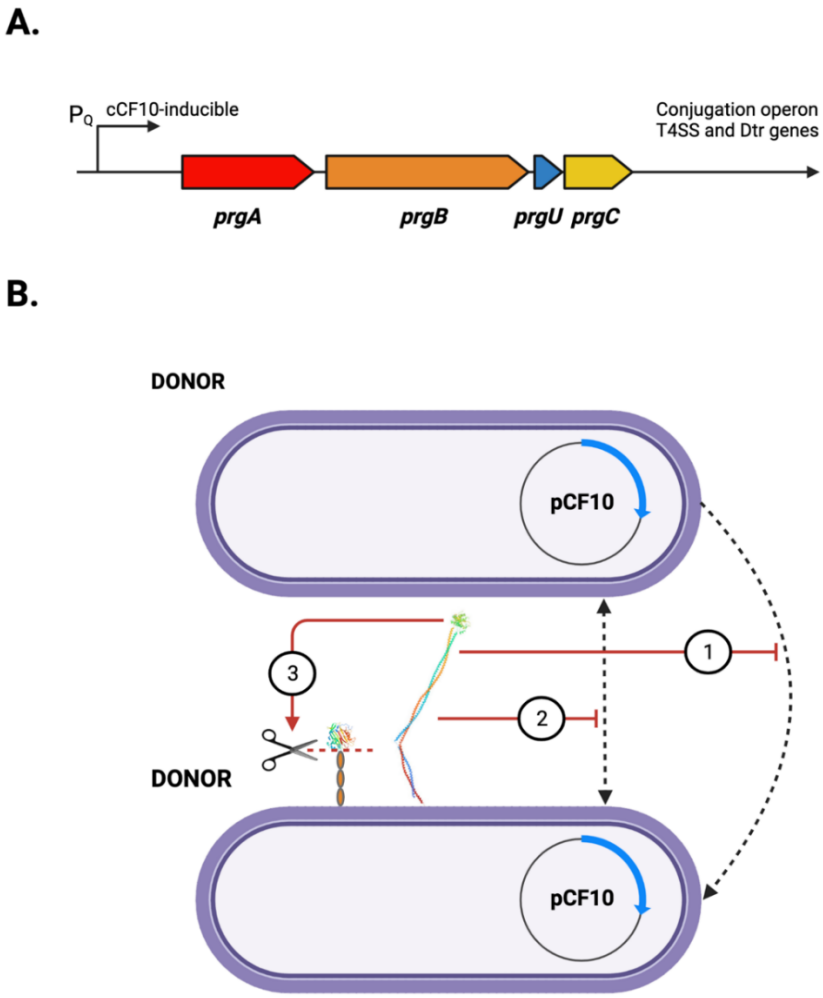


Figure 13. The PrgABC system of the enterococcal sex-pheromone responsive plasmid pCF10

A. Schematic representation of the attachment module of pCF10. The cluster encodes sortase-dependent surface proteins; PrgA, PrgB and PrgC, and the RNA-binding protein, PrgU. The P_Q cCF10 inducible promoter from which the cluster is transcribed is shown. T4SS and Dtr processing genes encoded downstream are indicated. **B.** Schematic adapted from [267], depicting the various functions of PrgA. Surface exclusion (1): PrgA inhibits the uptake of additional pCF10 plasmids in pCF10-containing cells. Decreased cellular adhesion (2): PrgA reduces cellular adhesion through steric hindrance. Proteolytic procession of PrgB (3): The CAP functional domain of PrgA mediated cleavage of PrgB leading to removal of the PrgB adhesion module, thus antagonizing PrgB-mediated aggregation.

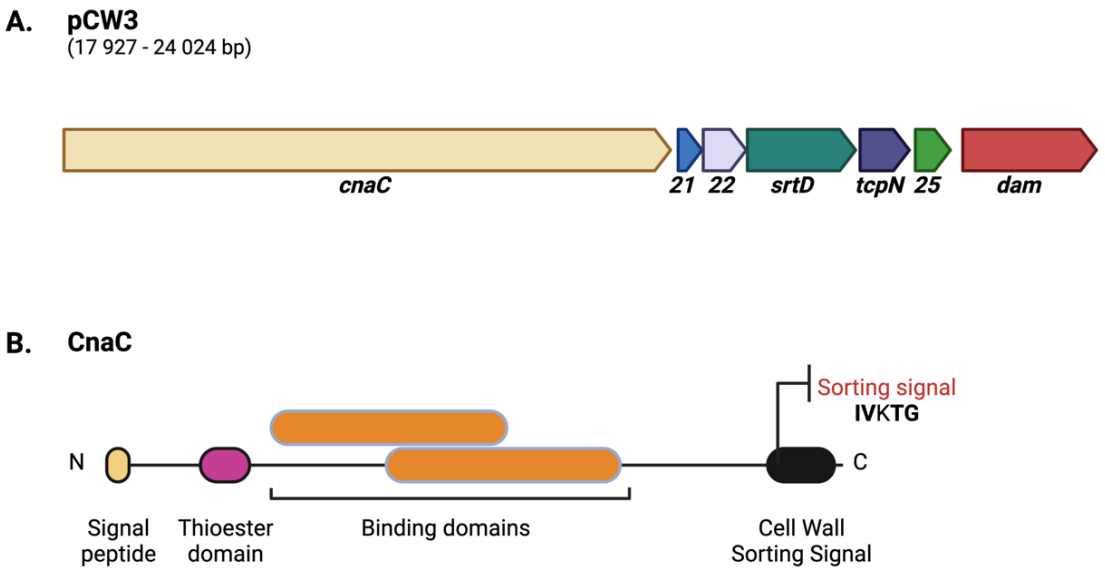


Figure 14. Schematic representation of the pCW3 *cnaC* region and CnaC$_{pCW3}$ domain organisation

A. Schematic representation of the pCW3 *cnaC* region, encoding, *cnaC*, *srtD*, *dam* and five hypothetical genes (*pcw324,* flanking the *srtD* gene, is later renamed *tcpN*). **B.** The predicted domain organization of the 1 187 aa CnaC protein. The signal peptide region is shown in yellow, a putative thioester domain in pink and the two overlapping predicted collagen binding domains are shown in orange. The location of the proposed sortase cleavage site 'IVKTG' is indicated within the cell well sorting signal shown in black.

Table 7. Defining features of the six different classes of sortases

Sortase Class	Cleavage site	Main Function	Specific bacterial genus
A	LPXTG	Surface protein anchoring	-
B	(N/S/P)PXTG	Heme uptake	-
C	LPXTG	Pili assembly	-
D	LPXTA	Spore formation	*Bacillus*
E	LAXTG	Aerial hyphae formation, Surface protein anchoring, Pilus attachment	-
F	LPXTG	Unknown	*Propionibacterium*

Conclusion

From a historical perspective, the study of plasmid biology was not only important for the knowledge that was acquired regarding these elements, but also played a prominent role in the biotechnological revolution [275]. Pioneering experiments on transformation of plasmid DNA and the development of plasmids as cloning vectors has underpinned a plethora of major discoveries in molecular genetics, and these tools are still extensively used in genetic research today [276], [277]. At present, plasmids are widely recognized as facilitators of genome restructuring, as drivers of genome plasticity and gene flow within bacterial communities [278]. Research has demonstrated that plasmid-borne genes can enhance the survival, durability and/or adaptation of their host bacterium to environmental or physiological changes, thus the ecological importance of plasmid-carriage is beyond doubt [279]–[281]. While there is a wealth of information regarding the biology and behaviour of model plasmids of G- bacteria, there is a considerable lack of knowledge regarding human-associated plasmids of G+ origin beyond those of the archetypal pathogen-associated and AMR-plasmid families.

The study of native bifidobacterial plasmids not only represents the potential to develop much needed molecular tools for use in bifidobacterial research but may also uncover novel plasmid-borne genes of potential biological importance, shed light on the host's evolution and ecology and/or reveal plasmid-mediated fitness effects on the host. These rather overlooked facets of plasmid biology could make a significant contribution to our understanding of the role(s) of plasmid-encoded functions within these bacteria and enable a more comprehensive interpretation of how bifidobacteria respond to changes in their environment within a community context.

In Chapter II we aim to elucidate the replication mechanisms of bifidobacterial megaplasmid pMP7017 and extrapolate this knowledge to develop a series of novel high-capacity cloning vectors. Chapter III focuses on the conjugative mechanisms of pMP7017 where we seek to delineate the genetic components of the plasmid's DNA secretion system and provide the first working model of conjugation for the *Bifidobacterium* genus. Chapter IV of this thesis comprises an investigation into the contribution of pMP7017-encoded cell-surface proteins in mediating adhesion to recipient cells during conjugal transfer. In chapter V we detail the *in silico* functional

annotation of pMP7017-encoded features and describe an in-depth investigation of the pMP7017 temporal transcriptome during exponential growth and further assess the impact of pMP7017-carriage on the host strain.

REFERENCES

[1] **H. Tissier**, 'Le bacterium coli et la reaction chromophile d'Escherich et le Bacterium coli', *Crit. Rev. Soc. Biol.*, no. 51, Art. no. 51, 1899.

[2] **J. A. Poupard, I. Husain, and R. F. Norris**, 'Biology of the bifidobacteria.', *Bacteriol. Rev.*, vol. 37, no. 2, pp. 136–165, Jun. 1973.

[3] **Dhanashree, S. Rajashekharan, B. Krishnaswamy, and R. Kammara**, 'Bifid Shape Is Intrinsic to Bifidobacterium adolescentis', *Front. Microbiol.*, vol. 8, p. 478, 2017, doi: 10.3389/fmicb.2017.00478.

[4] **J. F. Kim *et al.***, 'Genome Sequence of the Probiotic Bacterium Bifidobacterium animalis subsp. lactis AD011', *J. Bacteriol.*, vol. 191, no. 2, pp. 678–679, Jan. 2009, doi: 10.1128/JB.01515-08.

[5] **A. Mazé, M. O'Connell-Motherway, G. F. Fitzgerald, J. Deutscher, and D. van Sinderen**, 'Identification and characterization of a fructose phosphotransferase system in Bifidobacterium breve UCC2003', *Appl. Environ. Microbiol.*, vol. 73, no. 2, pp. 545–553, Jan. 2007, doi: 10.1128/AEM.01496-06.

[6] **S. M. Ryan, G. F. Fitzgerald, and D. van Sinderen**, 'Transcriptional regulation and characterization of a novel beta-fructofuranosidase-encoding gene from Bifidobacterium breve UCC2003', *Appl. Environ. Microbiol.*, vol. 71, no. 7, pp. 3475–3482, Jul. 2005, doi: 10.1128/AEM.71.7.3475-3482.2005.

[7] **K. Pokusaeva, G. F. Fitzgerald, and D. van Sinderen**, 'Carbohydrate metabolism in Bifidobacteria', *Genes Nutr.*, vol. 6, no. 3, Art. no. 3, Aug. 2011, doi: 10.1007/s12263-010-0206-6.

[8] **M. Ventura, F. Turroni, G. A. Lugli, and D. van Sinderen**, 'Bifidobacteria and humans: our special friends, from ecological to genomics perspectives', *J. Sci. Food Agric.*, vol. 94, no. 2, pp. 163–168, 2014, doi: 10.1002/jsfa.6356.

[9] **C. B. Wong, T. Odamaki, and J. Xiao**, 'Insights into the reason of Human-Residential Bifidobacteria (HRB) being the natural inhabitants of the human gut and their potential health-promoting benefits', *FEMS Microbiol. Rev.*, vol. 44, no. 3, pp. 369–385, May 2020, doi: 10.1093/femsre/fuaa010.

[10] **T. Odamaki *et al.***, 'Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan', *Sci. Rep.*, vol. 8, Jan. 2018, doi: 10.1038/s41598-017-18391-x.

[11] **F. Bottacini, M. Ventura, D. van Sinderen, and M. O'Connell Motherway**, 'Diversity, ecology and intestinal function of bifidobacteria', *Microb. Cell Factories*, vol. 13, no. 1, p. S4, Aug. 2014, doi: 10.1186/1475-2859-13-S1-S4.

[12] **Md. A. K. Azad, M. Sarker, T. Li, and J. Yin**, 'Probiotic Species in the Modulation of Gut Microbiota: An Overview', *BioMed Res. Int.*, vol. 2018, May 2018, doi: 10.1155/2018/9478630.

[13] **M. Derrien, A.-S. Alvarez, and W. M. de Vos**, 'The Gut Microbiota in the First Decade of Life', *Trends Microbiol.*, vol. 27, no. 12, pp. 997–1010, Dec. 2019, doi: 10.1016/j.tim.2019.08.001.

[14] **A. O'Callaghan and D. van Sinderen**, 'Bifidobacteria and Their Role as Members of the Human Gut Microbiota', *Front. Microbiol.*, vol. 7, p. 925, 2016, doi: 10.3389/fmicb.2016.00925.

[15] **S. J. Masoumi, D. Mehrabani, M. Saberifiroozi, M. R. Fattahi, F. Moradi, and M. Najafi**, 'The effect of yogurt fortified with Lactobacillus acidophilus and Bifidobacterium sp. probiotic in patients with lactose intolerance', *Food Sci. Nutr.*, vol. 9, no. 3, pp. 1704–1711, 2021, doi: 10.1002/fsn3.2145.

[16] **P. Vitellio, G. Celano, L. Bonfrate, M. Gobbetti, P. Portincasa, and M. De Angelis**, 'Effects of Bifidobacterium longum and Lactobacillus rhamnosus on Gut Microbiota in Patients with Lactose Intolerance and Persisting Functional Gastrointestinal Symptoms: A Randomised, Double-Blind, Cross-Over Study', *Nutrients*, vol. 11, no. 4, p. E886, Apr. 2019, doi: 10.3390/nu11040886.

[17] **A. Bordoni *et al.***, 'Cholesterol-lowering probiotics: in vitro selection and in vivo testing of bifidobacteria', *Appl. Microbiol. Biotechnol.*, vol. 97, no. 18, pp. 8273–8281, Sep. 2013, doi: 10.1007/s00253-013-5088-2.

[18] **F. Miremadi, M. Ayyash, F. Sherkat, and L. Stojanovska**, 'Cholesterol reduction mechanisms and fatty acid composition of cellular membranes of probiotic Lactobacilli and Bifidobacteria', *J. Funct. Foods*, vol. 9, pp. 295–305, Jul. 2014, doi: 10.1016/j.jff.2014.05.002.

[19] **I. Zanotti *et al.***, 'Evidence for cholesterol-lowering activity by Bifidobacterium bifidum PRL2010 through gut microbiota modulation', *Appl. Microbiol. Biotechnol.*, vol. 99, no. 16, pp. 6813–6829, Aug. 2015, doi: 10.1007/s00253-015-6564-7.

[20] **S. Fukuda *et al.***, 'Bifidobacteria can protect from enteropathogenic infection through production of acetate', *Nature*, vol. 469, no. 7331, Art. no. 7331, Jan. 2011, doi: 10.1038/nature09646.

[21] **L. Zhao *et al.***, 'Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes', *Science*, vol. 359, no. 6380, pp. 1151–1156, Mar. 2018, doi: 10.1126/science.aao5774.

[22] **E. Abdolalipour *et al.***, 'Evaluation of the antitumor immune responses of probiotic Bifidobacterium bifidum in human papillomavirus-induced tumor model', *Microb. Pathog.*, vol. 145, p. 104207, Aug. 2020, doi: 10.1016/j.micpath.2020.104207.

[23] **G. Longhi, D. van Sinderen, M. Ventura, and F. Turroni**, 'Microbiota and Cancer: The Emerging Beneficial Role of Bifidobacteria in Cancer Immunotherapy', *Front. Microbiol.*, vol. 11, p. 2188, 2020, doi: 10.3389/fmicb.2020.575072.

[24] **V. Matson *et al.***, 'The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients', *Science*, vol. 359, no. 6371, pp. 104–108, Jan. 2018, doi: 10.1126/science.aao3290.

[25] **A. Sivan *et al.***, 'Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy', *Science*, vol. 350, no. 6264, pp. 1084–1089, Nov. 2015, doi: 10.1126/science.aac4255.

[26] **S. Sun *et al.***, 'Bifidobacterium alters the gut microbiota and modulates the functional metabolism of T regulatory cells in the context of immune checkpoint blockade', *Proc. Natl. Acad. Sci.*, vol. 117, no. 44, pp. 27509–27515, Nov. 2020, doi: 10.1073/pnas.1921223117.

[27] **M. O'Connell Motherway *et al.***, 'A Bifidobacterial pilus-associated protein promotes colonic epithelial proliferation', *Mol. Microbiol.*, vol. 111, no. 1, pp. 287–301, 2019, doi: 10.1111/mmi.14155.

[28] **F. Turroni *et al.***, 'Role of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in modulating bacterium-host interactions', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 27, Art. no. 27, Jul. 2013, doi: 10.1073/pnas.1303897110.

[29] **A. Hickey *et al.***, 'Bifidobacterium breve Exopolysaccharide Blocks Dendritic Cell Maturation and Activation of CD4+ T Cells', *Front. Microbiol.*, vol. 12, p. 1441, 2021, doi: 10.3389/fmicb.2021.653587.

[30] **B. Sánchez *et al.***, 'The Effects of Bifidobacterium breve on Immune Mediators and Proteome of HT29 Cells Monolayers', *BioMed Res. Int.*, vol. 2015, p. 479140, 2015, doi: 10.1155/2015/479140.

[31] **D. Püngel *et al.***, 'Bifidobacterium breve UCC2003 Exopolysaccharide Modulates the Early Life Microbiota by Acting as a Potential Dietary Substrate', *Nutrients*, vol. 12, no. 4, p. 948, Mar. 2020, doi: 10.3390/nu12040948.

[32] **K. R. Hughes *et al.***, 'Bifidobacterium breve reduces apoptotic epithelial cell shedding in an exopolysaccharide and MyD88-dependent manner', *Open Biol.*, vol. 7, no. 1, p. 160155, Jan. 2017, doi: 10.1098/rsob.160155.

[33] **A. P. Allen *et al.***, 'Bifidobacterium longum 1714 as a translational psychobiotic: modulation of stress, electrophysiology and neurocognition in healthy volunteers', *Transl. Psychiatry*, vol. 6, no. 11, Art. no. 11, Nov. 2016, doi: 10.1038/tp.2016.191.

[34] **M. I. Pinto-Sanchez *et al.***, 'Probiotic Bifidobacterium longum NCC3001 Reduces Depression Scores and Alters Brain Activity: A Pilot Study in Patients With Irritable Bowel Syndrome', *Gastroenterology*, vol. 153, no. 2, pp. 448-459.e8, Aug. 2017, doi: 10.1053/j.gastro.2017.05.003.

[35] **J. Lopez-Tello *et al.***, 'Maternal gut microbiota Bifidobacterium promotes placental morphogenesis, nutrient transport and fetal growth in mice', *Cell. Mol. Life Sci.*, vol. 79, no. 7, p. 386, Jun. 2022, doi: 10.1007/s00018-022-04379-y.

[36] **J. Lederberg**, 'Cell genetics and hereditary symbiosis', *Physiol. Rev.*, vol. 32, no. 4, pp. 403–430, Oct. 1952, doi: 10.1152/physrev.1952.32.4.403.

[37] **F. Biet, Y. Cenatiempo, and C. Fremaux**, 'Identification of a Replicon from pTXL1, a Small Cryptic Plasmid from Leuconostoc mesenteroides subsp. mesenteroides Y110, and Development of a Food-Grade Vector', *Appl. Environ. Microbiol.*, vol. 68, no. 12, pp. 6451–6456, Dec. 2002, doi: 10.1128/AEM.68.12.6451-6456.2002.

[38] **T. M. Finan, B. Kunkel, G. F. De Vos, and E. R. Signer**, 'Second symbiotic megaplasmid in Rhizobium meliloti carrying exopolysaccharide and thiamine synthesis genes', *J. Bacteriol.*, vol. 167, no. 1, pp. 66–72, Jul. 1986, doi: 10.1128/jb.167.1.66-72.1986.

[39] **H. Kobori, C. W. Sullivan, and H. Shizuya**, 'Bacterial plasmids in antarctic natural microbial assemblages', *Appl. Environ. Microbiol.*, vol. 48, no. 3, pp. 515–518, Sep. 1984, doi: 10.1128/aem.48.3.515-518.1984.

[40] **W. Rozhon, E. Petutschnig, M. Khan, D. K. Summers, and B. Poppenberger**, 'Frequency and diversity of small cryptic plasmids in the genus Rahnella', *BMC Microbiol.*, vol. 10, no. 1, Art. no. 1, Dec. 2010, doi: 10.1186/1471-2180-10-56.

[41] **C. Rosenberg, F. Casse-Delbart, I. Dusha, M. David, and C. Boucher**, 'Megaplasmids in the plant-associated bacteria Rhizobium meliloti and Pseudomonas solanacearum', *J. Bacteriol.*, vol. 150, no. 1, pp. 402–406, Apr. 1982, doi: 10.1128/jb.150.1.402-406.1982.

[42] **E. Schwartz, Ed.**, *Microbial Megaplasmids*. Berlin Heidelberg: Springer-Verlag, 2009. doi: 10.1007/978-3-540-85467-8.

[43] **P. Stewart, P. A. Rosa, and K. Tilly**, 'Linear Plasmids in Bacteria: Common Origins, Uncommon Ends', in *Plasmid Biology*, John Wiley & Sons, Ltd, 2004, pp. 291–301. doi: 10.1128/9781555817732.ch13.

[44] **G. D. Solar and M. Espinosa**, 'Plasmid copy number control: an ever-growing story', *Mol. Microbiol.*, vol. 37, no. 3, pp. 492–500, 2000, doi: 10.1046/j.1365-2958.2000.02005.x.

[45] **R. G. Taylor, D. C. Walker, and R. R. McInnes**, 'E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing', *Nucleic Acids Res.*, vol. 21, no. 7, pp. 1677–1678, Apr. 1993, doi: 10.1093/nar/21.7.1677.

[46] **A. Wegrzyn, A. Czyz, M. Gabig, and G. Wegrzyn**, 'ClpP/ClpX-mediated degradation of the bacteriophage lambda O protein and regulation of lambda phage and lambda plasmid replication', *Arch. Microbiol.*, vol. 174, no. 1–2, pp. 89–96, 2000, doi: 10.1007/s002030000177.

[47] **G. Wegrzyn and A. Wegrzyn**, 'Stress responses and replication of plasmids in bacterial cells', *Microb. Cell Factories*, vol. 1, no. 1, Art. no. 1, May 2002, doi: 10.1186/1475-2859-1-2.

[48] **M. Jahn, C. Vorpahl, T. Hübschmann, H. Harms, and S. Müller**, 'Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR', *Microb. Cell Factories*, vol. 15, no. 1, Art. no. 1, Dec. 2016, doi: 10.1186/s12934-016-0610-8.

[49] **M. A. Providenti, J. M. O'Brien, R. J. Ewing, E. S. Paterson, and M. L. Smith**, 'The copy-number of plasmids and other genetic elements can be determined by SYBR-Green-based quantitative real-time PCR', *J. Microbiol. Methods*, vol. 65, no. 3, pp. 476–487, Jun. 2006, doi: 10.1016/j.mimet.2005.09.007.

[50] **R. Reyes-Lamothe *et al.***, 'High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division', *Nucleic Acids Res.*, vol. 42, no. 2, pp. 1042–1051, Jan. 2014, doi: 10.1093/nar/gkt918.

[51] **P. M. Bennett**, 'Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria', *Br. J. Pharmacol.*, vol. 153, no. Suppl 1, pp. S347–S357, Mar. 2008, doi: 10.1038/sj.bjp.0707607.

[52] **D. B. Clewell**, 'Antibiotic Resistance Plasmids in Bacteria', in *eLS*, John Wiley & Sons, Ltd, 2014. doi: 10.1002/9780470015902.a0001491.pub3.

[53] **K. Bush and G. H. Miller**, 'Bacterial enzymatic resistance: beta-lactamases and aminoglycoside-modifying enzymes', *Curr. Opin. Microbiol.*, vol. 1, no. 5, pp. 509–515, Oct. 1998, doi: 10.1016/s1369-5274(98)80082-9.

[54] **Q. Li, W. Chang, H. Zhang, D. Hu, and X. Wang**, 'The Role of Plasmids in the Multiple Antibiotic Resistance Transfer in ESBLs-Producing Escherichia coli Isolated From Wastewater Treatment Plants', *Front. Microbiol.*, vol. 10, 2019, Accessed: Jan. 23, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00633

[55] **C. M. Sedgley, D. B. Clewell, and S. E. Flannagan**, 'Plasmid pAMS1-Encoded, Bacteriocin-Related "Siblicide" in Enterococcus faecalis', *J. Bacteriol.*, vol. 191, no. 9, pp. 3183–3188, May 2009, doi: 10.1128/JB.00147-09.

[56] **S. Silver**, 'Bacterial resistances to toxic metal ions--a review', *Gene*, vol. 179, no. 1, pp. 9–19, Nov. 1996, doi: 10.1016/s0378-1119(96)00323-x.

[57] **C. Stephens and W. Murray**, 'Pathogen evolution: How good bacteria go bad', *Curr. Biol.*, vol. 11, no. 2, pp. R53–R56, Jan. 2001, doi: 10.1016/S0960-9822(01)00012-4.

[58] **J. Boyd, J. Williams, B. Curtis, C. Kozera, R. Singh, and M. Reith**, 'Three small, cryptic plasmids from Aeromonas salmonicida subsp. salmonicida A449', *Plasmid*, vol. 50, no. 2, pp. 131–144, Sep. 2003, doi: 10.1016/s0147-619x(03)00058-1.

[59] **G. del Solar, R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas**, 'Replication and Control of Circular Bacterial Plasmids', *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, pp. 434–464, Jun. 1998.

[60] **D. E. Bussiere and D. Bastia**, 'Termination of DNA replication of bacterial and plasmid chromosomes', *Mol. Microbiol.*, vol. 31, no. 6, pp. 1611–1618, 1999, doi: 10.1046/j.1365-2958.1999.01287.x.

[61] **G. del Solar, M. Moscoso, and M. Espinosa**, 'Rolling circle-replicating plasmids from gram-positive and gram-negative bacteria: a wall falls', *Mol. Microbiol.*, vol. 8, no. 5, pp. 789–796, May 1993, doi: 10.1111/j.1365-2958.1993.tb01625.x.

[62] **M. Espinosa, G. del Solar, F. Rojo, and J. C. Alonso**, 'Plasmid rolling circle replication and its control', *FEMS Microbiol. Lett.*, vol. 130, no. 2–3, pp. 111–120, Aug. 1995, doi: 10.1111/j.1574-6968.1995.tb07707.x.

[63] **H. te Riele, B. Michel, and S. D. Ehrlich**, 'Single-stranded plasmid DNA in Bacillus subtilis and Staphylococcus aureus.', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 83, no. 8, pp. 2541–2545, Apr. 1986.

[64] **A. Gruss and S. D. Ehrlich**, 'The family of highly interrelated single-stranded deoxyribonucleic acid plasmids.', *Microbiol. Rev.*, vol. 53, no. 2, pp. 231–241, Jun. 1989.

[65] **E. Scherzinger, V. Haring, R. Lurz, and S. Otto**, 'Plasmid RSF1010 DNA replication in vitro promoted by purified RSF1010 RepA, RepB and RepC proteins.', *Nucleic Acids Res.*, vol. 19, no. 6, pp. 1203–1211, Mar. 1991.

[66] **V. HARING *et al.***, 'Protein RepC is involved in copy number control of the broad host range plasmid RSF1010', *Protein RepC Involv. Copy Number Control Broad Host Range Plasmid RSF1010*, vol. 82, no. 18, pp. 6090–6094, 1985.

[67] **E. Spanier**, 'Promiscuous plasmids of gram-negative bacteria. London: Academic Press Ltd., 1989. 276 pp., Ł 29.50, ISBN 0-12-688480-3', *Acta Biotechnol.*, vol. 10, no. 6, pp. 563–563, 1990, doi: 10.1002/abio.370100621.

[68] **Y. Honda, H. Sakai, T. Komano, and M. Bagdasarian**, 'RepB' is required in trans for the two single-strand DNA initiation signals in oriV of plasmid RSF1010', *Gene*, vol. 80, no. 1, pp. 155–159, Aug. 1989, doi: 10.1016/0378-1119(89)90261-8.

[69] **Y. Honda, H. Sakai, H. Hiasa, K. Tanaka, T. Komano, and M. Bagdasarian**, 'Functional division and reconstruction of a plasmid replication origin: molecular dissection of the oriV of the broad-host-range plasmid RSF1010.', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 88, no. 1, pp. 179–183, Jan. 1991.

[70] **R. Kiewiet, S. Bron, K. de Jonge, G. Venema, and J. F. M. L. Seegers**, 'Theta replication of the lactococcal plasmid pWVO2', *Mol. Microbiol.*, vol. 10, no. 2, pp. 319–327, 1993, doi: 10.1111/j.1365-2958.1993.tb01958.x.

[71] **C. Bruand, E. Le Chatelier, S. D. Ehrlich, and L. Jannière**, 'A fourth class of theta-replicating plasmids: the pAM beta 1 family from gram-positive bacteria', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 24, pp. 11668–11672, Dec. 1993, doi: 10.1073/pnas.90.24.11668.

[72] **A. Kornberg and T. A. Baker**, *DNA Replication*. University Science Books, 2005.

[73] **T. Horii and T. Itoh**, 'Replication of ColE2 and ColE3 plasmids: the regions sufficient for autonomous replication', *Mol. Gen. Genet. MGG*, vol. 212, no. 2, pp. 225–231, May 1988, doi: 10.1007/BF00334689.

[74] **P. Wawrzyniak, G. Płucienniczak, and D. Bartosik**, 'The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins', *Front. Microbiol.*, vol. 8, 2017, doi: 10.3389/fmicb.2017.02353.

[75] **S. Austin, M. Ziese, and N. Sternberg**, 'A novel role for site-specific recombination in maintenance of bacterial replicons', *Cell*, vol. 25, no. 3, pp. 729–736, Sep. 1981, doi: 10.1016/0092-8674(81)90180-x.

[76] **D. Summers**, 'Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability', *Mol. Microbiol.*, vol. 29, no. 5, pp. 1137–1145, Sep. 1998, doi: 10.1046/j.1365-2958.1998.01012.x.

[77] **B. E. Funnell**, 'ParB Partition Proteins: Complex Formation and Spreading at Bacterial and Plasmid Centromeres', *Front. Mol. Biosci.*, vol. 3, p. 44, Aug. 2016, doi: 10.3389/fmolb.2016.00044.

[78] **K. Gerdes, J. Møller-Jensen, and R. B. Jensen**, 'Plasmid and chromosome partitioning: surprises from phylogeny', *Mol. Microbiol.*, vol. 37, no. 3, pp. 455–466, 2000, doi: 10.1046/j.1365-2958.2000.01975.x.

[79] **J. Møller-Jensen and K. Gerdes**, 'Plasmid segregation: spatial awareness at the molecular level', *J. Cell Biol.*, vol. 179, no. 5, pp. 813–815, Dec. 2007, doi: 10.1083/jcb.200710192.

[80] **M. Dmowski and G. Jagura-Burdzy**, 'Active stable maintenance functions in low copy-number plasmids of Gram-positive bacteria I. Partition systems', *Pol. J. Microbiol.*, vol. 62, no. 1, pp. 3–16, 2013.

[81] **M. Dmowski, I. Sitkiewicz, and P. Cegłowski**, 'Characterization of a Novel Partition System Encoded by the δ and ω Genes from the Streptococcal Plasmid pSM19035', *J. Bacteriol.*, vol. 188, no. 12, pp. 4362–4372, Jun. 2006, doi: 10.1128/JB.01922-05.

[82] **M. V. Francia, K. E. Weaver, P. Goicoechea, P. Tille, and D. B. Clewell**, 'Characterization of an active partition system for the Enterococcus faecalis pheromone-responding plasmid pAD1', *J. Bacteriol.*, vol. 189, no. 23, pp. 8546–8555, Dec. 2007, doi: 10.1128/JB.00719-07.

[83] **D. Popp *et al.***, 'Structure and filament dynamics of the pSK41 actin-like ParM protein: implications for plasmid DNA segregation', *J. Biol. Chem.*, vol. 285, no. 13, pp. 10130–10140, Mar. 2010, doi: 10.1074/jbc.M109.071613.

[84] **T. Tanaka**, 'Functional analysis of the stability determinant AlfB of pBET131, a miniplasmid derivative of bacillus subtilis (natto) plasmid pLS32', *J. Bacteriol.*, vol. 192, no. 5, pp. 1221–1230, Mar. 2010, doi: 10.1128/JB.01312-09.

[85] **R. A. Larsen, C. Cusumano, A. Fujioka, G. Lim-Fong, P. Patterson, and J. Pogliano**, 'Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis', *Genes Dev.*, vol. 21, no. 11, pp. 1340–1352, Jun. 2007, doi: 10.1101/gad.1546107.

[86] **N. Fraikin, F. Goormaghtigh, and L. V. Melderen**, 'Type II Toxin-Antitoxin Systems: Evolution and Revolutions', *J. Bacteriol.*, vol. 202, no. 7, Mar. 2020, doi: 10.1128/JB.00763-19.

[87] **J. Peltier *et al.***, 'Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in Clostridioides difficile', *Commun. Biol.*, vol. 3, no. 1, Art. no. 1, Nov. 2020, doi: 10.1038/s42003-020-01448-5.

[88] **G. Singh, M. Yadav, C. Ghosh, and J. S. Rathore**, 'Bacterial toxin-antitoxin modules: classification, functions, and association with persistence', *Curr. Res. Microb. Sci.*, vol. 2, p. 100047, Dec. 2021, doi: 10.1016/j.crmicr.2021.100047.

[89] **M. Sengupta and S. Austin**, 'Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria', *Infect. Immun.*, vol. 79, no. 7, pp. 2502–2509, Jul. 2011, doi: 10.1128/IAI.00127-11.

[90] **K. Gerdes *et al.***, 'Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon', *EMBO J.*, vol. 5, no. 8, pp. 2023–2029, Aug. 1986, doi: 10.1002/j.1460-2075.1986.tb04459.x.

[91] **J. Vogel, L. Argaman, E. G. H. Wagner, and S. Altuvia**, 'The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide', *Curr. Biol. CB*, vol. 14, no. 24, pp. 2271–2276, Dec. 2004, doi: 10.1016/j.cub.2004.12.003.

[92] **P. Bernard and M. Couturier**, 'Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes', *J. Mol. Biol.*, vol. 226, no. 3, pp. 735–745, Aug. 1992, doi: 10.1016/0022-2836(92)90629-x.

[93] **Y. Jiang, J. Pogliano, D. R. Helinski, and I. Konieczny**, 'ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase', *Mol. Microbiol.*, vol. 44, no. 4, pp. 971–979, 2002, doi: 10.1046/j.1365-2958.2002.02921.x.

[94] **Y. Zhang, J. Zhang, K. P. Hoeflich, M. Ikura, G. Qing, and M. Inouye**, 'MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli', *Mol. Cell*, vol. 12, no. 4, pp. 913–923, Oct. 2003, doi: 10.1016/s1097-2765(03)00402-7.

[95] **J. W. Cruz, F. P. Rothenbacher, T. Maehigashi, W. S. Lane, C. M. Dunham, and N. A. Woychik**, 'Doc toxin is a kinase that inactivates elongation factor Tu', *J. Biol. Chem.*, vol. 289, no. 11, pp. 7788–7798, Mar. 2014, doi: 10.1074/jbc.M113.544429.

[96] **E. Germain, D. Castro-Roa, N. Zenkin, and K. Gerdes**, 'Molecular mechanism of bacterial persistence by HipA', *Mol. Cell*, vol. 52, no. 2, pp. 248–254, Oct. 2013, doi: 10.1016/j.molcel.2013.08.045.

[97] **P. C. Fineran, T. R. Blower, I. J. Foulds, D. P. Humphreys, K. S. Lilley, and G. P. C. Salmond**, 'The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 3, pp. 894–899, Jan. 2009, doi: 10.1073/pnas.0808832106.

[98] **H. Masuda, Q. Tan, N. Awano, K.-P. Wu, and M. Inouye**, 'YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in Escherichia coli', *Mol. Microbiol.*, vol. 84, no. 5, pp. 979–989, Jun. 2012, doi: 10.1111/j.1365-2958.2012.08068.x.

[99] **X. Wang *et al.***, 'A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS', *Nat. Chem. Biol.*, vol. 8, no. 10, pp. 855–861, Oct. 2012, doi: 10.1038/nchembio.1062.

[100] **M. Markovski and S. Wickner**, 'Preventing bacterial suicide: A novel toxin-antitoxin strategy', *Mol. Cell*, vol. 52, no. 5, pp. 611–612, Dec. 2013, doi: 10.1016/j.molcel.2013.11.018.

[101] **O. Marimon *et al.***, 'An oxygen-sensitive toxin–antitoxin system', *Nat. Commun.*, vol. 7, p. 13634, Dec. 2016, doi: 10.1038/ncomms13634.

[102] **J. S. Choi *et al.***, 'The small RNA, SdsR, acts as a novel type of toxin in Escherichia coli', *RNA Biol.*, vol. 15, no. 10, pp. 1319–1335, Oct. 2018, doi: 10.1080/15476286.2018.1532252.

[103] **S. Arboleya *et al.***, 'Gene-trait matching across the Bifidobacterium longum pan-genome reveals considerable diversity in carbohydrate catabolism among human infant strains', *BMC Genomics*, vol. 19, Jan. 2018, doi: 10.1186/s12864-017-4388-9.

[104] **F. Bottacini *et al.***, 'Comparative genomics and genotype-phenotype associations in Bifidobacterium breve', *Sci. Rep.*, vol. 8, no. 1, p. 10633, Jul. 2018, doi: 10.1038/s41598-018-28919-4.

[105] **W. Mancino, G. A. Lugli, D. van Sinderen, M. Ventura, and F. Turroni**, 'Mobilome and Resistome Reconstruction from Genomes Belonging to Members of the Bifidobacterium Genus', *Microorganisms*, vol. 7, no. 12, Dec. 2019, doi: 10.3390/microorganisms7120638.

[106] **F. Bottacini *et al.***, 'Discovery of a Conjugative Megaplasmid in Bifidobacterium breve', *Appl. Environ. Microbiol.*, vol. 81, no. 1, pp. 166–176, Jan. 2015, doi: 10.1128/AEM.02871-14.

[107] **S. Guglielmetti, B. Mayo, and P. Álvarez-Martín**, 'Mobilome and genetic modification of bifidobacteria', *Benef. Microbes*, vol. 4, no. 2, pp. 143–166, Jun. 2013, doi: 10.3920/BM2012.0031.

[108] **M. J. Gibbs, V. V. Smeianov, J. L. Steele, P. Upcroft, and B. A. Efimov**, 'Two families of rep-like genes that probably originated by interspecies recombination are represented in viral, plasmid, bacterial, and parasitic protozoan genomes', *Mol. Biol. Evol.*, vol. 23, no. 6, pp. 1097–1100, Jun. 2006, doi: 10.1093/molbev/msj122.

[109] **K. O'Riordan and G. F. Fitzgerald**, 'Molecular characterisation of a 5.75-kb cryptic plasmid from Bifidobacterium breve NCFB 2258 and determination of mode of replication', *FEMS Microbiol. Lett.*, vol. 174, no. 2, pp. 285–294, May 1999, doi: 10.1111/j.1574-6968.1999.tb13581.x.

[110] **G. Muth, M. Farr, V. Hartmann, and W. Wohlleben**, 'Streptomyces ghanaensis Plasmid pSG5: Nucleotide Sequence Analysis of the Self-Transmissible Minimal Replicon and Characterization of the Replication Mode', *Plasmid*, vol. 33, no. 2, pp. 113–126, Mar. 1995, doi: 10.1006/plas.1995.1013.

[111] **A. Sangrador-Vegas, C. Stanton, D. van Sinderen, G. F. Fitzgerald, and R. P. Ross**, 'Characterization of plasmid pASV479 from Bifidobacterium pseudolongum subsp. globosum and its use for expression vector construction', *Plasmid*, vol. 58, no. 2, pp. 140–147, Sep. 2007, doi: 10.1016/j.plasmid.2007.02.004.

[112] **J.-H. Lee and D. J. O'Sullivan**, 'Sequence Analysis of Two Cryptic Plasmids from Bifidobacterium longum DJO10A and Construction of a Shuttle Cloning Vector', *Appl. Environ. Microbiol.*, vol. 72, no. 1, pp. 527–535, Jan. 2006, doi: 10.1128/AEM.72.1.527-535.2006.

[113] **M. Takahata *et al.***, 'Complete sequence analysis of two cryptic plasmids from Bifidobacterium kashiwanohense JCM 15439 (type strain) isolated from healthy infant feces', *Anim. Sci. J. Nihon Chikusan Gakkaiho*, vol. 85, no. 2, pp. 158–163, Feb. 2014, doi: 10.1111/asj.12095.

[114] **P. Alvarez-Martín, M. O'Connell-Motherway, D. van Sinderen, and B. Mayo**, 'Functional analysis of the pBC1 replicon from Bifidobacterium catenulatum L48', *Appl. Microbiol. Biotechnol.*, vol. 76, no. 6, pp. 1395–1402, Oct. 2007, doi: 10.1007/s00253-007-1115-5.

[115] **G.-S. Moon, U. Wegmann, A. P. Gunning, M. J. Gasson, and A. Narbad**, 'Isolation and characterization of a theta-type cryptic plasmid from Bifidobacterium longum FI10564', *J. Microbiol. Biotechnol.*, vol. 19, no. 4, pp. 403–408, Apr. 2009, doi: 10.4014/jmb.0806.378.

[116] **D. Matteuzzi, P. Brigidi, M. Rossi, and D. Di**, 'Characterization and molecular cloning of Bifidobacterium longum cryptic plasmid pMB1', *Lett. Appl. Microbiol.*, vol. 11, no. 4, pp. 220–223, Oct. 1990, doi: 10.1111/j.1472-765x.1990.tb00165.x.

[117] **M. Rossi, P. Brigidi, A. Gonzalez Vara y Rodriguez, and D. Matteuzzi**, 'Characterization of the plasmid pMB1 from Bifidobacterium longum and its use for shuttle vector construction', *Res. Microbiol.*, vol. 147, no. 3, pp. 133–143, Jan. 1996, doi: 10.1016/0923-2508(96)80213-0.

[118] **R. L. Dineen, C. Penno, P. Kelleher, M. J. B. Bourin, M. O'Connell-Motherway, and D. van Sinderen**, 'Molecular analysis of the replication functions of the bifidobacterial conjugative megaplasmid pMP7017', *Microb. Biotechnol.*, May 2021, doi: 10.1111/1751-7915.13810.

[119] **M. Cronin, M. Knobel, M. O'Connell-Motherway, G. F. Fitzgerald, and D. van Sinderen**, 'Molecular Dissection of a Bifidobacterial Replicon', *Appl. Environ. Microbiol.*, vol. 73, no. 24, Art. no. 24, Dec. 2007, doi: 10.1128/AEM.01630-07.

[120] **N. Corneau, E. Emond, and G. LaPointe**, 'Molecular characterization of three plasmids from Bifidobacterium longum', *Plasmid*, vol. 51, no. 2, pp. 87–100, Mar. 2004, doi: 10.1016/j.plasmid.2003.12.003.

[121] **M. S. Park, D. W. Shin, K. H. Lee, and G. E. Ji**, 'Sequence analysis of plasmid pKJ50 from Bifidobacterium longum', *Microbiol. Read. Engl.*, vol. 145 (Pt 3), pp. 585–592, Mar. 1999, doi: 10.1099/13500872-145-3-585.

[122] **N.-S. Park, D.-W. Shin, K.-H. Lee, and G.-E. Ji**, 'Characterization of Plasmid pKJ36 from Bifidobacterium longum and Construction of an E. coli-Bifidobacterium Shuttle Vector', *J. Microbiol. Biotechnol.*, vol. 10, no. 3, pp. 312–320, 2000.

[123] **M.-S. Park, H.-W. Moon, and G.-E. Ji**, 'Molecular Characterization of Plasmid from Bifidobacterium longum', *J. Microbiol. Biotechnol.*, vol. 13, no. 3, pp. 457–462, 2003.

[124] **P. Alvarez-Martín, J. Zycka-Krzesińska, J. Bardowski, and B. Mayo**, 'Sequence analysis of plasmid pSP02 from Bifidobacterium longum M62 and construction of pSP02-derived cloning vectors', *Plasmid*, vol. 69, no. 2, pp. 119–126, Mar. 2013, doi: 10.1016/j.plasmid.2012.11.006.

[125] **J.-H. Lee and D. J. O'Sullivan**, 'Sequence Analysis of Two Cryptic Plasmids from Bifidobacterium longum DJO10A and Construction of a Shuttle Cloning Vector', *Appl. Environ. Microbiol.*, vol. 72, no. 1, Art. no. 1, Jan. 2006, doi: 10.1128/AEM.72.1.527-535.2006.

[126] **G.-F. Fu, Y. Yin, B. Hu, and G.-X. Xu**, 'Bifidobacterium as a Delivery System of Functional Genes for Cancer Gene Therapy', in *Emerging Cancer Therapy*, John Wiley & Sons, Ltd, 2010, pp. 99–117. doi: 10.1002/9780470626528.ch5.

[127] **M. N. Huda *et al.***, 'Bifidobacterium Abundance in Early Infancy and Vaccine Response at 2 Years of Age', *Pediatrics*, vol. 143, no. 2, Art. no. 2, Feb. 2019, doi: 10.1542/peds.2018-1489.

[128] **M. O. Motherway, J. O'Driscoll, G. F. Fitzgerald, and D. V. Sinderen**, 'Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in Bifidobacterium breve UCC2003', *Microb. Biotechnol.*, vol. 2, no. 3, pp. 321–332, 2009, doi: 10.1111/j.1751-7915.2008.00071.x.

[129] **K. Sakaguchi, J. He, S. Tani, Y. Kano, and T. Suzuki**, 'A targeted gene knockout method using a newly constructed temperature-sensitive plasmid mediated homologous recombination in Bifidobacterium longum', *Appl. Microbiol. Biotechnol.*, vol. 95, no. 2, Art. no. 2, Jul. 2012, doi: 10.1007/s00253-012-4090-4.

[130] **M. O. Motherway, G. F. Fitzgerald, S. Neirynck, S. Ryan, L. Steidler, and D. van Sinderen**, 'Characterization of ApuB, an Extracellular Type II Amylopullulanase from Bifidobacterium breve UCC2003', *Appl. Environ. Microbiol.*, vol. 74, no. 20, pp. 6271–6279, Oct. 2008, doi: 10.1128/AEM.01169-08.

[131] **H. Sakurama *et al.***, 'Lacto-N-biosidase Encoded by a Novel Gene of Bifidobacterium longum Subspecies longum Shows Unique Substrate Specificity and Requires a Designated Chaperone for Its Active Expression', *J. Biol. Chem.*, vol. 288, no. 35, pp. 25194–25206, Aug. 2013, doi: 10.1074/jbc.M113.484733.

[132] **C. Hidalgo-Cantabrana *et al.***, 'A Single Mutation in the Gene Responsible for the Mucoid Phenotype of Bifidobacterium animalis subsp. lactis Confers Surface and Functional Characteristics', *Appl. Environ. Microbiol.*, vol. 81, no. 23, pp. 7960–7968, Dec. 2015, doi: 10.1128/AEM.02095-15.

[133] **Y. Hirayama *et al.***, 'Development of a Double-Crossover Markerless Gene Deletion System in Bifidobacterium longum: Functional Analysis of the α-Galactosidase Gene for Raffinose Assimilation', *Appl. Environ. Microbiol.*, vol. 78, no. 14, pp. 4984–4994, Jul. 2012, doi: 10.1128/AEM.00588-12.

[134] **L. Ruiz, M. O. Motherway, N. Lanigan, and D. van Sinderen**, 'Transposon Mutagenesis in Bifidobacterium breve: Construction and Characterization of a Tn5 Transposon Mutant Library for Bifidobacterium breve UCC2003', *PLOS ONE*, vol. 8, no. 5, Art. no. 5, May 2013, doi: 10.1371/journal.pone.0064699.

[135] **L. Ruiz *et al.***, 'The essential genomic landscape of the commensal Bifidobacterium breve UCC2003', *Sci. Rep.*, vol. 7, no. 1, Art. no. 1, Jul. 2017, doi: 10.1038/s41598-017-05795-y.

[136] **M. Sakanaka *et al.***, 'A Transposon Mutagenesis System for Bifidobacterium longum subsp. longum Based on an IS3 Family Insertion Sequence, ISBlo11', *Appl. Environ. Microbiol.*, vol. 84, no. 17, 01 2018, doi: 10.1128/AEM.00824-18.

[137] **S. M. Kelly *et al.***, 'Bifidobacterial biofilm formation is a multifactorial adaptive phenomenon in response to bile exposure', *Sci. Rep.*, vol. 10, p. 11598, Jul. 2020, doi: 10.1038/s41598-020-68179-9.

[138] **N. Lanigan, F. Bottacini, P. G. Casey, M. O'Connell Motherway, and D. van Sinderen**, 'Genome-Wide Search for Genes Required for Bifidobacterial Growth under Iron-Limitation', *Front. Microbiol.*, vol. 8, p. 964, 2017, doi: 10.3389/fmicb.2017.00964.

[139] **M. Pan *et al.***, 'Genomic and epigenetic landscapes drive CRISPR-based genome editing in Bifidobacterium', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 119, no. 30, p. e2205068119, Jul. 2022, doi: 10.1073/pnas.2205068119.

[140] **K. Yasui *et al.***, 'Improvement of bacterial transformation efficiency using plasmid artificial modification', *Nucleic Acids Res.*, vol. 37, no. 1, p. e3, Jan. 2009, doi: 10.1093/nar/gkn884.

[141] **A. O'Callaghan, F. Bottacini, M. O'Connell Motherway, and D. van Sinderen**, 'Pangenome analysis of Bifidobacterium longum and site-directed mutagenesis through by-pass of restriction-modification systems', *BMC Genomics*, vol. 16, Oct. 2015, doi: 10.1186/s12864-015-1968-4.

[142] **F. Bottacini *et al.***, 'Comparative genome and methylome analysis reveals restriction/modification system diversity in the gut commensal Bifidobacterium breve', *Nucleic Acids Res.*, vol. 46, no. 4, pp. 1860–1877, Feb. 2018, doi: 10.1093/nar/gkx1289.

[143] **E. C. Hoedt *et al.***, 'Broad Purpose Vector for Site-Directed Insertional Mutagenesis in Bifidobacterium breve', *Front. Microbiol.*, vol. 12, 2021, doi: 10.3389/fmicb.2021.636822.

[144] **K. Tanaka, K. Samura, and Y. Kano**, 'Structural and functional analysis of pTB6 from Bifidobacterium longum', *Biosci. Biotechnol. Biochem.*, vol. 69, no. 2, pp. 422–425, Feb. 2005, doi: 10.1271/bbb.69.422.

[145] **P. Álvarez-Martín, A. B. Flórez, A. Margolles, G. del Solar, and B. Mayo**, 'Improved Cloning Vectors for Bifidobacteria, Based on the Bifidobacterium catenulatum pBC1 Replicon', *Appl Env. Microbiol*, vol. 74, no. 15, Art. no. 15, Aug. 2008, doi: 10.1128/AEM.00074-08.

[146] **M. S. Park, H. W. Moon, and G. E. Ji**, 'Molecular Characterization of Plasmid from Bifidobacterium longum', vol. 13, no. 3, pp. 457–462, Jun. 2003.

[147] **L. Ruiz, M. O'Connell-Motherway, A. Zomer, C. G. de los Reyes-Gavilán, A. Margolles, and D. van Sinderen**, 'A bile-inducible membrane protein mediates bifidobacterial bile resistance', *Microb. Biotechnol.*, vol. 5, no. 4, pp. 523–535, Jul. 2012, doi: 10.1111/j.1751-7915.2011.00329.x.

[148] **R. T. Long *et al.***, 'Bifidobacterium as an oral delivery carrier of oxyntomodulin for obesity therapy: inhibitory effects on food intake and body weight in overweight mice', *Int. J. Obes. 2005*, vol. 34, no. 4, pp. 712–719, Apr. 2010, doi: 10.1038/ijo.2009.277.

[149] **A. Klijn, D. Moine, M. Delley, A. Mercenier, F. Arigoni, and R. D. Pridmore**, 'Construction of a Reporter Vector for the Analysis of Bifidobacterium longum Promoters', *Appl. Environ. Microbiol.*, vol. 72, no. 11, pp. 7401–7405, Nov. 2006, doi: 10.1128/AEM.01611-06.

[150] **X. Li *et al.***, 'Bifidobacterium adolescentis as a delivery system of endostatin for cancer gene therapy: selective inhibitor of angiogenesis and hypoxic tumor growth', *Cancer Gene Ther.*, vol. 10, no. 2, pp. 105–111, Feb. 2003, doi: 10.1038/sj.cgt.7700530.

[151] **M. Sakanaka *et al.***, 'Functional analysis of bifidobacterial promoters in Bifidobacterium longum and Escherichia coli using the α-galactosidase gene as a reporter', *J. Biosci. Bioeng.*, vol. 118, no. 5, pp. 489–495, Nov. 2014, doi: 10.1016/j.jbiosc.2014.05.002.

[152] **A. Mauras *et al.***, 'A New Bifidobacteria Expression SysTem (BEST) to Produce and Deliver Interleukin-10 in Bifidobacterium bifidum', *Front. Microbiol.*, vol. 9, 2018, doi: 10.3389/fmicb.2018.03075.

[153] **M. Cronin, A. Zomer, G. Fitzgerald, and D. van Sinderen**, 'Identification of iron-regulated genes of Bifidobacterium breve UCC2003 as a basis for controlled gene expression', *Bioengineered*, vol. 3, no. 3, pp. 159–169, May 2012, doi: 10.4161/bbug.18985.

[154] **P. Alvarez-Martin *et al.***, 'A Conserved Two-Component Signal Transduction System Controls the Response to Phosphate Starvation in Bifidobacterium breve UCC2003', *Appl. Environ. Microbiol.*, vol. 78, no. 15, pp. 5258–5269, Aug. 2012, doi: 10.1128/AEM.00804-12.

[155] **A. N. Shkoporov, B. A. Efimov, E. V. Khokhlova, J. L. Steele, L. I. Kafarskaia, and V. V. Smeianov**, 'Characterization of plasmids from human infant Bifidobacterium strains: Sequence analysis and construction of E. coli–Bifidobacterium shuttle vectors', *Plasmid*, vol. 60, no. 2, pp. 136–148, Sep. 2008, doi: 10.1016/j.plasmid.2008.06.005.

[156] **Z. Sun, C. Westermann, J. Yuan, and C. U. Riedel**, 'Experimental determination and characterization of the gap promoter of Bifidobacterium bifidum S17', *Bioengineered*, vol. 5, no. 6, pp. 371–377, 2014, doi: 10.4161/bioe.34423.

[157] **T. NAKAMURA *et al.***, 'Cloned Cytosine Deaminase Gene Expression of Bifidobacterium longum and Application to Enzyme/Pro-drug Therapy of Hypoxic Solid Tumors', *Biosci. Biotechnol. Biochem.*, vol. 66, no. 11, pp. 2362–2366, Jan. 2002, doi: 10.1271/bbb.66.2362.

[158] **G.-S. Moon, Y.-R. Pyun, M. S. Park, G. E. Ji, and W. J. Kim**, 'Secretion of Recombinant Pediocin PA-1 by Bifidobacterium longum, Using the Signal Sequence for Bifidobacterial α-Amylase', *Appl. Environ. Microbiol.*, vol. 71, no. 9, pp. 5630–5632, Sep. 2005, doi: 10.1128/AEM.71.9.5630-5632.2005.

[159] **M. S. Park, B. Kwon, J. J. Shim, C. S. Huh, and G. E. Ji**, 'Heterologous expression of cholesterol oxidase in Bifidobacterium longum under the control of 16S rRNA gene promoter of bifidobacteria', *Biotechnol. Lett.*, vol. 30, no. 1, pp. 165–172, Jan. 2008, doi: 10.1007/s10529-007-9514-3.

[160] **M. Cronin, R. D. Sleator, C. Hill, G. F. Fitzgerald, and D. van Sinderen**, 'Development of a luciferase-based reporter system to monitor Bifidobacterium breve UCC2003 persistence in mice', *BMC Microbiol.*, vol. 8, no. 1, Art. no. 1, Sep. 2008, doi: 10.1186/1471-2180-8-161.

[161] **S. Guglielmetti, A. Ciranna, D. Mora, C. Parini, and M. Karp**, 'Construction, characterization and exemplificative application of bioluminescent Bifidobacterium longum biovar longum', *Int. J. Food Microbiol.*, vol. 124, no. 3, pp. 285–290, Jun. 2008, doi: 10.1016/j.ijfoodmicro.2008.03.033.

[162] **Y. Wang, J. Y. Kim, M. S. Park, and G. E. Ji**, 'Novel Bifidobacterium promoters selected through microarray analysis lead to constitutive high-level gene expression', *J. Microbiol. Seoul Korea*, vol. 50, no. 4, pp. 638–643, Aug. 2012, doi: 10.1007/s12275-012-1591-x.

[163] **C. Montenegro-Rodríguez, A. Peirotén, A. Sanchez-Jimenez, J. L. Arqués, and J. M. Landete**, 'Analysis of gene expression of bifidobacteria using as the reporter an anaerobic fluorescent protein', *Biotechnol. Lett.*, vol. 37, no. 7, pp. 1405–1413, Jul. 2015, doi: 10.1007/s10529-015-1802-8.

[164] **N. Castro-Bravo, C. Hidalgo-Cantabrana, M. A. Rodriguez-Carvajal, P. Ruas-Madiedo, and A. Margolles**, 'Gene Replacement and Fluorescent Labeling to Study the Functional Role of Exopolysaccharides in Bifidobacterium animalis subsp. lactis', *Front. Microbiol.*, vol. 8, p. 1405, Jul. 2017, doi: 10.3389/fmicb.2017.01405.

[165] **J. M. Landete, S. Langa, C. Revilla, A. Margolles, M. Medina, and J. L. Arqués**, 'Use of anaerobic green fluorescent protein versus green fluorescent protein as reporter in lactic acid bacteria', *Appl. Microbiol. Biotechnol.*, vol. 99, no. 16, pp. 6865–6877, Aug. 2015, doi: 10.1007/s00253-015-6770-3.

[166] **T. Kozakai, Y. Shimofusa, I. Nomura, and T. Suzuki**, 'Construction of a reporter system for bifidobacteria using chloramphenicol acetyltransferase and its application for evaluation of promoters and terminators', *Biosci. Microbiota Food Health*, vol. 40, no. 2, pp. 115–122, 2021, doi: 10.12938/bmfh.2020-070.

[167] **T. Kozakai *et al.***, 'Structure of a Core Promoter in Bifidobacterium longum NCC2705', *J. Bacteriol.*, vol. 202, no. 7, Mar. 2020, doi: 10.1128/JB.00540-19.

[168] **S. McGrath, G. F. Fitzgerald, and D. van Sinderen**, 'Improvement and optimization of two engineered phage resistance mechanisms in Lactococcus lactis', *Appl. Environ. Microbiol.*, vol. 67, no. 2, pp. 608–616, Feb. 2001, doi: 10.1128/AEM.67.2.608-616.2001.

[169] **A. N. Shkoporov, B. A. Efimov, E. V. Khokhlova, L. I. Kafarskaia, and V. V. Smeianov**, 'Production of human basic fibroblast growth factor (FGF-2) in Bifidobacterium breve using a series of novel expression/secretion vectors', *Biotechnol. Lett.*, vol. 30, no. 11, pp. 1983–1988, Nov. 2008, doi: 10.1007/s10529-008-9772-8.

[170] **M. L. Reyes Escogido, A. De León Rodríguez, and A. P. Barba de la Rosa**, 'A novel binary expression vector for production of human IL-10 in Escherichia coli and Bifidobacterium longum', *Biotechnol. Lett.*, vol. 29, no. 8, pp. 1249–1253, Aug. 2007, doi: 10.1007/s10529-007-9376-8.

[171] **E. V. Khokhlova, B. A. Efimov, L. I. Kafarskaia, and A. N. Shkoporov**, 'Heterologous expression of secreted biologically active human interleukin-10 in Bifidobacterium breve', *Arch. Microbiol.*, vol. 192, no. 9, pp. 769–774, Sep. 2010, doi: 10.1007/s00203-010-0606-4.

[172] **A. Osswald, C. Westermann, Z. Sun, and C. U. Riedel**, 'A Phytase-Based Reporter System for Identification of Functional Secretion Signals in Bifidobacteria', *PLOS ONE*, vol. 10, no. 6, p. e0128802, Jun. 2015, doi: 10.1371/journal.pone.0128802.

[173] **S. Takei *et al.***, 'Oral administration of genetically modified Bifidobacterium displaying HCV-NS3 multi-epitope fusion protein could induce an HCV-NS3-specific systemic immune response in mice', *Vaccine*, vol. 32, no. 25, pp. 3066–3074, May 2014, doi: 10.1016/j.vaccine.2014.03.022.

[174] **S. Yamamoto *et al.***, 'Genetically modified Bifidobacterium displaying Salmonella-antigen protects mice from lethal challenge of Salmonella Typhimurium in a murine typhoid fever model', *Vaccine*, vol. 28, no. 41, pp. 6684–6691, Sep. 2010, doi: 10.1016/j.vaccine.2010.08.007.

[175] **J. Law, G. Buist, A. Haandrikman, J. Kok, G. Venema, and K. Leenhouts**, 'A system to generate chromosomal mutations in Lactococcus lactis which allows fast analysis of targeted genes', *J. Bacteriol.*, vol. 177, no. 24, pp. 7011–7018, Dec. 1995, doi: 10.1128/jb.177.24.7011-7018.1995.

[176] **F. Zuo, Z. Zeng, L. Hammarström, and H. Marcotte**, 'Inducible Plasmid Self-Destruction (IPSD) Assisted Genome Engineering in Lactobacilli and Bifidobacteria', *ACS Synth. Biol.*, Jul. 2019, doi: 10.1021/acssynbio.9b00114.

[177] **Y. Hirayama *et al.***, 'Development of a Double-Crossover Markerless Gene Deletion System in Bifidobacterium longum: Functional Analysis of the α-Galactosidase Gene for Raffinose Assimilation', *Appl. Environ. Microbiol.*, vol. 78, no. 14, pp. 4984–4994, Jul. 2012, doi: 10.1128/AEM.00588-12.

[178] **H. Altaib, Y. Ozaki, T. Kozakai, K. Sakaguchi, I. Nomura, and T. Suzuki**, 'Bifidobacterium-Escherichia coli Shuttle Vector Series pKO403, with Temperature-Sensitive Replication Origin for Gene Knockout in Bifidobacterium', *Microbiol. Resour. Announc.*, vol. 11, no. 1, pp. e00884-21, Jan. 2022, doi: 10.1128/mra.00884-21.

[179] **T. V. Plavec and A. Berlec**, 'Safety Aspects of Genetically Modified Lactic Acid Bacteria', *Microorganisms*, vol. 8, no. 2, Art. no. 2, Feb. 2020, doi: 10.3390/microorganisms8020297.

[180] **W. Dominguez and D. J. O'Sullivan**, 'Developing an efficient and reproducible conjugation-based gene transfer system for bifidobacteria', *Microbiol. Read. Engl.*, vol. 159, no. Pt 2, Art. no. Pt 2, Feb. 2013, doi: 10.1099/mic.0.061408-0.

[181] **A. R. Burmeister**, 'Horizontal Gene Transfer', *Evol. Med. Public Health*, vol. 2015, no. 1, pp. 193–194, Jul. 2015, doi: 10.1093/emph/eov018.

[182] **E. V. Koonin, K. S. Makarova, and L. Aravind**, 'Horizontal Gene Transfer in Prokaryotes: Quantification and Classification', *Annu. Rev. Microbiol.*, vol. 55, no. 1, pp. 709–742, 2001, doi: 10.1146/annurev.micro.55.1.709.

[183] **T. J. Treangen, O. H. Ambur, T. Tonjum, and E. P. Rocha**, 'The impact of the neisserial DNA uptake sequences on genome evolution and stability', *Genome Biol.*, vol. 9, no. 3, p. R60, 2008, doi: 10.1186/gb-2008-9-3-r60.

[184] **T. J. Treangen and E. P. C. Rocha**, 'Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes', *PLoS Genet.*, vol. 7, no. 1, p. e1001284, Jan. 2011, doi: 10.1371/journal.pgen.1001284.

[185] **F. Bottacini *et al.***, 'Comparative genomics of the Bifidobacterium breve taxon', *BMC Genomics*, vol. 15, no. 1, Mar. 2014, doi: 10.1186/1471-2164-15-170.

[186] **L. S. Frost**, 'Conjugation, Bacterial', in *Encyclopedia of Microbiology (Third Edition)*, M. Schaechter, Ed. Oxford: Academic Press, 2009, pp. 517–531. doi: 10.1016/B978-012373944-5.00007-9.

[187] **I. Chen, P. J. Christie, and D. Dubnau**, 'The Ins and Outs of DNA Transfer in Bacteria', *Science*, vol. 310, no. 5753, Art. no. 5753, Dec. 2005, doi: 10.1126/science.1114021.

[188] **M. Llosa, F. X. Gomis-Rüth, M. Coll, and F. de la Cruz Fd**, 'Bacterial conjugation: a two-step mechanism for DNA transport', *Mol. Microbiol.*, vol. 45, no. 1, pp. 1–8, Jul. 2002, doi: 10.1046/j.1365-2958.2002.03014.x.

[189] **A. S. Lang, O. Zhaxybayeva, and J. T. Beatty**, 'Gene transfer agents: phage-like elements of genetic exchange', *Nat. Rev. Microbiol.*, vol. 10, no. 7, pp. 472–482, Jun. 2012, doi: 10.1038/nrmicro2802.

[190] **G. P. Dubey and S. Ben-Yehuda**, 'Intercellular Nanotubes Mediate Bacterial Communication', *Cell*, vol. 144, no. 4, pp. 590–600, Feb. 2011, doi: 10.1016/j.cell.2011.01.015.

[191] **S. Fischer *et al.***, 'Indication of Horizontal DNA Gene Transfer by Extracellular Vesicles', *PLoS ONE*, vol. 11, no. 9, p. e0163665, Sep. 2016, doi: 10.1371/journal.pone.0163665.

[192] **Y. Kawamura, A. Sanchez Calle, Y. Yamamoto, T.-A. Sato, and T. Ochiya**, 'Extracellular vesicles mediate the horizontal transfer of an active LINE-1 retrotransposon', *J. Extracell. Vesicles*, vol. 8, no. 1, p. 1643214, Jul. 2019, doi: 10.1080/20013078.2019.1643214.

[193] **J. Lederberg and E. L. Tatum**, 'Gene Recombination in Escherichia Coli', *Nature*, vol. 158, no. 4016, pp. 558–558, Oct. 1946, doi: 10.1038/158558a0.

[194] **E. L. Wollman, F. Jacob, and W. Hayes**, 'Conjugation and genetic recombination in Escherichia coli K-12', *Cold Spring Harb. Symp. Quant. Biol.*, vol. 21, pp. 141–162, 1956, doi: 10.1101/sqb.1956.021.01.012.

[195] **T. Watanabe and T. Fukasawa**, 'Episome-mediated transfer of drug resistance in Enterobacteriaceae. I. Transfer of resistance factors by conjugation', *J. Bacteriol.*, vol. 81, pp. 669–678, May 1961, doi: 10.1128/jb.81.5.669-678.1961.

[196] **J. A. Heinemann and G. F. Sprague**, '[13] Transmission of plasmid DNA to yeast by conjugation with bacteria', in *Methods in Enzymology*, vol. 194, Academic Press, 1991, pp. 187–195. doi: 10.1016/0076-6879(91)94016-6.

[197] **N. Van Larebeke *et al.***, 'Acquisition of tumour-inducing ability by non-oncogenic agrobacteria as a result of plasmid transfer', *Nature*, vol. 255, no. 5511, pp. 742–743, Jun. 1975, doi: 10.1038/255742a0.

[198] **R. Aminov**, 'Horizontal Gene Exchange in Environmental Microbiota', *Front. Microbiol.*, vol. 2, p. 158, 2011, doi: 10.3389/fmicb.2011.00158.

[199] **C. Possoz, C. Ribard, J. Gagnat, J.-L. Pernodet, and M. Guérineau**, 'The integrative element pSAM2 from Streptomyces: kinetics and mode of conjugal transfer', *Mol. Microbiol.*, vol. 42, no. 1, pp. 159–166, 2001, doi: https://doi.org/10.1046/j.1365-2958.2001.02618.x.

[200] **L. Thoma and G. Muth**, 'Conjugative DNA-transfer in Streptomyces, a mycelial organism', *Plasmid*, vol. 87–88, pp. 1–9, Sep. 2016, doi: 10.1016/j.plasmid.2016.09.004.

[201] **G. S. Pettis**, 'Spreading the news about the novel conjugation mechanism in Streptomyces bacteria', *Environ. Microbiol. Rep.*, vol. 10, no. 5, pp. 503–510, 2018, doi: 10.1111/1758-2229.12659.

[202] **P. J. Christie**, 'Type IV secretion: the Agrobacterium VirB/D4 and related conjugation systems', *Biochim. Biophys. Acta*, vol. 1694, no. 1–3, pp. 219–234, Nov. 2004, doi: 10.1016/j.bbamcr.2004.02.013.

[203] **J. Zupan, C. A. Hackworth, J. Aguilar, D. Ward, and P. Zambryski**, 'VirB1* promotes T-pilus formation in the vir-Type IV secretion system of Agrobacterium tumefaciens', *J. Bacteriol.*, vol. 189, no. 18, pp. 6551–6563, Sep. 2007, doi: 10.1128/JB.00480-07.

[204] **J. Guglielmini, F. de la Cruz, and E. P. C. Rocha**, 'Evolution of conjugation and type IV secretion systems', *Mol. Biol. Evol.*, vol. 30, no. 2, pp. 315–331, Feb. 2013, doi: 10.1093/molbev/mss221.

[205] **M. McCallum *et al.***, 'Multiple conformations facilitate PilT function in the type IV pilus', *Nat. Commun.*, vol. 10, no. 1, p. 5198, Nov. 2019, doi: 10.1038/s41467-019-13070-z.

[206] **E. Sagulenko, V. Sagulenko, J. Chen, and P. J. Christie**, 'Role of Agrobacterium VirB11 ATPase in T-Pilus Assembly and Substrate Selection', *J. Bacteriol.*, vol. 183, no. 20, pp. 5813–5825, Oct. 2001, doi: 10.1128/JB.183.20.5813-5825.2001.

[207] **D. R. Byrd and S. W. Matson**, 'Nicking by transesterification: the reaction catalysed by a relaxase', *Mol. Microbiol.*, vol. 25, no. 6, pp. 1011–1022, Sep. 1997, doi: 10.1046/j.1365-2958.1997.5241885.x.

[208] **S. W. Matson, J. K. Sampson, and D. R. Byrd**, 'F plasmid conjugative DNA transfer: the TraI helicase activity is essential for DNA strand transfer', *J. Biol. Chem.*, vol. 276, no. 4, pp. 2372–2379, Jan. 2001, doi: 10.1074/jbc.M008728200.

[209] **D. R. Helinski and D. B. Clewell**, 'Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event', *Biochemistry*, vol. 9, no. 22, pp. 4428–4440, Oct. 1970, doi: 10.1021/bi00824a026.

[210] **D. B. Clewell and D. R. Helinski**, 'SUPERCOILED CIRCULAR DNA-PROTEIN COMPLEX IN Escherichia coli: PURIFICATION AND INDUCED CONVERSION TO AN OPEN CIRCULAR DNA FORM*', *PNAS*, 1969. https://www.pnas.org/doi/abs/10.1073/pnas.62.4.1159 (accessed Mar. 22, 2022).

[211] **M. P. Garcillán-Barcia, M. V. Francia, and F. de la Cruz**, 'The diversity of conjugative relaxases and its application in plasmid classification', *FEMS Microbiol. Rev.*, vol. 33, no. 3, pp. 657–687, May 2009, doi: 10.1111/j.1574-6976.2009.00168.x.

[212] **M. Chandler, F. de la Cruz, F. Dyda, A. B. Hickman, G. Moncalian, and B. Ton-Hoang**, 'Breaking and joining single-stranded DNA: the HUH endonuclease superfamily', *Nat. Rev. Microbiol.*, vol. 11, no. 8, pp. 525–538, Aug. 2013, doi: 10.1038/nrmicro3067.

[213] **W. Pansegrau, W. Schröder, and E. Lanka**, 'Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 7, pp. 2925–2929, Apr. 1993, doi: 10.1073/pnas.90.7.2925.

[214] **M. V. Francia, A. Varsaki, M. P. Garcillán-Barcia, A. Latorre, C. Drainas, and F. de la Cruz**, 'A classification scheme for mobilization regions of bacterial plasmids', *FEMS Microbiol. Rev.*, vol. 28, no. 1, pp. 79–100, Feb. 2004, doi: 10.1016/j.femsre.2003.09.001.

[215] **M. P. Garcillán-Barcia, S. Redondo-Salvo, L. Vielva, and F. de la Cruz**, 'MOBscan: Automated Annotation of MOB Relaxases', in *Horizontal Gene Transfer: Methods and Protocols*, F. de la Cruz, Ed. New York, NY: Springer US, 2020, pp. 295–308. doi: 10.1007/978-1-4939-9877-7_21.

[216] **J. Kopec, A. Bergmann, G. Fritz, E. Grohmann, and W. Keller**, 'TraA and its N-terminal relaxase domain of the Gram-positive plasmid pIP501 show specific oriT binding and behave as dimers in solution', *Biochem. J.*, vol. 387, no. 2, pp. 401–409, Apr. 2005, doi: 10.1042/BJ20041178.

[217] **R. Meyer**, 'Replication and conjugative mobilization of Broad Host-range IncQ plasmids', *Plasmid*, vol. 62, no. 2, pp. 57–70, Sep. 2009, doi: 10.1016/j.plasmid.2009.05.001.

[218] 'Bacteroidota Krieg et al., 2021'. https://www.gbif.org/species/144094785 (accessed Mar. 28, 2022).

[219] **M. Llosa, G. Grandoso, M. A. Hernando, and F. de la Cruz**, 'Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function', *J. Mol. Biol.*, vol. 264, no. 1, pp. 56–67, Nov. 1996, doi: 10.1006/jmbi.1996.0623.

[220] **C. de Antonio, M. E. Farías, M. G. de Lacoba, and M. Espinosa**, 'Features of the Plasmid pMV158-encoded MobM, a Protein Involved in its Mobilization††Dedicated to the late Brian M. Wilkins, an excellent scientist and gentleman.', *J. Mol. Biol.*, vol. 335, no. 3, pp. 733–743, Jan. 2004, doi: 10.1016/j.jmb.2003.11.017.

[221] **R. Pluta *et al.***, 'Structural basis of a histidine-DNA nicking/joining mechanism for gene transfer and promiscuous spread of antibiotic resistance', *Proc. Natl. Acad. Sci.*, vol. 114, no. 32, pp. E6526–E6535, Aug. 2017, doi: 10.1073/pnas.1702971114.

[222] **J.-H. Heilers, J. Reiners, E.-M. Heller, A. Golzer, S. H. J. Smits, and C. van der Does**, 'DNA processing by the MOBH family relaxase TraI encoded within the gonococcal genetic island', *Nucleic Acids Res.*, vol. 47, no. 15, pp. 8136–8153, Sep. 2019, doi: 10.1093/nar/gkz577.

[223] **M. V. Francia, D. B. Clewell, F. de la Cruz, and G. Moncalián**, 'Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases', *Proc. Natl. Acad. Sci.*, vol. 110, no. 33, pp. 13606–13611, Aug. 2013, doi: 10.1073/pnas.1310037110.

[224] **N. Soler *et al.***, 'Characterization of a relaxase belonging to the MOBT family, a widespread family in Firmicutes mediating the transfer of ICEs', *Mob. DNA*, vol. 10, no. 1, p. 18, May 2019, doi: 10.1186/s13100-019-0160-9.

[225] **J. A. Wisniewski, D. A. Traore, T. L. Bannam, D. Lyras, J. C. Whisstock, and J. I. Rood**, 'TcpM: a novel relaxase that mediates transfer of large conjugative plasmids from Clostridium perfringens', *Mol. Microbiol.*, vol. 99, no. 5, pp. 884–896, 2016, doi: 10.1111/mmi.13270.

[226] **C. Ambroset *et al.***, 'New Insights into the Classification and Integration Specificity of Streptococcus Integrative Conjugative Elements through Extensive Genome

Exploration', *Front. Microbiol.*, vol. 6, p. 1483, 2015, doi: 10.3389/fmicb.2015.01483.

[227] **R. Bantwal *et al.***, 'The peptidoglycan hydrolase TcpG is required for efficient conjugative transfer of pCW3 in Clostridium perfringens', *Plasmid*, vol. 67, no. 2, pp. 139–147, Mar. 2012, doi: 10.1016/j.plasmid.2011.12.016.

[228] **J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay**, 'Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold', *EMBO J.*, vol. 1, no. 8, pp. 945–951, 1982.

[229] **N. Whitaker, Y. Chen, S. J. Jakubowski, M. K. Sarkar, F. Li, and P. J. Christie**, 'The All-Alpha Domains of Coupling Proteins from the Agrobacterium tumefaciens VirB/VirD4 and Enterococcus faecalis pCF10-Encoded Type IV Secretion Systems Confer Specificity to Binding of Cognate DNA Substrates', *J. Bacteriol.*, vol. 197, no. 14, pp. 2335–2349, Jul. 2015, doi: 10.1128/JB.00189-15.

[230] **M. Llosa and I. Alkorta**, 'Coupling Proteins in Type IV Secretion', *Curr. Top. Microbiol. Immunol.*, vol. 413, pp. 143–168, 2017, doi: 10.1007/978-3-319-75241-9_6.

[231] **Y. Chen, X. Zhang, D. Manias, H.-J. Yeo, G. M. Dunny, and P. J. Christie**, 'Enterococcus faecalis PcfC, a Spatially Localized Substrate Receptor for Type IV Secretion of the pCF10 Transfer Intermediate', *J. Bacteriol.*, vol. 190, no. 10, pp. 3632–3645, May 2008, doi: 10.1128/JB.01999-07.

[232] **P. J. Christie**, 'The Mosaic Type IV Secretion Systems', *EcoSal Plus*, vol. 7, no. 1, 2016, doi: 10.1128/ecosalplus.ESP-0020-2015.

[233] **M. Llosa, S. Zunzunegui, and F. de la Cruz**, 'Conjugative coupling proteins interact with cognate and heterologous VirB10-like proteins while exhibiting specificity for cognate relaxosomes', *Proc. Natl. Acad. Sci.*, vol. 100, no. 18, pp. 10465–10470, Sep. 2003, doi: 10.1073/pnas.1830264100.

[234] **R. L. Segura *et al.***, 'The transmembrane domain of the T4SS coupling protein TrwB and its role in protein–protein interactions', *Biochim. Biophys. Acta BBA - Biomembr.*, vol. 1828, no. 9, pp. 2015–2025, Sep. 2013, doi: 10.1016/j.bbamem.2013.05.022.

[235] **A. Redzej *et al.***, 'Structure of a VirD4 coupling protein bound to a VirB type IV secretion machinery', *EMBO J.*, vol. 36, no. 20, pp. 3080–3095, Oct. 2017, doi: 10.15252/embj.201796629.

[236] **I. Hormaeche, I. Iloro, J. L. R. Arrondo, F. M. Goñi, F. de la Cruz, and I. Alkorta**, 'Role of the Transmembrane Domain in the Stability of TrwB, an Integral Protein Involved in Bacterial Conjugation *', *J. Biol. Chem.*, vol. 279, no. 12, pp. 10955–10961, Mar. 2004, doi: 10.1074/jbc.M310422200.

[237] **I. Hormaeche, I. Alkorta, F. Moro, J. M. Valpuesta, F. M. Goñi, and F. de la Cruz**, 'Purification and Properties of TrwB, a Hexameric, ATP-binding Integral Membrane Protein Essential for R388 Plasmid Conjugation *', *J. Biol. Chem.*, vol. 277, no. 48, pp. 46456–46462, Nov. 2002, doi: 10.1074/jbc.M207250200.

[238] **I. Hormaeche, R. L. Segura, A. J. Vecino, F. M. Goñi, F. de la Cruz, and I. Alkorta**, 'The transmembrane domain provides nucleotide binding specificity to the bacterial conjugation protein TrwB', *FEBS Lett.*, vol. 580, no. 13, pp. 3075–3082, 2006, doi: 10.1016/j.febslet.2006.04.059.

[239] **I. Álvarez-Rodríguez, B. Ugarte-Uribe, I. de la Arada, J. L. R. Arrondo, C. Garbisu, and I. Alkorta**, 'Conjugative Coupling Proteins and the Role of Their Domains in Conjugation, Secondary Structure and in vivo Subcellular Location', *Front. Mol. Biosci.*, vol. 7, 2020, doi: 10.3389/fmolb.2020.00185.

[240] **J. E. Gunton, M. W. Gilmour, K. P. Baptista, T. D. Lawley, and D. E. Taylor**, 'Interaction between the co-inherited TraG coupling protein and the TraJ membrane-associated protein of the H-plasmid conjugative DNA transfer system resembles chromosomal DNA translocases', *Microbiol. Read. Engl.*, vol. 153, no. Pt 2, pp. 428–441, Feb. 2007, doi: 10.1099/mic.0.2006/001297-0.

[241] **M. S. Gilmore, R. A. Segarra, M. C. Booth, C. P. Bogie, L. R. Hall, and D. B. Clewell**, 'Genetic structure of the Enterococcus faecalis plasmid pAD1-encoded cytolytic toxin system and its relationship to lantibiotic determinants', *J. Bacteriol.*, vol. 176, no. 23, pp. 7335–7344, Dec. 1994, doi: 10.1128/jb.176.23.7335-7344.1994.

[242] **J. A. Parsons, T. L. Bannam, R. J. Devenish, and J. I. Rood**, 'TcpA, an FtsK/SpoIIIE homolog, is essential for transfer of the conjugative plasmid pCW3 in Clostridium perfringens', *J. Bacteriol.*, vol. 189, no. 21, pp. 7782–7790, Nov. 2007, doi: 10.1128/JB.00783-07.

[243] **J. A. Steen, T. L. Bannam, W. L. Teng, R. J. Devenish, and J. I. Rood**, 'The Putative Coupling Protein TcpA Interacts with Other pCW3-Encoded Proteins To Form an Essential Part of the Conjugation Complex', *J. Bacteriol.*, vol. 191, no. 9, pp. 2926–2933, May 2009, doi: 10.1128/JB.00032-09.

[244] **L. Makart, A. Gillis, P. Hinnekens, and J. Mahillon**, 'A novel T4SS-mediated DNA transfer used by pXO16, a conjugative plasmid from Bacillus thuringiensis serovar israelensis', *Environ. Microbiol.*, vol. 20, no. 4, pp. 1550–1561, 2018, doi: 10.1111/1462-2920.14084.

[245] **C. E. Alvarez-Martinez and P. J. Christie**, 'Biological diversity of prokaryotic type IV secretion systems', *Microbiol. Mol. Biol. Rev. MMBR*, vol. 73, no. 4, pp. 775–808, Dec. 2009, doi: 10.1128/MMBR.00023-09.

[246] **P. Schumann**, '5 - Peptidoglycan Structure', in *Methods in Microbiology*, vol. 38, F. Rainey and A. Oren, Eds. Academic Press, 2011, pp. 101–129. doi: 10.1016/B978-0-12-387730-7.00005-X.

[247] **A. Susmitha, H. Bajaj, and K. Madhavan Nampoothiri**, 'The divergent roles of sortase in the biology of Gram-positive bacteria', *Cell Surf.*, vol. 7, p. 100055, Dec. 2021, doi: 10.1016/j.tcsw.2021.100055.

[248] **C. Heddle *et al.***, 'Host collagen signal induces antigen I/II adhesin and invasin gene expression in oral Streptococcus gordonii', *Mol. Microbiol.*, vol. 50, no. 2, pp. 597–607, Oct. 2003, doi: 10.1046/j.1365-2958.2003.03711.x.

[249] **C. M. Waters, H. Hirt, J. K. McCormick, P. M. Schlievert, C. L. Wells, and G. M. Dunny**, 'An amino-terminal domain of Enterococcus faecalis aggregation substance is required for aggregation, bacterial internalization by epithelial cells and binding to lipoteichoic acid', *Mol. Microbiol.*, vol. 52, no. 4, pp. 1159–1171, 2004, doi: 10.1111/j.1365-2958.2004.04045.x.

[250] **H. Tamura, A. Yamada, and H. Kato**, 'Molecular characterization of the dextran-binding lectin B gene dblB of Streptococcus criceti in Streptococcus mutans strain GS-5 with mutations in both gbpC and spaP genes', *Genes Genet. Syst.*, vol. 89, no. 2, pp. 41–50, 2014, doi: 10.1266/ggs.89.41.

[251] **J. L. Mieher *et al.***, 'Glucan Binding Protein C of Streptococcus mutans Mediates both Sucrose-Independent and Sucrose-Dependent Adherence', *Infect. Immun.*, vol. 86, no. 7, pp. e00146-18, Jun. 2018, doi: 10.1128/IAI.00146-18.

[252] **A. Schmitt *et al.***, 'PrgB Promotes Aggregation, Biofilm Formation, and Conjugation through DNA Binding and Compaction', *Mol. Microbiol.*, vol. 109, no. 3, pp. 291–305, Aug. 2018, doi: 10.1111/mmi.13980.

[253] **A. Mandlik, A. Swierczynski, A. Das, and H. Ton-That**, 'Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development', *Trends Microbiol.*, vol. 16, no. 1, pp. 33–40, Jan. 2008, doi: 10.1016/j.tim.2007.10.010.

[254] **T. J. Foster, J. A. Geoghegan, V. K. Ganesh, and M. Höök**, 'Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus', *Nat. Rev. Microbiol.*, vol. 12, no. 1, pp. 49–62, Jan. 2014, doi: 10.1038/nrmicro3161.

[255] **V. Krishnan**, 'Pilins in gram-positive bacteria: A structural perspective', *IUBMB Life*, vol. 67, no. 7, pp. 533–543, 2015, doi: 10.1002/iub.1400.

[256] **J. A. Geoghegan and T. J. Foster**, 'Cell Wall-Anchored Surface Proteins of Staphylococcus aureus: Many Proteins, Multiple Functions', *Curr. Top. Microbiol. Immunol.*, vol. 409, pp. 95–120, 2017, doi: 10.1007/82_2015_5002.

[257] **T. J. Foster**, 'The MSCRAMM Family of Cell-Wall-Anchored Surface Proteins of Gram-Positive Cocci', *Trends Microbiol.*, vol. 27, no. 11, pp. 927–941, Nov. 2019, doi: 10.1016/j.tim.2019.06.007.

[258] **Y. F. Dufrêne and A. Viljoen**, 'Binding Strength of Gram-Positive Bacterial Adhesins', *Front. Microbiol.*, vol. 11, 2020, doi: 10.3389/fmicb.2020.01457.

[259] **M. Kang *et al.***, 'Collagen-binding microbial surface components recognizing adhesive matrix molecule (MSCRAMM) of Gram-positive bacteria inhibit complement activation via the classical pathway', *J. Biol. Chem.*, vol. 288, no. 28, pp. 20520–20531, Jul. 2013, doi: 10.1074/jbc.M113.454462.

[260] **M. Bhatty *et al.***, 'Enterococcus faecalis pCF10-encoded surface proteins PrgA, PrgB (aggregation substance) and PrgC contribute to plasmid transfer, biofilm formation and virulence', *Mol. Microbiol.*, vol. 95, no. 4, pp. 660–677, Feb. 2015, doi: 10.1111/mmi.12893.

[261] **H. S. Manzer, A. H. Nobbs, and K. S. Doran**, 'The Multifaceted Nature of Streptococcal Antigen I/II Proteins in Colonization and Disease Pathogenesis', *Front. Microbiol.*, vol. 11, p. 602305, Nov. 2020, doi: 10.3389/fmicb.2020.602305.

[262] **M. Roky, J. O. Trent, and D. R. Demuth**, 'Identification of functional domains of the minor fimbrial antigen involved in the interaction of Porphyromonas gingivalis with oral streptococci', *Mol. Oral Microbiol.*, vol. 35, no. 2, pp. 66–77, 2020, doi: 10.1111/omi.12280.

[263] **D. Galli, F. Lottspeich, and R. Wirth**, 'Sequence analysis of Enterococcus faecalis aggregation substance encoded by the sex pheromone plasmid pAD1', *Mol. Microbiol.*, vol. 4, no. 6, pp. 895–904, 1990, doi: 10.1111/j.1365-2958.1990.tb00662.x.

[264] **D. Galli, A. Friesenegger, and R. Wirth**, 'Transcriptional control of sex-pheromone-inducible genes on plasmid pAD1 of Enterococcus faecalis and sequence analysis of a third structural gene for (pPD1 -encoded) aggregation substance', *Mol. Microbiol.*, vol. 6, no. 10, pp. 1297–1308, 1992, doi: 10.1111/j.1365-2958.1992.tb00851.x.

[265] **S. M. Kao, S. B. Olmsted, A. S. Viksnins, J. C. Gallo, and G. M. Dunny**, 'Molecular and genetic analysis of a region of plasmid pCF10 containing positive control genes and structural genes encoding surface proteins involved in pheromone-inducible conjugation in Enterococcus faecalis', *J. Bacteriol.*, vol. 173, no. 23, pp. 7650–7664, Dec. 1991, doi: 10.1128/jb.173.23.7650-7664.1991.

[266] **G. M. Dunny, D. L. Zimmerman, and M. L. Tortorello**, 'Induction of surface exclusion (entry exclusion) by Streptococcus faecalis sex pheromones: use of monoclonal antibodies to identify an inducible surface antigen involved in the

exclusion process', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 82, no. 24, pp. 8582–8586, Dec. 1985, doi: 10.1073/pnas.82.24.8582.

[267] **A. Schmitt *et al.***, 'Enterococcal PrgA Extends Far Outside the Cell and Provides Surface Exclusion to Protect against Unwanted Conjugation', *J. Mol. Biol.*, vol. 432, no. 20, pp. 5681–5695, Sep. 2020, doi: 10.1016/j.jmb.2020.08.018.

[268] **M. Bhatty *et al.***, 'PrgU: a suppressor of sex pheromone toxicity in Enterococcus faecalis', *Mol. Microbiol.*, vol. 103, no. 3, pp. 398–412, Feb. 2017, doi: 10.1111/mmi.13563.

[269] **T. L. Bannam, W. L. Teng, D. Bulach, D. Lyras, and J. I. Rood**, 'Functional Identification of Conjugation and Replication Regions of the Tetracycline Resistance Plasmid pCW3 from Clostridium perfringens', *J. Bacteriol.*, vol. 188, no. 13, pp. 4942–4951, Jul. 2006, doi: 10.1128/JB.00298-06.

[270] **S. A. Revitt-Mills, T. D. Watts, D. Lyras, V. Adams, and J. I. Rood**, 'The ever-expanding tcp conjugation locus of pCW3 from Clostridium perfringens', *Plasmid*, vol. 113, p. 102516, Jan. 2021, doi: 10.1016/j.plasmid.2020.102516.

[271] **E. M. Camacho, A. Serna, and J. Casadesús**, 'Regulation of conjugal transfer by Lrp and Dam methylation in plasmid R100', *Regulación de la transferencia conjugativa por Lrp y metilación Dam en el plásmido R100*, 2005, Accessed: Feb. 13, 2023. [Online]. Available: https://digital.csic.es/handle/10261/198698

[272] **E. M. Camacho and J. Casadesús**, 'Regulation of traJ transcription in the Salmonella virulence plasmid by strand-specific DNA adenine hemimethylation', *Mol. Microbiol.*, vol. 57, no. 6, pp. 1700–1718, 2005, doi: 10.1111/j.1365-2958.2005.04788.x.

[273] **M. Bhatty, J. A. Laverde Gomez, and P. J. Christie**, 'The Expanding Bacterial Type IV Secretion Lexicon', *Res. Microbiol.*, vol. 164, no. 6, 2013, doi: 10.1016/j.resmic.2013.03.012.

[274] **N. Goessweiner-Mohr, K. Arends, W. Keller, and E. Grohmann**, 'Conjugation in Gram-Positive Bacteria', *Microbiol. Spectr.*, vol. 2, no. 4, Aug. 2014, doi: 10.1128/microbiolspec.PLAS-0004-2013.

[275] **M. E. Tolmasky and J. C. Alonso, Eds.**, 'Plasmids: Biology and Impact in Biotechnology and Discovery', in *Plasmids: Biology and Impact in Biotechnology and Discovery*, John Wiley & Sons, 2015.

[276] **D. Hanahan**, 'Studies on transformation of Escherichia coli with plasmids', *J. Mol. Biol.*, vol. 166, no. 4, pp. 557–580, Jun. 1983, doi: 10.1016/S0022-2836(83)80284-8.

[277] **S. N. Cohen, A. C. Y. Chang, H. W. Boyer, and R. B. Helling**, 'Construction of Biologically Functional Bacterial Plasmids In Vitro', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 70, no. 11, pp. 3240–3244, Nov. 1973.

[278] **S. Köstlbacher, A. Collingro, T. Halter, D. Domman, and M. Horn**, 'Coevolving Plasmids Drive Gene Flow and Genome Plasticity in Host-Associated Intracellular Bacteria', *Curr. Biol.*, vol. 31, no. 2, pp. 346-357.e3, Jan. 2021, doi: 10.1016/j.cub.2020.10.030.

[279] **L. N. Lili, N. F. Britton, and E. J. Feil**, 'The persistence of parasitic plasmids', *Genetics*, vol. 177, no. 1, pp. 399–405, Sep. 2007, doi: 10.1534/genetics.107.077420.

[280] **F. R. Slater, M. J. Bailey, A. J. Tett, and S. L. Turner**, 'Progress towards understanding the fate of plasmids in bacterial communities', *FEMS Microbiol. Ecol.*, vol. 66, no. 1, pp. 3–13, Oct. 2008, doi: 10.1111/j.1574-6941.2008.00505.x.

[281] **D. J. Rankin, E. P. C. Rocha, and S. P. Brown**, 'What traits are carried on mobile genetic elements, and why?', *Heredity*, vol. 106, no. 1, pp. 1–10, Jan. 2011, doi: 10.1038/hdy.2010.24.

CHAPTER II

Molecular Analysis of the Replication Functions of the Bifidobacterial Conjugative Megaplasmid pMP7017

This chapter was published in:

R. L. Dineen, C. Penno, P. Kelleher, M. J. B. Bourin, M. O'Connell-Motherway, and D. van Sinderen, 'Molecular analysis of the replication functions of the bifidobacterial conjugative megaplasmid pMP7017', *Microb Biotechnol*, May 2021, doi: 10.1111/1751-7915.13810.

Dr. P. Kelleher performed HPLC analysis of Rep proteins.

ABSTRACT

pMP7017 is a conjugative megaplasmid isolated from the gut commensal *Bifidobacterium breve* JCM7017 and was shown to encode two putative replicases, designated here as RepA and RepB. In the current work, RepB was identified as the pMP7017 replicative initiator, as the *repB* gene and its surrounding region was shown to be sufficient to allow autonomous replication in two bifidobacterial species, *B. breve* and *Bifidobacterium longum* subsp. *longum*. RepB was shown to bind to repeat sequences downstream of its coding sequence and this region was determined to be essential for efficient replication. Based on our results, we hypothesize that pMP7017 is an iteron-regulated plasmid (IRP) under strict auto-regulatory control. Recombinantly produced and purified RepB was determined to exist as a dimer in solution, differing from replicases of other IRPs, which exist as a mix of dimers and monomers. Furthermore, a stable low-copy *Bifidobacterium-E. coli* shuttle vector, pRD1.3, was created which can be employed for cloning and expression of large genes, as was demonstrated by the cloning and heterologous expression of the 5.1 kb *apuB* gene encoding the extracellular amylopullulanase from *B. breve* UCC2003 into *B. longum* subsp. *longum* NCIMB8809.

INTRODUCTION

Bifidobacteria represent anaerobic, non-motile, high G+C% Gram-positive bacteria that belong to the phylum Actinobacteria. Bifidobacteria are considered key commensals in microbe-host interactions, purported to play a pivotal role in maintaining gastrointestinal health through various beneficial or probiotic activities [1], [2]. They have been reported to reduce the risk of microbial infections by enforcing epithelial barrier function through acetate production and secretion of lipophilic antimicrobials [3], [4]. In addition, certain bifidobacterial strains have been shown to interact with their host *in vivo* via extracellular surface structures [5]–[8] and through the production of tryptophan catabolites, such as indole-3-lactic acid via aryl hydrocarbon receptor stimulation [9], [10].

These reported health benefits have been the driving force behind a wave of research activities to study the biology and genetics of bifidobacteria and the way in which they interact with their host. Currently, knowledge on the health benefits of bifidobacteria is for a considerable part deduced from *in silico* analyses due to the recent surge of available bifidobacterial genome sequences, a 16.5-fold increase in three years from 62 [2] to 1,021 (https://www.ncbi.nlm.nih.gov/taxonomy, accessed May 2020) and to a lesser degree based on findings from functional analyses [11]–[15]. The paucity of functional data is partly due to the inherent resistance of *Bifidobacterium* to the artificial uptake of DNA due to natural barriers such as restriction-modification systems [16], but most significantly due to the scarcity of molecular techniques and tools suited to these bacteria [2], [17], [18]. Plasmids represent one of these crucial molecular tools, playing an integral role in the expansion of our understanding of gene function. This is in part due to their small size and relatively simple genetic organisation, in combination with the ease with which they can be isolated and manipulated *in vitro* without adversely affecting the host [19]. Plasmids are generally defined as autonomously replicating, extrachromosomal genetic elements, which are not essential to host survival, and which may carry a plethora of advantageous functions for rapid environmental adaptation [20]–[22]. Plasmids require one or more genes and/or cis-acting sequences, together termed the basic replicon, for replication and faithful transmission to each daughter cell during

cell division. The genetic composition of such a basic replicon can vary widely and depends on the mode of replication.

Plasmids were originally thought to be relatively rare in bifidobacteria [23], yet a total of 49 plasmids of bifidobacterial origin have been identified to date, isolated from 12 species and ranging in size from 1.8 kb in the case of plasmid pMB1 isolated from *Bifidobacterium longum* subsp. *longum* [24] to 190 kb in the case of pMP7017 from *Bifidobacterium breve* [25]. Furthermore, recent evidence indicates a higher prevalence of megaplasmids in this genus due to the identification of pMP7017 homologues in fifteen *B. longum* strains [26]. The word 'megaplasmid' was initially coined to describe a 450 kb plasmid identified in the Gram-negative rhizophere bacterium *Sinorhizobium meliloti* in 1981 [27], and in publications to date has been used to describe plasmids >100 kb [28], which appear in both phylogenetically and physiologically diverse groups of archaea and bacteria. While in-depth investigations have been conducted to characterise the replication functions of megaplasmids from rhizobia [29], [30] and *Vibrio cholera* chromosome II [31]–[34], relatively little is known about the replication mode of actinobacterial megaplasmids.

The 190,178 bp conjugative megaplasmid pMP7017 of *B. breve* JCM7017 represents the first and largest, conjugative megaplasmid identified among the actinobacterial gut commensals [25]. The report suggests that pMP7017 is the result of a co-integration of two plasmids, a notion based on the finding that there is a 4.8% deviation in G+C mol% from one arbitrary half to the other, while pMP7017 is also predicted to encode two replicative initiator proteins, denoted here as RepA and RepB. These putative replicases do not correspond or display homology to proteins from the RepABC family of plasmids, thus this nomenclature is purely for the purpose of differentiation. In the current report, we investigate the replication functions of the conjugative bifidobacterial megaplasmid pMP7017 in an effort to further our understanding of the replicative strategies of this and, by inference, other large plasmids. While research efforts to date have primarily focused on plasmid replication in Gram-negative and pathogenic Gram-positive bacteria, the current work serves to address the knowledge gap of these systems in Gram-positive commensals.

EXPERIMENTAL PROCEDURES

Bacterial strains, plasmids and cultivation conditions

Bacterial strains and plasmids are listed in **Supplementary Table S1**. Bifidobacterial strains were routinely cultivated in reinforced clostridial medium (RCM [Oxoid, Ltd., Basingstoke, Hampshire, United Kingdom]) or modified de Man, Rogosa, and Sharpe medium (mMRS) prepared from first principles [35]. The mMRS medium was supplemented with cysteine-HCl (0.05% [wt/vol]) and lactose (1.0% [wt/vol]) (Sigma) prior to inoculation. Bifidobacterial strains were cultivated in an anaerobic hood (Davidson and Hardy, Belfast, Ireland) for incubation at 37°C, maintaining an anaerobic environment. All *Escherichia coli* strains were cultivated exclusively in Luria-Bertani (LB) broth at 37°C with agitation [36]. Where applicable, growth media contained tetracycline (10 μg ml^{-1}), ampicillin (100 μg ml^{-1}), erythromycin (100 μg ml^{-1}), chloramphenicol (5 μg ml^{-1}) or kanamycin (50 μg ml^{-1}).

DNA manipulations and molecular cloning

Chromosomal DNA was isolated from *B. breve* JCM7017-199, a *B. breve* JCM7017-derivative strain containing a tetracycline resistance marker disrupting the open reading with locus tag pMP7017_199 of plasmid pMP7017, as previously described [25]. Plasmid DNA from *E. coli* was extracted using the High Pure Plasmid Isolation Kit (Sigma-Aldrich, United Kingdom). Standard oligonucleotides (**Supplementary Table S2**) used in this study were synthesized by Eurofins (Ebersberg, Germany). Both standard and colony PCRs were performed using Dreamtaq PCR 2x Mastermix (Thermo Fisher Scientific, Gloucester, United Kingdom), colony PCRs were carried out as described previously [37]. High-fidelity PCR was achieved using Q5 High-Fidelity 2X Master Mix (New England Biolabs Ltd, Hitchin, United Kingdom). PCR fragments were purified using the High Pure PCR Product Purification Kit (Sigma). Restriction endonucleases and T4 DNA ligase were used according to the manufacturer's instructions (Roche, Mannheim, Germany). Electroporation of plasmid DNA into *Bifidobacterium* spp. was conducted as previously described [38], while *E. coli* transformations were performed as described

by Sambrook *et al.,* [36]. All electroporation experiments were carried out using a GenePulser apparatus (Bio-Rad Laboratories, Richmond, Ca., USA).

Insertional mutagenesis

Insertional mutagenesis was performed based on a previously described protocol [13]. For this purpose, a 525 bp internal fragment of open reading frame (ORF) pMP7017_124 (*repA*) from codon positions 92 to 267 of the *repA* coding sequence and an internal 501 bp fragment of ORF pMP7017_146 (*repB*) from codon positions 101 to 268 of the *repB* coding sequence, were amplified by high fidelity PCR using B. *breve* JCM7017_199 total DNA as a template with primer combinations repAmut_F and repAmut_R, and repBmut_F and repBmut_R, respectively (**Supplementary Table S2**). The amplicons were ligated into the multiple cloning site of the integration vector pORI19 following HindIII and XbaI digestion. 5 µl of each ligation mixture was introduced into *E. coli* EC101 competent cells by electroporation. Recombinant *E. coli* derivatives were selected by blue-white screening on LB agar supplemented with 40 g ml^{-1} X-gal (5-bromo-4-chloro-3-indolyl-D-galactopyranoside) and 1 mM IPTG. Confirmation of recombination was achieved by restriction analysis prior to sub-cloning of the tetracycline resistance gene, *tet*(W), from plasmid pAM5 into the unique SacI restriction site of each plasmid. The resulting constructs designated pORI19_*repA* and pORI19_*repB*, were introduced into *B. breve* JCM7017 by electroporation and transformants were selected on tetracycline-containing media. Site-specific recombination in such tetracycline-resistant transformants was confirmed by colony PCR with primers tetW_F and tetW_R to verify the presence of the *tet*(W) gene. Total DNA was isolated from such *tet*(W)-positive transformants and used as a template for PCR with primer pair *repA*conf and TetR_final to confirm integration within the expected genetic position of the megaplasmid. The *B. breve* JCM7017_*repA* mutant was then verified by genome sequencing performed employing an Illumina MiSeq Sequencing Platform.

Plasmid Constructions

For the construction of plasmids pBif_*repA*, pBif_*repB*, pRD1.2, pRD1.3, pRD1.4, pRD1.5 and pRD1.6, various DNA fragments representing sections of the

repA or *repB* gene and surrounding sequences were obtained by PCR amplification using total DNA isolated from *B. breve* JCM7017 as a template in all reactions and primer combinations as listed in **Supplementary Table S2**. Recombinant DNA sequences for plasmids pRD1.7, pRD1.8, pRD1.9 and pRD1.10 were generated by overlap extension PCR containing two PCR segments with primer combinations outlined in **Supplementary Table S2**. Oligonucleotides were designed with an SphI restriction site incorporated at the 5' end to allow direct cloning of amplicons. The generated amplicons and the bifidobacterial replication probe-*E. coli* vector pBif [39] were digested with SphI and ligated. Ligation mixtures were dialyzed and transferred into *E. coli* XL1blue competent cells by electroporation. pBif consists of a chloramphenicol resistance cassette derived from plasmid pC194 [40] in combination with the pSC101 replication functions of pWSK29. Colony PCR was performed to confirm the presence of the *repA* or *repB* gene using primer pairs pMP_0124F and pMP_0124R, or pMP_0146F and pMP_0146R, designated to generate a 237 bp or 244 bp amplification product, respectively. Purified miniprep plasmid DNA of positive clones was used as a template for a second round of PCR with primer pRDconf_F in combination with the forward or reverse *repA* or *repB* cloning primers to determine the orientation of the gene, and plasmids containing *repA* or *repB* in the forward orientation were selected for further analysis.

Primer pair *apuB*F_Xba and *apuB*R_H3 was used to amplify a 5,437 bp region containing the *apuB* coding sequence and promoter region from *B. breve* UCC2003 genomic DNA. Both the resulting *apuB*-encompassing amplicon and pRD1.3 plasmid were digested with restriction enzymes Xba1 and HindIII prior to ligation and dialysis. 5 µl of ligation mixture was electroporated into *E. coli* XL1blue competent cells, the cells were resuspended in 950 µl of LB broth and incubated for 1 hr incubation at 37ºC shaking (160 rpm) prior to spread-plating on LB agar containing 5 µg ml^{-1} of chloramphenicol and 10 µg ml^{-1} of tetracycline. Colony PCR was performed with primers *apuB*F_Xba and *apuB*R_H3 to select colonies containing the insert, prior to plasmid preparation and sequencing. Plasmid pRD1.3_*apuB* was isolated and transformed into *B. longum* NCIMB8809 by electroporation.

Overexpression constructs pQE30_*repA* and pQE30_*repB* were produced by cloning each coding sequence in frame with the plasmid-encoded N-terminal hexa-histidine using primer pairs pQE30-124N_Sph1 for *repA*, and pQE30-124N_HindIII or pQE30-146N_Sph1 and pQE30-146N _HindIII for *repB*.

To construct plasmids pNZPP_*repA* and pNZPP_*repB*, 200 bp preceding the predicted translational start site of each gene was amplified by PCR with primer pairs 124PP_BgLII and 124PP_Pst1, or 146PP_BgLII and 146PP_Pst1, respectively. The resulting amplicons were cloned upstream of the promoterless β-glucuronidase gene in reporter plasmid pNZ272 and transformed into *L. lactis* NZ9000. The recombinant plasmids were then isolated and introduced into *B. breve* UCC2003 and *B. breve* UCC2003 pMP7017_199 by electroporation.

To create a C-terminal deletion in RepB, 669 bp of the *repB* gene excluding 297 bp of the 3'-end of this gene (representing the C-terminal amino acid positions 235 through to 332) was amplified by PCR using primers pQE146C-_SphIF and pQE146C-_H3R. The resulting amplicon was cloned into pQE30 in a manner to fuse a histidine tag to the N-terminus of the encoded, C-terminally truncated RepB protein (designated here as RepBΔC), resulting in the expression construct pQE30_RepBΔC. The sequence integrity of inserts in all recombinant plasmids was checked by Sanger sequence analysis (performed by MWG, Ebersberg, Germany).

Plasmid Stability

The stability of the pRD constructs was assayed by growing *B. breve* UCC2003 strains harbouring each of the individual constructs in nonselective media with successive subcultures at 12 hrs for approximately 132 generations. At regular time intervals, 100 µl of each culture was serially diluted to 10^{-9} and spot plated on both non-selective and selective agar plates. Colonies were counted and the CFU/ml obtained on both non-selective and selective plates was determined. Total DNA was isolated from five resistant colonies for each construct and plasmid content assessed by PCR with primers pRDconf_F and pRDconf_R.

Carbohydrate Utilization Assay

To assess the pullulan-dependent growth potential of strains *B. longum* NCIMB8809 and *B. longum* NCIMB8809 pRD1.3_*apuB,* mMRS medium was freshly prepared from first principles [41] and supplemented with 1% (v/v) of pullulan solution and 0.05% (v/v) of L-cysteine. This medium was inoculated with 1% (v/v) of stationary phase *B. longum* NCIMB8809 or *B. longum* NCIMB8809 pRD1.3_apuB

cells, from cultures cultivated overnight in 1% (v/v) glucose. The cultures were grown anaerobically at 37 °C and optical density was measured at OD_{600nm} after 24 hours using a UV-1280 spectrophotometer (Shimadzu Corporation, Kyoto, Japan). Analyses were carried out in triplicate.

Antibiotic resistance of vectors

The minimal inhibitory concentration (MIC) of chloramphenicol supported by the constructs in *B. breve* UCC2003 as a host was measured by the E-test method according to the manufacturer's instructions (AB *Biodisk*, Solna, Sweden). MIC assays were performed using Reinforced Clostridial Agar (RCA [Oxoid, Ltd., Basingstoke, Hampshire, United Kingdom]).

Determination of relative plasmid copy number

The relative copy number of pRD1-derived plasmids was assessed by quantitative real-time PCR using the culture and PCR conditions reported previously [42]. Amplification and detection were performed using a Fast real-time PCR system (Applied Biosystems, Foster City, CA) with QuantiNova SYBR Green PCR Kit (Qiagen). Primers 146A_F and 146A_R and, 146B_F and 146B_R (**Supplementary Table S2**) were designed based on the pRD1 *repB* sequence, resulting in amplicons of 120 bp and 150 bp, respectively. The 1-deoxy-D-xylulose 5-phosphate synthase gene (*dxsA*) (GenBank accession no. ABE95261) and the housekeeping gene *dnaA* of *B. breve* UCC2003 were used as single-copy chromosomal reference genes, respectively generating a 120 bp amplicon with primers *dxsA*_F and *dxsA*_R, and a 150 bp amplicon with primers UCC776B_F and UCC776B_R (**Supplementary Table S2**). The relative copy number of the derivatives was calculated using the formula $N_{relative} = 1+E^{-\Delta CT}$, where E is the amplification efficiency ΔCT is the difference between the threshold cycle number of the *dxsA*- and *dnaA*-targeting reactions and that of the *repB* target. Experiments were performed in triplicate; mean results are indicated.

Heterologous protein over-expression and purification

RepA, RepB, and RepBΔC were overexpressed and purified as N-terminal hexa-histidine fusion proteins using pQE30, an *E. coli* overexpression vector conferring an N-terminal six histidine tail to facilitate downstream purification. *E. coli* EC101 transformed with pQE30-*repA*, pQE30-*repB* or pQE30_RepBΔC was cultured in LB medium supplemented with 100 µg ml^{-1} ampicillin for growth at 37°C until the logarithmic phase (at OD_{600nm} of ~0.5-0.6) and induced by IPTG at a final concentration of 0.2 mM for 2 h at 37°C. Cells were harvested at 5,000 x *g* for 15 min at 4°C and the pellet re-suspended in 10 ml lysis buffer (50 mM Tris-HCl, 50 mM $CaCl_2$, 300 mM NaCl, 10 mM imidazole, pH 8.0). The cells were disrupted with 3 x 1 min rounds of bead-beating using the Mini-Beadbeater-16 (Biospec Products, Bartlesville, Oklahoma, USA) with 1 min intervals on ice. Cellular debris was removed by centrifugation at 12,000 × *g* for 30 min at 4°C. Proteins were purified from crude cell extracts via metal affinity chromatography.

Electrophoretic Mobility Shift Assays

Electrophoretic Mobility Shift Assays (EMSAs) were performed essentially as described previously [43]. Binding reactions were carried out in a final volume of 20 µl in the presence of poly[d(I-C)] in binding buffer (20 mM Tris-HCl [pH 7.0], 5 mM $MgCl_2$, 0.5 mM DTT, 1 mM EDTA, 200 mM KCl, 10% glycerol). Varying amounts of purified protein and the fixed concentration of probe were mixed on ice and subsequently incubated for 30 min at 37°C. Samples were loaded onto a 6 % non-denaturing Polyacrylamide (PAA) gel prepared in TAE buffer (40 mM Tris-acetate [pH 8.0], 2 mM EDTA) and run in a 0.5-2.0 x gradient of TAE at 100 V for 90 minutes in an Atto Mini PAGE system (Atto Bioscience and Biotechnology, Tokyo, Japan). Fluorescent signals were detected using the Odyssey Infrared Imaging System (Li-Cor Biosciences UK Ltd, Cambridge, UK) and captured using the supplied software (Odyssey V3.0). Band intensity of free probe and DNA-protein complex were calculated using the ImageJ software [44], [45]. Graphical representations and statistical analysis were performed using Prism5 (GraphPad Software, Inc., San Diego, CA). Details pertaining to each probe used for EMSAs are outlined in **Supplementary Table S3**.

Transcriptional start site determination by Primer Extension analysis

B. breve JCM7017 was grown in mMRS supplemented with 1% lactose to early exponential growth phase and total RNA isolated using a previously described method [46]. Primer extension analyses were performed by annealing 1 pmol of IRD700 labelled oligonucleotides to 15 μg of RNA as described previously [47]. Sequence ladders of the presumed promoter regions immediately upstream of *repA* or *repB* were amplified from genomic DNA with the relevant primers (**Supplementary Table S2**) using the Thermo Sequenase Primer Cycle Sequencing Kit (Thermo Scientific Inc., Waltham, Massachusetts, United States). Separation was achieved on a 6.5% Li-Cor Matrix KB Plus acrylamide gel. Signal detection and image capture were performed utilizing a Li-Cor sequencing instrument (Li-Cor Biosciences, Cambridge, United Kingdom).

GUS Assay

Intracellular β-glucuronidase (β-GUS) activity was determined essentially as described previously [40]. GusA activity was expressed in Miller Units and calculated using the equation: $1000^{*}((OD_{420} - (1.75^{*}OD_{550}))/ (t \times v \times OD_{600}))$, where t is the reaction time [mins], v is the cell volume [ml] and OD_{420}, OD_{550}, and OD_{600} are absorbance values at 420, 550, and 600 nm, respectively.

SEC RALS/LALS/RI Analysis

Size exclusion chromatography was carried out on an AKTA Pure HPLC system (GE Healthcare, Cork, Ireland) using a Superose 6 10/300 G/L column (GE Healthcare, Cork, Ireland) run in a buffer containing 50 mM Tris-HCl, 300 mM NaCl and 50 mM $CaCl_2$ at pH 7.5 with a flow rate of 0.5 mL/min. Proteins were injected at a final concentration of >0.8mg^{-ml}. Detection was performed using OmniSec REVEAL, a dual-angle light-scattering apparatus and refractometer (RALS/LALS/RI) (Malvern Instruments, Malvern, UK). Molecular weight calculations were performed with the OmniSec software (v10.4).

Nucleotide Sequence analysis and Comparative Analysis

Sequence data were obtained from the Artemis-mediated [48] genome annotations of the pMP7017 sequence [25]. Database searches were conducted using non-redundant sequences accessible at the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov) using BlastP or BlastN. Sequence verification and analysis were performed using the Seqman and Seqbuilder programs of the DNASTAR software package (DNASTAR, Madison, WI, USA v10.1.2). Amino acid sequences obtained from NCBI© were aligned using ClustalOmega© prior to gap removal, the finished alignment was generated using Jalview© in order to highlight conserved residues. Pfam [49] and NCBIs CDD search tool were used to identify functional domains. Structural analyses were performed using DALI [50] and Phyre2 [51], structural comparisons and overlays were achieved with The PyMOL Molecular Graphics System, Version 2.3.5 Schrödinger, LLC. The Genomic Nucleotide Skew application (Developed by TU Munich; maintained by Department of Computational Systems Biology of the University of Vienna, Austria) was used to plot the GC-skew across the pMP7017 sequence using a window size of 200 bp and a step size of 100 bp. Nucleotide sequences of *repB* from pMP7017 and *B. longum* homologues were extracted from the NCBI database and aligned with Clustal Omega [52], phylogeny was inferred using MEGAX [53] following gap removal. The Blast+ terminal [54] was used to determine percentage identity, the analysis was conducted in R ("R: The R Project for Statistical Computing," https://www.r-project.org) and the comparative figures produced using the genoPlotR package [56].

RESULTS

In silico analyses of the regions flanking *repA* and *repB*

As data pertaining to replication features of Gram-positive megaplasmids and bifidobacterial replicons are scarce, we focused our attention on two previously identified putative replicase genes [25] to assess the intrinsic features of pMP7017 replication. Inspection of the *repA* gene (nucleotide coordinates 97,215...96,166; GenBank accession no. KM406416) and surrounding sequences identified features characteristic of plasmid replicons within the 5' region of the *repA* coding sequence, such as an A+T rich region, a 35 bp palindrome and eight 15 bp direct repeats, designated here as A1 through to A8, with the consensus 5'-R**G**Y**GTCA**RW**ATTCTA**-3' (**Figure 1A**). An exploration of the non-coding region upstream of the *repB* gene (nucleotide coordinates 120,394...119,397) did not unveil any features typical of a plasmid origin. *In silico* scrutiny of the sequence located downstream of *repB* revealed a small 267 bp open reading frame (ORF) encoding a hypothetical, basic (pI 11.2) 10 kDa protein without any discerning motifs and referred to here as *hptX.* Furthermore, seventeen irregularly spaced 20-mer repeats, designated B1 through to B17 (**Figure 1B**), were identified conforming to the consensus 5'-hvr**TGG**m**CTATC**rs**TG**hmdh-3' with an additional sequence, B18, identified outside of the repeat region (nucleotide coordinates 115,314...115,334). Analysis of these repeats revealed that 20 mers B5 and B9 differ in their boundary sequences from the other repeats which appear to contain overlapping inversely orientated sequences, 5'-TTTTACACC-3', resembling those of putative DnaA boxes observed in the chromosomal origin of *Bifidobacterium* species [23].

When replication occurs bi-directionally from a single origin such as with theta-type replication, there are two replication forks of opposite polarity, which have been shown to exhibit a GC-skew [57]–[59]. Thus, in an effort to define the location of the replication origin and terminus the GC-skew properties across pMP7017 were plotted. A clear asymmetry was observed with two strong polarity switches at nucleotide positions 117,611 bp and 37,431 bp, presumably representing the approximate locations of the replication origin and terminus, respectively. This trend in disparity is reflected in a Z-curve analysis which predicts the *oriV* to be located at

position 120,395 bp. These findings, therefore, suggest that the replication origin is located in the vicinity of *repB*.

Analysis of the protein sequence of RepA, revealed a RepA_C domain (PF04796) at the N-terminus (interval: 57 – 215 aa, E-value: 1.6e-16). RepB also contains this domain at its C-terminus (PF04796, interval: 193 – 361 aa, E-value: 1.1e-12), which is part of the Rep_3 superfamily (CDD: cl19398, interval: 193-361 aa, E-value: 5.09e-08). Members of this family of replicases are found in a number of IRPs from Gram-negative bacteria such as RK2 [59], R6K [60], pSC101 [61], F [62] and pPS10 [61], and low GC, non-conjugative plasmids of Gram-positive origin such as pSK639 from *Streptococcus epidermidis* [63]. Homology modelling further supports this finding, as 57% of the total amino acid sequence of RepB (residues 102 –291) was modelled heuristically at > 90% confidence, using PDB templates of the C-terminus of IRP initiators RepE from F plasmid and the R6K π monomer, represented by c2z9oB [64] and c2nraC [65], respectively. RepB is therefore structurally most similar to replicases of IRPs, which are known to replicate by means of the theta-type mechanism.

The *repB* gene encodes the pMP7017 replication initiator protein

To establish the potential involvement of either (or both) of the putative replication initiator-encoding genes in pMP7017 replication, the *repA* and *repB* genes, in addition to the regions flanking their coding sequences, were individually cloned into the bifidobacterial replication probe vector pBif [39], generating pBif_*repA* and pBif_*repB*, respectively. Recombinant plasmids for each of the replication genes were successfully obtained in *E. coli* and were then introduced into *B. breve* UCC2003 by electroporation to assess their ability to independently facilitate replication. Failure to obtain *B. breve* UCC2003 pBif_*repA* transformants suggests that the 1,829 bp fragment encompassing *repA* and upstream repeat region was insufficient to allow autonomous replication (**Figure 1A**). In contrast, recovery of *B. breve* UCC2003 transformants containing pBif_*repB* demonstrated that this 3,258 bp region harbours all relevant features to facilitate autonomous replication in *Bifidobacterium* and thus is assumed to represent the basic pMP7017 replicon (**Figure 1B**). To corroborate this finding, we adopted a strategy of site-directed insertional mutagenesis to investigate the dispensability of either *repA* or *repB* for replication using a non-replicating vector

which integrates into the targeted gene via homologous recombination. Successful disruption of the *repA* coding sequence, as confirmed by Illumina sequencing, demonstrates that this gene is not essential for pMP7017 replication, while the inability to obtain a *repB* mutant despite multiple efforts indicates the indispensability of *repB*.

Iterons B12 - B15 and flanking AT-rich region are essential cis-acting elements for pMP7017 replication

To determine the minimal boundaries of the functional *repB*-associated replicon, successively shorter fragments of the 3,258 bp *repB*-containing replication region were cloned into the *Bifidobacterium* replication probe vector pBif and assayed for their capacity to support autonomous replication in *B. breve* UCC2003. In this manner, the minimal replicon was determined to be the 2,366 bp fragment of pRD1.4, which contains the *repB* promoter and coding sequence, *hptX* and seventeen 20-mer repeats. In order to determine the essentiality of individual iterons and the co-transcribed *hptX,* five additional constructs were made, pRD1.7 – pRD1.12, represented graphically in **Figure 1B**. Plasmid derivatives pRD1.7, pRD1.8, pRD1.9 and pRD1.10 were replication proficient in *B. breve* UCC2003, whereas derivatives pRD1.11 and pRD1.12 failed to replicate. It can be deduced from this data that the *hptX* coding sequence and iterons B2, B3, B4 and B9 are dispensable for replication, while iterons B12-B15 and its flanking AT-rich region represent essential *cis*-acting elements.

Iterons B2, B3 and B4 confer stability of the pMP7017 replicon

While the *hptX* gene, iterons B2, B3 and B4, and iteron B9 individually were shown to be non-essential for pMP7017 replication, these features may nonetheless play a role in other functions, such as plasmid segregation, stability and/or copy number. To ascertain the role of these elements, if any, in the stability of the replicon, *B. breve* UCC2003 harbouring each of the replication-proficient pRD derivatives was grown in the absence of selection pressure for 132 generations, then assayed for the presence of the plasmid as a measure of stability, presented in **Figure 1C**. This data shows that plasmids containing all putative iterons, pBif_*repB*, pRD1.2, pRD1.3,

pRD1.4 and pRD1.7, are quite stable as they are retained in the population for at least 132 generations without selective pressure. Instability was observed for plasmids pRD1.9, lacking two iterons (B2 and B3), and pRD 1.10 lacking three iterons (B2, B3 and B4) with total loss of the plasmid occurring at ≤ 52 and ≤ 35 generations, respectively. Furthermore, the absence of iteron B9 from plasmid pRD1.8 was shown to cause only a minimal loss of stability. The removal of the *hptX* gene appears to have no adverse effect on stability nor does the absence of the region located 5' of the *repB* promoter. Plasmids pRD1.2 and pRD1.3 were shown to be the most stable among the assessed plasmids as they were maintained in >85% of cells at 100 generations. Together these findings indicate that the iterons contribute to plasmid stability and highlight the 2,542 bp fragment of pRD1.3 as the basic replicon due to the greater stability of this plasmid derivative, apparently containing all features necessary to support replication and ensure stable maintenance.

Iterons B2, B3, B4 and B9 do not participate in copy control

The plasmid copy number (PCN) of the eight replicating pRD derivatives and megaplasmid pMP7017 was determined by the relative quantification method using single-copy housekeeping genes on the *B. breve* UCC2003 chromosome as a reference. As data displayed in **Figure 1D** illustrates, the PCN of all pRD1 plasmids approximate 1 copy per chromosome, with the exception of pRD1.3 with a somewhat higher value of 1.68 +/- 0.16. This data illustrates the copy number of all pRD1 derivates are comparable to that of pMP7017 from which the replicon derives, existing at approximately 1 copy/chromosomal equivalent in *B. breve* UCC2003, irrespective of the absence of the 5' region, *hptX* or iterons B2, B3, B4 and B9, thus suggesting these features do not play a major role in copy control. These non-essential iterons may nonetheless encode functions not explored in this work, considering their lack of participation in copy-control they will subsequently be referred to as direct repeats.

repA and *repB* are transcriptionally regulated in pMP7017

To gain insight into *repA* and *repB* transcription and possible transcriptional regulation, the presumed promoter regions and 5' UTRs of each *rep* gene were cloned upstream of the promoter-less β-glucuronidase (β-GUS)-encoding gene in the reporter plasmid pNZ272 [11], [66], generating plasmids pNZPP_*repA* and pNZPP_*repB*, respectively. These plasmids were individually introduced into *B. breve* UCC2003 and *B. breve* UCC2003 pMP7017_199, facilitating transcriptional quantification of these promoter regions in both the absence and presence of the megaplasmid. *B. breve* UCC2003 pMP7017_199 contains a tetracycline-tagged pMP7017 derivative with an insertion in ORF199 [25] encoding a surface-associated protein which displays no phenotypic deviation from the wild type and is used in the current work for the purpose of antibiotic-based plasmid selection. A bar chart of the data in **Figure S1A** illustrates a moderate level of transcriptional activity from both promoters in pNZ272 when present in *B. breve* UCC2003 pMP7017_199, while the intracellular β-GUS activity of the pNZ272 transcriptional fusions in a *B. breve* UCC2003 background was higher. *B. breve* UCC2003 pNZPP_*repA* revealed a marked increase in activity ~ 3.7-fold from its *B. breve* UCC2003 pMP7017_199 counterpart, while activity from the *repB* promoter increased ~1.6 fold in the absence of the megaplasmid. These findings indicate that the promoters of both *repA* and *repB* are repressed in strain *B. breve* UCC2003 pMP7017_199, suggesting that pMP7017 encodes trans-acting factors negatively regulating *repA* and *repB* transcription. To precisely define the location of the *repA* and *repB* promoters, the *repA* and *repB* transcription start sites (TSSs) were experimentally determined by primer extension (PE) analysis. A single extension product was obtained for *repA* corresponding to a position 71 nucleotides upstream of the predicted translational start site (**Figure S1B**). Investigation of the preceding sequence revealed hexamers resembling the consensus −10 and −35 recognition sequences of bifidobacterial vegetative promoters [67]. PE analysis for *repB* resulted in an extension product 131 nucleotides 5' of the predicted translational start site (**Figure S1C**) that is associated with a putative vegetative promoter sequence.

RepA binds to its corresponding promoter region

While it was unclear what role, if any, RepA plays in pMP7017 replication, mobility shift electrophoresis was performed to determine if the RepA protein interacts with probe 124P, containing the *repA* promoter and eight direct repeats A1-A8. The EMSA depicted in **Figure 2A** clearly shows the reduced mobility of DNA probe 124P upon addition of RepA, highlighting the affinity of RepA for this region. Assuming that the eight direct repeats contained within the probe are the recognition sequences of RepA, the observation of one strong shift despite eight potential binding sites suggests that RepA binding involves cooperativity, hence site saturation appears instantaneous rather than incremental. To further elucidate this interpretation, the fraction of bound probe (Φ) was calculated and plotted as a function of protein concentration to generate a specific binding curve. From the curve obtained we can see that RepA-124P complex formation is a non-linear function of RepA concentration consistent with cooperative interaction, further analysis revealed a Hill coefficient (nH) of 6.27 ± 1.36 reflecting marked positive cooperativity. Together this substantiates the notion that binding of RepA to each of the eight repeats within probe 124P does not occur independently. Furthermore, an apparent dissociation constant (K_d) of ~121.8 nM ± 8.0 nM was determined, from this the binding constant (K_a) was calculated to be 5.26 pM^{-1} suggesting a high affinity of RepA for its DNA substrate.

RepA exists predominantly as a homodimer in solution

The pronounced cooperativity observed in RepA binding to its cognate recognition sequence signaled a propensity for RepA-RepA interaction. To establish its oligomeric state in solution size exclusion chromatography was performed. From the chromatogram (**Figure S1D**) one main peak was observed at 17.87 retention volume (RV) (ml) representing a molecular weight of 82.49 kDa and 67.31% of the total protein, suggesting RepA exists predominantly as a homodimer. Two additional peaks can be observed at 164.8 kDa (17.23 RV; representing 16.98% of total protein) and 328 kDa (16.57 RV and representing 15.71% of the total protein) which we presume denote tetrameric and octameric forms of RepA, respectively, thus illustrating the multimerisation ability of this protein.

RepB does not display any affinity for its promoter or coding region

In an effort to ascertain where within the pMP7017 replicon RepB binds and to assess if the presumed RepB replication protein was able to physically interact with any of the identified DNA features flanking its coding sequence, nine IRD700-labelled probes were generated, designated here as A through to I, encompassing various segments of the replicon (**Supplementary Table S3**). EMSAs with probes A and B, encompassing the *repB* promoter-containing region, and probes C and D, spanning the *repB* coding sequence, revealed an absence of binding activity (**Figure S2**).

RepB binds specifically to iterons

In order to establish if RepB binds directly and specifically to the putative iterons identified 3' of its coding sequence, EMSAs were performed with DNA probes E through to I, each containing ≥1 20-mer repeat. EMSAs with each probe within this region resulted in clear mobility shifts, indicating that RepB is able to bind to the repeat-containing region downstream of the *repB* gene. EMSAs with short DNA fragments E and H (**Figure S3A-B**), each containing an individual 20-mer site, resulted in a single mobility shift interpreted as a single species of protein-DNA complex, RepB-E and RepB-H, respectively. To determine the quantitative parameters of these interactions, a titration curve was generated from the experimental data and fit to the Hill equation revealing a coefficient of ~1 in both cases with an apparent K_d of 254 ± 34 nM and 236.6 ± 66.5 nM for probe E and probe H, respectively. This result suggests an approximate K_a of 4.1 pM^{-1} ± 0.3 pM, reflecting high-affinity binding of RepB to the 20-mer sites B1 and B9. Additionally, the near equivalent estimations of dissociation suggest no predilection in terms of site orientation, while highlighting the interchangeability of residues 13 (A/G), 14 (C/G) and 18 (C/A) within the apparent recognition sequence, without any significant effect on affinity estimates.

RepB exists as a dimer in solution with no monomeric fraction

While quantitative evaluation of the binding isotherms allows for affinity approximation and estimation of the effective cooperativity behaviour, it was necessary to determine the oligomeric state of the protein to support our interpretation of the graphical presentations of the ligand-binding data. To achieve this size exclusion

chromatography was performed, revealing that when RepB was assayed at a high protein concentration (0.8 mg/ml) it exists as a homodimer in solution illustrated by one main peak at 18.75 RV in the chromatogram (**Figure S1E**) representing a molecular weight of 79.85 kDa and 97% of the total sample.

The binding behaviour of RepB to multi-site probes illustrates cooperative interactions

The EMSA (**Figure 2B**) employing DNA probe F containing presumed binding sites B2 and B3, clearly depicts two distinct electrophoretic species, which can be explained by assuming that the initial shift is the result of a single binding event, while the double shift occurs due to subsequent RepB binding to the second site, representing nucleoprotein complexes RepB-F and 2RepB-F, respectively. Analyses of the titration curve data generated from the binding reactions suggest positive cooperativity (nH = ~2), with an apparent dissociation constant of 13.7 nM ± 1.8 nM for RepB-F, and 21.8 nM ± 4.9 nM for 2RepB-F. Correspondingly, the EMSA displayed in **Figure S2C** demonstrates incremental binding of RepB to the 20-mer sites B2, B3 and B4 within probe G. Global sensitivity measures of the specific binding curves reveal positive cooperativity in the formation of complexes RepB-G (nH = 1.5 ± 0.34) and 2RepB-G (nH = 2.8 ± 0.62); conversely, a Hill coefficient of ~1 associated with 3RepB-G suggests independent binding at a third site. In consolidation with the results obtained from the EMSA employing DNA fragment F, it can be inferred that positive cooperativity occurs between sites B2 and B3 due to their relative proximity (3 bp intervening), while site B4 is bound independently. Although cooperativity has been indicated for contiguous tandem sites it was interesting to establish what binding behaviour would be observed with convergent proximal sites on opposing strands, a spatial arrangement observed for sites B16 and B17. EMSAs with probe I containing these sites display the formation of two protein-DNA complexes, RepB-I and 2RepB-I (**Figure 2C**). The persistence of nucleoprotein complex RepB-I, despite increasing RepB concentration, may be indicative of a faster K_{off} rate, potentially as a result of steric hindrance. Together these data illustrate an increased affinity of RepB for one recognition sequence in the presence of a second adjacent site, being reflected by an average ~20.7-fold reduction in the apparent K_d of the first binding event in addition to nH ≥ 2 suggesting two interacting sites.

GGnCnAnnR are critical bases within iterons for binding of RepB

To verify that RepB specifically binds to the putative iterons downstream of its coding sequence, thirteen core nucleotides (4 – 16) within the 20-mer motif were individually substituted by PCR site-directed mutagenesis, each probe (P1 – P18) was then assessed for its ability to bind RepB by EMSA. The results of this assay, displayed in **Figure 3**, revealed that substitutions of residues 5G>T (P2), 6G>T (P3), 8C>A (P5), 10A>C (P7) and 13G>T (P10), completely abolished the ability of RepB to bind, indicating that these residues are essential for RepB-iteron complex formation. Furthermore, the substitution of the remaining residues resulted in RepB-iteron binding events exhibiting a range of much-lowered affinities, which suggests a role in the stability of the complex. This data clearly shows that alterations within the motif impact on RepB-mediated DNA binding affinity and indicate that RepB specifically binds to GGnCnAnnR.

The RepB C-terminal domain is required for dimerization and DNA-binding

In silico structural homology modelling revealed a strong positive correlation of the C-terminus of the predicted 3D structure of RepB to the conserved iteron-binding domains of plasmid initiators RepE from F-plasmid [64] and π from R6K [68] (**Figure S4A**). Furthermore, structural analyses indicate RepB to be pseudo-symmetric with two putative DNA binding domains, a central winged helix-turn-helix 1 (wHTH1) and a C-terminal HTH2 flanked by a low complexity region rich in proline, serine and threonine residues (PST) (**Figure S4B**). Oligopeptides enriched in these amino acids are known to form flexible structures which participate in ligand binding under specific conditions [69], additionally, PST residues are known to serve as phosphorylation sites modulating protein activity [70], [71]. To define the role of the C-terminal domain in both DNA-binding and protein-protein interactions, a deletion mutant, RepB_ΔC, was created. RepB_ΔC was overexpressed and purified, followed by size exclusion chromatography to establish its oligomeric tendency in solution. From the chromatogram (**Figure S4C**) one peak at 18.7 RV (ml) representing 28.81 kDa and 100 % of the sample, indicating RepB_ΔC exists as a monomer in solution. Purified RepB_ΔC was then tested for its ability to bind iterons by EMSA.

The deletion variant showed no detectable binding to probe P17 as compared to the RepB control, revealing a functional requirement of the C-terminal 98 residues of the RepB protein in dimerization and DNA-binding (**Figure S4D**).

Highly conserved RepA and RepB homologues are present in all bifidobacterial megaplasmids

A comparative analysis of the RepA and RepB protein sequences in twelve recently identified bifidobacterial megaplasmids confirmed the presence of both RepA and RepB homologues in all plasmids. RepB is conserved at a high level (protein per cent identity range: 88.75 – 89.43%), while interestingly, RepA shows a higher level of conservation (protein per cent identity range: 89.97 – 96.56%). This indicates a potential functional importance of RepA despite its dispensability for pMP7017 replication.

The putative pMP7017 origin of replication is conserved between bifidobacterial replicons

A comparative analysis of the nucleotide sequence of the pMP7017 basic replicon between bifidobacterial megaplasmids, displayed in Supplementary data **Figure S5**, illustrates the preservation of the *repB* coding sequence and iteron-containing region. Each replicon retains a minimum of eleven iterons with between one and three additional iterons located outside the replicon (data not shown). Megaplasmids from strains *B. longum* MCC10043, *B. longum* MCC10070 and *B. longum* MCC10096 contain two genes, *orfA* and *orfB*, constituting an IS200 transposable element that is not present in the other plasmids. The *hptX* gene identified in the pMP7017 replicon is absent in all *B. longum* megaplasmids, supporting the experimental data that this hypothetical gene is not required for replication. In consideration of this data, the *hptX*, *orfA* and *orfB* coding sequences were removed and the nucleotide sequences aligned to better assess the iteron-containing region in terms of its conservation (**Figure S6**). The alignment displays the absence of sequences corresponding to pMP7017 20-mers B2, B6, B7 and B9 in one or more *B. longum* megaplasmids, while a moderate level of variability is observed in the intervening sequences between the first ten repeats (B1 – B10) which is not observed

between the remaining iterons. Furthermore, no sequence resembling the DnaA consensus motif was identified in the *B. longum* megaplasmids. Analysis of the curvature, bendability and GC content of this highly conserved region reveal a conspicuous peak of curvature (>7.5) from positions 118,659 bp to 118,686 bp, overlapping iterons B12, B13 and B14, followed by a region of high bendability in the AT-rich region with a GC content 14% less than the average from positions 118,772 bp to 118,264 bp.

pRD1.3: a novel bifidobacterial shuttle vector for cloning of large fragments

To verify the suitability of vectors containing pMP7017-derived replicons for cloning and expression purposes in bifidobacteria, the *apuB* gene from *B. breve* UCC2003, encoding an extracellular type II amylopullulanase, was directionally cloned into the multiple cloning site of the most stable pRD1 derivative, pRD1.3, resulting in plasmid pRD1.3_*apuB*. Both pRD1.3 and pRD1.3_*apuB* were introduced into *B. longum* NCIMB8809, a strain unable to metabolise pullulan [72]-[74]. *B. longum* NCIMB8809 pRD1.3 and *B. longum* NCIMB8809 pRD1.3_*apuB* were then assessed for their ability to utilise pullulan by measuring optical density after 24 hrs of incubation in MRS supplemented with the sugar. From **Figure 4A** it is obvious that *B. longum* NCIMB8809 pRD1.3 shows little to no growth (final OD_{600} 0.13 +/- 0.013), whereas *B. longum* NCIMB8809 pRD1.3_*apuB* displays substantial growth (final OD_{600} 1.80 +/- 0.16). The results show that pRD1.3_*apuB* confers pullulan-dependent growth in *B. longum* NCIMB8809, which validates the capacity of pRD1.3 as a novel bifidobacterial cloning and expression vector (**Figure 4B**). Plasmid pRD1.3 is presumed to carry all replication and stability functions of the pMP7017 megaplasmid, it may therefore be possible to employ pRD1.3 for cloning large (>25 kb and up to 100 kb) DNA fragments.

DISCUSSION

In the current work, *repA* was found to be nonessential for pMP7017 replication through successful site-directed mutagenesis, in addition to the inability of *repA* and surrounding regions to support autonomous replication in a replicase-deficient vector. As this gene is not critical for pMP7017 replication, *repA* may be a gene remnant encoding a dysfunctional replicase despite its high level of conservation among bifidobacterial megaplasmids. RepA was determined to exist predominantly as a dimer in solution with a tendency toward multimerization and was shown to bind its cognate promoter with high affinity.

We show that replication of the bifidobacterial megaplasmid pMP7017 depends on the plasmid-encoded initiator RepB, which exists as a dimer in solution at high concentrations with no observable monomeric fraction. RepB was demonstrated to be competent in iteron binding, displaying both iteron-dependent protein-protein interaction and iteron-independent interactions. The critical residues within the iteron sequence which facilitates the binding of RepB were established to be GGnCnAnnR, situated asymmetrically across the core of the 20-mer iteron. In all characterised IRPs, with the exception of the R6K π initiator, Rep proteins bind iterons as monomers, while Rep dimers are incompetent in iteron binding [75]. The study of Rep-iteron interactions is predominantly achieved with the use of dimerization-defective Rep mutants, through the addition of molecular chaperones in the binding reactions [76]–[78] or the use of chaotropic agents to induce dimer dissociation and facilitate the accumulation of iteron-binding active monomers. Interestingly, HIS-tagged WT RepB was shown to associate with a high affinity to iterons in EMSA in the absence of stimulation by host cell chaperones or the use of chaotropic agents. In addition, we did not observe a reduction in binding upon increasing RepB concentration nor did we observe a second shift representing a higher-order nucleoprotein complex in EMSA with single-site probes. Thus, it is tempting to speculate that the binding observed in this work is due to dimeric RepB. If this assumption is true, binding of iterons by dimeric RepB may inhibit replication as observed in the replication of plasmid R6K [75]. Dimeric RepB may then rely on the participation of temporally expressed cell cycle chaperones, such as DnaK/DnaJ for monomerization as described for many initiators of IRPs [64], [77], [79], [80].

The C-terminal defective mutant, RepB_ΔC, was determined to be both monomeric and incompetent in iteron binding. While this suggests that the 98 aa C-terminal end of RepB is required for oligomerisation and DNA-binding, the observed phenotype may also be due to protein instability or indeed signify a lack of affinity of RepB monomers for the 20-mer sites in favour of alternative sites. It has previously been reported that even a single point mutation within the C-terminal domain of structurally similar replication initiators, i.e. π from R6K plasmid and RepE from F plasmid, is enough to severely affect or abolish the ability of the protein to bind iterons [68], [81]. Thus, a more conservative approach to mutagenesis is necessary to substantiate this finding.

While the pMP7017 minimal replicon was defined as a 2,366 bp region (corresponding to pMP7017 nucleotide coordinates: 118,220…120,584) encompassing the *repB* coding sequence and seventeen repeats spanning 1,159 bps, it was established that not all features contained within the replicon are essential. Deletion of the *hptX* coding sequence did not have any negative effect on replication. In addition, it was demonstrated that 20-mers B2, B3, B4 and B9 are not essential. While repeats B2-B4 were shown to participate in the stability of the replicon, the removal of these sequences displayed no perceived effect on plasmid copy number suggesting the pMP7017 copy control functions are encoded by alternative iterons. Contrastingly, the removal of iterons B12-B17 or the AT-rich region located between iterons B15 and B16 abolished the plasmid's capacity for replication, indicating this region contains essential cis-acting elements and thus likely constitutes the pMP7017 origin of replication. This data is supported by a comparative analysis of the pMP7017 basic replicon between bifidobacterial megaplasmids which revealed a high degree of variability between and within repeats constituting the putative stability region, which is not observed in the highly conserved region containing the essential iterons B12-B17.

Although experimental validation is required to fully unravel the intricacies of pMP7017 replication, we here propose a mechanistic model for this bifidobacterial replicon (**Figure 5**). We hypothesize that the pMP7017 iterons constitute a framework for the binding of RepB monomers that interact to form a structure that disrupts the DNA helix, facilitating DNA unwinding at the flanking AT-rich region. Reminiscent of the layout of DnaA boxes within the *oriC* of bacterial genomes, RepB iterons are arranged with varying affinities, distances and orientations suggesting a dynamic

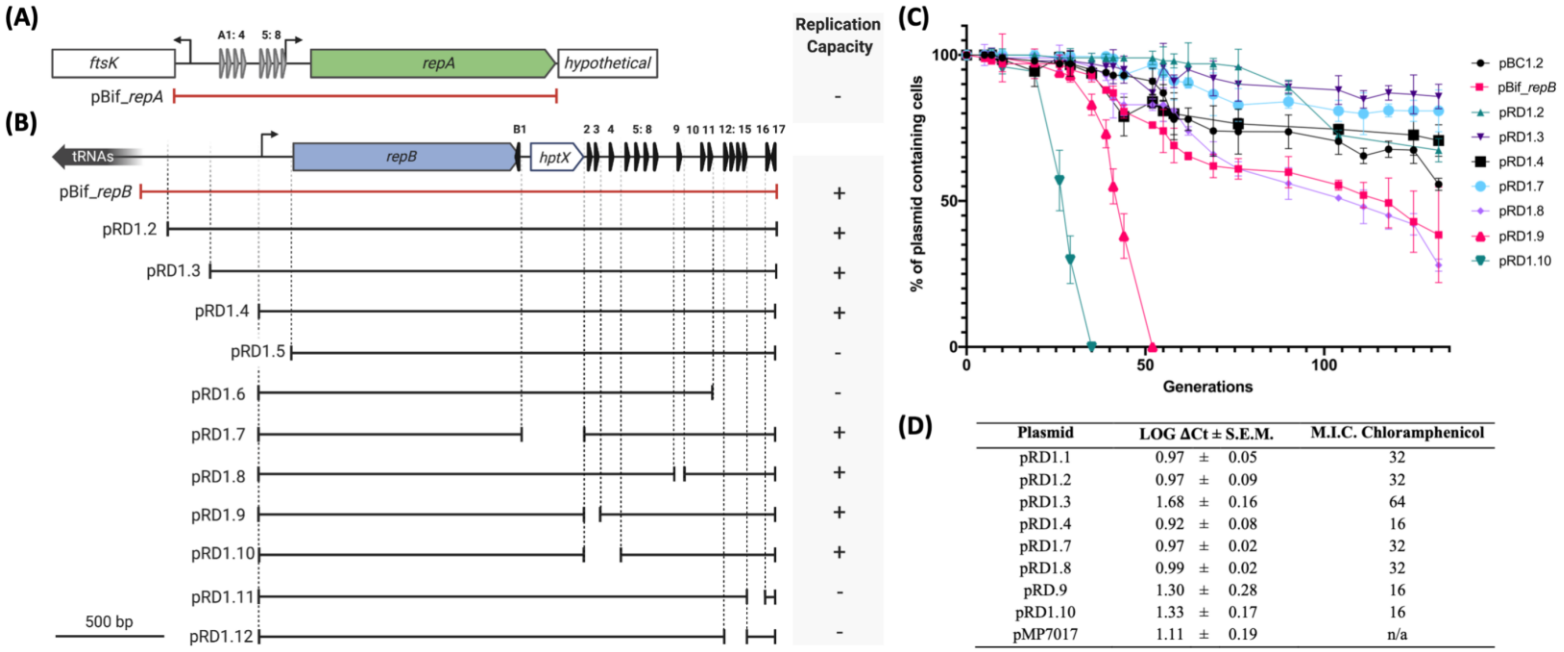
system regulating the formation of specific higher-order nucleoprotein complexes in the occlusion and stimulation of origin opening [82]. In addition to their importance for RepB assembly at the origin, the difference in sequence may constitute additional regulatory measures through weaker vs stronger binding sites. Furthermore, iteron-bound Rep protomers may inhibit origin melting as a result of coupling two origins in a process of replication inhibition, referred to as 'handcuffing'. This dual functionality of iterons is a copy control strategy commonly adopted by large, low copy number plasmids preventing plasmid over-replication through origin pairing [79].

As the availability of genetic tools for the study of bifidobacteria is limited, the characterisation of bifidobacterial replicons is necessary to develop new cloning and shuttle vectors. Considering the pMP7017 replicon is sufficient to drive replication of a plasmid of >190 kb, it can be inferred that pRD1.3 represents a novel low-copy shuttle vector with the capacity to allow cloning of large fragments in bifidobacteria without instability or toxicity issues.

AUTHOR CONTRIBUTIONS

R.D. and D.v.S. designed and directed the project and experiments. R.D. performed all experiments outlined in this work. C.P. aided primer extension analysis. P.K. performed HPLC analysis of Rep proteins. M.J.B.B. assisted in the comparative analysis of the pMP7017 replicon. R.D. and D.v.S. analysed and interpreted the data. R.D drafted the manuscript and figures, R.D., D.v.S. and M.O.C.M. provided commentary and edits to the manuscript and figures in preparation of the final version of the article.

FIGURES AND TABLES



(D)

Plasmid	LOG ΔCt ± S.E.M.	M.I.C. Chloramphenicol
pRD1.1	0.97 ± 0.05	32
pRD1.2	0.97 ± 0.09	32
pRD1.3	1.68 ± 0.16	64
pRD1.4	0.92 ± 0.08	16
pRD1.7	0.97 ± 0.02	32
pRD1.8	0.99 ± 0.02	32
pRD.9	1.30 ± 0.28	16
pRD1.10	1.33 ± 0.17	16
pMP7017	1.11 ± 0.19	n/a

Figure 1. pRD derivative plasmids; stability and copy number

(A) Representation of the genomic architecture of *repA* and flanking regions. Grey arrowheads represent eight direct repeats preceding and overlapping the *repA* promoter, A1-A8. The region tested for replication is indicated by a red line. **(B)** Genetic organisation of *repB* and surrounding regions. tRNAs are depicted as black arrows, while the seventeen 20mer repeats, B1-B17, are represented as black arrowheads. The region tested for replication is indicated by a red line. A

schematic representation of the pMP7017 replicon within each pRD1 derivative plasmid is depicted below. **(C)** Graph illustrating the segregational stability of the pRD1 derivatives over 132 generations, each data point is averaged from three independent replicates, standard deviations are indicated as error bars. **(D)** Table of pRD1 plasmid copy numbers and MIC of chloramphenicol resistance.

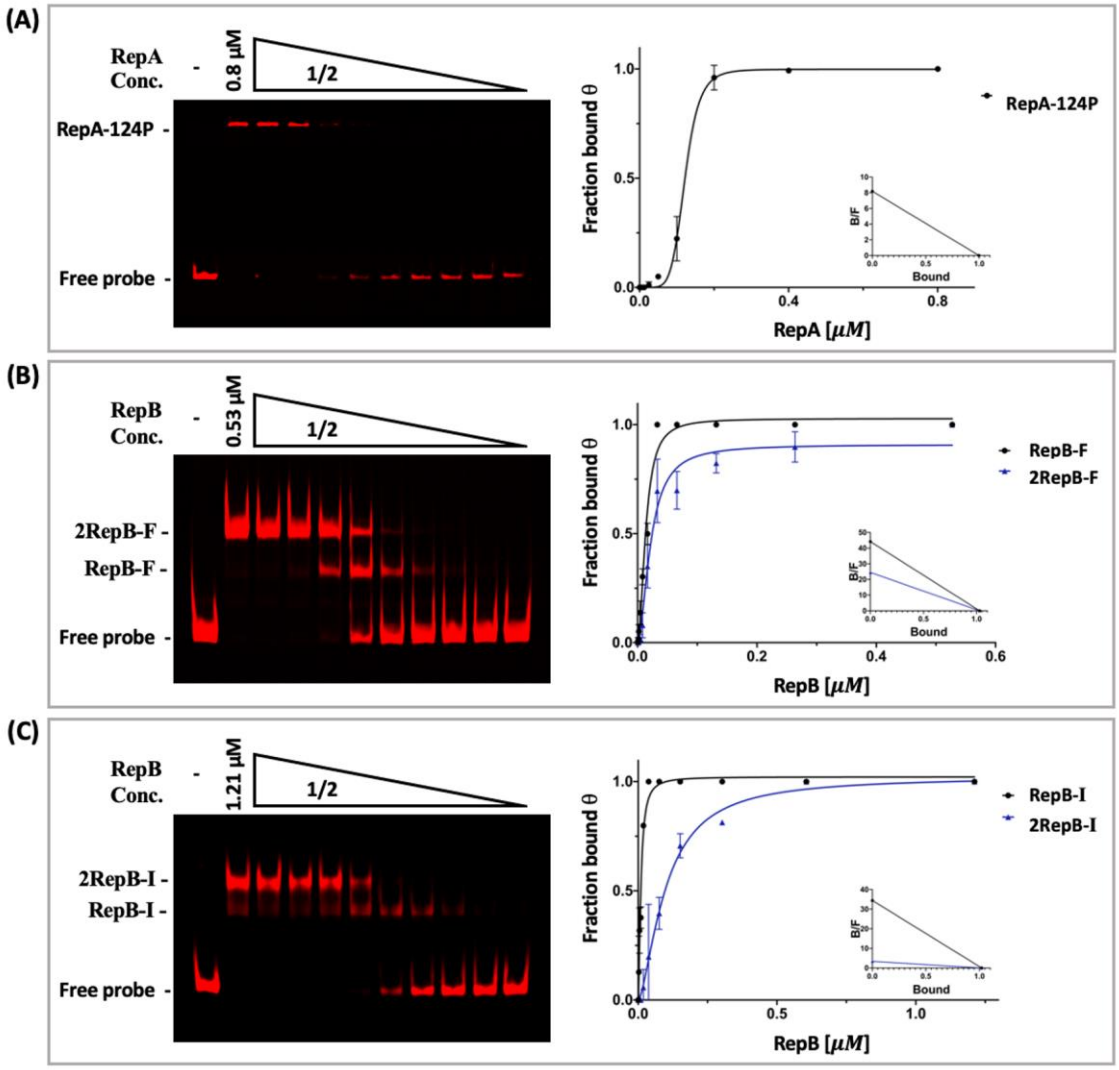


Figure 2. The binding behaviour of RepA to its corresponding promoter region and of RepB to multi-site probes in EMSA

(A) EMSA image displaying the interaction of RepA with promoter probe 124P. The binding isotherm and Scatchard plot (inset) of this data is presented to the right. **(B)** A representative EMSA gel displaying two clear shifts representing the binding of RepB to the 2-site probe, F, containing two tandemly arranged iterons. The binding isotherm and Scatchard plot (inset) of this data is presented to the right. **(C)** EMSA image of RepB binding to 2-site probe, I, containing two invertedly orientated iterons. The binding isotherm and Scatchard plot (inset) of this data is presented to the right.

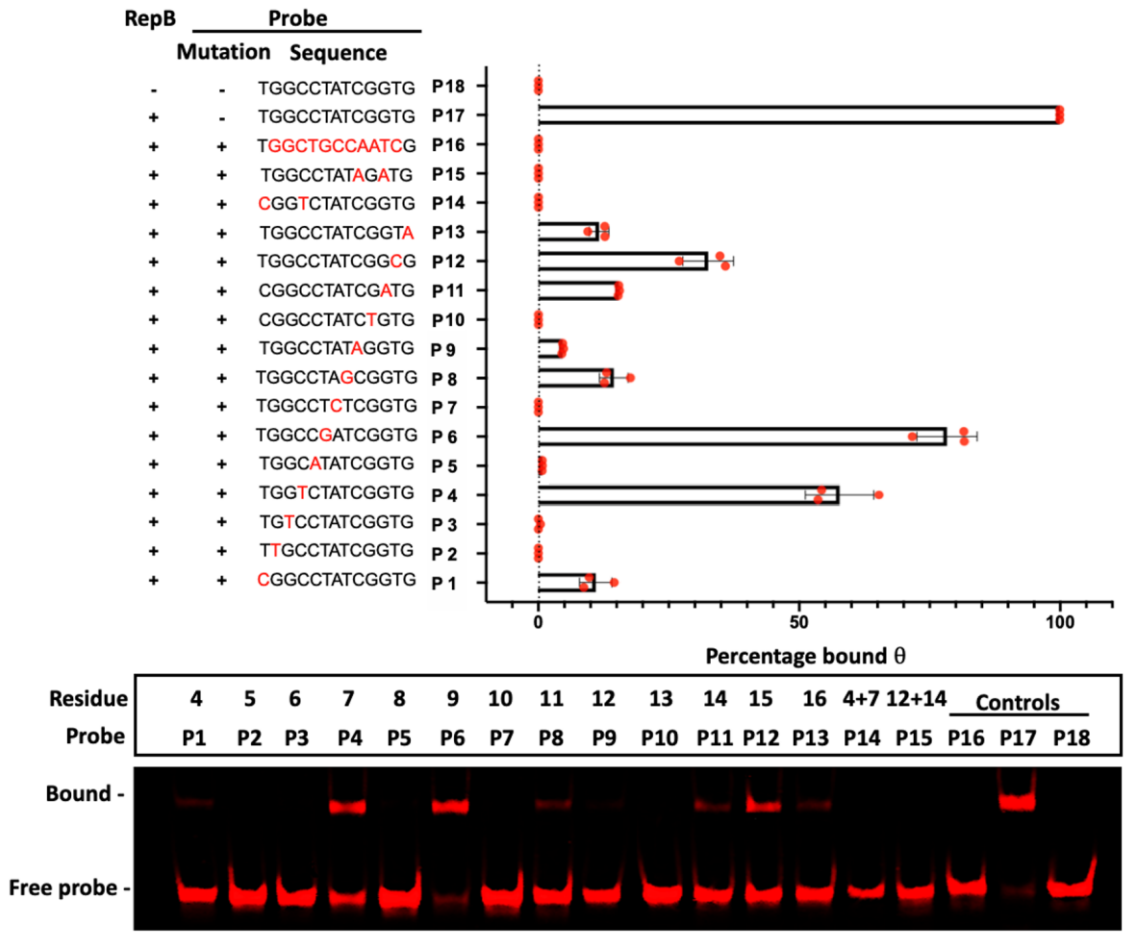


Figure 3. Identification of the critical nucleotide residues for the binding of RepB to iterons

A bar chart representing the percentage of probe bound by RepB, the presence or absence of RepB in the reaction is denoted by a + or – symbol, respectively. The sequence of each probe is given and the incorporated mutation/s are indicated in red text. An EMSA gel image representing a replicate from which the bar chart is based is displayed at the bottom.

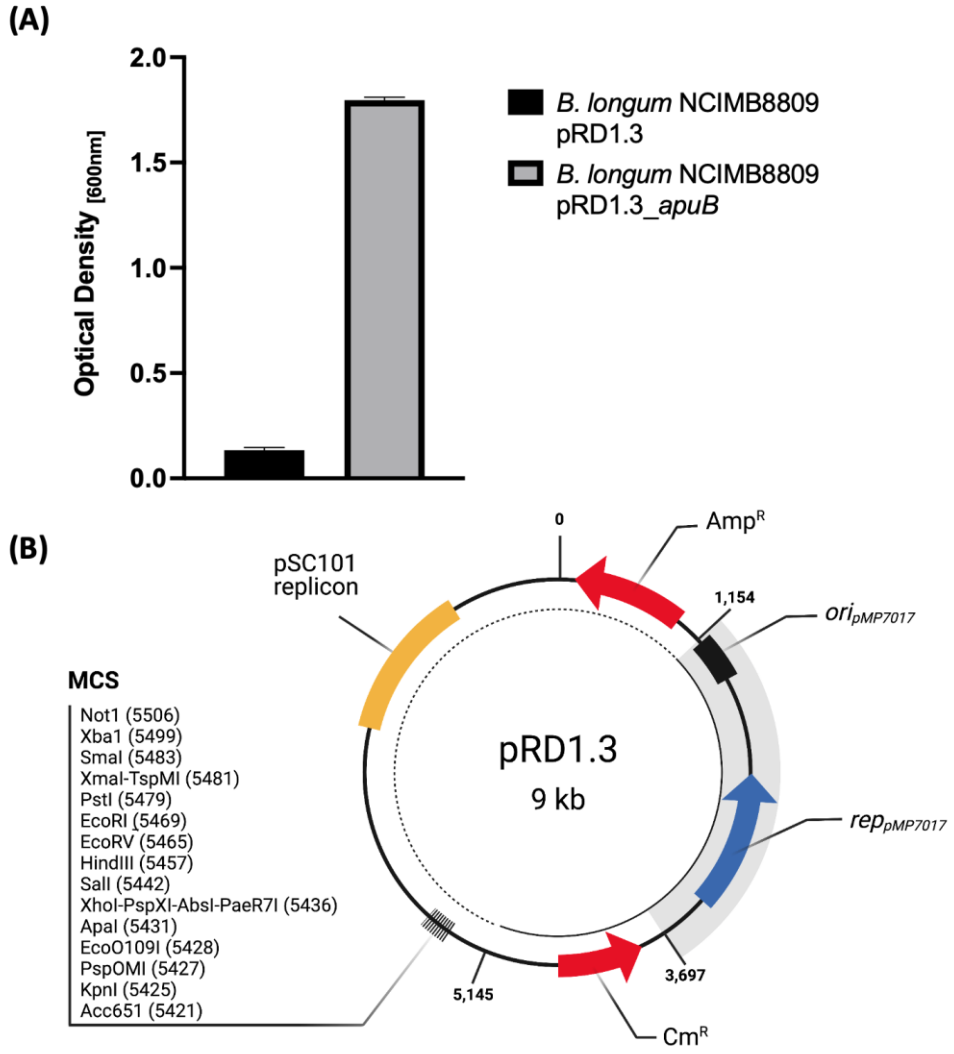


Figure 4. Development of a novel bifidobacterial shuttle vector, pRD1.3

(A) Bar chart displaying the final optical density values of *B. longum* NCIMB8809 and *B. longum* NCIMB8809 pRD1.3_*apuB* growth after 24 hr of incubation with 10% w/v pullulan as a sole carbon source. **(B)** Genetic map of plasmid pRD1.3 indicating the relevant features, the pMP7017 replicon is highlighted by a grey bar.

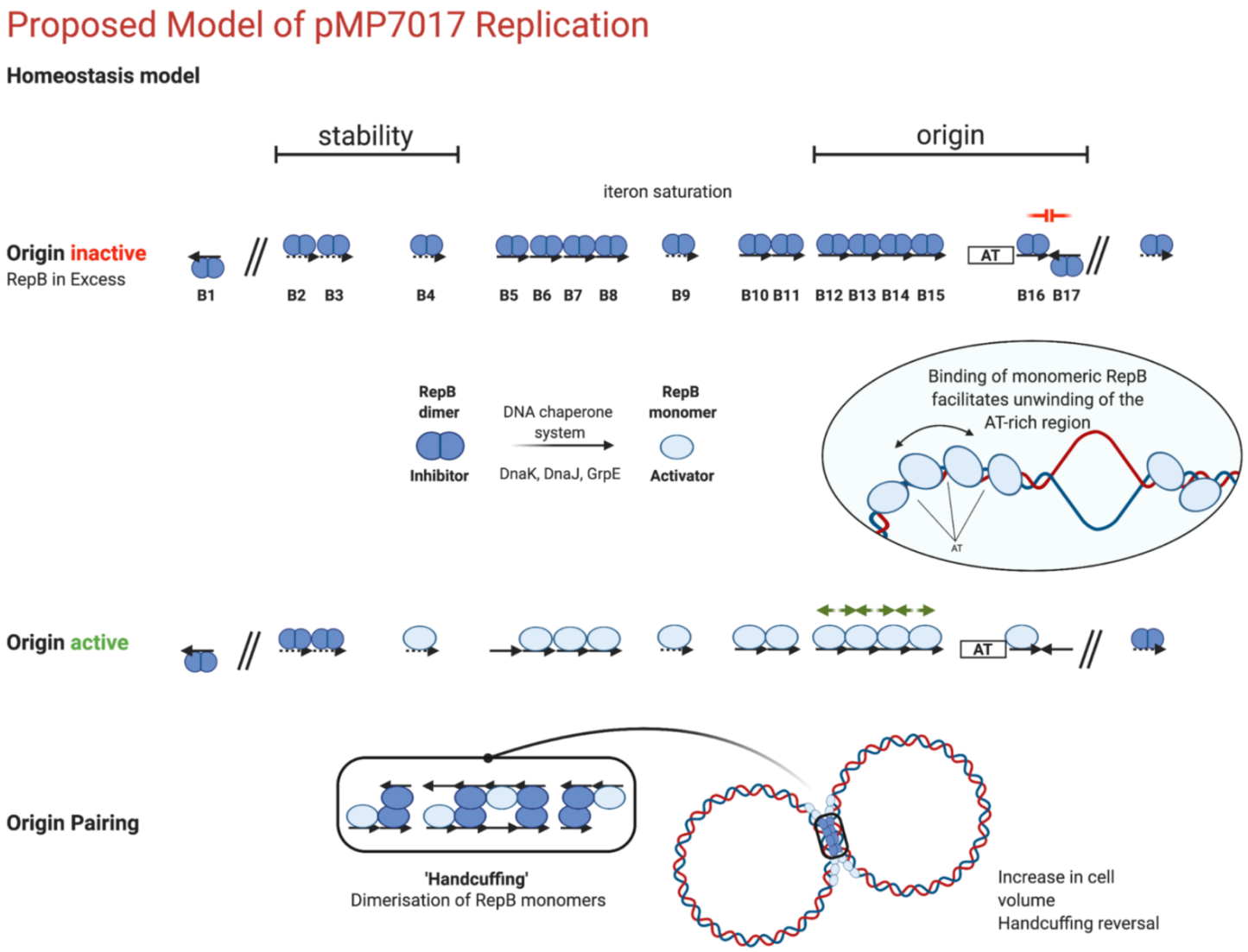


Figure 5. Working model of pMP7017 replication initiation

Iteron saturation by RepB dimers prevents origin firing. Cell-cycle chaperones facilitate the conversion of RepB dimers to active monomers. Cooperative interaction (represented by green arrows) of RepB monomers bound to tandemly arranged iterons facilitate the unwinding of the flanking AT-rich region. While iteron bound monomers on opposite template strands may participate in oligomerization to prevent over-replication in a process termed ‘handcuffing’. Iterons are depicted as solid black lines and non-essential direct repeats are shown as broken arrows.

SUPPLEMENTAL MATERIAL

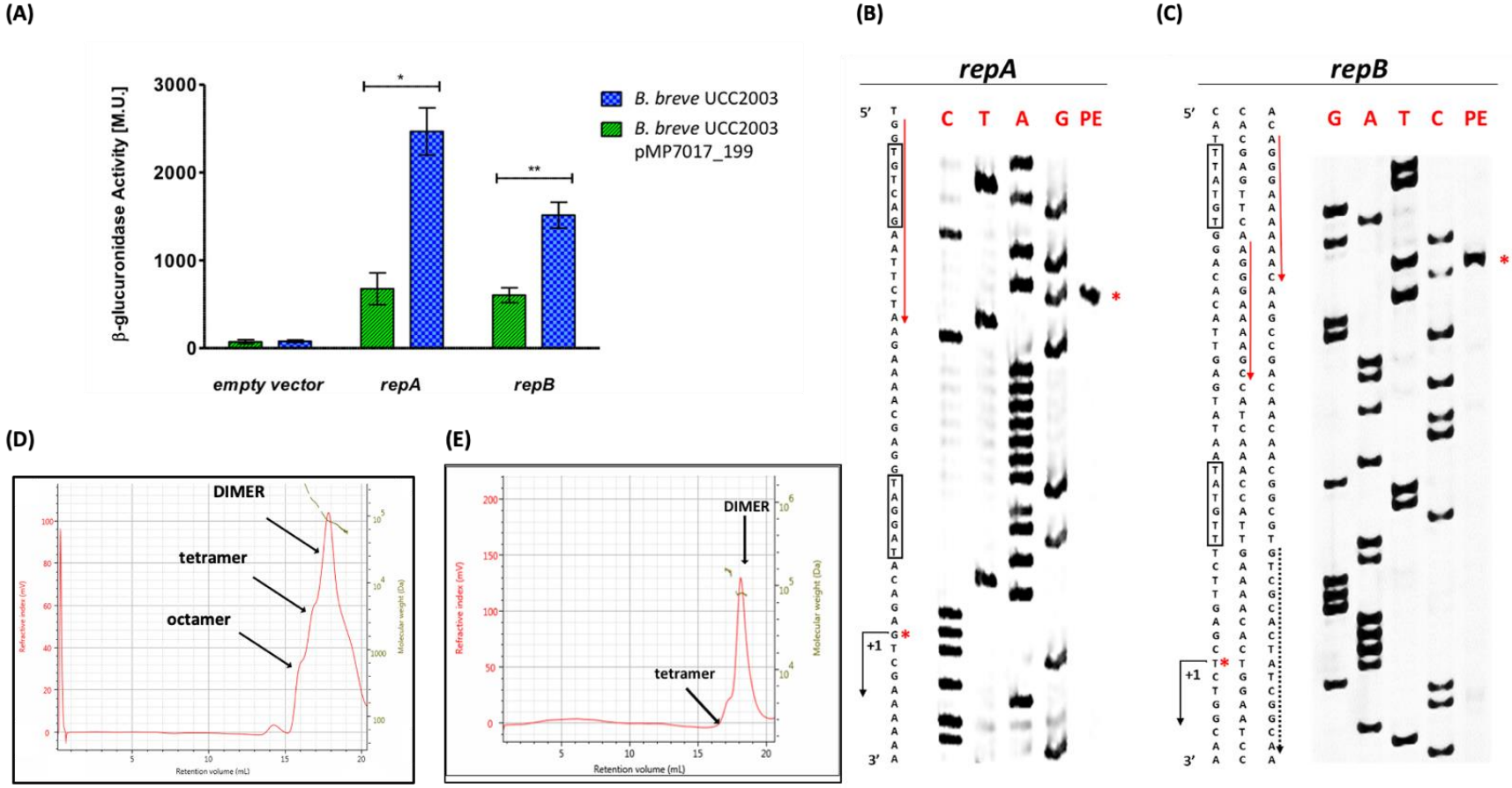


Figure S1. Promoter analysis of the *repA* and *repB* genes and HPLC chromatograms

(A) A bar chart of the β-galactosidase activity of the *repA* and *repB* promoter fusions in *B. breve* UCC2003 cells, in the presence or absence of pMP7017. **(B)** The *repA* promoter sequence is presented to the left of the primer extension gel for the *repA* promoter. Black boxes indicate the -10 and -35 hexamers, the overlapping A8 repeat sequence is indicated by a red arrow, while the experimentally determined transcriptional start site (TSS) is indicated by a red asterisk in both the sequence and the primer extension gel. **(C)** Primer extension analysis of the repB promoter, the solid red and black dotted arrows highlight potential regulator motifs. Additional features are as above. **(D)** Size exclusion chromatogram of RepA, the main peak represents the RepA dimer, with minor fractions of tetramers and octamers indicated. **(E)** Size exclusion chromatogram of purified RepB, the main peak represents the RepB dimer, with a minor tetrameric fraction indicated.

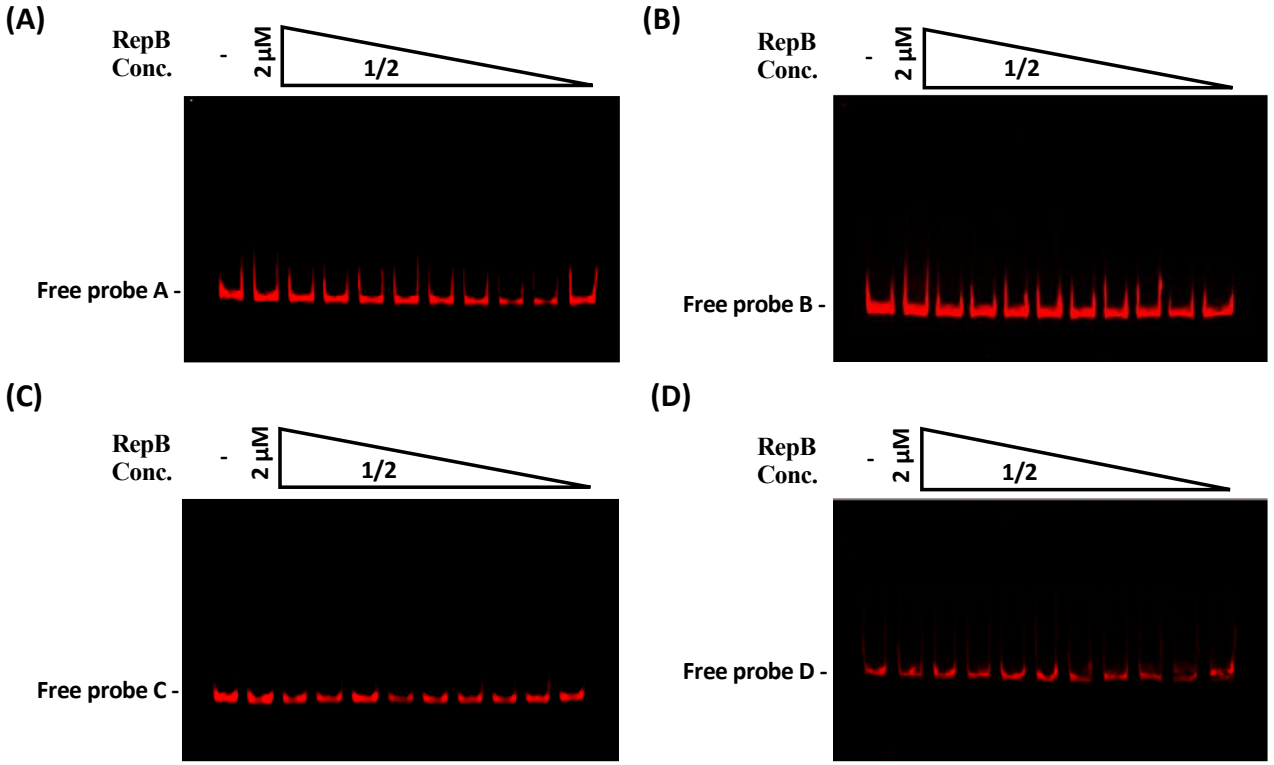


Figure S2. EMSAs representing the lack of affinity of RepB for its promoter and coding regions

(A) An EMSA gel revealing an absence of binding of RepB to probe A, containing the region preceding its promoter. **(B)** An EMSA gel illustrating an absence of affinity of RepB for probe B, containing the *repB* promoter. **(C)** A representative EMSA gel, illustrating the lack of affinity of RepB for probe C, containing the *repB* N-terminal coding sequence. **(D)** A representative EMSA gel, illustrating the lack of affinity of RepB for probe D, containing the *repB* C-terminal coding sequence

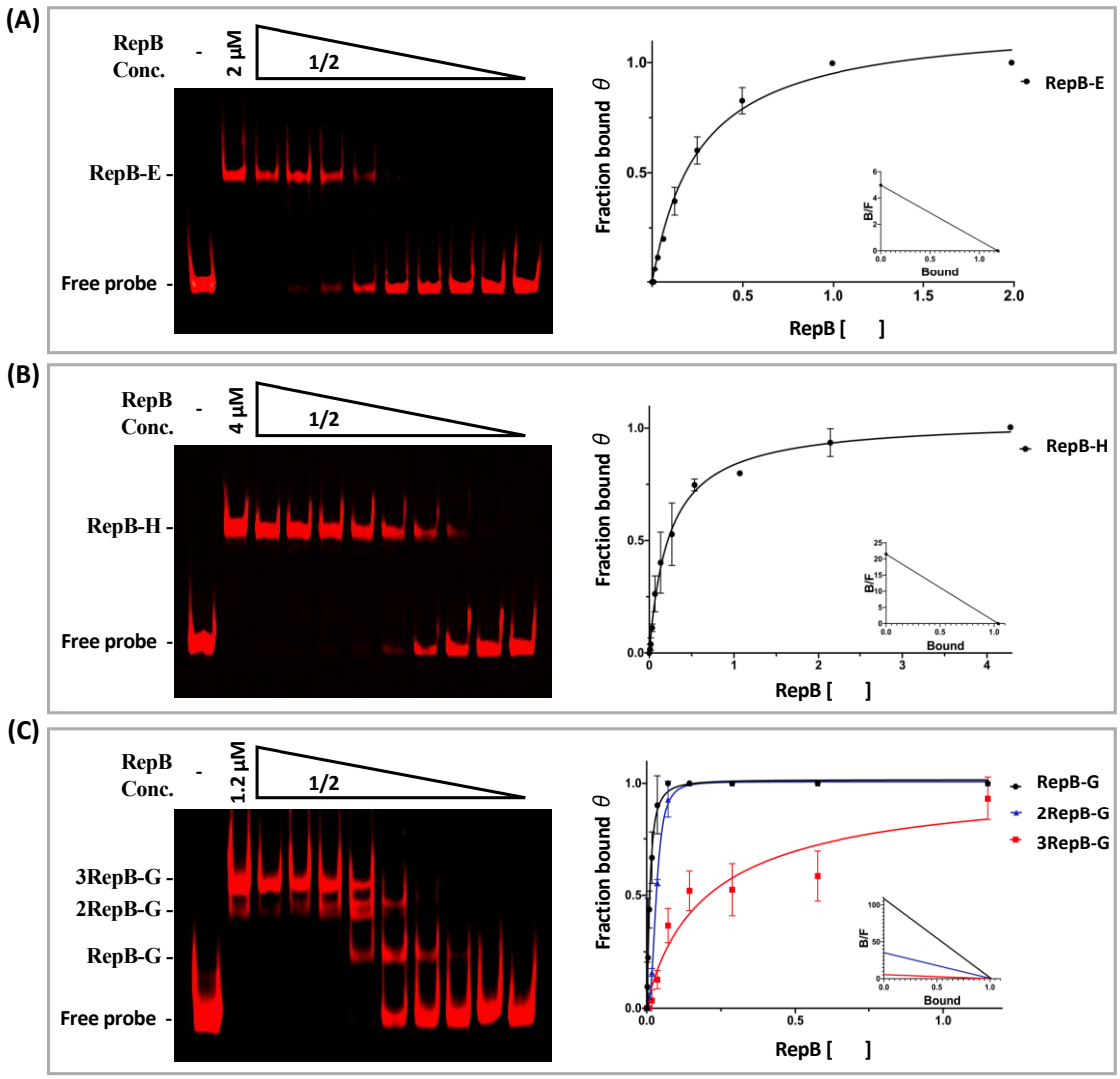


Figure S3. EMSA data illustrating the affinity of RepB for single-site probes, E and H, and probe G containing 3 putative iterons

(A) EMSA data of RepB binding to probe E containing a single iteron, illustrating a single shift representing RepB-E nucleoprotein complex. A binding isotherm of triplicate data is presented to the right indicating an apparent K_d of 254 ± 34 nM. **(B)** EMSA data of RepB binding to single-site probe, H. One shift is observed, representing a single nucleoprotein complex RepB-H. The binding isotherm is located to right, indicating an apparent K_d of 236.6 ± 66.5 nM. **(C)** EMSA data of RepB binding probe G containing three iterons. Three nucleoprotein complexes are observed representing the binding of RepB to one, two or three iterons, RepB-G, 2RepB-G and 3RepB-G, respectively. The binding isotherm of triplicate data is presented on the right indicating an apparent K_d of 10.7 ± 2.2 nM, 32.8 ± 1.5 nM and 216.5 ± 61.4 nM, respectively.

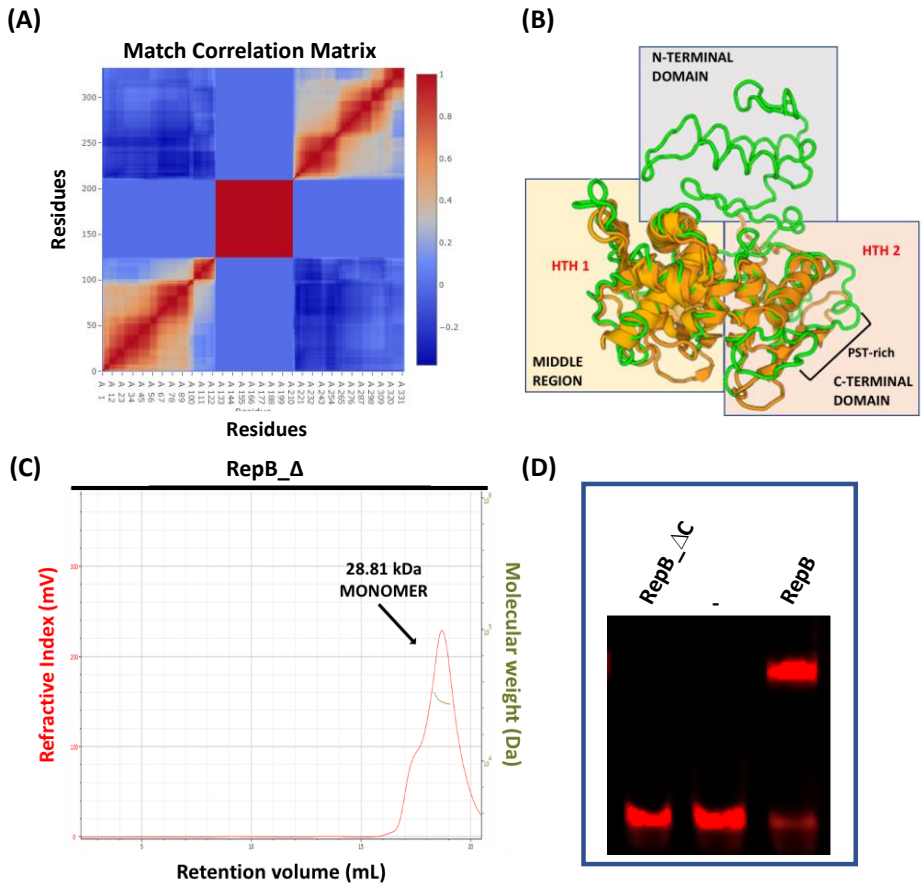


Figure S4. Heuristic modelling of RepB to replicases of IRPs, HPLC chromatogram of RepB_ΔC and EMSA displaying the lack of affinity of RepB_ΔC for probe, P17

(A) Anti-correlated blocks in the match correlation matrix of RepB indicate the presence of subdomains. **(B)** The predicted RepB structure (green) modelled against PDB files 2nraC (R6K π initiator) (orange) and 1REPC (RepE initiator from F plasmid) (light orange). **(C)** HPLC chromatogram of RepB_ΔC displaying one main peak representing a weight of 28.81 kDa suggesting RepB_ΔC exists as a monomer in solution at a concentration of 0.8mg/ml. **(D)** EMSA with RepB_ΔC and probe P17 containing a single iteron, no binding is observed with RepB_ΔC as compared to the positive control, WT HIS-tagged RepB (lane3).

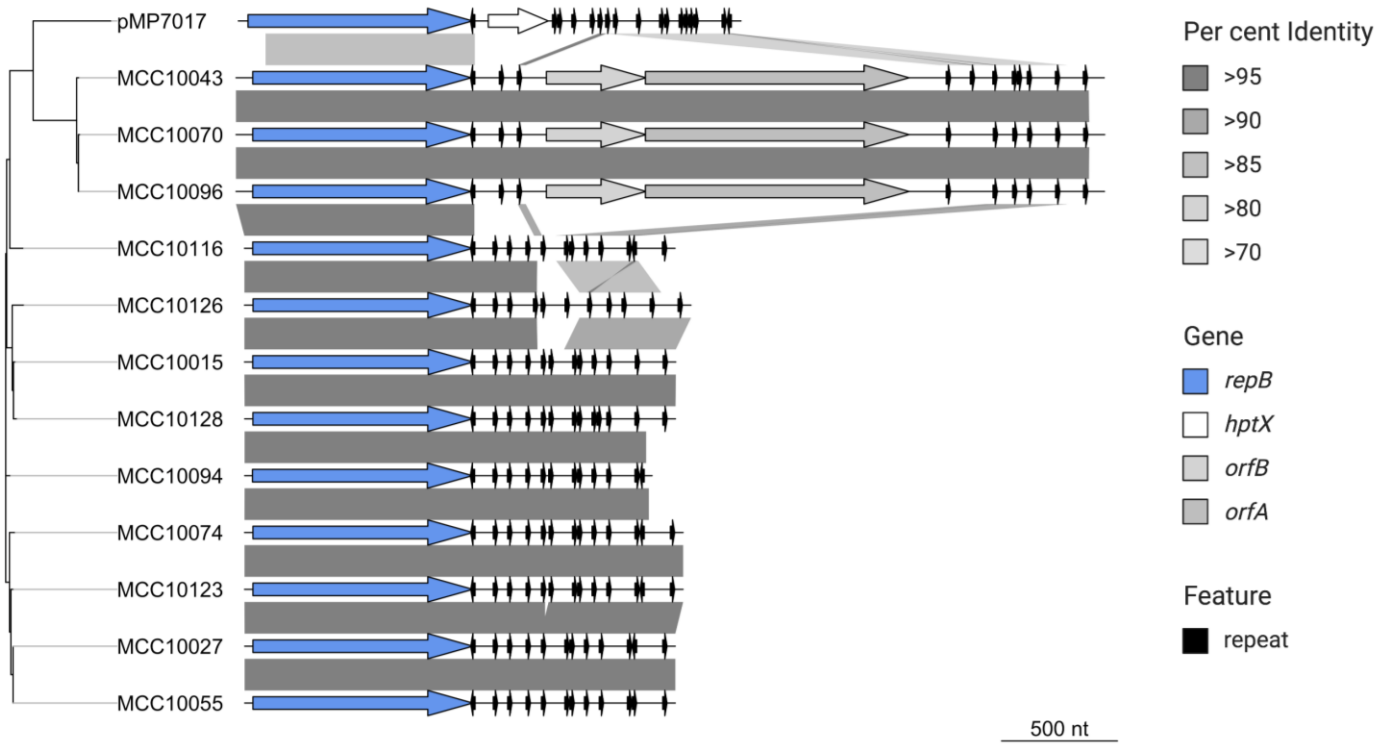


Figure S5. Comparative analysis of the pMP7017 basic replicon across bifidobacterial megaplasmids

Comparative analysis of the pMP7017 basic replicon. A neighbour-joining tree based on the nucleotide sequence of the replicon in each *Bifidobacterium* megaplasmid is located to the left. The genetic organization within each replicon is displayed and a description of each feature is given in the legend.

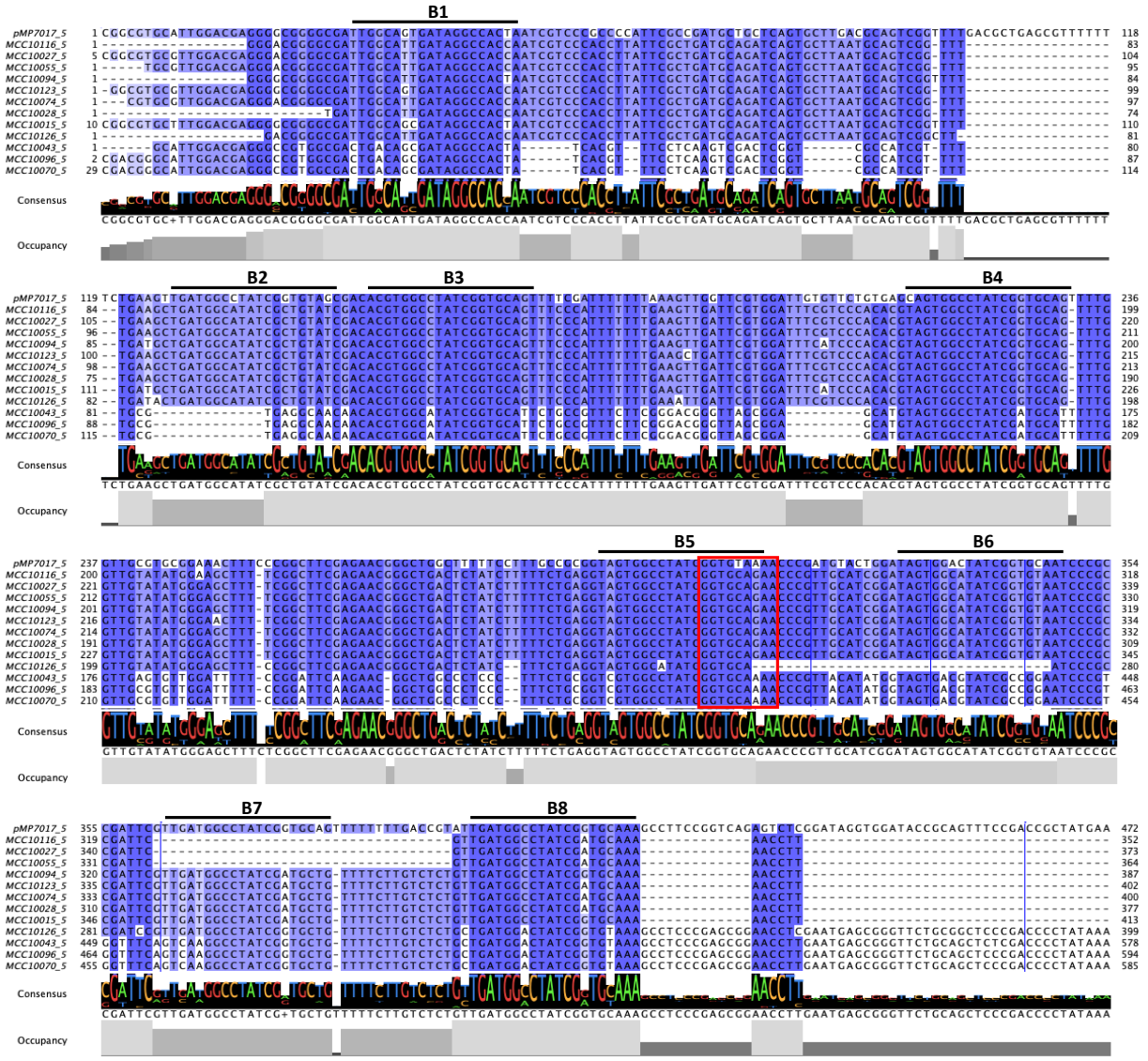


Figure S6 continued next page.

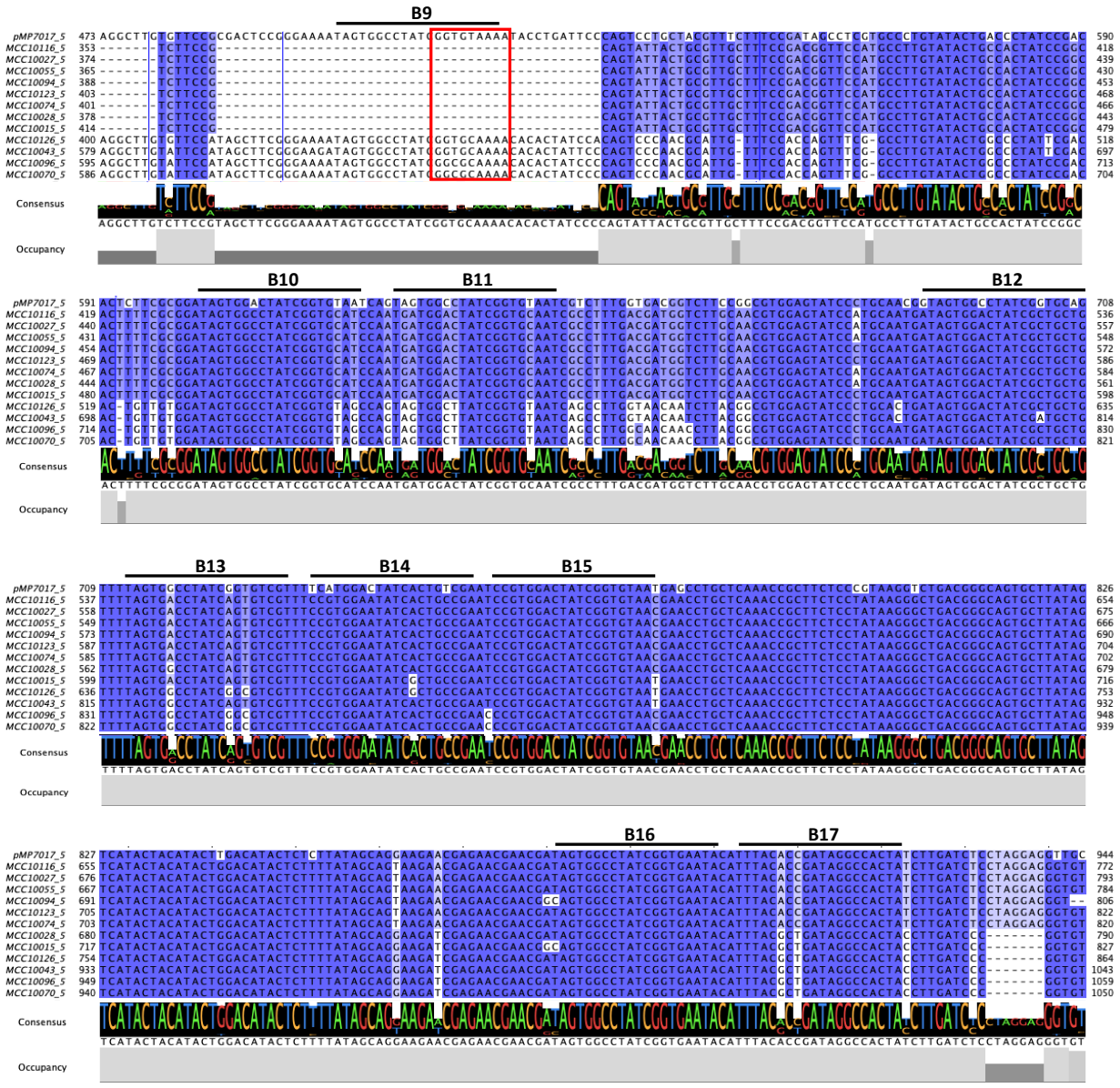


Figure S6. Alignment of the repeat contain region between 13 Bifidobacterial megaplasmids

Residues are highlighted by conservation, dark purple >80%, purple >70% and light purple >50%, with the consensus motif displayed below. Putative DnaA box elements are boxed in red.

Supplementary Table S1. Strains and plasmids used in this study

Strains or plasmid	Relevant genotype or phenotype	Source or reference
Strains		
Escherichia coli strains		
EC101	Cloning host, *repA*$^{+}$, Km^{r}	[82]
EC101 pORI19	EC101 harbouring pORI19	[82]
EC101 pORI19_*repA*	EC101 harbouring pORI19_*repA*	this study
EC101 pORI19_tet_*repA*	EC101 harbouring pORI19_tet_*repA*	this study
EC101 pORI19_*repB*	EC101 harbouring pORI19_*repB*	this study
EC101 pORI19_tet_*repB*	EC101 harbouring pORI19_tet_*repB*	this study
XL1 blue	RecA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tet^{r})]	Stratagene
XL1 blue pBif	XL1 blue harbouring pBif	APC- University College Cork
XL1 blue pBif_*repA*	EC101 harbouring pBif_*repA*	this study
XL1 blue pBif_*repB*	EC101 harbouring pBif_*repB*	this study
XL1 blue pRD1.2	EC101 harbouring pRD1.2	this study
XL1 blue pRD1.3	EC101 harbouring pRD1.3	this study
XL1 blue pRD1.4	EC101 harbouring pRD1.4	this study
XL1 blue pRD1.5	EC101 harbouring pRD1.5	this study
XL1 blue pRD1.6	EC101 harbouring pRD1.6	this study
XL1 blue pRD1.7	EC101 harbouring pRD1.7	this study
XL1 blue pRD1.8	EC101 harbouring pRD1.8	this study
XL1 blue pRD1.9	EC101 harbouring pRD1.9	this study
XL1 blue pRD1.10	EC101 harbouring pRD1.10	this study
XL1 blue pRD1.11	EC101 harbouring pRD1.11	this study
XL1 blue pRD1.12	EC101 harbouring pRD1.12	this study
EC101 pQE_*repA*	EC101 harbouring pQE_*repA*	this study
EC101 pQE_*repB*	EC101 harbouring pQE_*repB*	this study
EC101 pQE_*repB*△C	EC101 harbouring pQE_*repB*△C	this study

Strains or plasmid	Relevant genotype or phenotype	Source or reference
Lactococcus lactis strains		
NZ9000	MG1363 pepN::nisRK, host strain for nisin-inducible overexpression	[84]
NZ9000 pNZ272	NZ9000 harbouring pNZ272, Promoterless *gusA*, Cm^r	[65]
NZ9000 pNZ_*repA*	NZ9000 harbouring a pNZ272 derivative containing the *repA* promoter	this study
NZ9000 pNZ_*repB*	NZ9000 harbouring a pNZ272 derivative containing the *repB* promoter	this study
Bifidobacterium breve strains		
UCC2003	Human intestinal isolate, plasmid-free	APC - University College Cork
UCC2003 pMP7017_199	UCC2003 harbouring pMP7017_199, Tet^r	this study
JCM7017_199	pORI19-tet-199 insertion mutant of pMP7017 in JCM7017	[25]
JCM7017_*repA*	JCM7017 harbouring pMP7017-pORI_tet_*repA* insertion mutant	this study
UCC2003 pBif_*repB*	UCC2003 harbouring pBif_*repB*, Cm^r	this study
UCC2003 pRD1.2	UCC2003 harbouring pRD1.2, Cm^r	this study
UCC2003 pRD1.3	UCC2003 harbouring pRD1.3, Cm^r	this study
UCC2003 pRD1.4	UCC2003 harbouring pRD1.4, Cm^r	this study
UCC2003 pRD1.7	UCC2003 harbouring pRD1.7, Cm^r	this study
UCC2003 pRD1.8	UCC2003 harbouring pRD1.8, Cm^r	this study
UCC2003 pRD1.9	UCC2003 harbouring pRD1.9, Cm^r	this study
UCC2003 pRD1.10	UCC2003 harbouring pRD1.10, Cm^r	this study
Bifidobacterium longum subsp. *longum* strains		
NCIMB8809	Isolate from nursling stool	NCIMB
NCIMB8809 pRD1.3	NCIMB8809 harbouring pRD1.3	this study
NCIMB8809 pRD1.3_*apuB*	NCIMB8809 harbouring pRD1.3_*apuB*	this study
Plasmids		
pMP7017_199	pMP7017 ORF199 insertion mutant, Tet^r	[25]
pNZ272	Promoterless gusA, Cm^r	[66]

Strains or plasmid	Relevant genotype or phenotype	Source or reference
pNZ_*repA*	pNZ272 derivative containing the *repA* promoter, Cm^r	this study
pNZ_*repB*	pNZ272 derivative containing the *repB* promoter, Cm^r	this study
pQE30	Amp^r, IPTG-inducible vector	Qiagen
pQE30_*repA*	pQE30 derivative containing *repA,* Amp^r	this study
pQE30_*repB*	pQE30 derivative containing *repB,* Amp^r	this study
pQE30_*repB*△C	pQE30 derivative containing *repB* with C-terminal deletion (G235-Q332del)	this study
pAM5	pBC1-pUC19 Tet^r [*tet*(W)]	[39]
pORI19	Em^r , $repA^-$, ori^+ , cloning vector	[83]
pORI19_*repA*	pORI19 containing an internal 528 bp fragment of *repA*	this study
pORI19_tet_*repA*	pORI19_*repA* containing the Tet^r cassette from pAM5	this study
pORI19_*repB*	pORI19 containing an internal 501 bp fragment of *repB*	this study
pORI19_tet_*repB*	pORI19_*repB* containing the Tet^r cassette from pAM5	this study
pBif	Amp^r, Cm^r, 6.5 kbp	APC-University College Cork
pBif_*repA*	pBif containing 1,829 bp of the *repA* replication region	this study
pBif_*repB*	pBif containing 3,257 bp of the *repB* replication region	this study
pRD1.2	pBif containing 2,843 bp of the *repB* replication region	this study
pRD1.3	pBif containing 2,541 bp of the *repB* replication region	this study
pRD1.4	pBif containing 2,381 bp of the *repB* replication region	this study
pRD1.5	pBif containing 2,180 bp of the *repB* replication region	this study
pRD1.6	pBif containing 1,809 bp of the *repB* replication region	this study
pRD1.7	pBif containing 2,050 bp of the *repB* replication region	this study
pRD1.8	pBif containing 2,284 bp of the *repB* replication region	this study
pRD1.9	pBif containing 2,284 bp of the *repB* replication region	this study
pRD1.10	pBif containing 2,216 bp of the *repB* replication region	this study
pRD1.11	pBif containing 2,295 bp of the *repB* replication region	this study
pRD1.12	pBif containing 2,280 bp of the *repB* replication region	this study
pRD1.3_*apuB*	pRD1.3 containing the *apuB* gene and promoter from *B. breve* UCC2003	this study

Supplementary Table S2. List of primers used in this study

Purpose	Primer	Sequence (5' - 3')
Cloning of an internal 528 bp fragment of *repA* in pORI19	repAmut_R	aattaaAAGCTTgcatcggtgcgaggataata
	repAmut_F	catcacTCTAGAcagcttccaggaattctgcg
Cloning of an internal 501 bp fragment of *repB* in pORI19	repBmut_R	aattaaAAGCTTcgtccatgtagccaatccc
	repBmut_F	catcacTCTAGAgctgctgatggcatggatg
Confirmation of site-specific homologous recombination	*repA*conf	gttctgcttgaccggtcgaatc
	TetR_final	gttgcggacatacacat
Cloning 1,829 bp of the *repA* coding region in pBif	124MR_sph1_1	ccgtcaGCATGCctgctagttcttcttcgtgccatac
	124MR_sph1_2	cagtcaGCATGCcgagggttttcatgatggc
Cloning various lengths of the *repB* replication region in pBif to define the minimal replicon	MR_sph1_1	catcaaGCATGCctgttaatccgcctgtt
	MR_sph1_2	ctacaaGCATGCccagtccgattcaaagt
	MR_sph1_3	gttagaGCATGCatgcgggcgtgagttg
	MR_sph1_4	cttgttGCATGCggcaggaggtgttggtttc
	MR_sph1_5	gatcaaGCATGCctatgatgtcaaccatgac
	MR_sph1_6	tatcaaGCATGCgacggtcttccggcgtggag
	MR_sph1_R	ccggtaGCATGCtaggagatcaagatagt
Overlap extension PCR for the construction of pRD1.7, pRD1.8, pRD1.9, pRD1.10, pRD1.11 and pRD1.12	OEpRD1.7_R1	gcttcaatcgcgaacacgccagatcactgccaatcgccccgcccctc
	OEpRD1.7_F2	gaggggcggggcgattggcagtgatctggcgtgttcgcgattgaagc
	OEpRD1.8_R1	gtatacagggcacgaggctatcggcatagcggtcggaaactgcggtatc
	OEpRD1.8_F2	gataccgcagtttccgaccgctatgccgatagcctcgtgccctgtatac
	OEpRD1.9_R1	gttaatatctggcgtgttcgcgattgagttcgtggattgtgttctgtgagc
	OEpRD1.9_F2	gctcacagaacacaatccacgaactcaatcgcgaacacgccagatattaac
	OEpRD1.10_R1	ggaaaaagccagcccgttctcgaagcgcttcaatcgcgaacacgccaga
	OEpRD1.10_F2	tctggcgtgttcgcgattgaagcgcttcgagaacgggctggctttttcc
	OEpRD1.11_R1	ccgataggccactatcgttcgttcgggagaagcggtttgagcaggctc
	OEpRD1.11_F2	gagcctgctcaaaccgcttctcccgaacgaacgatagtggcctatcgg
	OEpRD1.12_R1	cgggagaagcggtttgagcaggctcccgttgcagggatactccacgccgg
	OEpRD1.12_F2	ccggcgtggagtatccctgcaacgggagcctgctcaaaccgcttctcccg
	MR4T.SphIF1	cttgttGCATGCGTGACATTAGAAAACCGACTGTAAAA AGTACAGTCGGttgacttcctgtttgggc

	MR8T.SphIR2	cttcgtGCATGCGTGACATTAGAAAACCGACTGTAAA AAGTACAGTCGGgagctgatggatgcaacctc
Confirmation of *repA* fragment integration in pBif	pMP_0124F	cgcgtcaatccagttcaccg
	pMP_0124R	tgcttgaacctgcgtcggaa
Confirmation of *repB* fragment integration in pBif	pMP_0146F	ggaacagttgcgccgacttc
	pMP_0146R	gccgctcagcagatagttgg
Confirmation of insert orientation in pBif	pRDconf_F	agtgccacctaaattgtaagcgtt
	pRDconf_R	gagcctcggaacccttaacaa
qPCR primers	146A_F	gctgatggcatggatggc
	146A_R	cggatagctggtacatgcttgg
	146B_F	cggaggccaatcatcgaaacg
	146B_R	caccatctgcatgttcacggaa
	dxsA_F	cacgaacccggacgaga
	dxsA_R	tatcgggcccctccatattg
	UCC776B_F	tgtcacctccgtcctgct
	UCC776B_R	ggcgttgatcttgttggtggac
Cloning of *repA* in pQE30 for protein overexpression	pQE30-124N_Sph1	ataattGCATGCaccaataataacgccatc
	pQE30-124N_HindIII	catcacAAGCTTacccaacaaaacccgaga
Cloning of *repB* in pQE30 for protein overexpression	pQE30-146N_Sph1	ataattGCATGCacttcctgtttgggca
	pQE30-146N_HindIII	catcacAAGCTTctgccaatcgccccgc
IRD700 3' labelled primers for EMSA	146B_F1	cgattggcagtgataggc
	146B_F2	gtatggttaatatctggcgt
	146B_F3	agagtctcggataggtggat
	146B_R1	aacgaatgagctgatggatg
	146B_R2	gaagagtgtcggatagggt
	146B_R3	tccgcacgcaaccaaaactg
	146proF	atgcgggcgtgagttgc
	146int_F	ttgacttcctgtttgggc
	146int_R	cttggcttcgtgatcgtc
Cloning of the *apuB* gene in pRD1.3	*apuB*F_Xba	catcacTCTAGActcgtgcaattcgtggtatc
	*apuB*R_H3	aattaaAAGCTTctactgttcggagccctc
Mutagenesis of the RepB recognition sequence;		

IRD labelled probe, P1	146RSM_F1	gtgagcagcggcctatcggtgcagttttggttg
IRD labelled probe, P2	146RSM_F2	gtgagcagttgcctatcggtgcagttttggttg
IRD labelled probe, P3	146RSM_F3	gtgagcagtgtcctatcggtgcagttttggttg
IRD labelled probe, P4	146RSM_F4	gtgagcagtggtctatcggtgcagttttggttg
IRD labelled probe, P5	146RSM_F5	gtgagcagtggcatatcggtgcagttttggttg
IRD labelled probe, P6	146RSM_F6	gtgagcagtggccgatcggtgcagttttggttg
IRD labelled probe, P7	146RSM_F7	gtgagcagtggcctctcggtgcagttttggttg
IRD labelled probe, P8	146RSM_F8	gtgagcagtggcctagcggtgcagttttggttg
IRD labelled probe, P9	146RSM_F9	gtgagcagtggcctataggtgcagttttggttg
IRD labelled probe, P10	146RSM_F10	gtgagcagtggcctatctgtgcagttttggttg
IRD labelled probe, P11	146RSM_F11	gtgagcagtggcctatcgatgcagttttggttg
IRD labelled probe, P12	146RSM_F12	gtgagcagtggcctatcggcgcagttttggttg
IRD labelled probe, P13	146RSM_F13	gtgagcagtggcctatcggtacagttttggttg
IRD labelled probe, P14	146RSM_F14	gtgagcagcggtctatcggtgcagttttggttg
IRD labelled probe, P15	146RSM_F15	gtgagcagtggcctatagatgcagttttggttg
IRD labelled probe, P16	146RSM_F16	ctgtgagcagtcactgccaatcgcagttttggttg
Non-mutated sequence, IRD labelled probe, P17	146RSM_F17	gtgagcagtggcctatcggtgcagttttggttg
	146RSM_R	ccacctatccgagactctg
Deletion of the C-terminal of RepB	pQE146C-_SphIF	ggaagtGCATGCtcaaccatgacaaaaaacagc
	pQE146C-_H3R	catcacAAGCTTctagctcagcagatagttggtg

Supplementary Table S3. Probe information for EMSAs

Probe	Forward primer	Reverse primer	Region Coordinates	Size (bp)	Region features/ Bind Sites	Probe conc. nM	Protein
124P	124PRO_F	124PRO_R	97,193...97,536	344	*repA* promoter repeats A1:A8	10	RepA
A	MR2	146PE_R1	120,589...121,076	488	Non-coding region 3'of the *repB* promoter	20	RepB
B	MR3	146int_R	120,446...120,774	433	*repB* promoter	16	RepB
C	146int_F	146A_R	119,981...120,614	634	*repB* promoter and coding sequence	10	RepB
D	MR_ptX_F	MR_ptX_R	119,075...119,343	269	hptX	8	RepB
E	146B_F1	BR_R1	119,345...119,409	65	B1	20	RepB
F	146B_F2	MR6	118,966...119,092	127	B2, B3	20	RepB
G	146B_F2	146_R3	118,921...119,092	172	B2, B3, B4	20	RepB
H	146B_F3	MR_7	118,595...118,740	146	B9	20	RepB
I	BR_R1P_F	146B_R1	118,206...118,370	165	B17, B18	20	RepB
P1	146RSM_F1	146B_R3	118,921...118,961	41	B3 (1T>C)	20	RepB
P2	146RSM_F2	146B_R3	118,921...118,961	41	B3 (2G>T)	20	RepB
P3	146RSM_F3	146B_R3	118,921...118,961	41	B3 (3G>T)	20	RepB
P4	146RSM_F4	146B_R3	118,921...118,961	41	B3 (4C>T)	20	RepB
P5	146RSM_F5	146B_R3	118,921...118,961	41	B3 (5C>A)	20	RepB
P6	146RSM_F6	146B_R3	118,921...118,961	41	B3 (6T>G)	20	RepB
P7	146RSM_F7	146B_R3	118,921...118,961	41	B3 (7A>C)	20	RepB
P8	146RSM_F8	146B_R3	118,921...118,961	41	B3 (8T>G)	20	RepB
P9	146RSM_F9	146B_R3	118,921...118,961	41	B3 (9C>A)	20	RepB
P10	146RSM_F10	146B_R3	118,921...118,961	41	B3 (10G>T)	20	RepB
P11	146RSM_F11	146B_R3	118,921...118,961	41	B3 (11G>A)	20	RepB
P12	146RSM_F12	146B_R3	118,921...118,961	41	B3 (12T>C)	20	RepB
P13	146RSM_F13	146B_R3	118,921...118,961	41	B3 (13G>A)	20	RepB
P14	146RSM_F14	146B_R3	118,921...118,961	41	B3 (1T>C, 4C>T)	20	RepB
P15	146RSM_F15	146B_R3	118,921...118,961	41	B3 (9C>A, 11G>A)	20	RepB

P16	146RSM_F16	146B_R3	118,921...118,961	41	B3 (2_12delins-TCACTGCCAATCG)	20	RepB
P17	146RSM_F17	146B_R3	118,921...118,961	41	B3	20	RepB
P17	146RSM_F17	146B_R3	118,921...118,961	41	B3	20	RepBΔC

REFERENCES

[1] **S. C. Leahy, D. G. Higgins, G. F. Fitzgerald, and D. van Sinderen**, 'Getting better with bifidobacteria', *J. Appl. Microbiol.*, vol. 98, no. 6, pp. 1303–1315, 2005, doi: 10.1111/j.1365-2672.2005.02600.x.

[2] **A. O'Callaghan and D. van Sinderen**, 'Bifidobacteria and Their Role as Members of the Human Gut Microbiota', *Front. Microbiol.*, vol. 7, Jun. 2016, doi: 10.3389/fmicb.2016.00925.

[3] **S. Fukuda *et al.***, 'Bifidobacteria can protect from enteropathogenic infection through production of acetate', *Nature*, vol. 469, no. 7331, Art. no. 7331, Jan. 2011, doi: 10.1038/nature09646.

[4] **V. Lievin**, 'Bifidobacterium strains from resident infant human gastrointestinal microflora exert antimicrobial activity', *Gut*, vol. 47, no. 5, Art. no. 5, Nov. 2000, doi: 10.1136/gut.47.5.646.

[5] **S. Fanning *et al.***, 'Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 6, Art. no. 6, Feb. 2012, doi: 10.1073/pnas.1115621109.

[6] **M. O'Connell Motherway *et al.***, 'A Bifidobacterial pilus-associated protein promotes colonic epithelial proliferation', *Mol. Microbiol.*, vol. 111, no. 1, pp. 287–301, 2019, doi: 10.1111/mmi.14155.

[7] **E. Schiavi *et al.***, 'The Surface-Associated Exopolysaccharide of Bifidobacterium longum 35624 Plays an Essential Role in Dampening Host Proinflammatory Responses and Repressing Local TH17 Responses', *Appl. Environ. Microbiol.*, vol. 82, no. 24, pp. 7185–7196, Dec. 2016, doi: 10.1128/AEM.02238-16.

[8] **F. Turroni *et al.***, 'Role of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in modulating bacterium-host interactions', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 27, Art. no. 27, Jul. 2013, doi: 10.1073/pnas.1303897110.

[9] **H. M. Roager and T. R. Licht**, 'Microbial tryptophan catabolites in health and disease', *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Aug. 2018, doi: 10.1038/s41467-018-05470-4.

[10] **C. B. Wong, A. Tanaka, T. Kuhara, and J.-Z. Xiao**, 'Potential Effects of Indole-3-Lactic Acid, a Metabolite of Human Bifidobacteria, on NGF-induced Neurite Outgrowth in PC12 Cells', *Microorganisms*, vol. 8, no. 3, Mar. 2020, doi: 10.3390/microorganisms8030398.

[11] **M. Cronin, M. Knobel, M. O'Connell-Motherway, G. F. Fitzgerald, and D. van Sinderen**, 'Molecular Dissection of a Bifidobacterial Replicon', *Appl. Environ. Microbiol.*, vol. 73, no. 24, Art. no. 24, Dec. 2007, doi: 10.1128/AEM.01630-07.

[12] **H. Koguchi *et al.***, 'Application of Recombinase-Based In Vivo Expression Technology to Bifidobacterium longum subsp. longum for Identification of Genes Induced in the Gastrointestinal Tract of Mice', *Microorganisms*, vol. 8, no. 3, Mar. 2020, doi: 10.3390/microorganisms8030410.

[13] **M. O'Connell Motherway, J. O'Driscoll, G. F. Fitzgerald, and D. Van Sinderen**, 'Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in Bifidobacterium breve UCC2003', *Microb. Biotechnol.*, vol. 2, no. 3, Art. no. 3, May 2009, doi: 10.1111/j.1751-7915.2008.00071.x.

[14] **L. Ruiz, M. O. Motherway, N. Lanigan, and D. van Sinderen**, 'Transposon Mutagenesis in Bifidobacterium breve: Construction and Characterization of a Tn5 Transposon Mutant Library for Bifidobacterium breve UCC2003', *PLOS ONE*, vol. 8, no. 5, Art. no. 5, May 2013, doi: 10.1371/journal.pone.0064699.

[15] **M. Sakanaka *et al.***, 'A Transposon Mutagenesis System for Bifidobacterium longum subsp. longum Based on an IS3 Family Insertion Sequence, ISBlo11', *Appl. Environ. Microbiol.*, vol. 84, no. 17, 01 2018, doi: 10.1128/AEM.00824-18.

[16] **F. Bottacini *et al.***, 'Comparative genome and methylome analysis reveals restriction/modification system diversity in the gut commensal Bifidobacterium breve', *Nucleic Acids Res.*, vol. 46, no. 4, pp. 1860–1877, Feb. 2018, doi: 10.1093/nar/gkx1289.

[17] **M. Cronin, M. Ventura, G. F. Fitzgerald, and D. van Sinderen**, 'Progress in genomics, metabolism and biotechnology of bifidobacteria', *Int. J. Food Microbiol.*, vol. 149, no. 1, Art. no. 1, Sep. 2011, doi: 10.1016/j.ijfoodmicro.2011.01.019.

[18] **M. Ventura, D. van Sinderen, G. F. Fitzgerald, and R. Zink**, 'Insights into the taxonomy, genetics and physiology of bifidobacteria', *Antonie Van Leeuwenhoek*, vol. 86, no. 3, Art. no. 3, Oct. 2004, doi: 10.1023/B:ANTO.0000047930.11029.ec.

[19] **G. del Solar, R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas**, 'Replication and Control of Circular Bacterial Plasmids', *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, pp. 434–464, Jun. 1998.

[20] **L. Dziewit *et al.***, 'Diversity and role of plasmids in adaptation of bacteria inhabiting the Lubin copper mine in Poland, an environment rich in heavy metals', *Front. Microbiol.*, vol. 6, Mar. 2015, doi: 10.3389/fmicb.2015.00152.

[21] **H. Heuer and K. Smalla**, 'Plasmids foster diversification and adaptation of bacterial populations in soil', *FEMS Microbiol. Rev.*, vol. 36, no. 6, pp. 1083–1104, Nov. 2012, doi: 10.1111/j.1574-6976.2012.00337.x.

[22] **P. A. Sobecky**, 'Plasmid ecology of marine sediment microbial communities', *Hydrobiologia*, vol. 401, no. 0, pp. 9–18, May 1999, doi: 10.1023/A:1003726024628.

[23] **J.-H. Lee and D. J. O'Sullivan**, 'Genomic Insights into Bifidobacteria', *Microbiol. Mol. Biol. Rev.*, vol. 74, no. 3, pp. 378–416, Sep. 2010, doi: 10.1128/MMBR.00004-10.

[24] **M. Rossi, P. Brigidi, A. Gonzalez Vara y Rodriguez, and D. Matteuzzi**, 'Characterization of the plasmid pMB1 from Bifidobacterium longum and its use for shuttle vector construction', *Res. Microbiol.*, vol. 147, no. 3, pp. 133–143, Jan. 1996, doi: 10.1016/0923-2508(96)80213-0.

[25] **F. Bottacini *et al.***, 'Discovery of a Conjugative Megaplasmid in Bifidobacterium breve', *Appl. Environ. Microbiol.*, vol. 81, no. 1, pp. 166–176, Jan. 2015, doi: 10.1128/AEM.02871-14.

[26] **T. Odamaki *et al.***, 'Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan', *Sci. Rep.*, vol. 8, Jan. 2018, doi: 10.1038/s41598-017-18391-x.

[27] **C. Rosenberg, P. Boistard, J. Dénarié, and F. Casse-Delbart**, 'Genes controlling early and late functions in symbiosis are located on a megaplasmid in Rhizobium meliloti', *Mol. Gen. Genet. MGG*, vol. 184, no. 2, pp. 326–333, Dec. 1981, doi: 10.1007/BF00272926.

[28] **E. Schwartz, Ed.**, *Microbial Megaplasmids*. Berlin Heidelberg: Springer-Verlag, 2009. doi: 10.1007/978-3-540-85467-8.

[29] **A. Lagares, J. Sanjuán, and M. Pistorio**, 'The Plasmid Mobilome of the Model Plant-Symbiont Sinorhizobium meliloti: Coming up with New Questions and Answers',

Microbiol. Spectr., vol. 2, no. 5, Oct. 2014, doi: 10.1128/microbiolspec.PLAS-0005-2013.

[30] **M. G. López-Guerrero *et al.***, 'Rhizobial extrachromosomal replicon variability, stability and expression in natural niches', *Plasmid*, vol. 68, no. 3, pp. 149–158, Nov. 2012, doi: 10.1016/j.plasmid.2012.07.002.

[31] **J. H. Baek and D. K. Chattoraj**, 'Chromosome I Controls Chromosome II Replication in Vibrio cholerae', *PLOS Genet.*, vol. 10, no. 2, p. e1004184, Feb. 2014, doi: 10.1371/journal.pgen.1004184.

[32] **F. de Lemos Martins, F. Fournes, M.-V. Mazzuoli, D. Mazel, and M.-E. Val**, 'Vibrio cholerae chromosome 2 copy number is controlled by the methylation-independent binding of its monomeric initiator to the chromosome 1 crtS site', *Nucleic Acids Res.*, vol. 46, no. 19, pp. 10145–10156, Nov. 2018, doi: 10.1093/nar/gky790.

[33] **F. Fournes, M.-E. Val, O. Skovgaard, and D. Mazel**, 'Replicate Once Per Cell Cycle: Replication Control of Secondary Chromosomes', *Front. Microbiol.*, vol. 9, 2018, doi: 10.3389/fmicb.2018.01833.

[34] **C. Stokke, T. Waldminghaus, and K. Skarstad**, 'Replication patterns and organization of replication forks in Vibrio cholerae', *Microbiol. Read. Engl.*, vol. 157, no. Pt 3, pp. 695–708, Mar. 2011, doi: 10.1099/mic.0.045112-0.

[35] **M. Egan *et al.***, 'Staying alive: growth and survival of Bifidobacterium animalis subsp. animalis under in vitro and in vivo conditions', *Appl. Microbiol. Biotechnol.*, vol. 102, no. 24, pp. 10645–10663, Dec. 2018, doi: 10.1007/s00253-018-9413-7.

[36] **J. Sambrook, E. F. Fritsch, and T. Maniatis**, 'Molecular cloning: a laboratory manual.', *Mol. Cloning Lab. Man.*, no. Ed. 2, Art. no. Ed. 2, 1989, Accessed: Mar. 19, 2019. [Online]. Available: https://www.cabdirect.org/cabdirect/abstract/19901616061

[37] **A. Mazé, M. O'Connell-Motherway, G. F. Fitzgerald, J. Deutscher, and D. van Sinderen**, 'Identification and characterization of a fructose phosphotransferase system in Bifidobacterium breve UCC2003', *Appl. Environ. Microbiol.*, vol. 73, no. 2, pp. 545–553, Jan. 2007, doi: 10.1128/AEM.01496-06.

[38] **M. Rossi, P. Brigidi, and D. Matteuzzi**, 'An efficient transformation system for Bifidobacterium spp.', *Lett. Appl. Microbiol.*, vol. 24, no. 1, pp. 33–36, 1997, doi: https://doi.org/10.1046/j.1472-765X.1997.00335.x.

[39] **P. Alvarez-Martín, M. O'Connell-Motherway, D. van Sinderen, and B. Mayo**, 'Functional analysis of the pBC1 replicon from Bifidobacterium catenulatum L48', *Appl. Microbiol. Biotechnol.*, vol. 76, no. 6, pp. 1395–1402, Oct. 2007, doi: 10.1007/s00253-007-1115-5.

[40] **A. Sangrador-Vegas, C. Stanton, D. van Sinderen, G. F. Fitzgerald, and R. P. Ross**, 'Characterization of plasmid pASV479 from Bifidobacterium pseudolongum subsp. globosum and its use for expression vector construction', *Plasmid*, vol. 58, no. 2, pp. 140–147, Sep. 2007, doi: 10.1016/j.plasmid.2007.02.004.

[41] **J. C. D. Man, M. Rogosa, and M. E. Sharpe**, 'A Medium for the Cultivation of Lactobacilli', *J. Appl. Bacteriol.*, vol. 23, no. 1, pp. 130–135, 1960, doi: 10.1111/j.1365-2672.1960.tb00188.x.

[42] **C. L. Lee, D. S. W. Ow, and S. K. W. Oh**, 'Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria', *J. Microbiol. Methods*, vol. 65, no. 2, Art. no. 2, May 2006, doi: 10.1016/j.mimet.2005.07.019.

[43] **L. W. Hamoen, A. F. V. Werkhoven, J. J. E. Bijlsma, D. Dubnau, and G. Venema**, 'The competence transcription factor of Bacillus subtilis recognizes short A/T-rich

sequences arranged in a unique, flexible pattern along the DNA helix', *Genes Dev.*, vol. 12, no. 10, pp. 1539–1550, May 1998, doi: 10.1101/gad.12.10.1539.

[44] **J. Schindelin *et al.***, 'Fiji: an open-source platform for biological-image analysis', *Nat. Methods*, vol. 9, no. 7, pp. 676–682, Jul. 2012, doi: 10.1038/nmeth.2019.

[45] **C. A. Schneider, W. S. Rasband, and K. W. Eliceiri**, 'NIH Image to ImageJ: 25 years of image analysis', *Nat. Methods*, vol. 9, pp. 671–675, Jun. 2012, doi: 10.1038/nmeth.2089.

[46] **O. P. Kuipers, M. M. Beerthuyzen, R. J. Siezen, and W. M. De Vos**, 'Characterization of the nisin gene cluster nisABTCIPR of Lactococcus lactis. Requirement of expression of the nisA and nisI genes for development of immunity', *Eur. J. Biochem.*, vol. 216, no. 1, pp. 281–291, Aug. 1993, doi: 10.1111/j.1432-1033.1993.tb18143.x.

[47] **M. Ventura, R. Zink, G. F. Fitzgerald, and D. van Sinderen**, 'Gene Structure and Transcriptional Organization of the dnaK Operon of Bifidobacterium breve UCC 2003 and Application of the Operon in Bifidobacterial Tracing', *Appl. Environ. Microbiol.*, vol. 71, no. 1, pp. 487–500, Jan. 2005, doi: 10.1128/AEM.71.1.487-500.2005.

[48] **K. Rutherford *et al.***, 'Artemis: sequence visualization and annotation', *Bioinforma. Oxf. Engl.*, vol. 16, no. 10, Art. no. 10, Oct. 2000.

[49] **S. El-Gebali *et al.***, 'The Pfam protein families database in 2019', *Nucleic Acids Res.*, vol. 47, no. D1, pp. D427–D432, Jan. 2019, doi: 10.1093/nar/gky995.

[50] **L. Holm**, 'DALI and the persistence of protein shape', *Protein Sci.*, vol. 29, no. 1, pp. 128–140, 2020, doi: https://doi.org/10.1002/pro.3749.

[51] **L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg**, 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nat. Protoc.*, vol. 10, no. 6, Art. no. 6, Jun. 2015, doi: 10.1038/nprot.2015.053.

[52] **F. Sievers *et al.***, 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Mol. Syst. Biol.*, vol. 7, p. 539, Oct. 2011, doi: 10.1038/msb.2011.75.

[53] **G. Stecher, K. Tamura, and S. Kumar**, 'Molecular Evolutionary Genetics Analysis (MEGA) for macOS', *Mol. Biol. Evol.*, vol. 37, no. 4, pp. 1237–1239, Apr. 2020, doi: 10.1093/molbev/msz312.

[54] **C. Camacho *et al.***, 'BLAST+: architecture and applications', *BMC Bioinformatics*, vol. 10, p. 421, Dec. 2009, doi: 10.1186/1471-2105-10-421.

[55] 'R: The R Project for Statistical Computing'. https://www.r-project.org/ (accessed Nov. 24, 2020).

[56] **L. Guy, J. Roat Kultima, and S. G. E. Andersson**, 'genoPlotR: comparative gene and genome visualization in R', *Bioinformatics*, vol. 26, no. 18, pp. 2334–2335, Sep. 2010, doi: 10.1093/bioinformatics/btq413.

[57] **K. Arakawa and M. Tomita**, 'Measures of compositional strand bias related to replication machinery and its applications', *Curr. Genomics*, vol. 13, no. 1, pp. 4–15, Mar. 2012, doi: 10.2174/138920212799034749.

[58] **G. Zhang and F. Gao**, 'Quantitative analysis of correlation between AT and GC biases among bacterial genomes', *PLoS ONE*, vol. 12, no. 2, Feb. 2017, doi: 10.1371/journal.pone.0171408.

[59] **H.-L. Zhao, Z.-K. Xia, F.-Z. Zhang, Y.-N. Ye, and F.-B. Guo**, 'Multiple Factors Drive Replicating Strand Composition Bias in Bacterial Genomes', *Int. J. Mol. Sci.*, vol. 16, no. 9, pp. 23111–23126, Sep. 2015, doi: 10.3390/ijms160923111.

[60] **M. Filutowicz *et al.***, 'Role of the π Initiation Protein and Direct Nucleotide Sequence Repeats in the Regulation of Plasmid R6K Replication', in *Plasmids in Bacteria*, D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson, and A. Hollaender, Eds. Boston, MA: Springer US, 1985, pp. 125–140. doi: 10.1007/978-1-4613-2447-8_13.

[61] **Y. Kazuo and Y. Mitsuyo**, 'The replication origin of pSC101: the nucleotide sequence and replication functions of the ori region', *Gene*, vol. 29, no. 1, pp. 211–219, Jul. 1984, doi: 10.1016/0378-1119(84)90181-1.

[62] **T. Murotsu, K. Matsubara, H. Sugisaki, and M. Takanami**, 'Nine unique repeating sequences in a region essential for replication and incompatibility of the mini-F plasmid', *Gene*, vol. 15, no. 2–3, pp. 257–271, Nov. 1981, doi: 10.1016/0378-1119(81)90135-9.

[63] **S. Apisiridej, A. Leelaporn, C. D. Scaramuzzi, R. A. Skurray, and N. Firth**, 'Molecular analysis of a mobilizable theta-mode trimethoprim resistance plasmid from coagulase-negative staphylococci', *Plasmid*, vol. 38, no. 1, pp. 13–24, 1997, doi: 10.1006/plas.1997.1292.

[64] **A. Nakamura, C. Wada, and K. Miki**, 'Structural basis for regulation of bifunctional roles in replication initiator protein', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 47, pp. 18484–18489, Nov. 2007, doi: 10.1073/pnas.0705623104.

[65] **M. K. Swan, D. Bastia, and C. Davies**, 'Crystal structure of π initiator protein–iteron complex of plasmid R6K: Implications for initiation of plasmid DNA replication', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 49, pp. 18481–18486, Dec. 2006, doi: 10.1073/pnas.0609046103.

[66] **C. Platteeuw, G. Simons, and W. M. de Vos**, 'Use of the Escherichia coli beta-glucuronidase (gusA) gene as a reporter gene for analyzing promoters in lactic acid bacteria.', *Appl. Environ. Microbiol.*, vol. 60, no. 2, Art. no. 2, Feb. 1994.

[67] **F. Bottacini *et al.***, 'Global transcriptional landscape and promoter mapping of the gut commensal Bifidobacterium breve UCC2003', *BMC Genomics*, vol. 18, Dec. 2017, doi: 10.1186/s12864-017-4387-x.

[68] **M. K. Swan, D. Bastia, and C. Davies**, 'Crystal structure of pi initiator protein-iteron complex of plasmid R6K: implications for initiation of plasmid DNA replication.', *Proc.Natl.Acad.Sci.Usa*, vol. 103, pp. 18481–18486, 2006, doi: 10.2210/pdb2nra/pdb.

[69] **N. Matsushima *et al.***, 'Flexible Structures and Ligand Interactions of Tandem Repeats Consisting of Proline, Glycine, Asparagine, Serine, and/or Threonine Rich Oligopeptides in Proteins', *Curr. Protein Pept. Sci.*, vol. 9, no. 6, pp. 591–610, Dec. 2008, doi: 10.2174/138920308786733886.

[70] **L. N. Johnson**, 'The regulation of protein phosphorylation', *Biochem. Soc. Trans.*, vol. 37, no. Pt 4, pp. 627–641, Aug. 2009, doi: 10.1042/BST0370627.

[71] **M. B. Yaffe and S. J. Smerdon**, 'PhosphoSerine/Threonine Binding Domains: You Can't pSERious?', *Structure*, vol. 9, no. 3, pp. R33–R38, Mar. 2001, doi: 10.1016/S0969-2126(01)00580-9.

[72] **M. O. Motherway, G. F. Fitzgerald, S. Neirynck, S. Ryan, L. Steidler, and D. van Sinderen**, 'Characterization of ApuB, an Extracellular Type II Amylopullulanase from Bifidobacterium breve UCC2003', *Appl. Environ. Microbiol.*, vol. 74, no. 20, pp. 6271–6279, Oct. 2008, doi: 10.1128/AEM.01169-08.

[73] **S. Arboleya *et al.***, 'Gene-trait matching across the Bifidobacterium longum pan-genome reveals considerable diversity in carbohydrate catabolism among human infant strains', *BMC Genomics*, vol. 19, Jan. 2018, doi: 10.1186/s12864-017-4388-9.

[74] **S. M. Ryan, G. F. Fitzgerald, and D. van Sinderen**, 'Screening for and Identification of Starch-, Amylopectin-, and Pullulan-Degrading Activities in Bifidobacterial Strains', *Appl. Environ. Microbiol.*, vol. 72, no. 8, pp. 5289–5296, Aug. 2006, doi: 10.1128/AEM.00257-06.

[75] **S. Kunnimalaiyaan, R. Krüger, W. Ross, S. A. Rakowski, and M. Filutowicz**, 'Binding modes of the initiator and inhibitor forms of the replication protein pi to the gamma ori iteron of plasmid R6K', *J. Biol. Chem.*, vol. 279, no. 39, pp. 41058–41066, Sep. 2004, doi: 10.1074/jbc.M403151200.

[76] **D. K. Chattoraj, R. Ghirlando, K. Park, J. A. Dibbens, and M. S. Lewis**, 'Dissociation kinetics of RepA dimers: implications for mechanisms of activation of DNA binding by chaperones', *Genes Cells*, vol. 1, no. 2, pp. 189–199, 1996, doi: https://doi.org/10.1046/j.1365-2443.1996.d01-235.x.

[77] **S. DasGupta, G. Mukhopadhyay, P. P. Papp, M. S. Lewis, and D. K. Chattoraj**, 'Activation of DNA binding by the monomeric form of the P1 replication initiator RepA by heat shock proteins DnaJ and DnaK', *J. Mol. Biol.*, vol. 232, no. 1, pp. 23–34, Jul. 1993, doi: 10.1006/jmbi.1993.1367.

[78] **M. Ishiai, C. Wada, Y. Kawasaki, and T. Yura**, 'Replication initiator protein RepE of mini-F plasmid: functional differentiation between monomers (initiator) and dimers (autogenous repressor)', *Proc. Natl. Acad. Sci.*, vol. 91, no. 9, pp. 3839–3843, Apr. 1994, doi: 10.1073/pnas.91.9.3839.

[79] **D. K. Chattoraj**, 'Control of plasmid DNA replication by iterons: no longer paradoxical', *Mol. Microbiol.*, vol. 37, no. 3, pp. 467–476, 2000, doi: 10.1046/j.1365-2958.2000.01986.x.

[80] **S. Zzaman, J. M. Reddy, and D. Bastia**, 'The DnaK-DnaJ-GrpE chaperone system activates inert wild type pi initiator protein of R6K into a form active in replication initiation', *J. Biol. Chem.*, vol. 279, no. 49, pp. 50886–50894, Dec. 2004, doi: 10.1074/jbc.M407531200.

[81] **H. Komori, F. Matsunaga, Y. Higuchi, M. Ishiai, C. Wada, and K. Miki**, 'Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 Å resolution', *EMBO J.*, vol. 18, no. 17, Art. no. 17, Sep. 1999, doi: 10.1093/emboj/18.17.4597.

[82] **B. Ekundayo and F. Bleichert**, 'Origins of DNA replication', *PLOS Genet.*, vol. 15, no. 9, p. e1008320, Sep. 2019, doi: 10.1371/journal.pgen.1008320.

[83] **J. Law, G. Buist, A. Haandrikman, J. Kok, G. Venema, and K. Leenhouts**, 'A system to generate chromosomal mutations in Lactococcus lactis which allows fast analysis of targeted genes', *J. Bacteriol.*, vol. 177, no. 24, pp. 7011–7018, Dec. 1995, doi: 10.1128/jb.177.24.7011-7018.1995.

[84] **P. G. de Ruyter, O. P. Kuipers, and W. M. de Vos**, 'Controlled gene expression systems for Lactococcus lactis with the food-grade inducer nisin', *Appl. Environ. Microbiol.*, vol. 62, no. 10, pp. 3662–3667, Oct. 1996, doi: 10.1128/AEM.62.10.3662-3667.1996.

CHAPTER III

Molecular Aspects of the Conjugative Transfer System of Megaplasmid pMP7017 from *Bifidobacterium breve* JCM7017

ABSTRACT

pMP7017 is a 190 kb megaplasmid harboured by the human isolate *Bifidobacterium breve* JCM7017 and represents the first and thus far only bifidobacterial plasmid identified with the ability to conjugate. To extend our understanding of the pMP7017 conjugal process, its conjugation efficiency was improved 359.5-fold through the cloning and expression of a recipient-derived methylase in the donor, apparently protecting pMP7017 from degradation by the corresponding restriction endonuclease upon entering the recipient cell. Furthermore, predictive functional analysis in combination with site-directed mutagenesis were employed to identify genes required for conjugal transfer, referred to here as *tra* genes. This analysis revealed various *tra* genes, which were shown to be organized into two unconnected gene clusters, designated *tra1* and *tra2*. The *tra* genes are predicted to encode proteins presumably constituting a T4SS as described for characterised conjugation systems of various Gram-positive bacteria. Comparative genomics of related plasmids revealed that these *tra* genes are present in all currently available pMP7017 homologues.

INTRODUCTION

Members of the *Bifidobacterium* genus belong to the phylum *Actinomycetota* and many bifidobacterial species can be found as ubiquitous inhabitants of the gastrointestinal tract (GIT) of mammals, including humans [1]. Typically, once an anaerobic environment is established within the gut of a healthy human neonate, particular bifidobacterial species quickly become highly abundant in this intestinal niche. While the composition of the gastrointestinal microbiota fluctuates greatly with age, *Bifidobacterium* populations generally persist throughout the human lifespan [2] and their relative abundance is estimated to range between 3% and 6% of the adult faecal microbiota [3], [4]. Bifidobacteria are commonly considered as gut symbionts with various purported health-promoting attributes [5]–[9], and for this reason these commensals have been commercially exploited as probiotic ingredients in a wide variety of functional foods [10]. Studies highlighting the efficacy of such probiotics in randomised human trials, such as the alleviation of clinical symptoms associated with Inflammatory Bowel Disease and related pathologies [11]–[13], have fuelled research efforts to elucidate the mechanisms by which these bacteria exert their purported beneficial properties and adaptive abilities [14], [15].

Restriction modification (RM) systems are abundant in bifidobacteria and constitute a substantial portion of the intra- and inter-species variation [16]. These prokaryotic defence mechanisms can be classified into four different types based on the composition of subunits, cleavage position, recognition sequence, substrate specificity and cofactor requirements [17]. Type II RM systems are comprised of a restriction endonuclease (REase) which typically exert their nuclease activity at defined recognition sites within an unmethylated dsDNA substrate, in combination with a cognate methyltransferase (MTase) that precludes this activity by protective methylation at the corresponding recognition sites [18]. These cellular defence mechanisms constitute a natural barrier to the acquisition of exogenous DNA, thus representing key obstacles in the genetic accessibility of these bacteria while also posing a significant hindrance to the promiscuity of a conjugative plasmid [19].

Megaplasmids are extrachromosomal genetic elements in the size range of 100 kb and larger [20], [21]. By definition, megaplasmids are not essential for the viability of their hosts under all growth conditions, but paradoxically many megaplasmids carry

the genetic information for the defining and characteristic traits of the organism in which they reside [22]. A 190 kb conjugative megaplasmid pMP7017 was previously identified in the human isolate *Bifidobacterium breve* JCM7017, representing the first and only bifidobacterial plasmid identified to date with conjugative ability [23], [24]. While plasmids of bifidobacterial origin were once thought to be rare, pMP7017 homologs have recently been discovered in various *Bifidobacterium longum* strains [25].

While there is a stark underrepresentation of fully resolved conjugal systems in Gram-positive (G^+) bacteria, characterised Type Four Secretion systems (T4SSs) from enterococcal sex-pheromone responsive plasmids pCF10, pIP501 and pAD1 [26], quorum sensing pLS20 from *Bacillus* [27], staphylococcal multi-resistance plasmid pSK41 and clostridial pCW3 [28] serve as models in monoderms. Characterised T4SSs of G^+ species differ from their Gram-negative (G^-) counterparts primarily due to the absence of outer membrane core complex components and the VirB11 ATPase [29]. While the overall composition and structure of G+ T4SSs deviates, genetic and structural data gathered for these systems have established that the transfer of ssDNA across the cytoplasmic membrane and cell wall in unicellular G^+ bacteria requires a minimum of four conserved subunits that are distant homologues of prototypical VirB/D4 T4SS components described for G^- bacteria, namely, homologs of the VirB6 mating channel component, VirB1 cell wall hydrolase, VirB4 and VirD4 ATPases [30], [31].

Conjugation is an energy demanding process, as highlighted by the presence of two ATPases that appear to be essential for the core functionality of a conjugative system. A membrane-associated VirD4-like or FtsK-like coupling protein (CP), located on the entry side of the donor cytoplasm, is involved in the recognition of the DNA substrate and is believed to play an active role in pumping the nucleoprotein complex through the secretion channel into the recipient cell, aided by a VirB4-like ATPase [30], [32]. Membrane-associated peptidoglycan (PG) hydrolases from G^+ bacteria such as (names are assigned as; Protein $_{\text{encoding mobile element}}$) $TraG_{pSK41}$, $PrgK_{pCF10}$, $TcpG_{pCW3}$, $Orf7_{pIP501}$ are believed to be responsible for local degradation of the peptidoglycan layer facilitating the formation of the secretion channel [32]-[35]. Unlike VirB1 from G^- T4SSs, G^+ VirB1-like PG hydrolases are essential components of the G^+ DNA secretion system and generally contain two catalytic domains, making them much larger than their G^- counterparts [37]. Bi-topic VirB8-like subunits such

as $TcpC_{pCW3}$ and $TraM_{pIP501}$ and poly-topic VirB6-like subunits such as $TcpH_{pCW3}$ and $Orf12_{pIP501}$ are believed to act as scaffolding proteins which form the core of the DNA transfer channel [32], [37], [38]. The genetic framework for this complex macromolecular machinery is typically represented as a single operon, or a set of adjacent transcriptional units (TUs).

In addition to the T4S macromolecular complex, transfer of plasmid DNA requires a strand-specific DNA processing complex. Biochemical processing of the conjugative element is initiated at an essential *cis*-acting region of DNA, ranging in size from 40 –500 bps, known as the origin of transfer (*oriT*). Within the *oriT* the mobilisation (Mob) protein exerts its strand- and site-specific endonuclease activity to generate a ssDNA molecule for transfer and subsequently pilots the covalently bound ssDNA to the secretion apparatus. Mob proteins in are highly diverse and can be categorised into nine known families: (1) MobV, MobB, MobP, MobQ, MobF and MobH families, which can be functionally grouped together as so-called HUH endonucleases, (2) MobT family relaxases resemble rolling circle initiators of replication, (3) the MobC family which represent PD-(D/E)XK restriction endonucleases, and (4) the MobM family, representing tyrosine recombinases [40].

During the current investigation we significantly increased the conjugation efficiency of pMP7017 through MTase expression in the donor prior to conjugal mating, thereby protecting the transferred DNA from degradation from cognate REases expressed in the recipient. This increase in efficiency enabled an investigation of conjugation-associated genes, through a systematic approach of mutational analysis, not previously attainable. This analysis allowed the genetic delineation of several required components of the pMP7017 conjugal system which are contained within two unconnected gene clusters, *tra1* and *tra2*. In addition, this work uncovered a number of genes encoding potential partitioning and stability functions that are required for efficient horizontal transmission of pMP7017. Collectively this data enabled the proposal of the first working model for conjugation in bifidobacteria and represents a significant advancement in our understanding of T4SSs in G+ commensals.

MATERIALS AND METHODS

Strains, plasmids and cultivation conditions

Bacterial strains and plasmids used in this study are listed in **Supplementary Table S1**. Bifidobacterial strains were initially cultivated from glycerol stocks in reinforced clostridial medium (RCM; Oxoid Ltd., Basingstoke, England) and subsequently sub-cultured into either filtered RCM or modified de Man, Rogosa, and Sharpe medium (mMRS) prepared from first principles [41], [42]. Prior to inoculation, mMRS medium was supplemented with cysteine-HCl (0.05%, wt/vol) and 1% (wt/vol) lactose (Sigma-Aldrich, Steinheim, Germany). Cultures were grown at 37°C in an anaerobic chamber (Davidson and Hardy, Belfast, Ireland). *Escherichia coli* strains were routinely cultivated in Luria-Bertani medium (LB) and grown at 37°C with agitation. Growth media included, where appropriate, the addition of the following antibiotics: tetracycline (10 μg ml^{-1}) and/or chloramphenicol (2 μg ml^{-1}) for bifidobacterial strains; tetracycline (10 μg ml^{-1}), chloramphenicol (5 μg ml^{-1}), ampicillin (100 μg ml^{-1}), erythromycin (100 μg ml^{-1}) and/or kanamycin (50 μg ml^{-1}) for *E. coli* cultures.

Conjugal Matings

Individual overnight cultures of donor and recipient strains were inoculated at 1 % (v/v) into mMRS broth supplemented with cysteine-HCl (0.05 % [wt/vol]) and lactose (1.0 % [wt/vol]), and grown to an optical density (OD_{600nm}) value 0.8-1.0 at 37°C under anaerobic conditions. Donor and recipient cells were combined in a 1:1 ratio, based on OD value, up to a total volume of 2 ml, briefly vortexed and concentrated by centrifugation at 2,200 x *g* for 5 min. The resultant cell pellet was resuspended in 200 μl of RCM and spread plated onto non-selective agar. RCA plates were incubated at 37°C overnight under anaerobic conditions. Following incubation, the donor/recipient conjugation mixture was recovered from the plates using a cell scraper (Sarstedt, Germany) and resuspended in 2 ml of phosphate-buffered saline and serially diluted. Donor and recipient cells were enumerated based on selective carbohydrate utilization by the donor or recipient strain on mMRS agar, while transconjugants were isolated based on the selective carbohydrate for recipient cells

in combination with tetracycline (10 μg ml^{-1}) to select for the presence of a pMP7017-derivative that had been marked with a tetracycline resistance gene [23]. Fifty selected putative transconjugant colonies were restreaked twice on fresh mMRS plates supplemented with ribose and 10 mg ml^{-1} of tetracycline. For validation purposes, ten suspected transconjugants from each experimental group were grown in mMRS broth with tetracycline at a concentration of 10 mg ml^{-1} and the appropriate carbohydrate, after which DNA was isolated and PCR performed with primers based on genes unique to the donor strain, recipient strain and the megaplasmid. Those which were PCR negative for the donor strain, and PCR positive for the recipient and megaplasmid were counted. Mutant donor strains that did not yield transconjugants within the limit of detection were considered to be unable to perform conjugation.

DNA Manipulations

Total genomic DNA was isolated from bifidobacterial strains as previously described [43]. Minipreparations of plasmid DNA were isolated from *E. coli* with modifications for *Bifidobacterium* as detailed previously [44]–[46]. Procedures for DNA manipulations were conducted essentially as described in a previous publication [47]. Single-stranded DNA oligonucleotide primers used in this study were synthesised by Eurofins (Ebersberg, Germany) and are listed in **Supplementary Table S2**. Both standard and colony PCRs [48] were routinely carried out with DreamTaq Green PCR Master Mix (Thermo Scientific, Waltham, United States), while high-fidelity PCR was achieved using Q5 High-Fidelity 2X Master Mix (New England Biolabs, Herefordshire, United Kingdom). All reactions were carried out in a Life Technologies Proflex PCR System (Thermo Scientific, Waltham, United States). PCR fragment purification was achieved using the Roche High Pure PCR purification kit (Roche Diagnostics, Basel, Switzerland). Verification of the expected PCR product size was assessed by 1% agarose gel electrophoresis containing Invitrogen SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, Waltham, Massachusetts, United States) prior to visualisation on a UV transilluminator. T4 DNA ligase and restriction endonucleases were obtained from Roche Diagnostics (Basel, Switzerland) and used according to the supplier's instructions. Introduction of plasmid DNA into *E. coli* and transformation of bifidobacterial [49] strains was achieved by electroporation according to published protocols. Sequencing was performed at Eurofins (Ebersberg,

Germany) on all plasmid constructs to validate the desired orientation of the cloned fragment and verify sequence integrity.

Methylase Cloning

For the construction of pBbrU a 1.6 kb DNA fragment comprising the Bbr_1119 coding sequence encoding methylase BbrUII.M and presumed promoter region was amplified by high-fidelity PCR using primer pair UCCPst1MF and UCCPst1MR (**Supplementary Table S2**) containing HindIII and NotI recognition sites, respectively, to facilitate forced cloning into pBC1.2. Both the resulting PCR product and pBC1.2 cloning vector DNA were digested with HindIII and NotI prior to PCR purification and ligation at a 2:1 ratio of insert: vector. The ligation mixture was used to transform *E. coli* XL1-blue competent cells with selection for tetracycline and ampicillin as described previously [48]. Colony PCR was performed on resistant isolates to confirm the presence of the BbrUII.M-encoding gene using the above mentioned UCCPst1MF and UCCPst1MR primers and the PCR product of positive clones sent for sequencing with Eurofins Genomics. 10 µl of total chromosomal DNA isolated from PCR positive strains were incubated at 37°C overnight with 1 µl of PstI in a total reaction volume of 30 µl. 15 µl of the reactions were subsequently run on an agarose gel to assess the level of DNA degradation as compared to the controls. Strains whose DNA did not display degradation were protected from restriction due to methylation by BbrUII.M, thus demonstrating the functionality of the cloned methylase.

Generation of pMP7017 Insertion Mutants

An internal ~500 bp fragment of pMP7017-associated genes with the following locus-tags; B7017_p0031, B7017_p0051, B7017_p0053, B7017_p0056, B7017_p0057, B7017_p0059, B7017_p0060, B7017_p0061, B7017_p0062, B7017_p0063, B7017_p0064, B7017_p0108, B7017_p0114, B7017_p0115, B7017_p0116, B7017_p0119, B7017_p0122, B7017_p0124, B7017_p0125, B7017_p0126, B7017_p0127, B7017_p0134, B7017_p0193 and B7017_p0204 were amplified using gene-specific primers listed in **Supplementary Table S2**, with a HindIII site incorporated into the forward primer and an XbaI site incorporated into

the reverse primer. The approach taken in the construction of the insertion mutants is detailed in a previous publication [47]. Site-specific recombination in tetracycline-resistant isolates was assessed by an initial screen for *tet*(W) gene integration via colony PCR with the primer pair TetWF and TetWR. Confirmation of integration at the correct genetic locus was established with primers designed upstream of the integration site (**Supplementary Table S2**) in combination with either TetWF or TetWR, depending on the orientation of the *tet*(W) gene as ascertained by PCR and sequencing by Eurofins Genomics. The strain names of resulting insertion mutants were based on the locus tag name of the corresponding disrupted gene and are outlined in **Supplementary Table S1**. The resulting strains harbouring pMP7017- derivatives with specific gene disruptions were then used as donors and assessed for their conjugation ability employing *B. breve* UCC2003 as the recipient strain. Sequencing of pMP7017-derivatives was performed on an Illumina MiSeq Sequencing platform, to verify integration at the correct locus.

Plasmid vector constructions

To generate the plasmid constructs for genetic complementation, the coding sequences and presumed promoter regions of, B7017_p0062, B7017_p00114 and B7017_p0125 were amplified from *B. breve* JCM7017 total DNA by PCR using the primer pairs outlined in **Supplementary Table S2**. The resulting amplicons were ligated into the ClaI and PspXI sites in plasmid pBbrU following restriction and purification. 4 µl of each ligation mixture was introduced into *E. coli* XL1-blue compentent cells by electroporation. Plasmid preparations of *E. coli* XL1-blue cells harbouring the constructs, pBbrU_62C, pBbrU_114C and pBbrU_125C, respectively, were sequenced using primer pair, M13 and M21 to confirm the presence and fidelity of the insert. The constructs where then introduced into *B. breve* JCM7017 strains harbouring pMP7017 with the corresponding mutation.

To identify the *ori*T-encompassing DNA region, various non-coding regions in proximity to genes identified in this work to be essential for conjugation (intergenic regions located between: B7017_p0062 and B7017_p0063, B7017_p0113 and B7017_p0114, B7017_p0124 and B7017_p0125, B7017_p0128 and B7017_p0129), in addition to large non-coding regions identified throughout the plasmid (intergenic regions located between: B7017_p0033 and B7017_p0034, B7017_p0046 and

B7017_p0047, B7017_p0048 and B7017_p0049, B7017_p0066 and B7017_p0067, B7017_p0076 and B7017_p0077, B7017_p0079 and B7017_p0080, B7017_p0130 and B7017_p0131, B7017_p0144 and B7017_p0146, B7017_p0197 and B7017_p0198) were amplified using the primers given in **Supplementary Table S2**. The resulting amplicons were cloned into the ClaI and PspXI sites of plasmid pBbrU. The constructs were recovered from *E. coli* XL1blue cells by plasmid preparation and introduced into *B. breve* JCM7017-199 by electroporation, resulting in strains *B. breve* JCM7017-199 pBC_oriT1 through to *B. breve* JCM7017-199 pBC_oriT13 (**Supplementary Table S1**).

In silico structural analyses and heuristic modelling

Sequence alignments were performed with Clustal Omega [50] and visualised using Jalview [51]. SignalP [52] was used to identify signal peptides, TMHMM [53] was employed to identify transmembrane regions, Pfam [54] and CDD [55] were used to identify conserved functional regions. The protein sequence and structural analysis were performed using HHpred (Homology detection & structure prediction by HMM-HMM comparison) using the default settings [56], in addition to Phyre2 [57] and RaptorX web server [58].

Comparative genomics

Nucleotide sequences of pMP7017 conjugal clusters and its identified *B. longum* homologues were extracted from the NCBI database and aligned with Clustal Omega [59], phylogeny was inferred using MEGAX [60] following gap removal. The Blast+ terminal [61] was used to determine percentage identity, the analysis was conducted in R and the comparative figures produced using the genoPlotR package [62], [63].

RESULTS

Recipient expressed RM systems constitute a significant barrier to pMP7017 acquisition

While intraspecies conjugal transfer of pMP7017 has previously been demonstrated, the conjugation efficiencies reported were relatively low precluding the genetic delineation of conjugation-associated genes [23]. The efficiencies obtained in this research ranged from 5.1×10^{-9} with recipient *B. breve* UCC2003 to 1.33×10^{-6} with recipient strain *B. breve* JCM7019, representing a 3.9-thousand-fold difference in transfer efficiency. This prompted an investigation into the intra-species genomic diversity which may account for this variance and an evaluation of the conjugation protocol to improve the transfer efficiency of pMP7017.

Comprehensive comparative genomics and methylome analysis has previously highlighted significant diversity in the *Bifidobacterium breve* accessory genome whereby RM systems constitute a substantial portion of the variable gene content [16]. In the aforementioned publication, *B. breve* UCC2003 was demonstrated to express two active Type II RM systems M.BbrUIII/BbrUIII and M.BbrUII/BbrUII, with equivalent specificities to PstI (5'-GTCG^{6m}AC-3') and SalI (5'-GT^{5m}CGAC-3'), respectively. Although *B. breve* JCM7019 was not included in these analyses, it is predicted to encode a single Type II RM system, M.Bbr7019ORF15P/ Bbr7019ORF16P, with an equivalent specificity to BglII (5'-AG^{6m}ATCT-3') [64]. To preliminarily assess if a correlation could be observed between conjugation efficiency and recipient-expressed RM systems we compared the number of RM recognition sites on pMP7017 corresponding to chromosomally-encoded RM systems of *B. breve* UCC2003 vs *B. breve* JCM7019, revealing 45 PstI and 42 SalI sites vs 7 BglII sites, respectively (**Figure 1A**). As such, a cumulative effect based on the number of corresponding RM sites may have caused the discrepancy in efficiency observed between the two *B. breve* recipient strains.

To experimentally validate this assumption, the Pst1 isoschizomer, BbrUIII methylase-encoding gene from *B. breve* UCC2003 was expressed in the donor strain *B. breve* JCM7017-199 on plasmid pBbrU prior to mating, facilitating methylation at the 45 PstI recognition sites of pMP7017. Conjugal mating experiments of *B. breve* JCM7017-199 pBbrU with recipient strain *B. breve* UCC2003 resulted in ~359.5-fold

increase in conjugation efficiency as compared to the unmethylated donor *B. breve* JCM7017-199 (based on the mean of quadruplicate experiments on solid medium) (**Figure 1B**). These findings illustrate that active RM systems in the recipient constitute a significant barrier to the establishment of pMP7017 by conjugal transfer in bifidobacteria. Furthermore, relevant MTase expression in the donor cell prior to mating allows for a substantial increase in conjugation efficiency, enabling the molecular investigation of pMP7017-encoded genes which had not previously been possible.

Functional predictions of pMP7017-encoded proteins reveal two candidate T4SS clusters, *tra1* and *tra2*

To identify candidate components of the pMP7017 conjugal system to be targeted for mutagenesis, we first thoroughly assessed all pMP7017 coding sequences for the presence of known functional domains and conserved motifs associated with DNA secretion system components. Further to this, we determined structural homologies and membrane topologies to enable more comprehensive functional predictions. From this analysis seven genes encoding putative ATPases, two predicted PG hydrolase-specifying genes and several additional suspected conjugation-associated genes were identified, discussed in more detail below (**Table 1**). As the majority of these genes do not have specified gene names we will refer to each targeted gene in this section by a derivative of their locus tag (where the prefix "B7017_" will be removed, for example, the gene corresponding to locus tag B7017_p0053 will be written as p0053 when referring to the gene and P0053 when referring to the protein product).

Five of the identified ATPases, i.e. P0053, P0062, P0114, P0126 and P0204, are presumed cytosolic proteins due to the absence of predicted TMDs, thus represent VirB4-like candidates. P0053 and P0062 each contain a MinD/ParA functional domain. MinD/ParA-like proteins are P-loop NTPases that function in orchestrating the subcellular organization of bacterial cells, such as plasmid segregation and the positioning of macromolecular complexes, the latter of which includes spatial coordination of the tDNA substrate-T4S channel docking, as observed for VirC1 ATPases of G- T4SSs [65], [66]. P0114 displays significant structural similarity to the VirB4 family of ATPases including TrwK from the T4SS of plasmid R388 and

contains an N-terminal AAA_10 domain and a centrally located CagE_TrbE_VirB domain, both common among VirB4 and VirB4-like ATPases of T4SSs [67]. P0126 and P0204 contain an FtsK/SpoIIIE family domain and an AAA_16 domain, respectively. Furthermore, P0126 displays structural similarity to the pCW3 conjugation coupling protein, TcpA, from *C. perfringens*. Two putative membrane-anchored ATPases representing potential conjugal CPs were also identified in our analysis, P0059 and P0125, with three and six predicted TMDs, respectively. While P0059 does not present with a recognizable ATPase functional domain, it displays significant structural homology to the IcmO (DotL) TraG-family coupling protein of the *Legionella* Icm/Dot T4BSS [68]. P0125 contains an FtsK/SpoIIIE domain and displays structural similarity at its C-terminus to the TcpA coupling protein from the T4SS encoded by the clostridial plasmid pCW3 (PDB accession: 6O1X, unpublished).

Two putative PG hydrolases were additionally identified, corresponding to genes p0116 and p0193. P0116 is predicted to contain a C-terminal "cysteine, histidine-dependent amidohydrolases/peptidases" (CHAP) domain and centrally located phage tail lysozyme domain, while P0193 contains a single C-terminal CHAP domain.

This analysis further highlights two distinct loci encoding several gene products that display structural similarity to those of G+ T4SS components and cluster with candidate conjugation-associated ATPases, these loci will henceforth be referred to as *tra1* and *tra2* (**Figure 2**).

The *tra1* locus was found to encompass several genes with putative functions in DNA processing; promoting stability and faithful transmission of plasmid DNA (**Table 1**). p0051 is predicted to specify a monotopic protein displaying structural similarity to the VirB8 homolog, TcpC, of the pCW3 conjugal system of *C. perfringens* [38]. The product encoded by p0056 contains eight predicted TMDs and displays structural similarity to the VirB6 homolog, TrwI, from the T4SS of conjugative plasmid R388 from *E. coli*. The deduced product of p0057 represents a 134 aa polytopic membrane protein with two predicted TMDs and shares structural similarity to the PrgI family of proteins found encoded in T4SS operons and are postulated to interact with VirB4 ATPases [69]. Although P0060 encoded by *tra1* is currently annotated as hypothetical, *in silico* analyses reveals significant structural similarity to the nucleoid associated protein (NAP), mycobacterial integration host factor (mIHF) [70]. p0064 encodes a predicted cytosolic protein that shows structural

similarity to RctB, the primary initiator of replication of of *Vibrio cholerae*'s secondary chromosome. In addition to the aforementioned genes with predicted links to conjugation, the genes corresponding to p0057, p0061 and p0063, which encode hypothetical proteins, were additionally selected for site-directed mutagenesis due to their genomic location within the *tra1* cluster and flanking suspected conjugation-associated genes.

Locus *tra2*, containing two of the aforestated ATPase-encoding genes, p0114 and p0125, also encompasses genes with diverse putative functions and located in varying orientations. p0119 of locus *tra2* encodes a 153 aa polytopic membrane protein with four predicted TMDs, representing a potential VirB6-like candidate. p0122 encodes a protein which is structurally similar to the *E. faecalis* pIP501-encoded T4SS protein TraN, an accessory protein of the $TraA_{pIP501}$ relaxase and contains a HTH_17 domain, common among DNA-binding transcriptional regulators.

P0124 displays structural similarity to initiators of replication such as RctB and encompasses a RepA_C domain. p0124 flanks the candidate coupling ATPase-encoding gene p0125 and was in previous work shown to be dispensable for pMP7017 replication [24]. P0127 displays weak structural homology to PG endopeptidase RipA and contains a C-terminal PAP_PilO functional domain. PilO proteins are involved in Type IV pilus biogenesis and are essential for its functioning in various characterised T4SSs of G- origin [71]. p0134 encodes a putative cell division protein as it contains a DivIVA domain. Mutation of *divIVA* has been shown to significantly affect invasiveness and cell-cell interactions in *Listeria monocytogenes*, decrease adhesion and attenuate pathogenesis in *Streptococcus suis* and negatively affect virulence in *E. faecalis*. These effects have mainly been attributed to its role in biofilm formation as well as indirect effects due to the altered activity of its indicated interactive partners [72]–[74].

Despite a manual search of all pMP7017 protein sequences using the minimum conserved amino acid signature for each characterised MOB group, a clear relaxase candidate was not identified in our analysis [75]–[77]. In light of this, and considering the diversity of relaxases in G+ bacteria, genes containing nuclease-related functional domains in the vicinity of the two putative conjugal clusters, p0031 and p0108 were also considered for mutagenesis.

Two independent gene clusters encode essential components of the pMP7017 conjugal system

Following individual mutation of 24 genes (**Figure 2**) with predicted or suspected functions in conjugation, plasmid pBbrU was introduced into all mutant derivatives to increase the conjugation frequency in the experimental recipient-donor set-up. Each pBbrU-harbouring recombinant strain was subsequently used as a donor in mating assays with recipient strain *B. breve* UCC2003 and the effect of each mutation on conjugation efficiency was assessed relative to the control donor (*B. breve* JCM7017-199). For clarity mutant donor strains were named in the following manner; *B. breve* JCM7017 harbouring pMP7017 with an insertion mutation at locus p0053 and pBbrU will be referred to as *B. breve* JCM7017-53.

Subsequent to this examination, mutant donor strains exhibiting a statistically significant deviation in conjugation efficiency as compared to the control strain were differentiated by the level of impact their individual mutation effected on conjugation efficiency. Impact levels were categorised as; not deviated (no statistical significance), moderately deviated (± 10 -99.9 fold change (fc)), major impact (±100 –1,000 fc), or essential (where no transconjugants recovered within the limit of detection, 10^{-10}).

No transconjugants were recovered within the limit of detection in matings with mutant donors; *B. breve* JCM7017-56, *B. breve* JCM7017-57, *B. breve* JCM7017-62, *B. breve* JCM7017-114, *B. breve* JCM7017-115, *B. breve* JCM7017-116 or *B. breve* JCM7017-125. Thus genes; p0056, p0057, p0062, p0114, p0115, p0116 and p0125 are deemed to be essential for conjugation (**Figure 3**). The plasmids conjugative ability suffered a major decrease in efficiency as a result of p0051 and p0059 mutation, while moderate effects were observed for mutants *B. breve* JCM7017-60, *B. breve* JCM7017-61, *B. breve* JCM7017-63, *B. breve* JCM7017-64, *B. breve* JCM7017-126 and *B. breve* JCM7017-127. Mutation of eleven genes; p0031, p0053, p0063, p0064, p0108, p0119, p0122, p0124 p0134, p0193 and p0204 caused no statistically significant deviation in conjugation efficiency as compared to the reference strain. This data suggests that both the *tra1* and *tra2* loci contain conjugation-associated genes.

To validate the involvement of both gene clusters we then endeavoured to verify the essentiality of the presumed energetic components of the pMP7017 T4SS, encoded by genes p0062, p0114 and p0125. To achieve this, mutants *B. breve*

JCM7017-62, *B. breve* JCM7017-114 and *B. breve* JCM7017-125 were supplied with their corresponding coding sequence in plasmids pBbrU_62C, pBbrU_114C and pBbrU_125C, respectively. Reintroduction of intact copies of genes, p0062 and p0125 in their corresponding conjugation-deficient mutant strains restored the plasmid conjugative ability, thus these two genes are considered essential for conjugation (**Figure 4**). Genetic complementation of p0114 was unsuccessful in restoring a conjugation positive phenotype, however we cannot exclude the possibility of *cis*-mediated effects on p0116 encoded downstream within the same presumed transcriptional unit (TU) confounding the assignment of causality.

Mobilisation of plasmid pBbrU_oriT3 by pMP7017 indicates putative pMP7017 *oriT*

In order to identify the $oriT_{pMP7017}$ we investigated the non-coding regions flanking the identified conjugation-associated genes, in addition to a number of large non-coding regions across the plasmid for the presence of inverted repeats (IRs) (repeat length: > 10 bp, mismatches: ≤ 1, distance < 100 bps). From this analysis twelve segments of non-coding stretches with varying lengths (212 bps – 1,175 bps) were identified, containing ≤ 1 IR. These segments were cloned into the non-mobilizable low-copy vector pBbrU generated in this work, and the resulting constructs (designated pBbrU_oriT1: pBbrU_oriT13) were introduced into reference strain *B. breve* JCM7017-199, and their comobilization assessed in mating assays as previously described. No transconjugants were obtained with donor strain *B. breve* JCM7017 pMP7017-199 containing non-mobilisable plasmids pBbrU_oriT1, pBbrU_oriT2 and pBbrU_oriT4: pBbrU_oriT13, while plasmid pBbrU_oriT3, containing the 720 bp intergenic region between genes p0062 and p0063 (nucleotide position: 52,651 bp…53,370 bp) was shown to exhibit a co-mobility frequency of 2.94% (**Table 2**). While this is significantly lower than anticipated for a region containing the *cis*-acting *oriT* sequence as compared to assays of a similar nature performed with other plasmids, it is necessary to point out that we selected for the presence of the pBbrU_oriT constructs via antibiotic selection in addition to the selection for pMP7017 allowing the measurement of its acquisition, which may result in a lower rate of mobilisation [78]–[80].

Essential genes of the pMP7017 conjugal system are conserved among related megaplasmids

Due to the recent discovery of *B. longum* subsp. *longum* megaplasmids, we aimed to assess the presence and diversity of pMP7017 conjugation-associated gene clusters among related plasmids [25]. In addition to fourteen *B. longum* subsp. *longum* megaplasmids, two fully sequenced pMP7017 homologs, P1 and apr34_1788, identified in the NCBI database were also included in this analysis. Comparative analysis of pMP7017 was performed via all-against-all, bi-directional BLASTP alignment, and clustering implemented in the MCL pipeline. While the sequences of the *B. longum* genomes containing potential pMP7017-homologous megaplasmids are in draft status, the presence of these gene clusters within a single contig allows us to include these sequences in the dataset.

From this analysis, it was revealed that all conjugation-associated genes are present in fifteen pMP7017 homologs, while p0051 is absent from the *B. longum* subsp. *longum* strain MCC10015 megaplasmid (**Figure 5**). Genes within the *tra1* locus are present and highly conserved in all plasmids, while more variability is observed between *tra2* clusters. Evidence of gene loss and/or gene acquisition events is apparent in *tra2*, with the absence of potential gene pair p0122 and p0123 from ten pMP7017 homologs, in addition to the absence of a transposase-specifying p0118 and p0132 from two and one plasmids, respectively. Considering their lack of conservation among related plasmids, the gene products of p0122, p0123 and p0132 do not seem to be critical components of the conjugal system.

Furthermore, comparative analysis revealed that the presumed VirB4-specifying p0114 is well conserved, although it is much smaller than its *B. longum* counterparts, all of which contain an additional 1,644 bp centrally located intron, encoding an endonuclease domain (Hom_end, Pfam: PF05204, interval, E-value: 5.61^{-06}) not present in $p0114_{pMP7017}$. Despite a relatively low level of sequence conservation, all pMP7017 homologs encode proteins with structural features and functional domains homologous to the putative PG hydrolase p0116, VirB8-specifying p0051 and putative channel component-encoding p0127. In addition, p0125 encoding the presumed FtsK-like CP is well conserved among the *B. longum* subsp. *longum* megaplasmids analysed.

While not essential for conjugation or replication, p0124 encoding the RepA-like protein, is present and highly conserved between all bifidobacterial megaplasmids [24].

DISCUSSION

Horizontal gene transfer mediates rapid evolution in bacteria and the efficiency of genetic transfer can be influenced by a number of biotic and abiotic factors. The genetic intractability of bifidobacteria has at least partly been attributed to RM systems which pose a fundamental barrier hindering *in vitro* gene manipulation techniques used to study them. Circumvention of host RM systems has been a strategy utilized in the transformation of bifidobacteria through the cloning of MTases [47], [81]–[83] or through the removal of RM targets in synthetically engineered plasmid sequences [84]. In this work we established that *in vivo* methylation of pMP7017 significantly increases conjugation efficiency, illustrating the sensitivity of pMP7017 during transfer to RM systems expressed in the recipient while simultaneously raising the transfer efficiency to a sufficiently high level to enable the molecular investigations of plasmid-encoded genes.

In total 24 insertion-based pMP7017 mutants were generated which were tested for their ability to conjugate following methylation. Thirteen of the twenty-four mutations were shown to be associated with a defective conjugation phenotype. Seven mutations in genes p0056, p0057, p0062, p0114, p0115, p0116 and p0125 completely abolished the plasmids conjugative ability, while an additional six mutants p0051, p0059, p0060, p0061, p0126 and p0127 displayed a statistically significant reduction in efficiency as compared to the reference strain. In contrast to other characterized systems where conjugation genes tend to be co-orientated and encoded in a single operon or a set of contiguous operons, or, though less frequently observed, split in two independent genetic locations encoding defined functions such as DNA transfer and mating pair formation, our data indicates the concerted action of two unconnected genomic loci in the conjugative transfer of pMP7017 [85].

Based on our observations of the mutagenesis data, in combination with a rigorous evaluation of functional domains associated with T4SS proteins and structural homologies to components of characterised conjugal systems, where possible a putative role has been proposed for gene products presenting with a conjugation-defective phenotype upon mutation. However, it is important to note that

complementation studies are required for all conjugation-associated genes identified in this study to validate their proposed role in pMP7017 conjugation.

We propose from this work that p0125 acts as an FtsK-type coupling protein such like $TcpA_{pCW3}$, and suggest that $p0116_{pMP7017}$ encodes a VirB1-like PG hydrolase required for local opening of the peptidoglycan layer (**Figure 6**) [36]. Furthermore, based on the conjugation-defective phenotype observed upon mutation of p0056, in addition to its predicted membrane topology and structural relatedness to VirB6-like scaffolding proteins, we infer that p0056 encodes the VirB6-like protein of the pMP7017 translocon. By a similar reasoning, we propose that the protein products of p0051 and p0057 specify VirB8-like and VirB3-like components, respectively. A characteristic feature of many G^+ T4SSs is a soluble VirB4-like ATPase which has been postulated to functionally interact with a membrane-associated protein encoded upstream and often overlapping the virB4-like gene, as observed with PrgI and PrgJ from pCF10 [29]. In the pMP7017 system however, the presumed VirB4-like specifying gene, p0114, is not flanked by a gene encoding a TM-encompassing protein. Thus, the presumed VirB4-like protein of pMP7017 may associate with the secretion channel via interactions with the VirB1-like protein product of the proximal p0116 gene, as described for the pIP501 T4SS [33], or indeed interact with the DNA substrate and P0125 coupling protein as suggested for the *E. faecalis* VirB4 homologue, $PrgJ_{pCF10}$ [69].

Segregation and conjugation are intrinsically linked to ensure faithful transmission and transfer of low copy plasmids [86]. As such functional associations between conjugation complexes and segregation systems have been observed for many characterised T4SSs. The *par*MRC system of *E. coli* plasmid R1 contributes to the assembly of the T4SS and facilitates initiation of conjugative transfer through its interactions with relaxase $TraI_{R1}$ and coupling protein $TraD_{R1}$ [87]. The ParABS system of the Gonococcal Genetic Island of *N. gonorrhoeae* is essential for T4SS-mediated DNA secretion [88]. Additionally, deletion of ATPase-encoding gene *stbB* from the highly conserved stability locus, *stbABC*, of plasmid R388 was shown to abolish the plasmid's conjugative ability [89]. Collectively these data may explain the lack of conjugative ability observed upon mutation of p0062, encoding the MinD/ParA type ATPase, in addition to the significant reduction in efficiency observed with mutant *B. breve* JCM7017-59.

Considering *tra1* encodes proteins with known functions in plasmid partition and stability, our initial inclination was to assume that this locus encodes segregation and stability determinants and that *tra2* encodes T4SS functions for pMP7017 conjugal transfer. While this may be a valid deduction, it appears that *tra2* does not encode all expected components of a G+ T4SS. Furthermore, it was established in this work that the sole VirB6 candidate of *tra2*, p0119, is dispensable for conjugation and a gene encoding an alternative VirB6-like candidate was not identified within the *tra2* locus nor its surrounding region. Furthermore, a comobility frequency of 2.94% was observed for plasmid pBbrU_oriT3 containing the 720 bp intergenic region between genes p0062 and p0063 suggesting it constitutes the pMP7017 origin of transfer. With this in mind, we propose a model where the pMP7017 conjugal system comprises an amalgamation of two once functionally independent T4SSs, which have adapted and evolved into coalescence. While the fundamental evidence in support of this postulation derives from the data obtained in this work, we have also considered potential inferential evidence which may corroborate this assumption.

It was previously proposed that pMP7017 is a co-integrant due to a ~10% deviation in GC content from one half of the plasmid to the other [23]. While it was recently deduced that locus p0146 encodes the essential replicase $RepB_{pMP7017}$, as indicated from our *in silico* analyses, pMP7017 contains two additional replicase-specifying genes, p0064 and p0124, located within *tra1* and *tra2*, respectively. Indeed, multiple-replicon plasmids have garnered a lot of recent attention, these plasmids result from the hybridisation of two or more plasmids [90], [91]. As multiple plasmid-encoded replicases is an indicator of cointegration, this may support the idea that these clusters originated in this manner. Although without knowledge of the evolutionary dynamics of pMP7017 this is purely speculative.

Based on the data obtained in this work, two alternative models for the conjugal transfer of pMP7017 may be proposed, either *tra1* encodes segregation and stability determinants and *tra2* T4SS functions, or both clusters encode components which constitute the pMP7017 T4SS. However, we would argue that neither model is solely sufficient to account for the data obtained in this work in its entirety and believe both suggestions are compatible and coexist. We therefore propose that the transfer apparatus of the low copy megaplasmid pMP7017 has evolved via an amalgamation of two independent T4SSs integrated with an active segregation system to optimize both its lateral and vertical modes of dissemination.

The mutational mapping of the *tra* loci of pMP7017 achieved in this study has enabled the genetic delineation of conjugation-associated genes of this and related plasmids, which are prevalent among the demonstrated bifidobacterial conjugative megaplasmids. While experimental validation of the role of each conjugation-associated gene product and subsequent mechanistic and structural studies are required to disentangle the information derived in this work, the research described herein should therefore be considered as a springboard for further functional investigations to define the pMP7017 T4SS architecture, its bases for substrate recruitment and translocation route for DNA substrates.

FIGURES AND TABLES

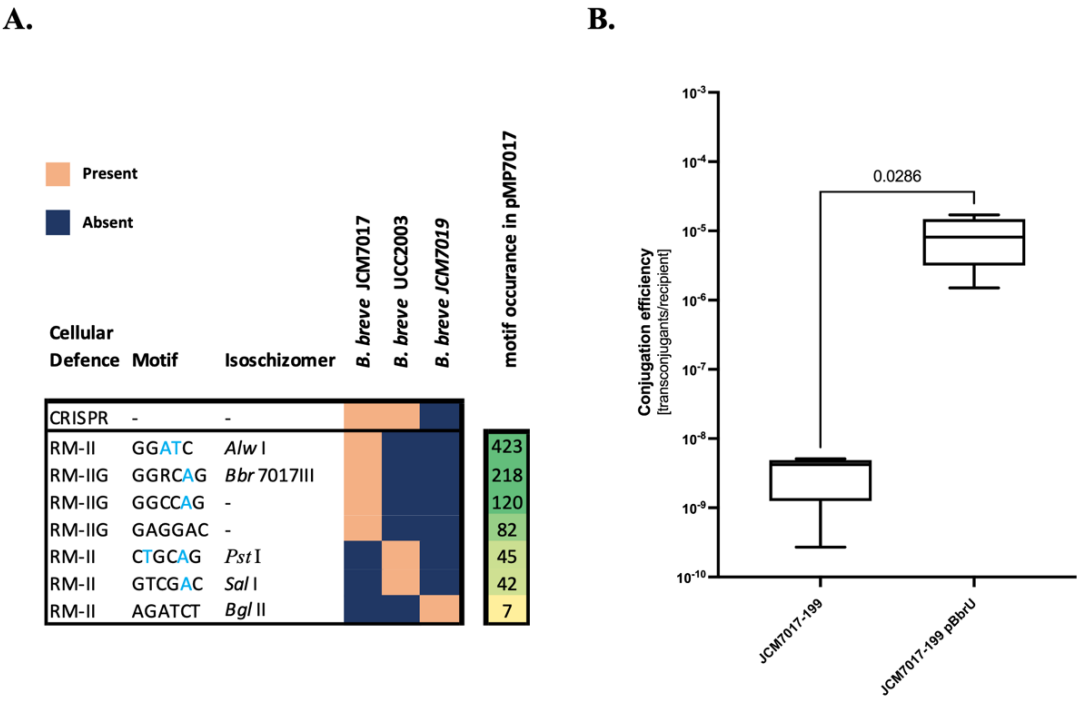


Figure 1. Effect of methylation on pMP7017 conjugation efficiency

A. Methylome comparison of *B. breve* JCM7017, *B. breve* UCC2003 and *B. breve* JCM7019, detailing the occurrence of motifs corresponding to their respective RM systems in pMP7017. **B.** Boxplot illustrating the 359.5-fold increase in conjugation efficiency upon methylation of pMP7017 prior to mating with donor strain *B. breve* JCM7017-199 pBbrU (JCM7017-199 pBbrU) expressing the BbrUII.M (RM-II) methylase, as compared to the unmethylated donor *B. breve* JCM7017-199 (JCM7017-199). Conjugation efficiency is calculated as number of transconjugants/ number of recipient cells, based on triplicate solid plate mating with recipient strain *B. breve* UCC2003.

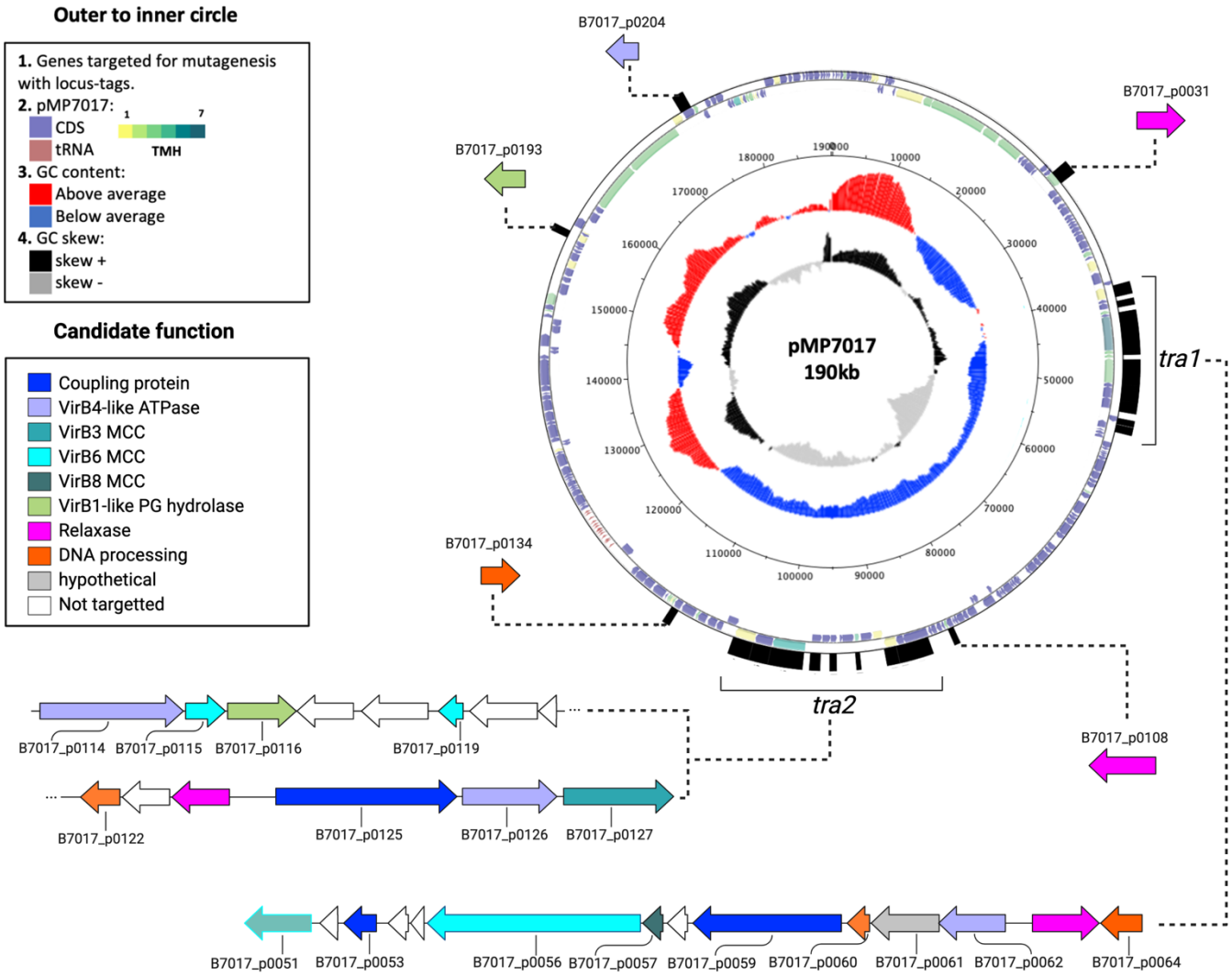


Figure 2. Genome map of pMP7017 indicating genes targeted for mutagenesis

Genetic map of pM7017 showing the organization of the 24 genes selected for mutagenesis, labelled with their locus-tag and are coloured by predicted function as indicated in the legend above. Genes within the pMP7017 candidate conjugation clusters, *tra1* and *tra2*, are indicated.

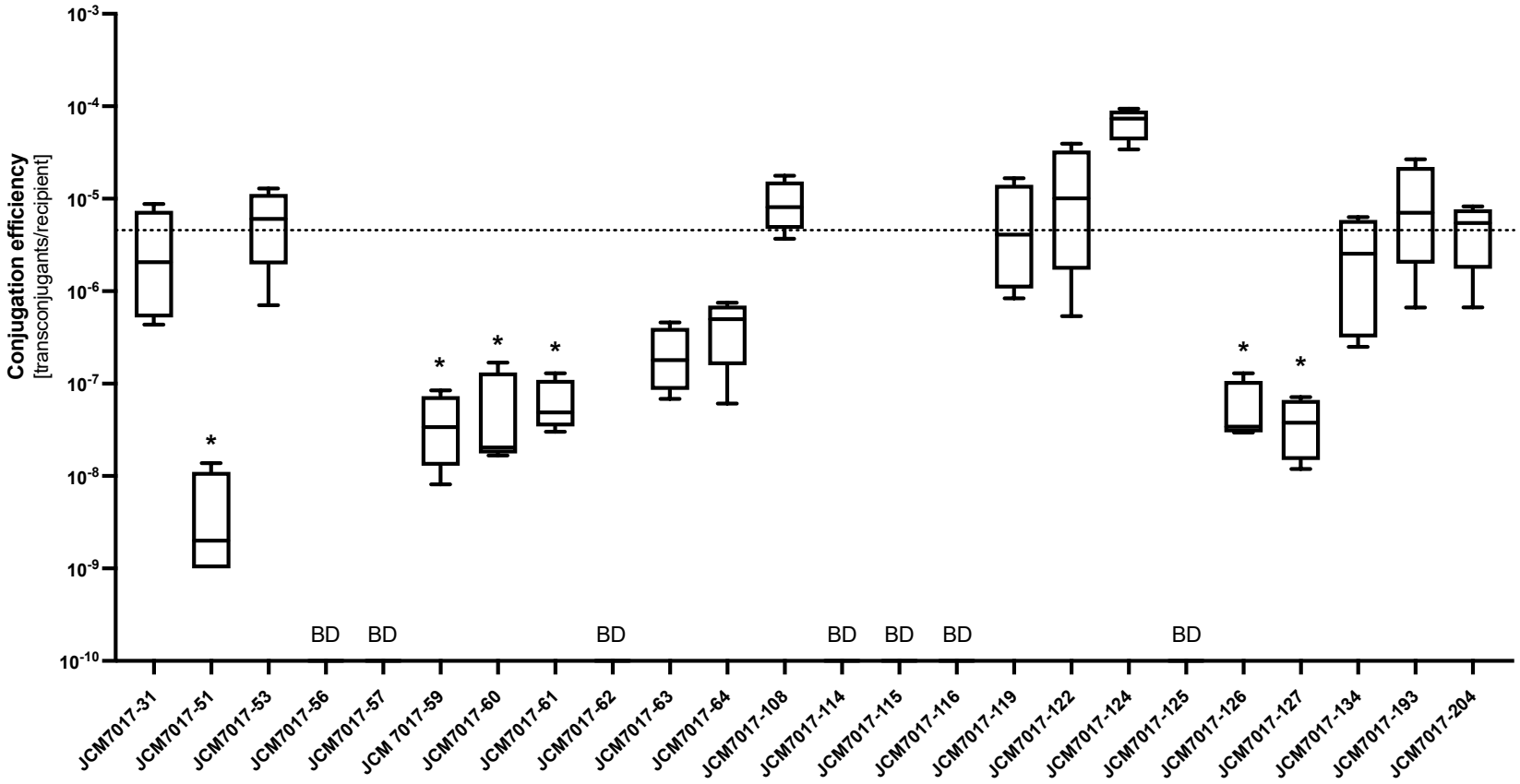


Figure 3. Mutational analysis of the pMP7017 candidate conjugation-associated genes

Conjugation efficiencies after site-directed mutagenesis of individual genes within pMP7017, mating's were conducted using the recombinant donor strain *B. breve* JCM7017 with the indicated mutation and recipient strain, *B. breve* UCC2003, in all cases. Efficiencies presented are based on quadruplicate mating

experiments. The horizontal dotted line represents the mean conjugation efficiency obtained with reference donor strain *B. breve* JCM7017-199. A P-value < 0.05 obtained in the Mann-Whitney U two tailed non-parametric test was considered significant and is represented in the plot with a single asterisk "*". Assays producing no transconjugants within the limit of detection are indicated with "BD" for Below Detection.

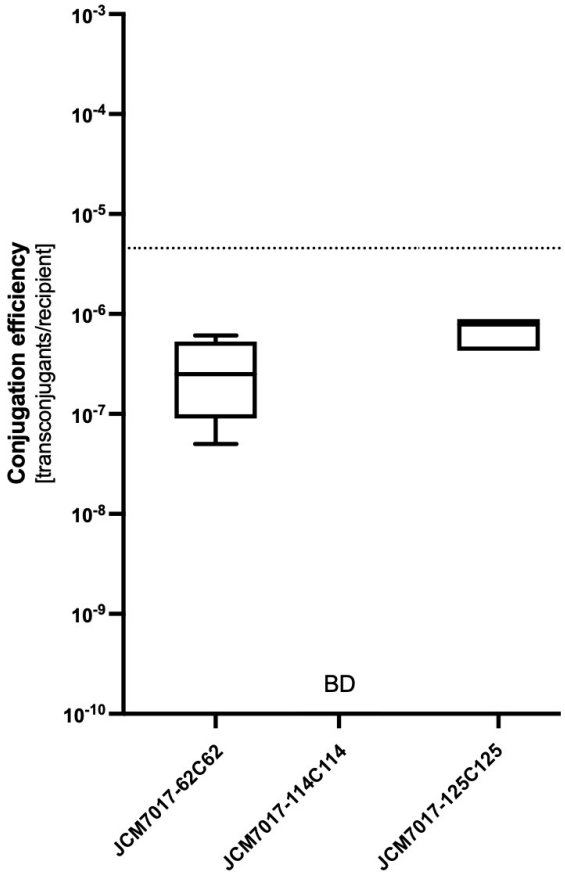


Figure 4. Genetic complementation of essential conjugation-associated ATPase-specifying gene mutations in pMP7017

Conjugal mating's were conducted as described previously, data based on quadruplicate mating experiments. Dotted line is the mean of conjugation efficiencies with donor, reference strain *B. breve* JCM7017-199. BD = below detection.

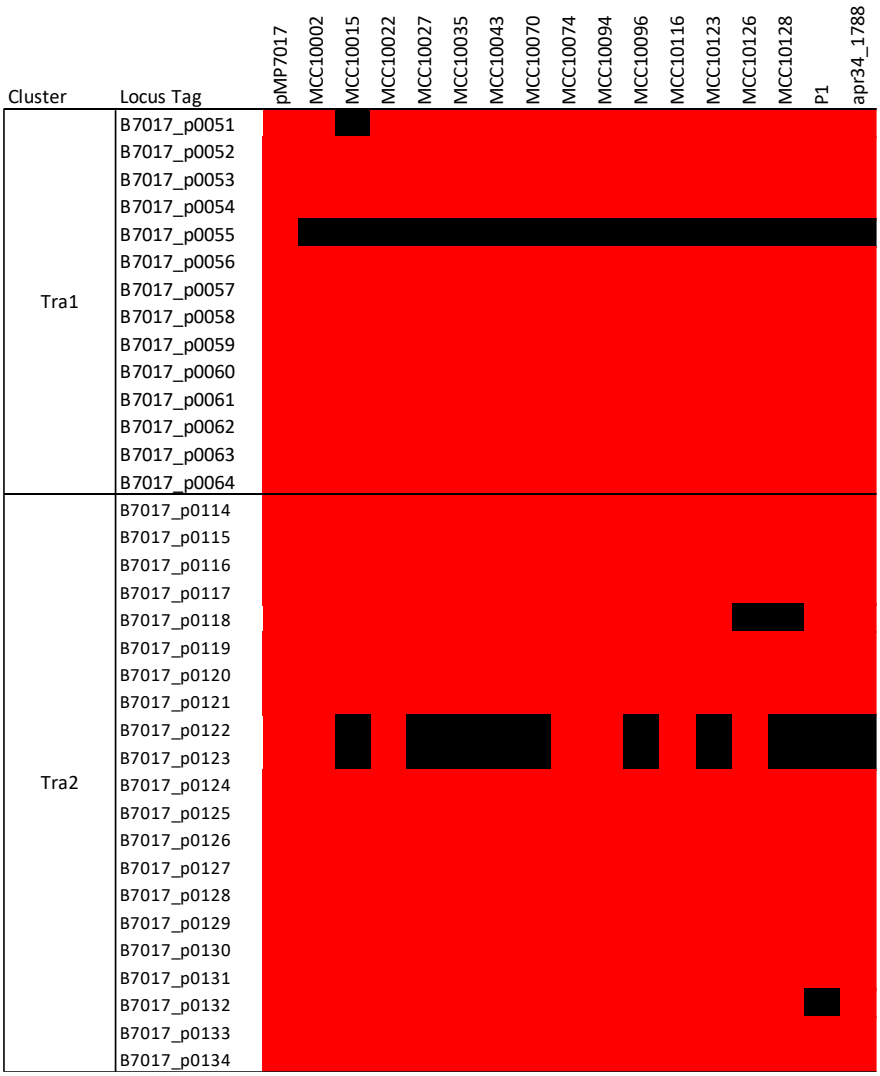


Figure 5. Presence/absence of conjugation-associated gene clusters of pMP7017-related megaplasmids

Gene locus tags are based on those of pMP7017 and gene clusters are indicated based on the assigned *tra* loci (this study). Red and black signify gene presence or absence, respectively.

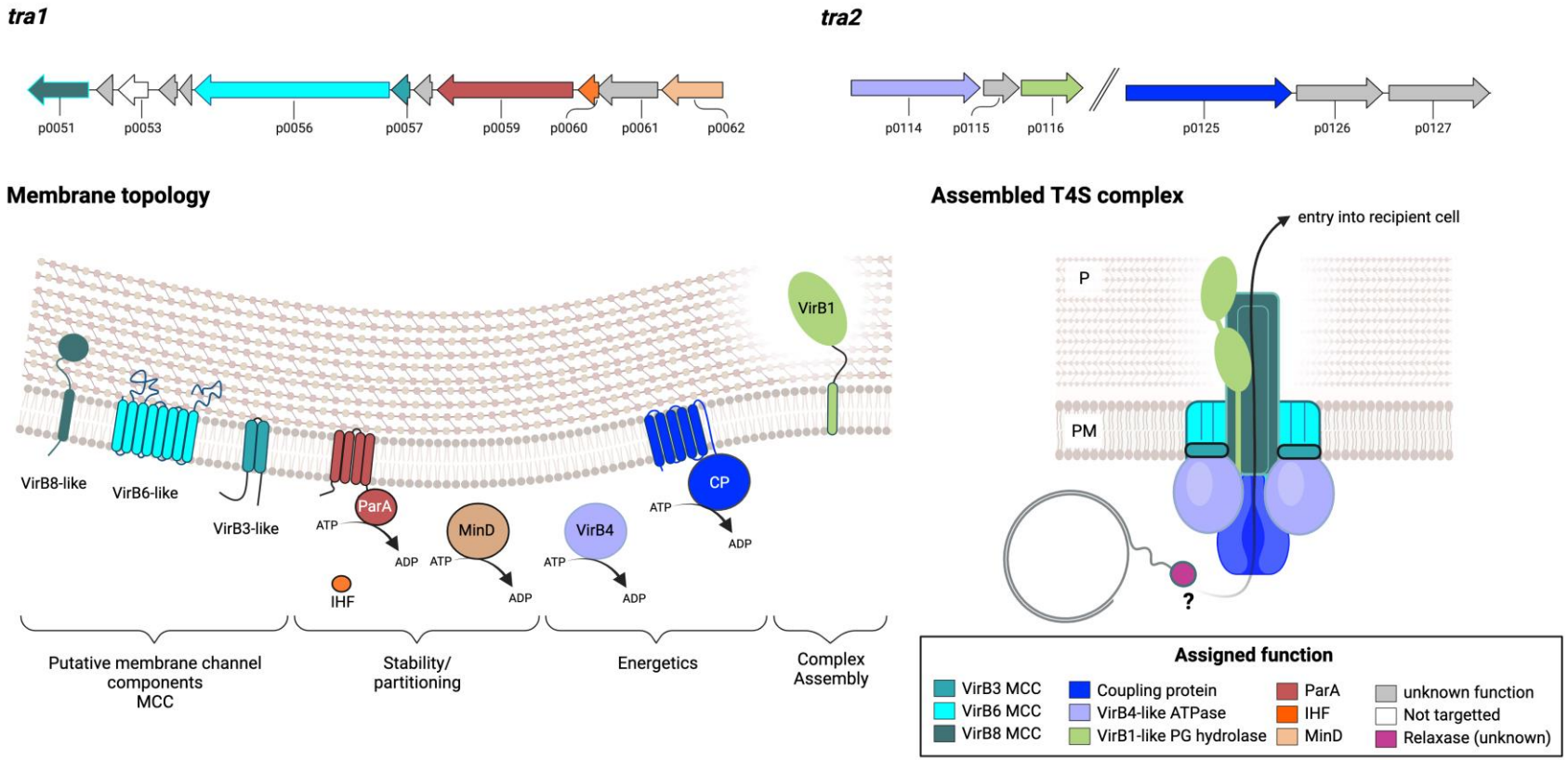


Figure 6. Proposed functional model of the pMP7017 conjugal system

Schematic representation of the putative role which gene products from the pMP7017 conjugation-associated clusters may play during the transfer of ssDNA during conjugation (based on the mutational analysis and predicted functional analysis conducted in this work, however, it should be emphasized that the function of these proteins requires experimental validation). Colour coding is consistent with the legend throughout.

Table 1. Functional analysis of pMP7017-encoded candidate conjugal genes

Putative function	Locus-tag	Current annotation	Length (aa)	TMDs	Software	Accession	Description	Confidence (%)	Probability (%)	E-value
ATPase	B7017_p0053	ParA	217	0	Phyre2	c2bekB	Segregation protein	100	-	-
					HHpred	4DZZ_A	ParF	-	99.93	9.70^{-23}
					Pfam	PF01656	CobQ/MinD/ParA	-	-	3.00^{-14}
	B7017_p0059	AAA-like domain protein	873	3	Phyre2	c4d2iB	HerA	99.3	-	-
					HHpred	7OVB_A	IcmO (DotL)	-	99.96	1.50^{-25}
	B7017_p0062	ParA-type ATPase	384	0	Phyre2	c2bekB	Segregation protein	100	-	-
					HHpred	4V03_B	site-determining protein	-	99.91	7.90^{-22}
					Pfam	PF01656	CobQ/MinD/ParA	-	-	1.50^{-04}
	B7017_p0114	FtsK/SpoIIIE family protein	885	0	Phyre2	c7o41B	Trwk	100	-	-
					HHpred	7O43_B	TrwK	-	100	4.20^{-49}
					HHpred	PF03135	CagE_TrbE_VirB	-	-	1.10^{-09}
					Pfam	PF12846	AAA_10	-	-	1.80^{-09}
	B7017_p0125	FtsK/SpoIIIE family protein	1140	6	Phyre2	c2iusB	FtsK	100	-	-
					HHpred	2IUU_D	FtsK	-	99.92	2.10^{-23}
					HHpred	6O1W_B	TcpA	-	99.92	1.60^{-22}
					Pfam	PF01580	FtsK/SpoIIIE	-	-	9.70^{-09}
	B7017_p0126	FtsK/SpoIIIE domain protein	617	0	Phyre2	c2iuuE	FtsK	100	-	-
					HHpred	2IUU_D	FtsK	-	99.96	3.20^{-27}
					HHpred	6O1W_B	TcpA	-	99.94	8.20^{-24}
					Pfam	PF01580	FtsK/SpoIIIE	-	-	1.10^{-02}
	B7017_p0204	AAA ATPase	388	0	Phyre2	c1w5sB	orc2	99.9	-	-
					HHpred	2QEN_A	WALKER-TYPE ATPASE	-	99.96	6.10^{-26}
					Pfam	PF13191	AAA_16	-	-	3.50^{-11}
PG hydrolase	B7017_p0116	hypothetical protein	415	1	Phyre2	c3ct5A	morphogenesis protein	100	-	-
					HHpred	2K3A_A	CHAP	-	98.56	1.20^{-06}
					Pfam	PF18013	Phage tail lysozyme	-	-	1.20^{-28}
					Pfam	PF05257	CHAP	-	-	7.80^{-10}

	B7017_p0193	hypothetical protein	248	1	Phyre2	c5t1qB	n-acetylmuramoyl-l-alanine amidase	100	-	-
					HHpred	2K3A_A	CHAP	-	99.51	5.50^{-12}
					Pfam	PF05257	CHAP	-	-	1.60^{-20}
DNA processing	B7017_p0031	FtsK solute binding protein	473	0	Phyre2	c3w0fA	endonuclease 8-like	88.3	-	-
					HHpred	2JRP_A	Putative cytoplasmic protein	-	97.85	2.20^{-05}
					Pfam	PF08378	NERD	-	-	6.90^{-10}
					Pfam	PF13240	zinc-ribbon domain	-	-	6.60^{-08}
	B7017_p0064	hypothetical protein	259	0	HHpred	5UBF_B	RctB replication initiator protein	-	98.73	4.80^{-06}
	B7017_p0108	exonuclease	198	0	Phyre2	d2guia1	3'-5' exonuclease	100	-	-
					HHpred	5M1S_D	DNA polymerase III subunit epsilon	-	99.91	2.00^{-21}
					Pfam	PF00929	Exonuclease	-	-	6.20^{-32}
	B7017_p0124	RepA protein	364	0	Phyre2	c2nraC_	Pi protein	96.1	-	-
					HHpred	5UBF_B	RctB	-	99.72	3.10^{-15}
					Pfam	PF04796	RepA_C	-	-	1.70^{-16}
Transcription regulation	B7017_p0060	hypothetical protein	145	0	Phyre2	c7jrph_	cytochrome bc1	52.7	-	-
					HHpred	6TOB_A	Integration host factor MIHF	-	99.32	1.30^{-11}
	B7017_p0122	helix-turn-helix DNA binding protein	228	0	Phyre2	c4j2nA_	gp37	78.7	-	-
					HHpred	4P0Z_A	TraN pIP501	-	97.91	2.80^{-05}
					Pfam	PF12728	HTH_17	-	-	6.70^{-04}
VirB6-like	B7017_p0056	hypothetical protein	1252	8	Phyre2	c7o3vJ_	TrwI protein	85.8	-	-
					HHpred	7O3V_I	TrwI protein	-	96.47	4.40^{-02}
	B7017_p0119	hypothetical protein	153	4	Phyre2	c7q21B_	cytochrome bc1	33.2	-	-
cell division	B7017_p0134	DivIVA domain protein	262	0	HHpred	4UG1_B	CELL CYCLE PROTEIN GPSB	-	98.93	7.50^{-09}
					Pfam	PF05103	DivIVA protein	-	-	5.00^{-03}
					Pfam	PF04136	Sec34-like family	-	-	7.70^{-03}
Unknown	B7017_p0051		1252	1	Phyre2	c3ub1C	orf13-like protein	47.2	-	-

	hypothetical protein			HHpred	3UB1	ORF13-like, NTF2-LIKE	-	95.57	0.64
B7017_p0056	hypothetical protein	1252	8	Phyre2	c7o3vJ_	TrwI protein	85.8	-	-
				HHpred	7O3V_I	TrwI protein	-	96.47	4.40^{-02}
B7017_p0057	hypothetical protein	134	2	HHpred	7KDP_B	Envelope glycoprotein B	-	73.05	1.00^{+01}
				HHpred	PF12666	PrgI	-	93.81	1.30^{+00}
				Pfam	PF09972	DUF2207	-	-	6.30^{-03}
B7017_p0061	hypothetical protein	399	0	Phyre2	c3dnxA	spo1766	66.3	-	-
B7017_p0063	hypothetical protein	201	0	Phyre2	d1zj8a1	Duplicated SiR/NiR-like domains 1 and 3	30	-	-
				Pfam	PF01580	DUF3141			1.50^{-01}
B7017_p0115	hypothetical protein	247	0	Phyre2	c4rlrA	cytochrome c, 1 heme-binding site	33.8	-	-
B7017_p0127	hypothetical protein	711	1	HHpred	4Q4G_X	RipA	-	95.55	5.30^{+00}
				Pfam	PF06864	PilO	-	-	3.20^{-03}

Predictions are based on amino-acid and protein sequence comparison using BLASTP and Pfam v32.0, TMHMM for the prediction of transmembrane helices in proteins, Phyre2 and HHpred were used to assign function based on structure prediction.

Table 2. Co-mobilisation assay

Transconjugant	Intergenic Region	CFU/ml	% CFU/ml
B. breve UCC2003-199-oriT1	B7017_p0124 - B7017_p0125	1.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT2	B7017_p0144 - B7017_p0146	0.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT3	B7017_p0062 - B7017_p0063	2.06+02	2.9
B. breve UCC2003 pMP7017-199 pBbrU_oriT4	B7017_p0128 - B7017_p0129	1.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT5	B7017_p0076 - B7017_p0077	1.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT6	B7017_p0066 - B7017_p0077	0.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT7	B7017_p0079 - B7017_p0080	1.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT8	B7017_p0113 - B7017_p0114	0.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT9	B7017_p0046 - B7017_p0047	0.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT10	B7017_p0033 - B7017_p0034	0.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT11	B7017_p0048 - B7017_p0049	0.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT12	B7017_p0130 - B7017_p0131	1.00	0.0
B. breve UCC2003 pMP7017-199 pBbrU_oriT13	B7017_p0197 - B7017_p0198	1.00	0.0
B. breve UCC2003 pMP7017-199	n/a	7.00+03	100

Supplementary Table S1. Strains and plasmids used in this study

Strain or plasmid	Relevant characteristics	Reference or Source
Strains		
Escherichia coli strains		
EC101	Cloning host, repA$^+$ (Km^r)	[92]
EC101 pORI19	EC101 harbouring pORI19 (Km^r)	[82]
EC101 pORI19-31	EC101 harbouring pORI19_31	this study
EC101 pORI19-tet-31	EC101 harbouring pORI19_tet_31	this study
EC101 pORI19-51	EC101 harbouring pORI19_51	this study
EC101 pORI19-tet-51	EC101 harbouring pORI19_tet_51	this study
EC101 pORI19-53	EC101 harbouring pORI19-53	this study
EC101 pORI19-tet-53	EC101 harbouring pORI19-tet-53	this study
EC101 pORI19-56	EC101 harbouring pORI19-56	this study
EC101 pORI19-tet-56	EC101 harbouring pORI19-tet-56	this study
EC101 pORI19-57	EC101 harbouring pORI19-57	this study
EC101 pORI19-tet-57	EC101 harbouring pORI19-tet-57	this study
EC101 pORI19-59	EC101 harbouring pORI19-59	this study
EC101 pORI19-tet-59	EC101 harbouring pORI19-tet-59	this study
EC101 pORI19-60	EC101 harbouring pORI19-60	this study
EC101 pORI19-tet-60	EC101 harbouring pORI19-tet-60	this study
EC101 pORI19-61	EC101 harbouring pORI19-61	this study
EC101 pORI19-tet-61	EC101 harbouring pORI19-tet-61	this study
EC101 pORI19-62	EC101 harbouring pORI19-62	this study
EC101 pORI19-tet-62	EC101 harbouring pORI19-tet-62	this study
EC101 pORI19-63	EC101 harbouring pORI19-63	this study
EC101 pORI19-tet-63	EC101 harbouring pORI19-tet-63	this study
EC101 pORI19-64	EC101 harbouring pORI19-64	this study
EC101 pORI19-tet-64	EC101 harbouring pORI19-tet-64	this study
EC101 pORI19-108	EC101 harbouring pORI19-108	this study
EC101 pORI19-tet-108	EC101 harbouring pORI19-tet-108	this study
EC101 pORI19-114	EC101 harbouring pORI19-114	this study

EC101 pORI19-tet-114	EC101 harbouring pORI19-tet-114	this study
EC101 pORI19-115	EC101 harbouring pORI19-115	this study
EC101 pORI19-tet-115	EC101 harbouring pORI19-tet-115	this study
EC101 pORI19-116	EC101 harbouring pORI19-116	this study
EC101 pORI19-tet-116	EC101 harbouring pORI19-tet-116	this study
EC101 pORI19-119	EC101 harbouring pORI19-119	this study
EC101 pORI19-tet-119	EC101 harbouring pORI19-tet-119	this study
EC101 pORI19-122	EC101 harbouring pORI19-122	this study
EC101 pORI19-tet-122	EC101 harbouring pORI19-tet-122	this study
EC101 pORI19-124	EC101 harbouring pORI19-124	[24]
EC101 pORI19-tet-124	EC101 harbouring pORI19-tet-124	[24]
EC101 pORI19-125	EC101 harbouring pORI19-125	this study
EC101 pORI19-tet-125	EC101 harbouring pORI19-tet-125	this study
EC101 pORI19-126	EC101 harbouring pORI19-126	this study
EC101 pORI19-tet-126	EC101 harbouring pORI19-tet-126	this study
EC101 pORI19-127	EC101 harbouring pORI19-127	this study
EC101 pORI19-tet-127	EC101 harbouring pORI19-tet-127	this study
EC101 pORI19-134	EC101 harbouring pORI19-134	this study
EC101 pORI19-tet-134	EC101 harbouring pORI19-tet-134	this study
EC101 pORI19-193	EC101 harbouring pORI19-193	this study
EC101 pORI19-tet-193	EC101 harbouring pORI19-tet-193	this study
EC101 pORI19-204	EC101 harbouring pORI19-204	this study
EC101 pORI19-tet-204	EC101 harbouring pORI19-tet-204	this study
XL1 blue	(Tet^r, Amp^r)	Stratagene
XL1 blue pBbrU	XL1 blue harbouring pBbrU (Tet^r, Amp^r, Cm^r)	this study
Bifidobacterium breve strains		
JCM7017-199	JCM7017 harbouring pMP7017_199, (Tet^r)	[23]
UCC2003	Human intestinal isolate, plasmid-free	APC
UCC2003-199	UCC2003 harbouring pMP7017-199, (Tet^r)	this study
JCM7017-31	JCM7017 harbouring pORI19-tet-31 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-51	JCM7017 harbouring pORI19-tet-51 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-53	JCM7017 harbouring pORI19-tet-53 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-56	JCM7017 harbouring pORI19-tet-56 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-57	JCM7017 harbouring pORI19-tet-57 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study

JCM7017-59	JCM7017 harbouring pORI19-tet-59 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-60	JCM7017 harbouring pORI19-tet-60 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-61	JCM7017 harbouring pORI19-tet-61 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-62	JCM7017 harbouring pORI19-tet-62 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-63	JCM7017 harbouring pORI19-tet-63 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-64	JCM7017 harbouring pORI19-tet-64 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-108	JCM7017 harbouring pORI19-tet-108 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-114	JCM7017 harbouring pORI19-tet-114 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-115	JCM7017 harbouring pORI19-tet-115 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-116	JCM7017 harbouring pORI19-tet-116 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-119	JCM7017 harbouring pORI19-tet-119 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-122	JCM7017 harbouring pORI19-tet-122 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-124	JCM7017 harbouring pORI19-tet-124 insertion mutant of pMP7017 (Tet^r, Cm^r)	[24]
JCM7017-125	JCM7017 harbouring pORI19-tet-125 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-126	JCM7017 harbouring pORI19-tet-126 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-127	JCM7017 harbouring pORI19-tet-127 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-134	JCM7017 harbouring pORI19-tet-134 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-193	JCM7017 harbouring pORI19-tet-193 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-204	JCM7017 harbouring pORI19-tet-204 insertion mutant of pMP7017 (Tet^r, Cm^r)	this study
JCM7017-62C62	JCM7017-62 harbouring pBbrU_62C (Tet^r)	this study
JCM7017-114C114	JCM7017-114 harbouring pBbrU_114C (Tet^r)	this study
JCM7017-125C125	JCM7017-125 harbouring pBbrU_125C (Tet^r)	this study
JCM7017-199 pBbrU_oriT1	JCM7017-199 harbouring pBbrU_oriT1 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT2	JCM7017-199 harbouring pBbrU_oriT2 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT3	JCM7017-199 harbouring pBbrU_oriT3 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT4	JCM7017-199 harbouring pBbrU_oriT4 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT5	JCM7017-199 harbouring pBbrU_oriT5 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT6	JCM7017-199 harbouring pBbrU_oriT6 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT7	JCM7017-199 harbouring pBbrU_oriT7 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT8	JCM7017-199 harbouring pBbrU_oriT8 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT9	JCM7017-199 harbouring pBbrU_oriT9 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT10	JCM7017-199 harbouring pBbrU_oriT10 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT11	JCM7017-199 harbouring pBbrU_oriT11 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT12	JCM7017-199 harbouring pBbrU_oriT12 (Tet^r, Cm^r)	this study
JCM7017-199 pBbrU_oriT13	JCM7017-199 harbouring pBbrU_oriT13 (Tet^r, Cm^r)	this study

Plasmids

pORI19	Em^r, repA-, Ori+, cloning vector	[92]
pBC1.2	pBC1-pSC101, (Cm^r, Amp^r)	[93]
pAM5	pBC1-puC19, (Tet^r)	[93]
pBbrU	pBC1.2 containing BbrUII.M from UCC2003, (Cm^r, Amp^r)	this work
pORI19-31	Internal 522 bp fragment of B7017_p0031 in pORI19	
pORI-tet-31	pORI19-31 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-51	Internal 541 bp fragment of B7017_p0051 in pORI19	
pORI-tet-51	pORI19-51 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-53	Internal 462 bp fragment of B7017_p0053 in pORI19	this study
pORI19-tet-53	pORI19-53 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-56	Internal 506 bp fragment of B7017_p0056 in pORI19	this study
pORI19-tet-56	pORI19-56 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-57	Internal 376 bp fragment of B7017_p0057 in pORI19	this study
pORI19-tet-57	pORI19-57 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-59	Internal 536 bp fragment of B7017_p0059 in pORI19	this study
pORI19-tet-59	pORI19-59 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-60	Internal 419 bp fragment of B7017_p0060 in pORI19	this study
pORI19-tet-60	pORI19-60 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-61	Internal 534 bp fragment of B7017_p0061 in pORI19	this study
pORI19-tet-61	pORI19-61 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-62	Internal 515 bp fragment of B7017_p0062 in pORI19	this study
pORI19-tet-62	pORI19-62 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-63	Internal 513 bp fragment of B7017_p0063 in pORI19	this study
pORI19-tet-63	pORI19-63 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-64	Internal 500 bp fragment of B7017_p0064 in pORI19	this study
pORI19-tet-64	pORI19-64 containing the Tet^r cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-108	Internal 510 bp fragment of B7017_p0108 in pORI19	this study
pORI19-tet-108	pORI19-108 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-114	Internal 577 bp fragment of B7017_p0114 in pORI19	this study
pORI19-tet-114	pORI19-114 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-115	Internal 490 bp fragment of B7017_p0115 in pORI19	this study
pORI19-tet-115	pORI19-115 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study

pORI19-116	Internal 474 bp fragment of B7017_p0116 in pORI19	this study
pORI19-tet-116	pORI19-116 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-119	Internal 450 bp fragment of B7017_p0119 in pORI19	this study
pORI19-tet-119	pORI19-119 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-122	Internal 510 bp fragment of B7017_p0122 in pORI19	this study
pORI19-tet-122	pORI19-122 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-124	Internal 528 bp fragment of B7017_p0124 in pORI19	[24]
pORI19-tet-124	pORI19-124 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	[24]
pORI19-125	Internal 531 bp fragment of B7017_p0125 in pORI19	this study
pORI19-tet-125	pORI19-125 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-126	Internal 587 bp fragment of B7017_p0126 in pORI19	this study
pORI19-tet-126	pORI19-126 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-127	Internal 508 bp fragment of B7017_p0127 in pORI19	this study
pORI19-tet-127	pORI19-127 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-134	Internal 507 bp fragment of B7017_p0134 in pORI19	this study
pORI19-tet-134	pORI19-134 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-193	Internal 456 bp fragment of B7017_p0193 in pORI19	this study
pORI19-tet-193	pORI19-193 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pORI19-204	Internal 485 bp fragment of B7017_p0204 in pORI19	this study
pORI19-tet-204	pORI19-204 containing the *tet*(W) cassette from pAM5 (Em^r, Tet^r)	this study
pBbrU_62C	pBbrU containing B7017_p0062 and native promoter (Cm^r, Amp^r)	this study
pBbrU_114C	pBbrU containing B7017_p0114 and native promoter (Cm^r, Amp^r)	this study
pBbrU_125C	pBbrU containing B7017_p0125 and native promoter (Cm^r, Amp^r)	this study
pBC_oriT1	pBbrU containing the intergenic region B7017_p0124- B7017_p0125 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT2	pBbrU containing the intergenic region B7017_p0144- B7017_p0146 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT3	pBbrU containing the intergenic region B7017_p062- B7017_p063 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT4	pBbrU containing the intergenic region B7017_p0128-129 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT5	pBbrU containing the intergenic region B7017_p076- B7017_p077 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT6	pBbrU containing the intergenic region B7017_p066- B7017_p067 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT7	pBbrU containing the intergenic region B7017_p079- B7017_p080 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT8	pBbrU containing the intergenic region B7017_p0113- B7017_p0114 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT9	pBbrU containing the intergenic region B7017_p046- B7017_p047 from pMP7017 (Cm^r, Amp^r)	this study

pBC_oriT10	pBbrU containing the intergenic region B7017_p033- B7017_p034 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT11	pBbrU containing the intergenic region B7017_p048- B7017_p049 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT12	pBbrU containing the intergenic region B7017_p0130- B7017_p0131 from pMP7017 (Cm^r, Amp^r)	this study
pBC_oriT13	pBbrU containing the intergenic region B7017_p0197- B7017_p0198 from pMP7017 (Cm^r, Amp^r)	this study

Supplementary Table S2. Oligonucleotide primers used in the study

Purpose	Primer	Sequence (5' - 3')
B. breve UCC2003 transconjugant confirmation	Bbr_1118F	gcacttccaaggcttcgtcg
	Bbr_1118R	cgagagaccatccgacgctt
B. breve JCM7017 transconjugant confirmation	0735F	cagaacatgccggtcgagga
	0735R	gcattctcaggacgaagcgc
pMP7017 transconjugant confirmation	pMP_20F	atggactcaggccgcatcat
	pMP_20R	agccagcttctcattcgcct
B. longum subsp. *longum* NCIMB8809 transconjugant confirmation	8809_1648F	aggaactcgcacgccaggcc
	8809_1648R	tgacctcgatgagactgccggcg
Cloning of…		
Bbr_01121 (BbrUII.M) in pBC1.2	UCCPst1MF	aattaa**GATATC**ggcaacgccgagaatatc
	UCCPst1MR	catcacTCTAGActcgatcagctggcttttgtc
Bbl.M607 EcoRII methylase in pNZ8048	8809EcoF	tacgca**CTGCAG**gacgccaccgaatgttca
	8809EcoR	gagcgg**CCATGG**gatcctgaggaattcgtc
522 bp fragment of p0031 in pORI19	0031H3F	aattaaAAGCTTaatcatcacgctcgccatcg
	0031XbaR	catcacTCTAGAtgcactgccgtatcgtcgaa
541 bp fragment of p0051 in pORI19	51mut_H3	aattaaAAGCTTtggaaatggaagccgcaccg
	51mut_Xba	gttgacTCTAGAtggttggacgggaataagggcc
462 bp fragment of p0053 in pORI19	53mut_H3	aattaaAAGCTTcccgcatggctcagatcgc
	53mut_Xba	catcacTCTAGAacctgatacctccccaaccacg
506 bp fragment of p0056 in pORI19	56H3F	aattaaAAGCTTtgccgaatataggacgccg
	56XbaR	catcacTCTAGAacgtgttgaacctctgccga
376 bp fragment of p0057 in pORI19	57mut_H3	aattaaAAGCTTtcaagctggtgaggtcttcc
	57mut_Xba	catcacTCTAGAcggaaacagttgcatcctcggg
536 bp fragment of p0059 in pORI19	59XbaR	gtagtgTCTAGAagcctggtgatgctgttcgt
	59H3F	aattaaAAGCTTtccaagtatccggagcgtcc
419 bp fragment of p0060 in pORI19	60mut_H3	aattaaAAGCTTtctcatctcctgtctcgtag
	60mut_Xba	catcacTCTAGAcaacttcgcatccaaccagc
534 bp fragment of p0061 in pORI19	61mut_H3	aattaaAAGCTTgattgcggaaattgggcgcg
	61mut_Xba	catcacTCTAGAcacgtcggactcagcctcgc
515 bp fragment of p0062 in pORI19	62H3F	aattaaAAGCTTggaaccgtgtccttgatgcc
	62XbaR	catcacTCTAGAggcaagaccgtatcctccaga
513 bp fragment of p0063 in pORI19	63mut_H3	aattaaAAGCTTgcgaaccggagcgttgaacc
	63mut_Xba	catcacTCTAGAgcccaagtcaacaccgtcgc
500 bp fragment of p0064 in pORI19	F64_sphI	catcaaGCATGCcctgccaatgaggaaccag
	F64_H3	aattaaAAGCTTggaggatgacaggtagatc
510 bp fragment of p0108 in pORI19	108_H3_F	aattaaAAGCTTccacgacgcataccaccactg
	108_Xba_R	catcacTCTAGActcctgctgcgcgtcggaa
577 bp fragment of p0114 in pORI19	114H3F	aattaaAAGCTTcgaggatgcagtcaccgcta
	114XbaR	catcacTCTAGAcgtctcggccgcttatggaa
490 bp fragment of p0115 in pORI19	115_H3_F	aattaaAAGCTTacgccgagatcagaagaagacga
	115_Xba_R	catcacTCTAGAagtcttccgccgccatccac
474 bp fragment of p0116 in pORI19	116H3F	aattaaAAGCTTagcagcagcacggtcaac
	116XbaR	catgtgTCTAGAatattcgttccagtgtcc
450 bp fragment of p0119 in pORI19	pMP119_H3_R	aattaaAAGCTTagtctttatgcgttgcg
	pMP119_Xba_F	catcacTCTAGAtgcttactctgcttggc
510 bp fragment of p0122 in pORI19	122mut_H3	aattaaAAGCTTgcccatagaacctgccgagc
	122mut_Xba	catcacTCTAGAccctccaccacaagcagccc
	repAmut_R	aattaaAAGCTTgcatcggtgcgaggataata

528 bp fragment of p0124 in pORI19	repAmut_F	catcacTCTAGAcagcttccaggaattctgcg
531 bp fragment of p0125 in pORI19	125H3F	aattaaAAGCTTcgtcgcgtctactggtggat
	125XbaR	catcacTCTAGAgcgaggcttgccatctgaac
587 bp fragment of p0126 in pORI19	126H3F	aattaaAAGCTTaaccagccgcaataccaagg
	126XbaR	catcacTCTAGAgccttgtcctcatcgtcgca
508 bp fragment of p0127 in pORI19	127H3F	aattaaAAGCTTgcataccacgcctacaacgc
	127XbaR	cactacTCTAGAcgtccgtgaccttgccgata
507 bp fragment of p0134 in pORI19	134H3F	aatattAAGCTTtacgacacgcagcaggttga
	134XbaR	cttcacTCTAGAtgcagcatactcatccgcct
456 bp fragment of p0193 in pORI19	pMP193H3_R	aattaaAAGCTTcagtccgagctcatgccg
	pMP193Xba_F	aattaaAAGCTTgcgcataaggcatcgaag
485 bp fragment of p0204 in pORI19	pMP204_H3_R	aattaaAAGCTTatcaattgcaccaccgtt
	pMP204Xba_F	catcacTCTAGAcactgctcctactcgcca
tet(W) amplification	tetW_F	tcagctgtcgacatgctcatgtacggtaag
	tetW_R	gcgacggtcgaccattaccttctgaaacat
Confirmation of site-specific homologous recombination	Conf31	cgtcattggattcctggcga
	51conf	ggcaccccagtcgatgattttc
	Conf53	tcccgctggtagatgacgtt
	Conf56	ctgctgctgagtccactg
	Conf57	cgcataccggggctgatgag
	Conf59	tgggtgagtacggtcttg
	Conf60	gccgattccgtcacccatgc
	Conf61	cactttctgcagcacccggt
	Conf62	atagttctgcaatccttc
	Conf63	gggcggcgcgaaaagttcag
	F64_conf	gacatcgacagtccggaatc
	108conf	aaccatgcgagcctgttgttc
	Conf114	tgagattcactggtcggctaca
	Conf115n	acggttctaaccagccgatac
	Conf116	gtgagcacgcctgatac
	Conf119n	gtatcccgccacaaacactc
	Conf122	acgacacacctcattcaaccga
	Conf124	gttctgcttgaccggtcgaatc
	Conf125	ttcttcctgttctccgcgct
	Conf126	gcgccgatggacatggattc
	Conf127	cctgatgaggaagcacagg
	Conf134	ccaccaatgttcgatgtc
	Conf193n	cgtgtcatcaacgtgacctg
	204conf	cttcggcgtgcctactgactg
p0062 in pBbrU	pMP62compF	catcacATCGATgggtagcgtaatatcaag
	pMP62compR	gagtccCCTCGAGGgtgcgatggaaaggtctaaca
p0114 in pBbrU	pMP114compF	catcac**ATCGAT**aaacatgaaatgactcttta
	pMP114compR	gagtcc**CCTCGAGG**actctcgccgaacatatctt
p0125 in pBbrU	pMP125compF	catcac**ATCGAT**ctgcgacttacagggatgcaa
	pMP125compR	gagtcc**CCTCGAGG**ctaccagttgacttcttcgat
Intergenic region 144-146 into pBbrU to generate construct pBC_oriT2	7017ori1_R	gtcaacCCTCGAGGacgattcagactagcttttg
	7017ori2_F	tgagtaATCGATgatgcaacctcctaggagatc
	7017ori2_R	gcatccCCTCGAGGtaggccactaatcgtcccgc
Intergenic region 62-63 into pBbrU to generate construct pBC_oriT3	7017ori3_F	tctcacATCGATgcttgttccggcgtttatgc
	7017ori3_R	cagtatCCTCGAGGctgacaatcaagactctag
intergenic region 128-129 into pBbrU to generate construct pBC_oriT4	7017ori4_F	catcacATCGATccttgaacgcgacggcctc
	7017ori4_R	gagtccCCTCGAGGcagttggcgcaggcgtctg

intergenic region 76-77 into pBbrU	7017ori5_F	catcacATCGATggtacaggaggaggatttc
to generate construct pBC_oriT5	7017ori5_R	tagaggCCTCGAGGgattccaaccgcacgtgatg
intergenic region 66-67 into pBbrU	7017ori6_F	catcacATCGATgatgcattctccgttttcg
to generate construct pBC_oriT6	7017ori6_R	gtcaacCCTCGAGGgttgtctgaaggtcatggtc
intergenic region 79-80 into pBbrU	7017oriT7_F	catcacATCGATgccgtttcgcttgagcag
to generate construct pBC_oriT7	7017oriT7_R	gtcaacCCTCGAGGgatgaagaacgtcggagcgg
intergenic region 113-114 into pBbrU to generate construct pBC_oriT8	7017ori8_F	gtacacATCGATgctctgcaatccgaagtc
	7017ori8_R	agtagtCCTCGAGGctggccgggctgatgaag
intergenic region 46-47 into pBbrU	7017ori9_F	catcacATCGATgtgtggtttcggtcaatg
to generate construct pBC_oriT9	7017ori9_R	ttcaacCCTCGAGGaaatcggagacgttggc
intergenic region 33-34 into pBbrU	7017ori10_F	tgagtaATCGATcgtgttcgcccaagagtc
to generate construct pBC_oriT10	7017ori10_R	gcagaaCCTCGAGGgaacaggagagagatagc
intergenic region 48-49 into pBbrU	7017ori11_F	gtagtaATCGATgcgttgctgttgatgaag
to generate construct pBC_oriT11	7017ori11_R	gtcaacCCTCGAGGtgaccgcctgagttcgtc
intergenic region 130-131 into pBbrU to generate construct pBC_oriT12	7017ori12_F	tgagtaATCGATgttaatctgagtgcctag
	7017ori12_R	gcatccCCTCGAGGtgtcggttccgctccac
intergenic region 197-198 into pBbrU to generate construct pBC_oriT13	7017ori13_F	tgagtaATCGATcagcacgcggagaatatc
	7017ori13_R	gcatccCCTCGAGGcgtttccgacattgatcg

REFERENCES

[1] **S. C. Leahy, D. G. Higgins, G. F. Fitzgerald, and D. van Sinderen**, 'Getting better with bifidobacteria', *J. Appl. Microbiol.*, vol. 98, no. 6, Art. no. 6, 2005, doi: 10.1111/j.1365-2672.2005.02600.x.

[2] **T. Odamaki *et al.***, 'Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study', *BMC Microbiol.*, vol. 16, no. 1, p. 90, May 2016, doi: 10.1186/s12866-016-0708-5.

[3] **M. Saxelin, S. Tynkkynen, T. Mattila-Sandholm, and W. M. de Vos**, 'Probiotic and other functional microbes: from markets to mechanisms', *Curr. Opin. Biotechnol.*, vol. 16, no. 2, pp. 204–211, Apr. 2005, doi: 10.1016/j.copbio.2005.02.003.

[4] **M. Ventura, D. van Sinderen, G. F. Fitzgerald, and R. Zink**, 'Insights into the taxonomy, genetics and physiology of bifidobacteria', *Antonie Van Leeuwenhoek*, vol. 86, no. 3, Art. no. 3, Oct. 2004, doi: 10.1023/B:ANTO.0000047930.11029.ec.

[5] **Y. Guo *et al.***, 'Prophylactic Effects of Bifidobacterium adolescentis on Anxiety and Depression-Like Phenotypes After Chronic Stress: A Role of the Gut Microbiota-Inflammation Axis', *Front. Behav. Neurosci.*, vol. 13, Jun. 2019, doi: 10.3389/fnbeh.2019.00126.

[6] **J. A. Krumbeck *et al.***, 'Probiotic Bifidobacterium strains and galactooligosaccharides improve intestinal barrier function in obese adults but show no synergism when used together as synbiotics', *Microbiome*, vol. 6, no. 1, p. 121, Jun. 2018, doi: 10.1186/s40168-018-0494-4.

[7] **X. Ling, P. Linglong, D. Weixia, and W. Hong**, 'Protective Effects of Bifidobacterium on Intestinal Barrier Function in LPS-Induced Enterocyte Barrier Injury of Caco-2 Monolayers and in a Rat NEC Model', *PLOS ONE*, vol. 11, no. 8, p. e0161635, Aug. 2016, doi: 10.1371/journal.pone.0161635.

[8] **D. R. Michael *et al.***, 'A randomised controlled study shows supplementation of overweight and obese adults with lactobacilli and bifidobacteria reduces bodyweight and improves well-being', *Sci. Rep.*, vol. 10, no. 1, p. 4183, Mar. 2020, doi: 10.1038/s41598-020-60991-7.

[9] **L. Ruiz, S. Delgado, P. Ruas-Madiedo, B. Sánchez, and A. Margolles**, 'Bifidobacteria and Their Molecular Communication with the Immune System', *Front. Microbiol.*, vol. 8, 2017, doi: 10.3389/fmicb.2017.02345.

[10] **D. M. Linares *et al.***, 'Lactic Acid Bacteria and Bifidobacteria with Potential to Design Natural Biofunctional Health-Promoting Dairy Foods', *Front. Microbiol.*, vol. 8, 2017, doi: 10.3389/fmicb.2017.00846.

[11] **D. Jakubczyk, K. Leszczyńska, and S. Górska**, 'The Effectiveness of Probiotics in the Treatment of Inflammatory Bowel Disease (IBD)—A Critical Review', *Nutrients*, vol. 12, no. 7, p. 1973, Jul. 2020, doi: 10.3390/nu12071973.

[12] **M. J. Saez-Lara, C. Gomez-Llorente, J. Plaza-Diaz, and A. Gil**, 'The Role of Probiotic Lactic Acid Bacteria and Bifidobacteria in the Prevention and Treatment of Inflammatory Bowel Disease and Other Related Diseases: A Systematic Review of Randomized Human Clinical Trials', *BioMed Res. Int.*, vol. 2015, 2015, doi: 10.1155/2015/505878.

[13] **S. Singh *et al.***, 'Anti-inflammatory Bifidobacterium strains prevent dextran sodium sulfate induced colitis and associated gut microbial dysbiosis in mice', *Sci. Rep.*, vol. 10, no. 1, p. 18597, Oct. 2020, doi: 10.1038/s41598-020-75702-5.

[14] **T. M. G. Dronkers, A. C. Ouwehand, and G. T. Rijkers**, 'Global analysis of clinical trials with probiotics', *Heliyon*, vol. 6, no. 7, p. e04467, Jul. 2020, doi: 10.1016/j.heliyon.2020.e04467.

[15] **A. O'Callaghan and D. van Sinderen**, 'Bifidobacteria and Their Role as Members of the Human Gut Microbiota', *Front. Microbiol.*, vol. 7, p. 925, 2016, doi: 10.3389/fmicb.2016.00925.

[16] **F. Bottacini *et al.***, 'Comparative genome and methylome analysis reveals restriction/modification system diversity in the gut commensal Bifidobacterium breve', *Nucleic Acids Res.*, vol. 46, no. 4, pp. 1860–1877, Feb. 2018, doi: 10.1093/nar/gkx1289.

[17] **R. J. Roberts *et al.***, 'A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes', *Nucleic Acids Res.*, vol. 31, no. 7, pp. 1805–1812, Apr. 2003, doi: 10.1093/nar/gkg274.

[18] **K. Mierzejewska, M. Bochtler, and H. Czapinska**, 'On the role of steric clashes in methylation control of restriction endonuclease activity', *Nucleic Acids Res.*, vol. 44, no. 1, pp. 485–495, Jan. 2016, doi: 10.1093/nar/gkv1341.

[19] **C. M. Thomas and K. M. Nielsen**, 'Mechanisms of, and barriers to, horizontal gene transfer between bacteria', *Nat. Rev. Microbiol.*, vol. 3, no. 9, pp. 711–721, Sep. 2005, doi: 10.1038/nrmicro1234.

[20] **D. Bröker, M. Arensk̈otter, A. Legatzki, D. H. Nies, and A. Steinbüchel**, 'Characterization of the 101-Kilobase-Pair Megaplasmid pKB1, Isolated from the Rubber-Degrading Bacterium Gordonia westfalica Kb1', *J. Bacteriol.*, vol. 186, no. 1, pp. 212–225, Jan. 2004, doi: 10.1128/JB.186.1.212-225.2004.

[21] **J. Zheng, D. Peng, L. Ruan, and M. Sun**, 'Evolution and dynamics of megaplasmids with genome sizes larger than 100 kb in the Bacillus cereus group', *BMC Evol. Biol.*, vol. 13, p. 262, Dec. 2013, doi: 10.1186/1471-2148-13-262.

[22] **E. Schwartz, Ed.**, *Microbial Megaplasmids*. Berlin Heidelberg: Springer-Verlag, 2009. doi: 10.1007/978-3-540-85467-8.

[23] **F. Bottacini *et al.***, 'Discovery of a Conjugative Megaplasmid in Bifidobacterium breve', *Appl. Environ. Microbiol.*, vol. 81, no. 1, pp. 166–176, Jan. 2015, doi: 10.1128/AEM.02871-14.

[24] **R. L. Dineen, C. Penno, P. Kelleher, M. J. B. Bourin, M. O'Connell-Motherway, and D. van Sinderen**, 'Molecular analysis of the replication functions of the bifidobacterial conjugative megaplasmid pMP7017', *Microb. Biotechnol.*, May 2021, doi: 10.1111/1751-7915.13810.

[25] **T. Odamaki *et al.***, 'Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan', *Sci. Rep.*, vol. 8, Jan. 2018, doi: 10.1038/s41598-017-18391-x.

[26] **A. J. Sterling, W. J. Snelling, P. J. Naughton, N. G. Ternan, and J. S. G. Dooley**, 'Competent but complex communication: The phenomena of pheromone-responsive plasmids', *PLOS Pathog.*, vol. 16, no. 4, p. e1008310, Apr. 2020, doi: 10.1371/journal.ppat.1008310.

[27] **W. J. J. Meijer *et al.***, 'Multiple Layered Control of the Conjugation Process of the Bacillus subtilis Plasmid pLS20', *Front. Mol. Biosci.*, vol. 8, 2021, doi: 10.3389/fmolb.2021.648468.

[28] **S. A. Revitt-Mills, T. D. Watts, D. Lyras, V. Adams, and J. I. Rood**, 'The ever-expanding tcp conjugation locus of pCW3 from Clostridium perfringens', *Plasmid*, vol. 113, p. 102516, Jan. 2021, doi: 10.1016/j.plasmid.2020.102516.

[29] **C. E. Alvarez-Martinez and P. J. Christie**, 'Biological diversity of prokaryotic type IV secretion systems', *Microbiol. Mol. Biol. Rev. MMBR*, vol. 73, no. 4, pp. 775–808, Dec. 2009, doi: 10.1128/MMBR.00023-09.

[30] **N. Goessweiner-Mohr, K. Arends, W. Keller, and E. Grohmann**, 'Conjugative type IV secretion systems in Gram-positive bacteria', *Plasmid*, vol. 70, no. 3, pp. 289–302, Nov. 2013, doi: 10.1016/j.plasmid.2013.09.005.

[31] **J. Sorokina *et al.***, 'VirB4- and VirD4-Like ATPases, Components of a Putative Type 4C Secretion System in Clostridioides difficile', *J. Bacteriol.*, vol. 203, no. 21, pp. e00359-21, Oct. 2021, doi: 10.1128/JB.00359-21.

[32] **I. Álvarez-Rodríguez, B. Ugarte-Uribe, I. de la Arada, J. L. R. Arrondo, C. Garbisu, and I. Alkorta**, 'Conjugative Coupling Proteins and the Role of Their Domains in Conjugation, Secondary Structure and in vivo Subcellular Location', *Front. Mol. Biosci.*, vol. 7, 2020, doi: 10.3389/fmolb.2020.00185.

[33] **M. Y. Abajy *et al.***, 'A type IV-secretion-like system is required for conjugative DNA transport of broad-host-range plasmid pIP501 in gram-positive bacteria', *J. Bacteriol.*, vol. 189, no. 6, pp. 2487–2496, Mar. 2007, doi: 10.1128/JB.01491-06.

[34] **R. Bantwal *et al.***, 'The peptidoglycan hydrolase TcpG is required for efficient conjugative transfer of pCW3 in Clostridium perfringens', *Plasmid*, vol. 67, no. 2, pp. 139–147, Mar. 2012, doi: 10.1016/j.plasmid.2011.12.016.

[35] **N. Firth *et al.***, 'Analysis of a transfer region from the staphylococcal conjugative plasmid pSK41', *Gene*, vol. 136, no. 1, pp. 13–25, Dec. 1993, doi: 10.1016/0378-1119(93)90442-6.

[36] **J. A. Laverde Gomez, M. Bhatty, and P. J. Christie**, 'PrgK, a Multidomain Peptidoglycan Hydrolase, Is Essential for Conjugative Transfer of the Pheromone-Responsive Plasmid pCF10', *J. Bacteriol.*, vol. 196, no. 3, pp. 527–539, Feb. 2014, doi: 10.1128/JB.00950-13.

[37] **N. Goessweiner-Mohr, K. Arends, W. Keller, and E. Grohmann**, 'Conjugation in Gram-Positive Bacteria', *Microbiol. Spectr.*, vol. 2, no. 4, Aug. 2014, doi: 10.1128/microbiolspec.PLAS-0004-2013.

[38] **C. J. Porter *et al.***, 'The conjugation protein TcpC from Clostridium perfringens is structurally related to the type IV secretion system protein VirB8 from Gram-negative bacteria', *Mol. Microbiol.*, vol. 83, no. 2, pp. 275–288, 2012, doi: https://doi.org/10.1111/j.1365-2958.2011.07930.x.

[39] **W. L. Teng, T. L. Bannam, J. A. Parsons, and J. I. Rood**, 'Functional Characterization and Localization of the TcpH Conjugation Protein from Clostridium perfringens', *J. Bacteriol.*, vol. 190, no. 14, pp. 5075–5086, Jul. 2008, doi: 10.1128/JB.00386-08.

[40] **M. P. Garcillán-Barcia, S. Redondo-Salvo, L. Vielva, and F. de la Cruz**, 'MOBscan: Automated Annotation of MOB Relaxases', in *Horizontal Gene Transfer: Methods and Protocols*, F. de la Cruz, Ed. New York, NY: Springer US, 2020, pp. 295–308. doi: 10.1007/978-1-4939-9877-7_21.

[41] **M. Egan *et al.***, 'Staying alive: growth and survival of Bifidobacterium animalis subsp. animalis under in vitro and in vivo conditions', *Appl. Microbiol. Biotechnol.*, vol. 102, no. 24, pp. 10645–10663, Dec. 2018, doi: 10.1007/s00253-018-9413-7.

[42] **J. C. D. Man, M. Rogosa, and M. E. Sharpe**, 'A Medium for the Cultivation of Lactobacilli', *J. Appl. Bacteriol.*, vol. 23, no. 1, pp. 130–135, 1960, doi: 10.1111/j.1365-2672.1960.tb00188.x.

[43] **D. G. Anderson and L. L. McKay**, 'Simple and rapid method for isolating large plasmid DNA from lactic streptococci.', *Appl. Environ. Microbiol.*, vol. 46, no. 3, pp. 549–552, Sep. 1983.

[44] **P. Álvarez-Martín, A. B. Flórez, A. Margolles, G. del Solar, and B. Mayo**, 'Improved Cloning Vectors for Bifidobacteria, Based on the Bifidobacterium catenulatum pBC1 Replicon', *Appl Env. Microbiol*, vol. 74, no. 15, Art. no. 15, Aug. 2008, doi: 10.1128/AEM.00074-08.

[45] **D. J. O'sullivan and T. R. Klaenhammer**, 'Rapid Mini-Prep Isolation of High-Quality Plasmid DNA from Lactococcus and Lactobacillus spp', *Appl. Environ. Microbiol.*, vol. 59, no. 8, pp. 2730–2733, Aug. 1993, doi: 10.1128/AEM.59.8.2730-2733.1993.

[46] **J. Sambrook, E. F. Fritsch, and T. Maniatis**, 'Molecular cloning: a laboratory manual.', *Mol. Cloning Lab. Man.*, no. Ed. 2, Art. no. Ed. 2, 1989, Accessed: Mar. 19, 2019. [Online]. Available: https://www.cabdirect.org/cabdirect/abstract/19901616061

[47] **M. O'Connell Motherway, J. O'Driscoll, G. F. Fitzgerald, and D. Van Sinderen**, 'Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in Bifidobacterium breve UCC2003', *Microb. Biotechnol.*, vol. 2, no. 3, Art. no. 3, May 2009, doi: 10.1111/j.1751-7915.2008.00071.x.

[48] **J. Sambrook, E. Fritsch, and T. Maniatis**, *Molecular cloning: A laboratory manual: Vol. 2*, 2. ed. S.l.: Cold Spring Harbor, 1989.

[49] **A. Mazé, M. O'Connell-Motherway, G. F. Fitzgerald, J. Deutscher, and D. van Sinderen**, 'Identification and characterization of a fructose phosphotransferase system in Bifidobacterium breve UCC2003', *Appl. Environ. Microbiol.*, vol. 73, no. 2, pp. 545–553, Jan. 2007, doi: 10.1128/AEM.01496-06.

[50] **F. Madeira *et al.***, 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic Acids Res.*, vol. 47, no. W1, pp. W636–W641, Jul. 2019, doi: 10.1093/nar/gkz268.

[51] **A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton**, 'Jalview Version 2—a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, vol. 25, no. 9, pp. 1189–1191, May 2009, doi: 10.1093/bioinformatics/btp033.

[52] **J. J. Almagro Armenteros *et al.***, 'SignalP 5.0 improves signal peptide predictions using deep neural networks', *Nat. Biotechnol.*, vol. 37, no. 4, Art. no. 4, Apr. 2019, doi: 10.1038/s41587-019-0036-z.

[53] **A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer**, 'Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes', *J. Mol. Biol.*, vol. 305, no. 3, pp. 567–580, Jan. 2001, doi: 10.1006/jmbi.2000.4315.

[54] **S. El-Gebali *et al.***, 'The Pfam protein families database in 2019', *Nucleic Acids Res.*, vol. 47, no. D1, pp. D427–D432, Jan. 2019, doi: 10.1093/nar/gky995.

[55] **A. Marchler-Bauer and S. H. Bryant**, 'CD-Search: protein domain annotations on the fly', *Nucleic Acids Res.*, vol. 32, no. Web Server issue, pp. W327–W331, Jul. 2004, doi: 10.1093/nar/gkh454.

[56] **J. Söding, A. Biegert, and A. N. Lupas**, 'The HHpred interactive server for protein homology detection and structure prediction', *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W244–W248, Jul. 2005, doi: 10.1093/nar/gki408.

[57] **L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg**, 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nat. Protoc.*, vol. 10, no. 6, Art. no. 6, Jun. 2015, doi: 10.1038/nprot.2015.053.

[58] **M. Källberg *et al.***, 'Template-based protein structure modeling using the RaptorX web server', *Nat. Protoc.*, vol. 7, no. 8, pp. 1511–1522, Jul. 2012, doi: 10.1038/nprot.2012.085.

[59] **F. Sievers *et al.***, 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Mol. Syst. Biol.*, vol. 7, p. 539, Oct. 2011, doi: 10.1038/msb.2011.75.

[60] **G. Stecher, K. Tamura, and S. Kumar**, 'Molecular Evolutionary Genetics Analysis (MEGA) for macOS', *Mol. Biol. Evol.*, vol. 37, no. 4, pp. 1237–1239, Apr. 2020, doi: 10.1093/molbev/msz312.

[61] **C. Camacho *et al.***, 'BLAST+: architecture and applications', *BMC Bioinformatics*, vol. 10, p. 421, Dec. 2009, doi: 10.1186/1471-2105-10-421.

[62] **L. Guy, J. Roat Kultima, and S. G. E. Andersson**, 'genoPlotR: comparative gene and genome visualization in R', *Bioinformatics*, vol. 26, no. 18, pp. 2334–2335, Sep. 2010, doi: 10.1093/bioinformatics/btq413.

[63] 'R: The R Project for Statistical Computing'. https://www.r-project.org/ (accessed Nov. 24, 2020).

[64] 'Official REBASE Homepage | The Restriction Enzyme Database | NEB'. http://rebase.neb.com/rebase/ (accessed Jul. 19, 2023).

[65] **J. Lutkenhaus**, 'The ParA/MinD family puts things in their place', *Trends Microbiol.*, vol. 20, no. 9, pp. 411–418, Sep. 2012, doi: 10.1016/j.tim.2012.05.002.

[66] **K. Atmakuri, E. Cascales, O. T. Burton, L. M. Banta, and P. J. Christie**, 'Agrobacterium ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions', *EMBO J.*, vol. 26, no. 10, pp. 2540–2551, May 2007, doi: 10.1038/sj.emboj.7601696.

[67] **A. Peña, J. Ripoll-Rozada, S. Zunzunegui, E. Cabezón, F. de la Cruz, and I. Arechaga**, 'Autoinhibitory Regulation of TrwK, an Essential VirB4 ATPase in Type IV Secretion Systems', *J. Biol. Chem.*, vol. 286, no. 19, pp. 17376–17382, May 2011, doi: 10.1074/jbc.M110.208942.

[68] **H. Kim *et al.***, 'Structural basis for effector protein recognition by the Dot/Icm Type IVB coupling protein complex', *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, May 2020, doi: 10.1038/s41467-020-16397-0.

[69] **F. Li, C. Alvarez-Martinez, Y. Chen, K.-J. Choi, H.-J. Yeo, and P. J. Christie**, 'Enterococcus faecalis PrgJ, a VirB4-Like ATPase, Mediates pCF10 Conjugative Transfer through Substrate Binding', *J. Bacteriol.*, vol. 194, no. 15, pp. 4041–4051, Aug. 2012, doi: 10.1128/JB.00648-12.

[70] **N. T. Odermatt *et al.***, 'Structural and DNA binding properties of mycobacterial integration host factor mIHF', *J. Struct. Biol.*, vol. 209, no. 3, p. 107434, Mar. 2020, doi: 10.1016/j.jsb.2019.107434.

[71] **D. Sakai and T. Komano**, 'Genes Required for Plasmid R64 Thin-Pilus Biogenesis: Identification and Localization of Products of the pilK, pilM, pilO, pilP, pilR, and pilT Genes', *J. Bacteriol.*, vol. 184, no. 2, pp. 444–451, Jan. 2002, doi: 10.1128/JB.184.2.444-451.2002.

[72] **S. Halbedel, B. Hahn, R. A. Daniel, and A. Flieger**, 'DivIVA affects secretion of virulence-related autolysins in Listeria monocytogenes', *Mol. Microbiol.*, vol. 83, no. 4, pp. 821–839, Feb. 2012, doi: 10.1111/j.1365-2958.2012.07969.x.

[73] **H. Ni *et al.***, 'Streptococcus suis DivIVA Protein Is a Substrate of Ser/Thr Kinase STK and Involved in Cell Division Regulation', *Front. Cell. Infect. Microbiol.*, vol. 8, 2018, Accessed: Sep. 14, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcimb.2018.00085

[74] **D. Sharma and A. U. Khan**, 'Role of cell division protein divIVA in Enterococcus faecalis pathogenesis, biofilm and drug resistance: A future perspective by in silico approaches', *Microb. Pathog.*, vol. 125, pp. 361–365, Dec. 2018, doi: 10.1016/j.micpath.2018.10.001.

[75] **D. L. Guzmán-Herrador and M. Llosa**, 'The secret life of conjugative relaxases', *Plasmid*, vol. 104, p. 102415, Jul. 2019, doi: 10.1016/j.plasmid.2019.102415.

[76] **K. P. Nowak, A. Sobolewska-Ruta, A. Jagiełło, A. Bierczyńska-Krzysik, P. Kierył, and P. Wawrzyniak**, 'Molecular and Functional Characterization of MobK Protein-A Novel-Type Relaxase Involved in Mobilization for Conjugational Transfer of Klebsiella pneumoniae Plasmid pIGRK', *Int. J. Mol. Sci.*, vol. 22, no. 10, p. 5152, May 2021, doi: 10.3390/ijms22105152.

[77] **N. Soler *et al.***, 'Characterization of a relaxase belonging to the MOBT family, a widespread family in Firmicutes mediating the transfer of ICEs', *Mob. DNA*, vol. 10, no. 1, p. 18, May 2019, doi: 10.1186/s13100-019-0160-9.

[78] **D. Ceccarelli, A. Daccord, M. René, and V. Burrus**, 'Identification of the Origin of Transfer (oriT) and a New Gene Required for Mobilization of the SXT/R391 Family of Integrating Conjugative Elements', *J. Bacteriol.*, vol. 190, no. 15, Art. no. 15, Aug. 2008, doi: 10.1128/JB.00150-08.

[79] **A. Hegyi, M. Szabó, F. Olasz, and J. Kiss**, 'Identification of oriT and a recombination hot spot in the IncA/C plasmid backbone', *Sci. Rep.*, vol. 7, no. 1, Art. no. 1, Sep. 2017, doi: 10.1038/s41598-017-11097-0.

[80] **C. A. Lee and A. D. Grossman**, 'Identification of the Origin of Transfer (oriT) and DNA Relaxase Required for Conjugation of the Integrative and Conjugative Element ICEBs1 of Bacillus subtilis', *J. Bacteriol.*, vol. 189, no. 20, pp. 7254–7261, Oct. 2007, doi: 10.1128/JB.00932-07.

[81] **A. O'Callaghan, F. Bottacini, M. O'Connell Motherway, and D. van Sinderen**, 'Pangenome analysis of Bifidobacterium longum and site-directed mutagenesis through by-pass of restriction-modification systems', *BMC Genomics*, vol. 16, Oct. 2015, doi: 10.1186/s12864-015-1968-4.

[82] **Z. Sun, A. Baur, D. Zhurina, J. Yuan, and C. U. Riedel**, 'Accessing the inaccessible: molecular tools for bifidobacteria', *Appl. Environ. Microbiol.*, vol. 78, no. 15, pp. 5035–5042, Aug. 2012, doi: 10.1128/AEM.00551-12.

[83] **K. Yasui *et al.***, 'Improvement of bacterial transformation efficiency using plasmid artificial modification', *Nucleic Acids Res.*, vol. 37, no. 1, p. e3, Jan. 2009, doi: 10.1093/nar/gkn884.

[84] **E. C. Hoedt *et al.***, 'Broad Purpose Vector for Site-Directed Insertional Mutagenesis in Bifidobacterium breve', *Front. Microbiol.*, vol. 12, 2021, doi: 10.3389/fmicb.2021.636822.

[85] **J. Silbert, V. de Lorenzo, and T. Aparicio**, 'Refactoring the Conjugation Machinery of Promiscuous Plasmid RP4 into a Device for Conversion of Gram-Negative Isolates to Hfr Strains', *ACS Synth. Biol.*, vol. 10, no. 4, pp. 690–697, Apr. 2021, doi: 10.1021/acssynbio.0c00611.

[86] **C. Guynet and F. de la Cruz**, 'Plasmid segregation without partition', *Mob. Genet. Elem.*, vol. 1, no. 3, pp. 236–241, 2011, doi: 10.4161/mge.1.3.18229.

[87] **C. J. Gruber *et al.***, 'Conjugative DNA Transfer Is Enhanced by Plasmid R1 Partitioning Proteins', *Front. Mol. Biosci.*, vol. 3, 2016, doi: 10.3389/fmolb.2016.00032.

[88] **E. Pachulec *et al.***, 'Functional Analysis of the Gonococcal Genetic Island of Neisseria gonorrhoeae', *PLOS ONE*, vol. 9, no. 10, p. e109613, Oct. 2014, doi: 10.1371/journal.pone.0109613.

[89] **C. Guynet, A. Cuevas, G. Moncalián, and F. de la Cruz**, 'The stb operon balances the requirements for vegetative stability and conjugative transfer of plasmid R388', *PLoS Genet.*, vol. 7, no. 5, p. e1002073, May 2011, doi: 10.1371/journal.pgen.1002073.

[90] **C. Coluzzi, M. P. Garcillán-Barcia, F. de la Cruz, and E. P. C. Rocha**, 'Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids', *Mol. Biol. Evol.*, vol. 39, no. 6, p. msac115, Jun. 2022, doi: 10.1093/molbev/msac115.

[91] **X. Wang *et al.***, 'Multiple-Replicon Resistance Plasmids of Klebsiella Mediate Extensive Dissemination of Antimicrobial Genes', *Front. Microbiol.*, vol. 12, 2021, Accessed: Sep. 21, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2021.754931

[92] **J. Law, G. Buist, A. Haandrikman, J. Kok, G. Venema, and K. Leenhouts**, 'A system to generate chromosomal mutations in Lactococcus lactis which allows fast analysis of targeted genes', *J. Bacteriol.*, vol. 177, no. 24, pp. 7011–7018, Dec. 1995, doi: 10.1128/jb.177.24.7011-7018.1995.

[93] **P. Alvarez-Martín, M. O'Connell-Motherway, D. van Sinderen, and B. Mayo**, 'Functional analysis of the pBC1 replicon from Bifidobacterium catenulatum L48', *Appl. Microbiol. Biotechnol.*, vol. 76, no. 6, pp. 1395–1402, Oct. 2007, doi: 10.1007/s00253-007-1115-5.

CHAPTER IV

Conjugal Transfer of Bifidobacterial Megaplasmid pMP7017 Depends on Cell Wall Associated Structures

ABSTRACT

Plasmid pMP7017 is a 190 kb megaplasmid harboured by the human isolate *Bifidobacterium breve* JCM7017. pMP7017 is the only bifidobacterial plasmid identified to date that has been demonstrated to possess the ability to conjugate. Plasmid-encoded cell-surface adhesins and pili have been identified as mediators of cell-cell contact facilitating horizontal transfer of conjugative plasmids in bacteria. Consequently, we endeavoured to investigate the pMP7017-encoded cell surface proteome for potential mediators of cell-cell adhesion that are required for pMP7017 conjugation.

Functional analysis revealed the presence of several adhesin gene candidates, including two genes encoding large cell surface-associated proteins and a sortase-dependent pilus gene cluster. Site-directed mutagenesis and subsequent genetic complementation enabled the identification of a putative cell adhesion locus encoding a class C sortase and three pilin subunits, two of which were shown to be essential for pMP7017 conjugative transfer. We therefore propose that a pMP7017-harbouring donor cell specifies the formation of a pilus structure which promotes adhesion to a recipient cell thereby establishing formation of a mating pair. Comparative genomics between pMP7017 homologs revealed that the adhesion cluster is variable in both its structural organization and encoded pilin subunits, suggestive of varying recipient specificity determinants.

INTRODUCTION

Bifidobacteria are Gram-positive (G^+) autochthonous colonic commensals and members of the *Actinomycetota* phylum that harbour Guanine and Cytosine (G+C)-rich genomes [1]. Several strains within this genus are considered probiotic microorganisms and evidence of their health-promoting attributes, such as prevention and treatment of disease, alleviation of intestinal and extraintestinal pathologies, have been well documented [2]–[6]. Bifidobacteria are highly adapted to the intestinal environment with some species encoding transporters that actively extrude bile salts or alternatively mediate bile tolerance, a crucial property underpinning their robustness and functioning as probiotics [7].

The G^+ cell wall is a rigid, three-dimensional structure, composed of long covalently cross-linked polysaccharide chains of alternating *N*-acetylglucosamine and *N*-acetylmuramic acid connected by β-1,4-O-glycosidic bonds [8]. While this structure primarily serves to maintain cellular integrity and provide protection against environmental stressors, peptidoglycan acts as a scaffold for polar lipids, teichoic acids, capsular polysaccharides and cell-surface appendages in the form of pili and protein adhesins. While monoderm-specific pili of G^+ bacteria were first observed for the pathogen *Corynebacterium renale* in 1967, the existence of sortase-dependent pili in bifidobacteria was only experimentally determined in recent years [9]. Since then, functional analyses have indicated strong links between sortase-dependent pili and the modulation of bacterium–host interactions [10]–[12]. They have been shown to trigger immunomodulatory responses; for example, increased TNFα and decreased IL10 levels have been observed in the murine cecum mucosa in the presence of a *Lactococcus lactis* strain expressing pili from *B. bifidum*, a response not observed in the WT strain [10]. In addition, these filamentous protein structures have been indicated to play a primary role in the adherence of bifidobacteria to the intestinal surface and mediate bacterial cell-cell interactions [11].

G^+ sortase-dependent pili assemble as long thin fibres on the surface of the cell, being ~2-5 microns in length and with a diameter of ~3 nm. These fibres are composed of repeating subunits of major pilins that form the shaft or stalk of the structure, in addition to a minor pilin, forming the adhesive tip of the filament. Additional minor pilin proteins may be present in the pilus structure, either forming

the base of the filament, referred to as a basal pilin or decorating the stalk of the filament, then called auxiliary pilins [13]. Pilin proteins are translated as precursors with an N-terminal signal peptide motif to facilitate Sec-dependent secretion and a C-terminal cell wall-sorting signal (CWSS) containing a five amino acid (aa) residue sortase-specific cleavage sequence (LPxTG, or a variant thereof) preceding a 30 –40 aa residue hydrophobic region and a string of positively charged residues required for sortase recognition. Genetic analysis of the major features associated with pilin subunits of the genus *Bifidobacterium* indicate a conserved pilin motif (TVxxK) required for polymerization to the adjoining pilin subunits, and as such these TVxxK-like pilin motifs are a key feature of both major and basal pilins. Furthermore, a second motif known as the E box (yxxx**ET**xxxx**G**y), containing a conserved glutamate residue, is believed to be necessary for the incorporation of minor pilin subunits and the stabilization of intramolecular bonds [9], [14], [15]. Sortases are membrane-anchored cysteine transpeptidases that exhibit specificity towards the CWSS of secreted pilin peptides, processing of this signal results in the attachment of the pilin to an amino group on the cell surface. Sortases can be divided into six classes (class A through to F) based on their primary sequences, main function and cleavage site [16]. Pilus assembly in G^+ bacteria is biphasic, requiring both a pilin-specific sortase and a housekeeping sortase, to promote pilus subunit polymerization and facilitate cell wall anchoring in the final stage of pilus biogenesis, respectively [17].

As G^+ bacteria appear to lack the dynamic type IV conjugative pili which facilitate cell-cell contact, it has been proposed that G^+ systems instead depend on surface-associated proteins to promote adhesion, thus facilitating the formation of the mating pair [18]. These adhesins may be in the form of aggregation substances such as PrgB from Enterococcal plasmid pCF10 [19] or pili, as demonstrated by the reduction in conjugation efficiency in *srtD* and *cnaB* mutants of the streptococcal plasmid pCW3 [20]. In addition, cell-surface proteins at this initial stage of the conjugal process may participate in surface exclusion, prohibiting conjugation in cells which have already acquired the conjugative plasmid [21], [22].

The 190 kb plasmid pMP7017 from the human commensal *B. breve* JCM7017 represents the first megaplasmid identified in *Bifidobacterium*. In addition, pMP7017 remains the sole plasmid within this genus which has been demonstrated to be conjugative [23]. Considering recent advances linking plasmid-expressed surface adhesins to the promotion of conjugal transfer in G^+ systems, we endeavoured in this

work to investigate pMP7017-encoded surface-proteins and assess their impact on the horizontal transfer of pMP7017 through a strategy of site-directed mutagenesis and subsequent genetic complementation [18]–[20], [24].

EXPERIMENTAL PROCEDURES

Functional Analysis, Protein Modelling and Quality Assessment

SignalP was used to identify signal peptides and TMHMM was employed for the identification of transmembrane regions [25] [26]. Pfam and NCBI's CD-Search were used to identify conserved functional regions, implementing an E-value threshold of <0.0001 for significance [27] [28]. Three-dimensional structures of putative pMP7017-encoded surface proteins were predicted using the AlphaFold v2.0 (AF) algorithm as implemented on the Colab server [29]. Predictions were performed without considering any homologous experimental template and with three recycles. Out of the five predicted models computed for each protein, only the top ranked model for each structure was considered for further analysis and presented in this work. The reliability of the AF predictions was assessed by analyzing the Predicted Aligned Error (PAE) and Local Distance Difference Test (LDDT) scores reported for each structure. Structural comparison was performed using the DALI server using default settings. Structural similarity was evaluated with the DALI Z-score, ranging from 1-20, which can be defined as the sum of equivalent residue-wise Cα -Cα distances among the three-dimensional co-ordinates of each protein pair [30]. Z-scores ≤2 are considered spurious, while scores ≥8 are considered to represent likely similar structures [31]. The Root Mean Square Deviation (RMSD) was also considered, where the calculated values represent the average deviation between the corresponding atoms of the two compared proteins. The TM-score obtained from pairwise structural comparison using the alignment tool of the Protein Data Bank server was also evaluated for each pairwise comparison. The TM-score is a metric for assessing the topological similarity of protein structures ranging from 0-1, where 1 is a perfect match [32]. Structures with a Z score ≥8 and a TM-score ≥0.2 were considered for further analysis.

Strains, plasmids and cultivation conditions

Bacterial strains and plasmids used in this study are listed in **Table 1** Bifidobacterial strains were initially cultivated from glycerol stocks in reinforced clostridial medium (RCM; Oxoid Ltd., Basingstoke, England) and subsequently sub-cultured into either filtered RCM or modified de Man, Rogosa, and Sharpe medium

(mMRS) prepared from first principles [33], [34]. Prior to inoculation, mMRS medium was supplemented with cysteine-HCl (0.05%, wt/vol) and 1% (wt/vol) lactose (Sigma-Aldrich, Steinheim, Germany). Cultures were grown at 37°C in an anaerobic chamber (Davidson and Hardy, Belfast, Ireland). *Escherichia coli* strains were routinely cultivated in Luria-Bertani medium (LB) and grown at 37°C with agitation. Growth media included the addition of the following antibiotics where appropriate: tetracycline (10 μg ml^{-1}) and/or chloramphenicol (2 μg ml^{-1}) for bifidobacterial strains; tetracycline (10 μg ml^{-1}), chloramphenicol (5 μg ml^{-1}), ampicillin (100 μg ml^{-1}), erythromycin (100 μg ml^{-1}) and/or kanamycin (50 μg ml^{-1}) for *E. coli* cultures; chloramphenicol (5 μg ml^{-1}) for *Lactococcus lactis*.

DNA Manipulations

Total genomic DNA was isolated from bifidobacterial strains as previously described [35]. Minipreparations of plasmid DNA were isolated from *E. coli* with modifications for *Bifidobacterium* and *Lactococcus* as detailed previously [36]–[38]. Procedures for DNA manipulations were conducted essentially as described in a previous publication [39]. Single-stranded DNA oligonucleotide primers used in this study were synthesised by Eurofins (Ebersberg, Germany) and are listed in **Table 2**. Both standard and colony PCRs [40] were routinely carried out with DreamTaq Green PCR Master Mix (Thermo Scientific, Waltham, United States), while high-fidelity PCR was achieved using Q5 High-Fidelity 2X Master Mix (New England Biolabs, Herefordshire, United Kingdom). All reactions were carried out in a Life Technologies Proflex PCR System (Thermo Scientific, Waltham, United States). PCR fragment purification was achieved using the Roche High Pure PCR purification kit (Roche Diagnostics, Basel, Switzerland). Verification of the expected PCR product size was assessed by 1% agarose gel electrophoresis containing Invitrogen SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, Waltham, Massachusetts, United States) prior to visualisation on a UV transilluminator. T4 DNA ligase and restriction endonucleases were obtained from Roche Diagnostics (Basel, Switzerland) and used according to the supplier's instructions. Introduction of plasmid DNA into *E. coli* and transformation of bifidobacterial [41] and lactococcal [42] strains was achieved by electroporation according to published protocols. Sequencing was performed at Eurofins (Ebersberg,

Germany) on all plasmid constructs to validate the desired orientation of the cloned fragment and verify sequence integrity.

Generation of pMP7017 Insertion Mutants

An internal ~500 bp fragment for each of the following pMP7017-associated genes with locus-tags B7017_p0020, B7017_p0021, B7017_p0022, B7017_p0023, B7017_p0024 and B7017_p0200 was amplified using gene-specific primers listed in **Table 2**, with a HindIII site incorporated into the forward primer and an XbaI site incorporated into the reverse primer. The approach taken in the construction of the insertion mutants is detailed in a previous publication [39]. Site-specific recombination in tetracycline-resistant isolates was assessed by an initial screen for *tet*(W) gene integration via colony PCR with the primer pair *tetW*_F and *tetW*_R. Confirmation of integration at the correct genetic locus was established with primers designed upstream of the integration site, in combination with either TetWF or TetWR, depending on the orientation of the *tet*(W) gene as ascertained by sequencing. The resulting strains harbouring pMP7017-derived plasmids were then used as donors and assessed for their conjugation ability employing *B. breve* UCC2003 as the recipient strain. Sequencing of pMP7017-derivatives was performed on an Illumina MiSeq Sequencing platform to verify integration at the expected locus.

Conjugal matings

Individual overnight cultures of donor and recipient strains were inoculated at 1% (v/v) into mMRS broth supplemented with cysteine-HCl (0.05% [wt/vol]) and lactose (1.0% [wt/vol]) and grown to an optical density (OD_{600nm}) value 0.8-1.0 at 37°C under anaerobic conditions. Donor and recipient cells were combined in a 1:1 ratio, based on OD value, up to a total volume of 2 ml, briefly vortexed and concentrated by centrifugation at 2,200 x *g* for 5 min. For solid plate mating, the resultant cell pellet was resuspended in 200 µl of RCM and spread plated onto non-selective agar. RCA plates were incubated at 37°C overnight under anaerobic conditions. Following incubation, the donor/recipient conjugation mixture was recovered from the plates using a cell scraper (Sarstedt, Germany) and resuspended in 2 ml of phosphate-buffered saline and serially diluted. Donor and recipient cells were

enumerated based on selective carbohydrate utilization by the donor or recipient strain on mMRS agar, while transconjugants were isolated based on the selective carbohydrate for recipient cells in combination with tetracycline (10 μg ml^{-1}) to select for the presence of a pMP7017-derivative that had been marked with a tetracycline resistance gene [23]. Putative transconjugant colonies were restreaked twice on fresh mMRS plates supplemented with ribose and 10 mg ml^{-1} of tetracycline. For validation purposes, five suspected transconjugants from each experimental group were grown in mMRS broth with tetracycline at a concentration of 10 mg ml^{-1} and ribose as the sole carbon source, after which DNA was isolated and PCR performed with primers based on genes unique to the donor strain, recipient strain and the megaplasmid. Those which were PCR negative for the donor strain, and PCR positive for the recipient and megaplasmid were counted. Mutant donor strains that did not yield transconjugants within the limit of detection (10^{-10}) were considered to be unable to perform conjugation.

Plasmid vector constructions

To generate the plasmid constructs for genetic complementation, the coding sequences and presumed promoter regions of *pilA-srtC*, and *pilC* were amplified from *B. breve* JCM7017 total DNA by PCR using the primer pairs outlined in **Table 2**. The resulting amplicons were ligated into the ClaI and PspXI sites in plasmid pBbrU following restriction and purification. 4 μl of each ligation mixture was transformed into *E. coli* XL1-blue competent cells by electroporation. Plasmid preparations of constructs were sequenced using primer pair, M13rev-28 and M13uni-21, to confirm the presence and fidelity of the inserts. Plasmids with the correct sequence were then introduced into *B. breve* JCM7017 strains harbouring pMP7017 with the corresponding mutation by electroporation, generating strains, *B. breve* JCM7017-*srtC*-C, *B. breve* JCM7017-*pilA*-C and *B. breve* JCM7017-*pilC*-C.

To construct plasmids pNZ_P*spd1*, pNZ_P*srtC,* pNZ_P*pilA*, pNZ_P*pilB* and pNZ_P*pilC*, ~200 bp preceding the predicted translational start site of each gene and presumed to contain the promoter-encompassing region of the corresponding gene was amplified by PCR with primer pairs outline in **Table 2**, respectively. The resulting amplicons were cloned upstream of the promoterless β-glucuronidase gene in reporter plasmid pNZ272 and transformed into *L. lactis* NZ9000. The recombinant plasmids

were then isolated and introduced into *B. breve* UCC2003 and *B. breve* UCC2003 pMP7017_199 by electroporation.

GUS Assay

Intracellular β-glucuronidase (β-GUS) activity was determined essentially as described previously [43]. GusA activity was expressed in Miller Units (M.U.) and calculated using the equation: $1000^{*}((OD_{420} - (1.75^{*}OD_{550}))/(t \times v \times OD_{600}))$, where t is the reaction time [mins], v is the cell volume [ml] and OD_{420}, OD_{550}, and OD_{600} are absorbance values at 420, 550, and 600 nm, respectively.

Comparative genomics

Nucleotide sequences of pMP7017 conjugal clusters and *B. longum* homologues were extracted from the NCBI database and aligned with Clustal Omega [44], phylogeny was inferred using MEGAX [45] following gap removal. The Blast+ terminal [46] was used to determine percentage identity, the analysis was conducted in R using the genoPlotR package [47], [48].

RESULTS

Functional and structural analysis of pMP7017-encoded surface proteins reveals several candidate adhesins

Since conjugative pili, which establish the physical connection between mating pairs in characterised conjugation systems of diderms, have not been described for G^+ bacteria, it has been proposed that they instead rely on surface proteins for adhesion to recipient cells [18]. To identify the protein(s) responsible for pMP7017 adhesion to recipient cells, we first screened pMP7017 for candidate genes encoding putative cell surface-associated proteins. In this manner we identified six genes encoding sortase-dependent proteins (SDPs) (locus tags; B7017_p0020, B7017_p0022, B7017_p0023, B7017_p0024, B7017_p0199 and B7017_p0200) and a presumed sortase-encoding gene (locus tag; B7017_p0021), henceforth referred to as *spd1*, *pilA*, *pilB*, *pilC*, *spd2* and *srtC,* respectively. Following their identification, the amino acid sequence of each pMP7017-encoded candidate adhesin was interrogated for potential functional domains and the 3D structures predicted to establish a role based on structural homologies where possible.

spd1 encodes a 102.3 kDa protein precursor, with a C-terminal CWSS containing a class E sorting motif (**Figure 1A**). Spd1 contains a centrally located CAP domain (CAP, Pfam: PF00188). While little is known about the function of the bacterial CAP domain, characterised members of the eukaryotic CAP superfamily display a large degree of functional diversity, from promoting cell-cell adhesion during fertilisation in mammals to their role in sterol export and virulence in fungi [49], [50]. Structurally, Spd1 is predicted to form a 2-helix coiled-coil structure with a central globular domain containing a 3-layer α-β-α fold cavity which harbours the catalytic Tetrad found in other CAP domains (**Figure 1B**). Furthermore, Spd1 displays a significant degree of structural similarity (Z-score: 16.5, TM-score: 0.64) to the surface exclusion protein PrgA of the conjugative enterococcal plasmid pCF10 ($PrgA_{pCF10}$) (**Table 3**). $PrgA_{pCF10}$ has been indicated to promote surface exclusion via spatial interference and to lower aggregation and mating efficiency by proteolytic cleavage of the aggregation substance $PrgB_{pCF10}$, facilitated by the serine-dependent proteolytic activity of its CAP domain which is protruding an estimated ~40 nm outside the cell wall [22].

The *srtC* gene encodes a 41.1 kDa polytopic protein with a centrally located Sortase functional domain (SrtC, Pfam: PF04203) and displays structural similarity to class C pilus-specific cysteine-transpeptidases, such as SrtC1 from *Lactobacillus rhamnosus* GG (PDB: 7CG9).

pilA encodes a large pro-protein of 223.3 kDa containing a C-terminal class C sorting motif while also encompassing an E box motif ^{1284}**ET**IATT**G**^{1290} for the incorporation of adjoining pilin subunits yet lacking a pilin motif, features that are consistent with those of a tip pilin (**Figure 2A**). PilA is predicted to contain a C-terminal Leucine Rich Repeat domain (LRR, Pfam: PF09479) with eight repeat units, spanning 831 aa residues. Additionally, this region is enriched in Glycine-Tryptophan (GW) dipeptide sequences. GW domains have been demonstrated to bind polyanions, such as lipoteichoic acids present on the cell surface of bacteria, where the strength of attachment is proportional to the number of GW domains present in the protein [51]. Structural modelling of the PilA C-proximal LRR domain produced a predicted elongated structure containing multiple subunits flexibly linked to each other which conform to a variation of the β-grasp fold, resembling the B-repeats of the host cell invasion protein InlB from *Listeria monocytogenes* (PDB: 2Y5P) [52] (**Figure 2B**). The β2 strand of the InlB B-repeat has been implicated to be of functional importance for protein–protein interactions and host cell receptor binding [52], [53]. Four structural subdomains are observed within the globular N-terminal putative adhesion domain of PilA (CBM1-4, numbered N- to C-terminal), each presenting with 6-8 opposing antiparallel β-sheets conforming to an overall jelly roll-like fold (**Figure 2C**). Subdomains CBM1 and CBM3 show structural homology to the carbohydrate binding modules (CBMs) of various carbohydrate-active enzymes (**Figure S1**). CBMs are noncatalytic modules that bring their appended enzymes in immediate proximity to their target ligand potentiating the activity of the enzyme [54]. CBM2 shares a high degree of structural homology to the variable domain of the Streptococcal agglutinin receptor, SspB (PDB: 2WD6-A), and the adhesin domain of the PrgB aggregation substance from the conjugative enterococcal plasmid pCF10 (PDB: 6GED-A) (**Figure 2D**).

The *pilB* gene encodes a 62.5 kDa pro-protein with a class C sorting motif, and is predicted to contain typical attributes of a major pilin: a centrally located pilin motif ^{249}**TV**PN**K**^{253}, a C-terminal E-box motif ^{479}YTVV**ET**REPA**G**Y^{490} located within the SpaH/EbpB family LPXTG-anchored major pilin domain (iso_D2_wall_anc super

family, CDD: NF033902) and Prealbumin-like fold domain (SpaA, Pfam: PF17802). PilB shows structural similarity to a number of major pilins of characterized sortase-dependent pilus structures, including the major pilin from *Clostridium perfringens* (PDB: 6IXY), the T6 major pilin from *Streptococcus pyogenes* (PDB: 4P0D) and major pilin SPB1 from *Streptococcus agalactiae* (PDB: 4UZG) (**Figure 3**).

pilC encodes a 96.6 kDa pro-protein containing both a class E sorting motif and pilin motif ^{412}**TV**ED**K**^{416}, with no discerning functional domains. Structural homology searches utilising the predicted PilC structure were below the designated thresholds for consideration.

spd2 encodes a 228.2 kDa pro-protein with a class E sorting motif and a C-terminal RGD motif implicated in binding of host cell integrins (**Figure 4A**) [55]. Structurally the distal end of the putative adhesion domain of Spd2 displays significant similarity to the Class II thioester domains (TED) domains of characterised cell surface adhesions from G+ bacteria (**Figure 4B**) (PDB: 6FWV-B, 6FX6-A). Recently, the conserved Class II TED containing adhesin, p34 from plasmid pLS20 of *Bacillus subtilis,* has been demonstrated to be essential for efficient conjugative transfer of pLS20 [56].

pMP7017 encodes an additional surface protein, locus tag B7017_p0199, however, mutation of B7017_p0199 in previous works established that the protein product of this gene has no perceived effect on the plasmid's conjugative ability, as such, the mutant strain *B. breve* JCM7017-199 will be utilised in this work as a positive control in mating experiments [23].

Sortase-dependent pili encoded by pMP7017 are essential for its conjugative transfer

To determine if any of the identified SDPs or associated sortase play a role in adhesion to facilitate horizontal transfer of pMP7017, site-directed mutagenesis (SDM) was performed targeting each gene individually. Subsequently, the generated mutant strains were used as donors in mating assays with recipient strain *B. breve* UCC2003 to assess the mutated plasmid's conjugative ability. Matings conducted with donor strains, *B. breve* JCM7017-*srtC*, *B. breve* JCM7017-*pilA*, and *B. breve* JCM7017-*pilC* produced no transconjugants within the limit of detection, thus implicating these genes as potential mediators of mating pair stabilization in the

conjugal transfer of pMP7017. Matings with mutant donor *B. breve* JCM7017-*pilB*, resulted in an average of 13.3 fold (P-value: 0.0286) increase conjugation efficiency (**Figure 5A**). Mutation of *spd1* or *sdp2* effected no significant change in conjugation efficiency and was comparable to that of the positive control, donor *B. breve* JCM7017-199.

To determine if the conjugation deficient phenotype observed with mutants, *B. breve* JCM7017-*srtC*, *B. breve* JCM7017-*pilA*, and *B. breve* JCM7017-*pilC* is directly due to their indicated mutations, genetic complementation was subsequently undertaken. *B. breve* JCM7017-*pilC* was supplied with plasmid pC24, containing the undisrupted *pilC* gene and corresponding promoter generating strain *B. breve* JCM7017-*pilC*-C. Due to the absence of motifs resembling a canonical bifidobacterial promoter sequence upstream of *srtC*, both mutant strains *B. breve* JCM7017-*srtC* and *B. breve* JCM7017-*pilA* were individually supplied with plasmid pC21-22 harbouring the presumed *srtC-pilA* transcriptional unit (TU) in addition to the region upstream of *pilA* containing its predicted promoter, generating strains *B. breve* JCM7017-*srtC*-C and *B. breve* JCM7017-*pilA*-C, respectively. Following the introduction of *in trans* genetic complements *srtC*, *pilA* and *pilC* mutant strains were tested for their ability to conjugate, as described in the materials and methods. While all three mutants containing their corresponding complement displayed a conjugative positive phenotype comparable to the control (**Figure 5B**), only the protein products of genes *srtC* and *pilC* can be considered essential factors for conjugal transfer of pMP7017. Complementation of the *pilA* mutant with the presumed *srtC-pilA* TU does not conclusively confirm its essentiality for the process, as SDM of *pilA* may indirectly disrupt the expression of the adjacent *srtC* gene, the potential for such polar effects may therefore confound the classification of *pilA* as (non-)essential. As such, an alternative method to definitively establish its involvement in mediating adhesion is required. Despite this, we believe the data collectively highlights the involvement of the *spd1-srtC-pilA-pilB-pilC* cluster in the facilitation of cell-cell contact which promotes the conjugal transfer of pMP7017 and will be referred to as the adhesion locus henceforth (**Figure 6A**).

The pMP7017 adhesion locus contains multiple promoters of varying strengths with potential stoichiometric implications

To verify the perceived transcriptional landscape of the adhesion locus and gain insight into the transcriptional activities of the putative promoters, the genomic regions preceding the *spd1*, *srtC*, *pilA, pilB* and *pilC* genes were individually cloned upstream of the promoter-less β-glucuronidase (β-GUS)-encoding gene in reporter plasmid pNZ272 [57], [58], generating plasmids pNZ_P*spd1*, pNZ_P*srtC,* pNZ_P*pilA*, pNZ_P*pilB* and pNZ_P*pilC*, respectively. These plasmids were individually introduced into *B. breve* UCC2003 and *B. breve* UCC2003 harbouring pMP7017-199, facilitating analysis of the constructed promoter fusions in both the presence and absence of the megaplasmid.

Transcriptional activity was observed with plasmid constructs containing regions upstream of *spd1, pilA, pilB* and *pilC*, however β- glucuronidase activity above that of the empty vector was not observed for the construct containing the region upstream of *srtC*, thus supporting the notion that *srtC* and *pilA* form a single TU, while *spd1, pilB* and *pilC* are transcribed as monocistronic units (**Figure 6**). Furthermore, these data indicate varying levels of promoter strength which may have stoichiometric implications, as differential expression may be reflective of the individual requirement for specific pilin subunit levels. As similar activity levels were observed in both the presence and absence of pMP7017, it appears that the genes within the pMP7017 adhesion locus are not transcriptionally regulated under the conditions tested.

Investigation of the regions 3' of each gene revealed putative Rho-independent transcriptional termination sequences 20 bps downstream of *pilB* and 104 bps 3' of *pilC*, while sequences resembling putative Rho-independent transcriptional termination signals were not observed in the regions 3' of genes *spd1, srtC* or *pilA* (**Figure 6C**).

Comparative analysis reveals marked variation in adhesion determinants

As the pMP7017 conjugation system represents the first such system of bifidobacterial origin to be characterised to date, we aimed to assess the presence and diversity of the pMP7017 adhesion locus among recently identified pMP7017 homologs from *B. longum* subsp. *longum* [59]. As these plasmids have not yet been

named, we will refer to each plasmid by the name of the strain from which they originate.

Both *spd1* encoding the predicted surface exclusion factor and *pilC* encoding the putative basal pilin are present in all adhesion loci analysed, delineating each cluster 3' and 5', respectively (**Figure 7**). While a *srtC* homolog is present in all loci, the size of the gene in addition to its location within the cluster varies. Interestingly, the highest level of diversity is observed for *pilA* encoding the putative tip-pilin, which varies significantly in size from 3,555 bp (MCC10043) to 7,884 bp (MCC10126). Despite this, *pilA* homologs encode functionally comparable proteins, as they each contain C-terminal Leucine Rich Repeat domains (LRR, Pfam: PF09479) and have retained structural similarity at their N-terminus to the CBMs of PilA from pMP7017.

The MCC10002 and MCC10096 adhesion loci encode a 281.1 kDa SDP in lieu of a *pilA* homolog, referred to here as *pilD*, in addition, neither cluster encode a PilB counterpart. The deduced PilD protein contains nine N-terminal FctA superfamily domains (Pfam: PF12892) and a centrally located SdrD B-like domain (Pfam: PF17210) (**Supplementary Table S1**). Characterised FctA domain-containing proteins of streptococcal and *Lactobacillus* origin are adhesive glycoproteins most commonly associated with cell–host adhesion through the binding of host fibronectin and collagen [60], [61]. The B-repeat of SdrD-like proteins forms an EF-hand motif with two stretches of acidic residues that bind calcium ions resulting in enhanced host cell adhesion and bacterial cell-cell aggregation [62]–[64].

While a homologous *pilA* gene is not present within the MCC10070 adhesion locus, this cluster encodes an alternative putative tip-pilin, referred to here as PilE, which is substantially smaller than its $PilA_{pMP7017}$ counterpart. PilE contains a single Listeria-Bacteroides repeat domain (Pfam: PF09479) in addition to an N-terminal SGNH hydrolase domain (CDD: cl01053) and a choice-of-anchor A domain (CDD: cl18868). SGNH hydrolases are hydrolytic enzymes with versatile functional roles, including proteases, thioesterases, acetylhydrolases, acetylesterases and various carbohydrate-modifying enzymes with broad substrate specificities [65], [66].

Furthermore, seven of the clusters analysed contain an additional pilin-encoding gene with no discernible functional domains, referred to here as *pilF*. As PilF contains a **TV**xx**K** pilin motif and a class E sorting signal yet no E box motif, it is likely that PilF represents an alternative basal pilin.

DISCUSSION

A prevalent way for bacteria to accomplish adhesion to recipient cells is by means of external appendages such as pili, which at their distal end contain an adhesin molecule, indicated to play an essential role in cell-cell interactions. Indeed, aggregation experiments with an *L. lactis* strain which heterologously expressed pili from *B. bifidum* PRL2010 revealed a 13- and 21-fold increase in aggregation levels as compared to their non-piliated counterparts [10]. As adhesins and surface structures appear to be a frequent feature encoded by G^+ conjugative plasmids, the current view is that these cell-surface proteins mediate cell-cell contact, facilitating horizontal DNA transfer in G^+ bacteria [67]–[69]. However, despite recent advances, publications pertaining to this proposed link are conflicting, therefore the essentiality of these plasmid-encoded adhesins with respect to conjugation remains speculative [18], [20], [24].

In the current work we established through site-directed mutagenesis and subsequent genetic complementation that the protein products of the pMP7017 adhesion locus *srtC*, *pilA* and *pilC* are essential factors for the horizontal transfer of pMP7017. Mutation of *spd1* did not illicit a conjugation-deficient phenotype, while interestingly, disruption of *pilB* resulted in a moderate, yet statistically significant increase in conjugation efficiency compared to the reference strain. While additional structural and functional information is necessary to determine an accurate model of the pilus architecture, based on the *in silico* data outlined in this work we propose that the proteins products of *pilA*, *pilB* and *pilC* form a heteropolymeric PilABC pilus, requiring the class C pilin-specific sortase SrtC for pilus polymerisation and a house keeping sortase for PG anchoring (**Figure 8A**).

Due to the absence of a VYPK-like pilin motif, which precludes the possibility of insertion of PilA subunits between adjacent major pilins, we propose that PilA forms the tip of the pilus which may function as an adhesin due to the presence of two CBMs (CBM1 and CBM3). Adhesins, also referred to as lectins, are non-enzymatic carbohydrate binding proteins that are well established mediators of bacterial aggregation via lectin-saccharide interactions [70]–[75]. Furthermore, PilA contains an additional N-terminal structural subdomain, CBM2, that displays significant

structural homology to the adhesion domains of the enterococcal aggregation substance $PrgB_{pCF10}$ and the variable adhesion domain of streptococcal SspB. Therefore, binding of PilA to extracellular DNA, as observed for $PrgB_{pCF10}$, may further promote coaggregation facilitating mating pair formation. Due to the significant structural homology of PilB with major pilins of G^+ pilus structures, in addition to the presence of a SpaA domain and all necessary conserved motifs associated with a major functionality, we propose that PilB is the major pilin forming the shaft of the PilABC structure. Although PilC does not present with any discernible functional domains or structural similarities, we propose that PilC forms the base of the structure relying on the catalytic activity of a class E housekeeping sortase to attach the preformed pilus to the PG.

With this structural organisation in mind, in the absence of the pilin-specific sortase SrtC, neither PilA nor PilB would be polymerized and are likely retained in the membrane following Sec-dependent export and folding (**Figure 8B.I**). Consequent to the absence of the pilus structure *srtC* mutant donors would be unable to mediate cell-cell contact necessary for mating pair formation, consistent with the conjugation-deficient phenotype observed in the *srtC* mutant strain. Disruption of *pilA,* encoding the presumed tip pilin, would allow formation of a PilBC pilus; however, a structure devoid of the presumed adhesion would be unable to facilitate adhesion to recipient cells thus resulting in the observed conjugation deficient phenotype (**Figure 8B.II**).

We have demonstrated in this work that the presumed basal pilin, PilC, is essential for adhesion and by extension conjugative transfer of pMP7017. It has been demonstrated that *Actinomyces oris* deficient of SrtA, which is responsible for anchoring the FimA-CafA pilus structure to the cell wall, fails to mediate coaggregation or interspecies interactions [76]. Through meticulous experimental validation, the authors of the aforementioned publication determined that this phenotype is attributed to the extended distance of the adhesive tip pilin due to the hyper-extension of pilus polymers resulting from continued polymerisation of the *fimA* major pilin. Similarly, complementation of the *spa*A gene, encoding the major pilin of the *Corynebacterium diphtheriae* SpaABC pilus, in the *spaA* mutant strain causes the formation of hyper-extended pilus fibers, furthermore, a comparable phenotype was observed upon the overproduction of the SpaD major pilin subunit of the *C. diphtheriae* SpaDEF pilus. With a similar rationale, the absence of the presumed basal pilin to terminate pilus biogenesis may result in aberrant pilus length which is

likely to disrupt the spatial positioning of the adhesive tip, thus abolishing the capacity of the structure to establish and maintain close cell-cell contact (**Figure 8C.III**).

Immunogold electron microscopy of the *spaA* deletion mutant encoding the major pilin of the *Corynebacterium diphtheriae* SpaABC pilus revealed both minor pilins, SpaB and SpaC, were present on the cell surface [77]. Furthermore, *E. coli* cells lacking the PapA major pilin of the pyelonephritis-associated pilus continued to adhere to human epithelial cells [78]. While the effects of pilus length on cell-to-cell interactions is not fully understood, it has been suggested that long pili promote long-distance contacts, while short pili and cell wall-linked pilins enable short-range interactions [79]. Considering this, as an increase in conjugation efficiency was observed upon mutation of *pilB* it is tempting to speculate that a structure devoid of the presumed major pilin, may enable a more intimate association between donor and recipient cells, thus enhancing short-range intercellular connections which afford an increased likelihood of mating pair formation and potentially providing greater stability between mating cells (**Figure 8B.IV**). However, the importance of PilB for adhesion to recipient cells of different strain backgrounds and the impact of its absence on conjugation efficiencies under different mating conditions needs to be explored.

Although mutation of *spd1* did not negatively impact conjugation efficiency, due to the proximity of this gene to the essential *srtC*, *pilA* and *pilC* genes and its structural homology to the recently characterised surface exclusion protein $PrgA_{pCF10}$ [22], it is possible that this gene encodes surface exclusion functions preventing donor-donor interaction, as such, a negative impact on the conjugative ability of pMP7017 upon mutation of *spd1* would not be expected.

The ability of surface adhesions to establish contact between donor and recipient cells can be considered the first rate-limiting step in the conjugative process and likely represents a key determinant of plasmid host range specificity. Comparative analysis of the adhesion cluster between pMP7017 homologues of bifidobacterial origin revealed a high degree of genetic diversity. Despite this, in the absence of a homologous pilus gene, such as that observed for the megaplasmids harboured by *B. longum* subsp. *longum* strains MCC100128, MCC10096 and MCC10070, their clusters appear to encode analogous proteins indicating assemblies of functionally similar subunits, possibly retaining their ability to adhere to recipient cells. Furthermore, the high sequence divergence between clusters may reflect rapid

evolution driving strain-specific adaptation, consequently it is tempting to speculate that these differences may be attributed to alternative binding specificities.

In conclusion, due to the essentiality of *srtC* and *pilC* for the horizontal transfer of pMP7017, we would propose from the findings in this work that the gene products, *pilA*, *pilB* and *pilC* form an adhesion complex facilitating attachment to recipient cells, polymerised by their associated pilus-specific sortase, SrtC, which promotes plasmid transfer by stabilising the mating pair. While Spd1 may potentiate surface exclusion, mechanistically similar to that observed for PrgA of pCF10.

FIGURES AND TABLES

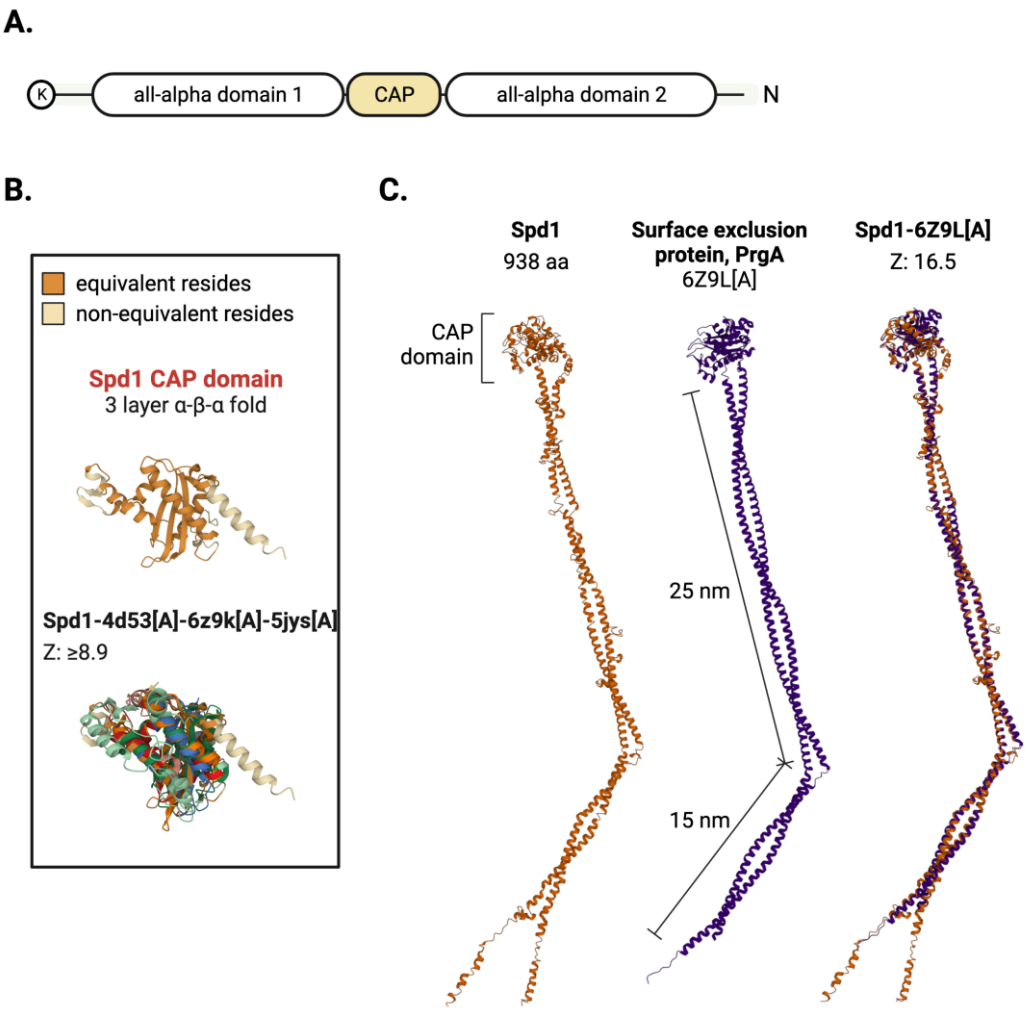


Figure 1. Domain architecture and structural homology of Spd1

A. Representation of the domain architecture of Spd1. The protein is anchored to the peptidoglycan via it's C-terminal sortase anchoring motif, represented here as an encircled 'K'. The CAP domain is centrally located with the protein and is flanked by two all-alpha domains which together form the coiled-coil stalk of the predicted 3D-structure. **B.** Inset, the α-β-α fold of Spd1 containing the cavity catalytic Tetrad. Superposition of the Spd1 CAP domain with the CAP domains of proteins originating from prokaryotes and simple eukaryotes, highlights the structural conservation of the fold despite significant diversification in sequence and function. Modelled structures include; the outer surface CAP superfamily protein BB0689 from *Borrelia burgdorferi* (PDB: 4d53-A), the CAP domain of the surface exclusion protein $PrgA_{pCF10}$ from *Enterococcus faecalis* (PDB: 6z9k-A) and the CAP domain of Pry1 from *Saccharomyces cerevisiae* S288C (PDB: 5jys-A). **C.** Mapping of the predicted

structure of Spd1 to the experimentally determined structure of the complete $PrgA_{pCF10}$ protein (PDB: 6z9L-A) demonstrates their tertiary structural similarity.

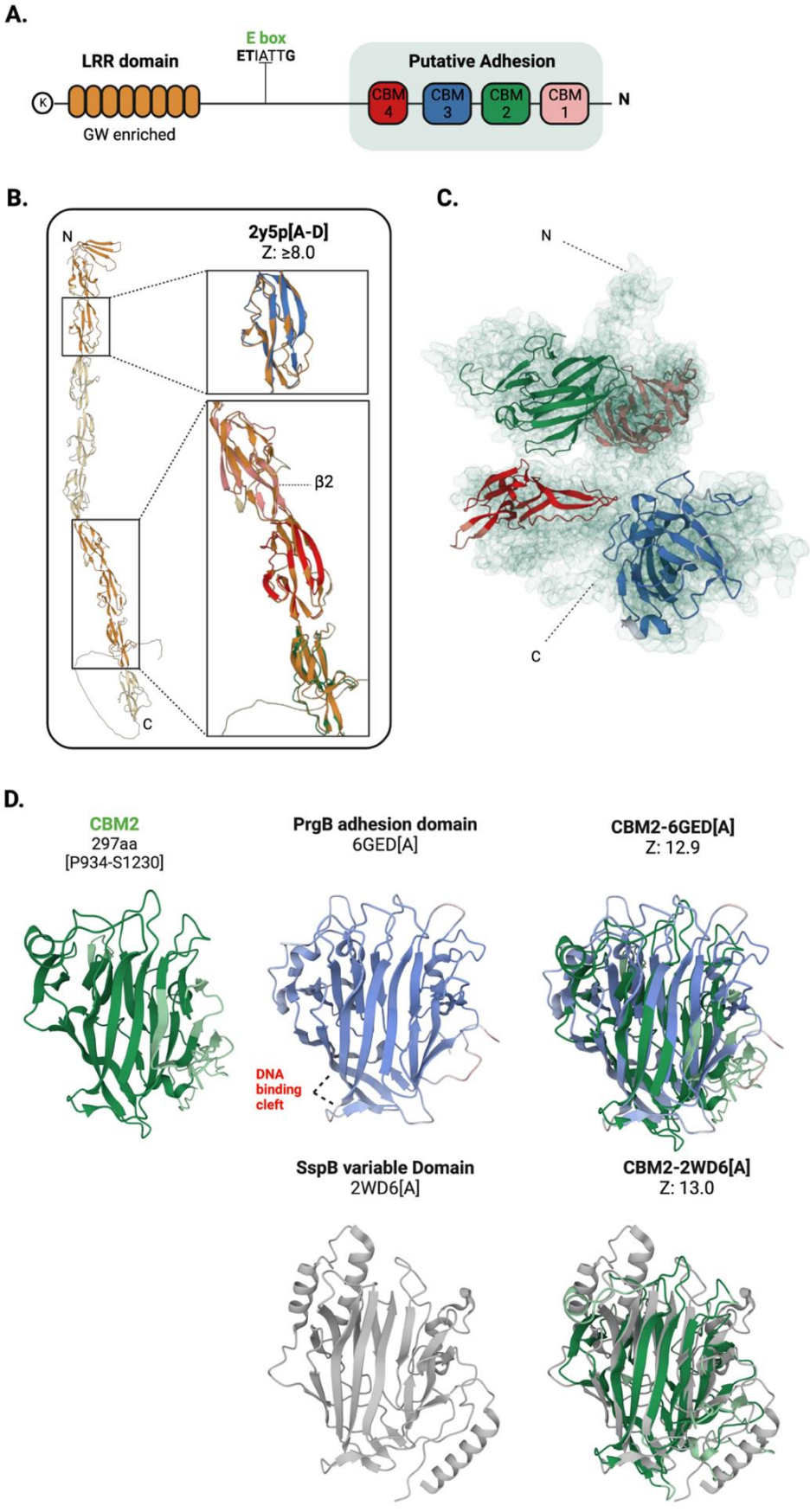


Figure 2. Domain architecture and structural homology of PilA

A. Representation of the domain architecture of PilA. The protein is anchored to the peptidoglycan via its C-terminal sortase anchoring motif (prior to polymerisation), represented here as an encircled 'K'. Features of the C-proximal Leucine Rich Repeat domain and the structural subdomains CBM1-4 of the putative N-terminal adhesion domain are indicated. **B.** Schematic representation of structure of the predicted C-terminal half of PilA illustrating an elongated structure with eight β-grasp fold-like subunits. Inset, superposition of the PilA CTD with the B-repeats of the host cell invasion protein InlB from *Listeria monocytogenes* (PDB: 2Y5P), the functionally important β2 strand is indicated. **C.** Schematic representation of the predicted globular N-terminal putative adhesion domain of PilA. Colour coding relates to what is depicted in (**A**). **D.** Mapping of the predicted structure of PilA-CBM2 to the experimentally determined structures of the adhesin domain of $PrgB_{pCF10}$ (PDB: 6GED-A) and the variable domain of SspB (PDB: 2WD6-A) demonstrates their tertiary structural similarity.

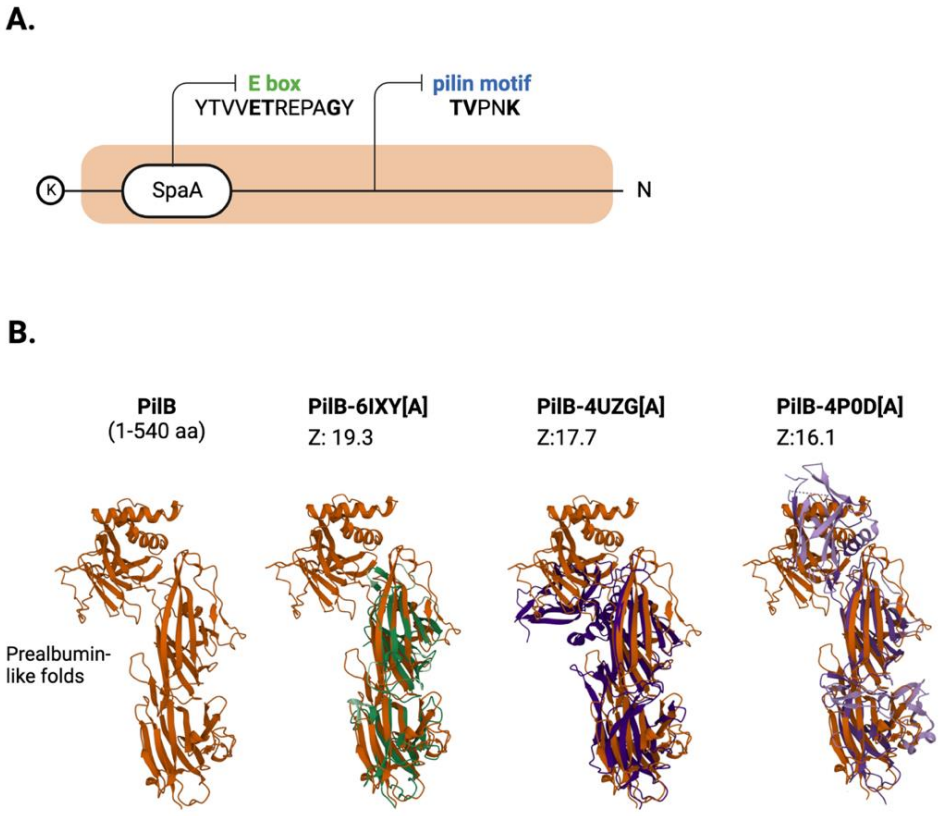


Figure 3. Domain architecture and structural homology of PilB

A. Representation of the domain architecture of PilB. The protein is anchored to the peptidoglycan via its C-terminal sortase anchoring motif (prior to polymerisation), represented here as an encircled 'K'. The N-proximal SpaA functional domain is depicted with the location of the E-box and pilin motifs indicated. **B.** Mapping of the predicted structure of PilB to the experimentally determined major pilin structures of Clostridial (PDB: 6IXY) and Streptococcal origin (PDB: 4UZG, 4P0D) demonstrates their tertiary structural similarity.

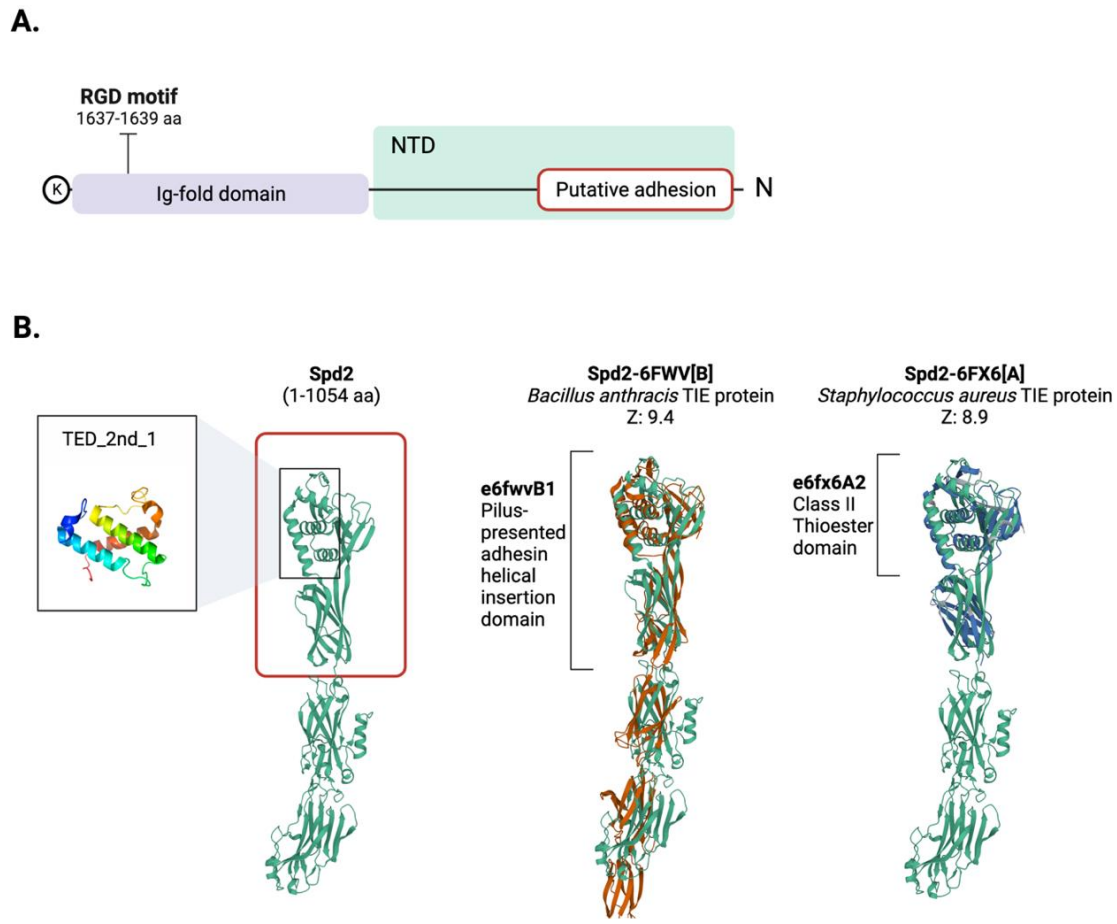


Figure 4. Domain architecture and structural homology of Spd2

A. Representation of the domain architecture of Spd2. The protein is anchored to the peptidoglycan via it's C-terminal sortase anchoring motif, represented here as an encircled 'K'. The C-terminal Ig-like fold domain and location of the RGD motif are indicated. The putative adhesion domain within the N-proximal region is depicted. **B.** Mapping of the predicted structure of PilB to the experimentally determined Class II thioester domains (TED) of the *Bacillus anthracis* TIE protein (PDB: 6FWV) and the *Staphylococcus aureus* SaTIE-TED protein (PDB: 6FX6) demonstrates their tertiary structural similarity.

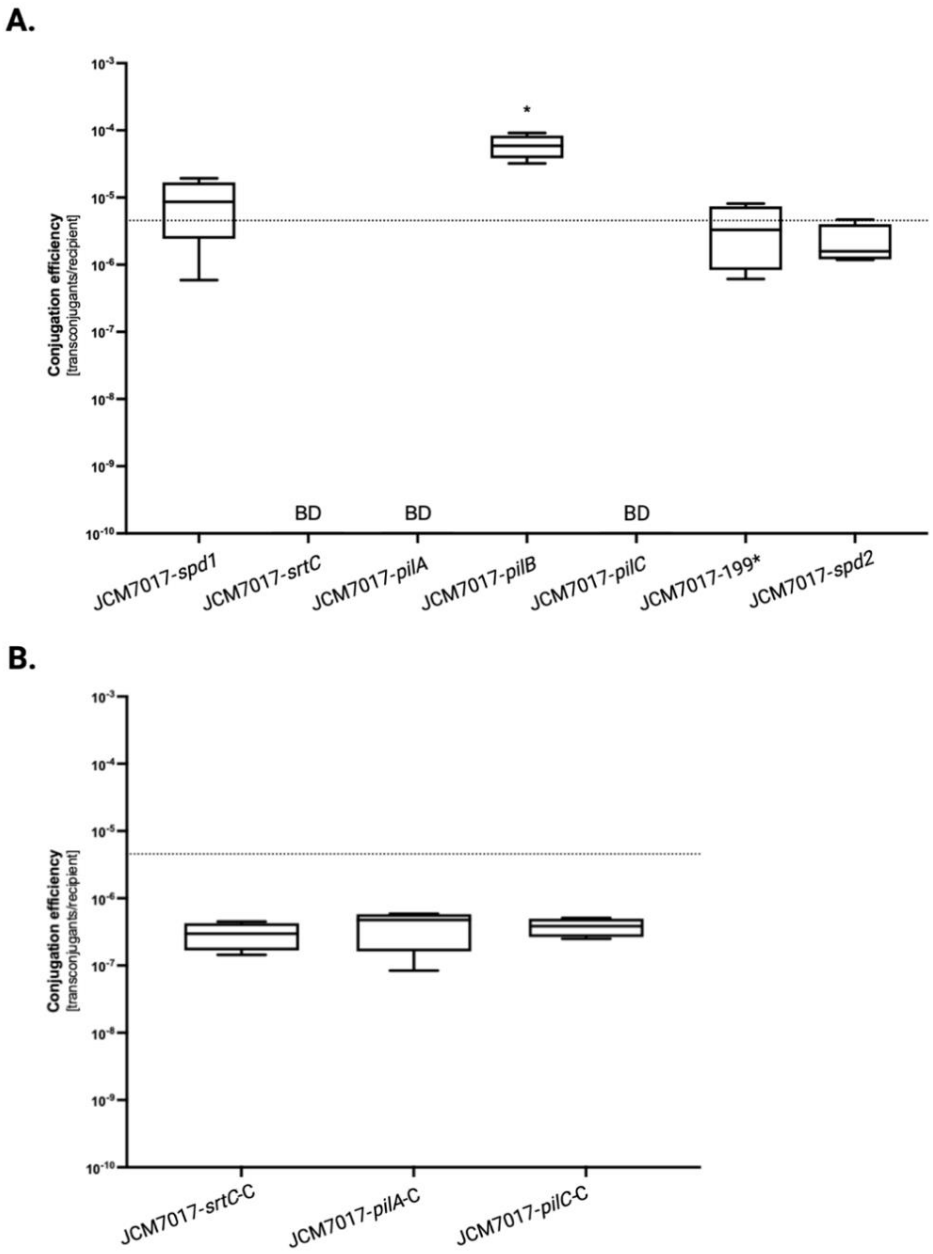
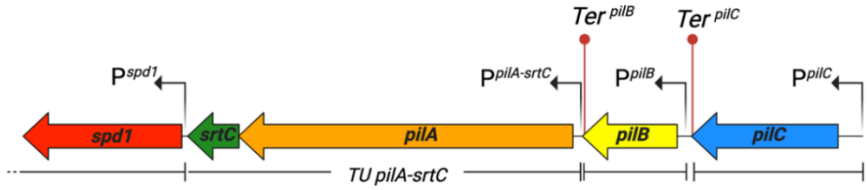


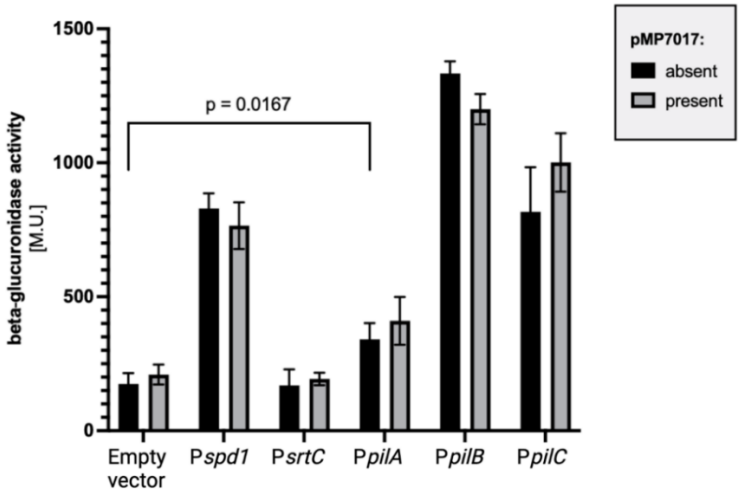
Figure 5. Mutation and complementation analysis of pMP7017 genes encoding surface proteins

A. Conjugation efficiencies of donor strains harbouring *spd1, srtC, pilA, pilB, pilC* and *spd2* mutations with recipient strain *B. breve* UCC2003. BD = below the limit of detection (10^{-10}). Conjugation between the *B. breve* JCM7017-199 donor and *B. breve* UCC2003 was used as a control. A p-value < 0.05 was considered significant and is represented with a single asterisk "∗". B. Transfer efficiencies of the *srtC, pilA* and *pilC* mutant strains containing their corresponding genetic complement cloned in plasmid pBbrU. *B. breve* UCC2003 was used as a recipient in all mating assays.

A.



B.



C.

Terminator	Start 5'	End 3'	MFE (kcal/mol)	Probability (%)
pilB	17 529	17 451	42.5	89.6
pilC	19 509	19 450	20.5	96.8

Figure 6. Transcriptional profile of the pMP7017-adhesion locus

A. Genetic organisation of the pMP7017 adhesion locus. Promoters are indicated as black bent arrows and terminators as red lollipops. **B.** Beta-glucuroridase activity per OD600, given in miller units (MU) of the enzyme activity was quantified by pNPG assay detailed in the experimental procedures, and the data are mean ± SD from three biological replicates.

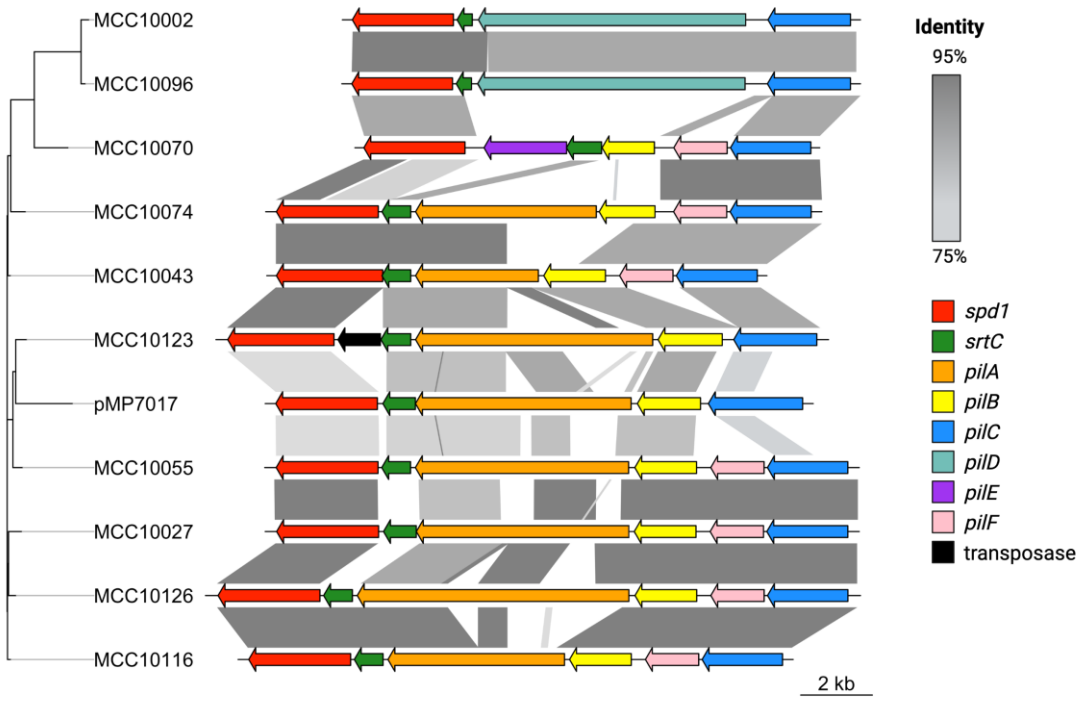


Figure 7. Schematic comparative representation of the *spd1-srtC-pil*ABC putative adhesion loci among pMP7017-homologs from various strains of *B. longum* subsp. *longum*

Each arrow indicates an open reading frame, the size of which is proportional to the length of the arrow. Colouring of the arrows represents gene function based on distinct functional domains and/or structural homologies of the encoded proteins, as outlined in the text and detailed in Supplementary Table S1. The percentage of pairwise average nucleotide identity between adjacent clusters is indicated by the grey blocks, where the strength of the colour positively correlates with an increase in identity. The *spd1* and *pilC* genes delineate the cluster in all cases and are well conserved. A *srtC* homolog is present in all loci, however, the size of the gene varies, as well as its location within the cluster. The *pilA* gene encoding the putative tip-pilin displays the most diversity and a homolog is not present in three of the loci analysed, and where present the size of the gene varies substantially between clusters. The MCC10002, MCC10096 clusters encode a functionally comparable alternative tip pilin, PilD. Similarly, the MCC10070 locus contains an alternative presumed adhesion, *pilE*. In seven clusters a gene encoding an additional presumed basal pilin is present, *pilF*. While a homologous *pilB* gene is absent from the MCC10002 and MCC10096 clusters.

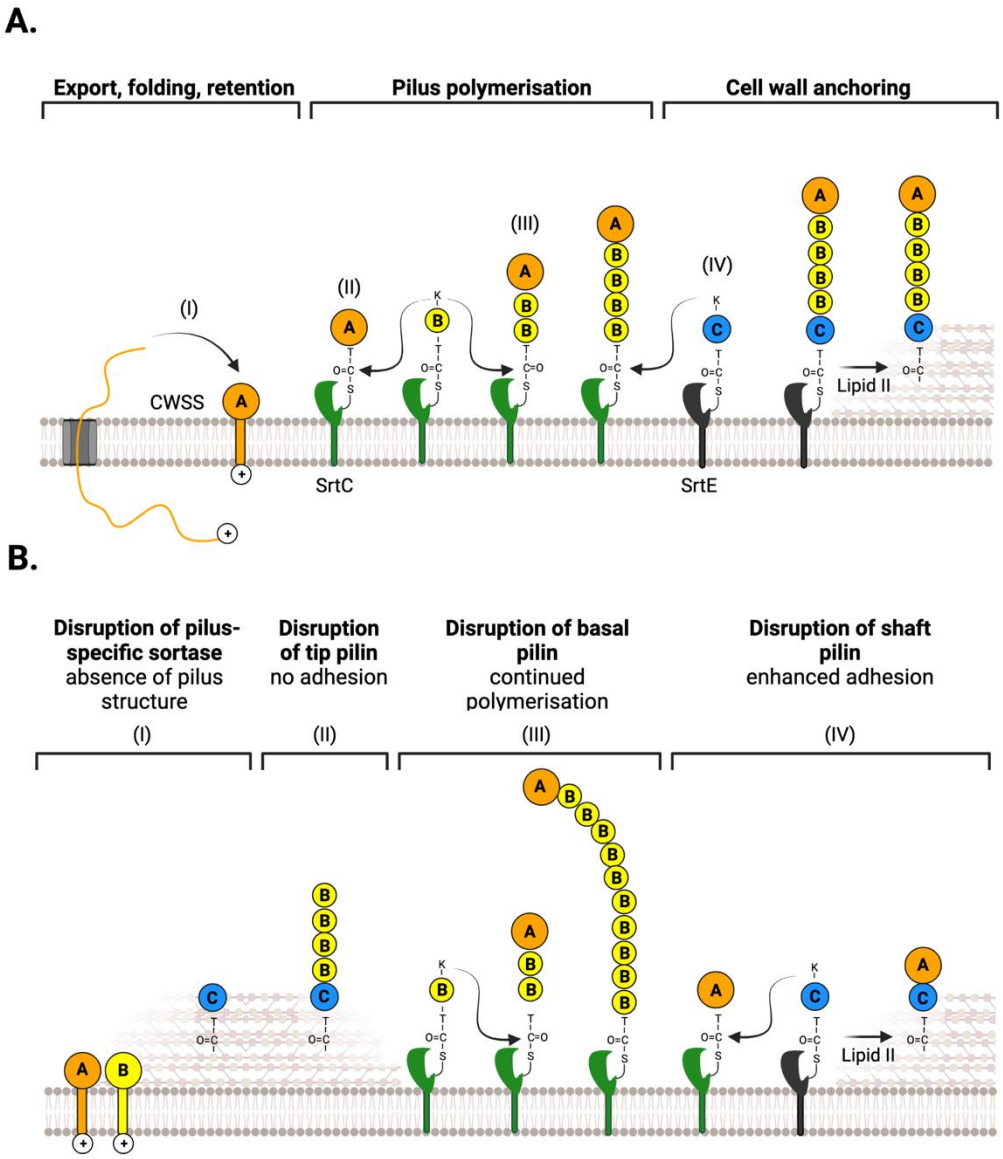


Figure 8. Mechanism of polymerisation and cell-wall anchoring of the PilABC structure and proposed phenotypes resulting from site-directed mutagenesis of individual pilin subunits

A. Proposed mechanism of cell wall anchoring and pilus assembly for the pMP7017 adhesion locus. Prior to pilus polymerisation both the sortase and substrate are both membrane bound. The Class C pilus-specific sortase, SrtC, forms individual acyl-enzyme intermediates with the presumed tip pilin, PilA, and presumed major pilin, PilB. This occurs via nucleophilic attack by the active site cysteine residue of the sortase on the threonine's carbonyl carbon atom within the sorting signal motif (**LP**x**TG**) of the pilins. (II) A lysine residue (K) within the pilin motif (**TV**pn**K**) of PilB then performs a nucleophilic attack on the thioacyl intermediate of the adjacent PilA, forming a new PilA–PilB isopeptide bond that covalently links the pilin

proteins. (III) This process is repeated to create an isopeptide-linked pilus shaft that contains multiple pilin proteins. (IV) Recognition of the sorting signal motif (**LA**x**TG**) of the presumed basal pilin, PilC, by the housekeeping sortase, SrtE, results in a SrtE-PilC thioacyl intermediate. Similar to step (II) the lysine residue (K) within the pilin motif (**TV**ed**K**) of PilC then performs a nucleophilic attack on the thioacyl intermediate of the adjacent isopeptide-linked PilA–PilB structure. The crossbridge peptide from a lipid II molecule nucleophilically attacks the thioacyl intermediate and a new peptide bond is then formed between the lipid II molecule and structure to produce a protein–lipid II intermediate that is incorporated into cell wall by the transglycosylation and transpeptidation reactions. **B.** See text for details.

Table 1. Strains and plasmids used in this study

Strains or plasmid	Relevant genotype or phenotype	Source or reference
Strains		
Escherichia coli strains		
EC101	Cloning host, *repA*$^{+}$, (Km^{r})	[42]
EC101 pORI19	EC101 harbouring pORI19, (Km^{r}, Em^{r})	[42]
EC101 pORI19_*spd1*	EC101 harbouring pORI19_*spd1*(Km^{r}, Em^{r})	this study
EC101 pORI19_tet_*spd1*	EC101 harbouring pORI19_tet_*spd1*, (Km^{r}, Em^{r}, Tet^{r})	this study
EC101 pORI19_*srtC*	EC101 harbouring pORI19_*srtC*, (Km^{r}, Em^{r})	this study
EC101 pORI19_tet_*srtC*	EC101 harbouring pORI19_tet_*srtC*, (Km^{r}, Em^{r}, Tet^{r})	this study
EC101 pORI19_*pilA*	EC101 harbouring pORI19_*pilA*, (Km^{r}, Em^{r})	this study
EC101 pORI19_tet_*pilA*	EC101 harbouring pORI19_tet_*pilA*, (Km^{r}, Em^{r}, Tet^{r})	this study
EC101 pORI19_*pilB*	EC101 harbouring pORI19_*pilB*, (Km^{r}, Em^{r})	this study
EC101 pORI19_tet_*pilB*	EC101 harbouring pORI19_tet_*pilB*, (Km^{r}, Em^{r}, Tet^{r})	this study
EC101 pORI19_*pilC*	EC101 harbouring pORI19_*pilC*, (Km^{r}, Em^{r})	this study
EC101 pORI19_tet_*pilC*	EC101 harbouring pORI19_tet_*pilC*, (Km^{r}, Em^{r}, Tet^{r})	this study
EC101 pORI19_*spd2*	EC101 harbouring pORI19_*spd2*, (Km^{r}, Em^{r})	this study
EC101 pORI19_tet_*spd2*	EC101 harbouring pORI19_tet_*spd2*, (Km^{r}, Em^{r}, Tet^{r})	this study
XL1 blue	RecA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tet^{r})]	Stratagene
XL1 blue pC21-22	XL1 blue harbouring pC21-22, (Tet^{r}, Cm^{r})	this study
XL1 blue pC23	XL1 blue harbouring pC23, (Tet^{r}, Cm^{r})	this study
XL1 blue pC24	XL1 blue harbouring pC24, (Tet^{r}, Cm^{r})	this study
Bifidobacterium breve strains		
JCM7017_199	JCM7017 harbouring pMP7017-pORI_tet_199 insertion mutant, (Tet^{r})	[23]

UCC2003	Human intestinal isolate, plasmid-free	APC - University College Cork
JCM7017-*spd1*	JCM7017 harbouring pMP7017-pORI19-tet-*spd1* insertion mutant, (Tet^r)	this study
JCM7017-*srtC*	JCM7017 harbouring pMP7017-pORI19-tet-*srtC* insertion mutant, (Tet^r)	this study
JCM7017-*pilA*	JCM7017 harbouring pMP7017-pORI19-tet-*pilA* insertion mutant, (Tet^r)	this study
JCM7017-*pilB*	JCM7017 harbouring pMP7017-pORI19-tet-*pilB* insertion mutant, (Tet^r)	this study
JCM7017-*pilC*	JCM7017 harbouring pMP7017-pORI19-tet-*pilC* insertion mutant, (Tet^r)	this study
JCM7017-*spd2*	JCM7017 harbouring pMP7017-pORI19-tet-*spd2* insertion mutant, (Tet^r)	this study
JCM7017-*srtC*-C	JCM7017-*srtC* harbouring pC21-22, (Tet^r, Cm^r)	this study
JCM7017-*pilA*-C	JCM7017-*pilA* harbouring pC21-22, (Tet^r, Cm^r)	this study
JCM7017-*pilC*-C	JCM7017-*pilC* harbouring pC24, (Tet^r, Cm^r)	this study
JCM7017-62C62	JCM7017-62 harbouring pC24, (Tet^r, Cm^r)	this study
JCM7017-P*spd1*	JCM7017-114 harbouring pP*spd1*, (Tet^r, Cm^r)	this study
JCM7017-199 pNZ_*Pspd1*	JCM7017-199 harbouring pNZ_*Pspd1*, (Cm^r)	this study
JCM7017-199 pNZ_*PsrtC*	JCM7017-199 harbouring pNZ_*Pspd1*, (Cm^r)	this study
JCM7017-199 pNZ_*PpilA*	JCM7017-199 harbouring pNZ_*Pspd1*, (Cm^r)	this study
JCM7017-199 pNZ_*PpilB*	JCM7017-199 harbouring pNZ_*Pspd1*, (Cm^r)	this study
JCM7017-199 pNZ_*PpilC*	JCM7017-199 harbouring pNZ_*Pspd1*, (Cm^r)	this study
Lactococcus lactis strains		
NZ9000	MG1363 pepN::nisRK, host strain for nisin-inducible overexpression	[80]
NZ9000 pNZ272	NZ9000 harbouring pNZ272, Promoterless *gusA*, (Cm^r)	[58]
NZ9000 pNZ_*Pspd1*	NZ9000 harbouring pNZ272 derivative containing the *spd1* promoter, (Cm^r)	this study
NZ9000 pNZ_*PsrtC*	NZ9000 harbouring pNZ272 derivative containing the *srtC* promoter, (Cm^r)	this study
NZ9000 pNZ_*PpilA*	NZ9000 harbouring pNZ272 derivative containing the *pilA* promoter, (Cm^r)	this study
NZ9000 pNZ_*PpilB*	NZ9000 harbouring pNZ272 derivative containing the *pilB* promoter, (Cm^r)	this study

NZ9000 pNZ_*PpilC*	NZ9000 harbouring pNZ272 derivative containing the *pilC* promoter, (Cm^r)	this study

Plasmids

pORI19	*repA*$^-$, ori$^+$, cloning vector, (Em^r)	[42]
pAM5	pBC1-puC19, (Tet^r)	[81]
pORI19_*spd1*	Internal bp fragment of B7017_p0020 cloned in pORI19 (Em^r)	this study
pORI-tet-B7017_*spd1*	*tet*(W) from pAM5 cloned in pORI19_*spd1* (Em^r, Tet^r)	this study
pORI19_*srtC*	Internal bp fragment of B7017_p0021 cloned in pORI19 (Em^r)	this study
pORI-tet-B7017_*srtC*	*tet*(W) from pAM5 cloned in pORI19_*srtC*(Em^r, Tet^r)	this study
pORI19_*pilA*	Internal bp fragment of B7017_p0022 cloned in pORI19 (Em^r)	this study
pORI-tet-B7017_*pilA*	*tet*(W) from pAM5 cloned in pORI19_*pilA* (Em^r, Tet^r)	this study
pORI19_*pilB*	Internal bp fragment of B7017_p0023 cloned in pORI19 (Em^r)	this study
pORI-tet-B7017_*pilB*	*tet*(W) from pAM5 cloned in pORI19_*pilB* (Em^r, Tet^r)	this study
pORI19_*pilC*	Internal bp fragment of B7017_p0024 cloned in pORI19 (Em^r)	this study
pORI-tet-B7017_*pilC*	*tet*(W) from pAM5 cloned in pORI19_*pilC* (Em^r, Tet^r)	this study
pORI19_*spd2*	Internal bp fragment of B7017_p0200 cloned in pORI19 (Em^r)	this study
pORI-tet-B7017_*spd2*	*tet*(W) from pAM5 cloned in pORI19_*spd2* (Em^r, Tet^r)	this study
pBbrU	pBC1.2 containing BbrUII.M from UCC2003, (Cm^r, Amp^r)	this study
pC21-22	pBbrU containing B7017_p0022-B7017_p0021 from pMP7017, (Cm^r, Amp^r)	this study
pC24	pBbrU containing B7017_p0024 from pMP7017, (Cm^r, Amp^r)	this study
pNZ272	Promoterless *gusA*, (Cm^r)	[58]
pNZ_P*spd1*	pNZ272 containing the B7017_p0020 promoter, (Cm^r)	this study
pNZ_P*srtC*	pNZ272 containing the B7017_p0021 promoter, (Cm^r)	this study
pNZ_P*pilA*	pNZ272 containing the B7017_p0022 promoter, (Cm^r)	this study
pNZ_P*pilB*	pNZ272 containing the B7017_p0023 promoter, (Cm^r)	this study
pNZ_P*pilC*	pNZ272 containing the B7017_p0024 promoter, (Cm^r)	this study

Table 2. List of primers used in this study

Purpose	Primer	Sequence (5' - 3')
B. breve UCC2003 strain confirmation	Bbr_1118F	gcacttccaaggcttcgtcg
	Bbr_1118R	cgagagaccatccgacgctt
B. breve JCM7017 strain confirmation	0735F	cagaacatgccggtcgagga
	0735R	gcattctcaggacgaagcgc
pMP7017 transconjugant confirmation	pMP_114F	caacagccgcaccatcgac
	pMP_114R	actgctccgcgtatggttct
522 bp fragment of *spd1* in pORI19	20mut_H3	aattaaAAGCTTccttccttgacgcggatctcg
	20mut_Xba	catcacTCTAGAggtgaaggccgcccagtcg
605 bp fragment of *srtC* in pORI19	21mut_H3	aattaaAAGCTTggtcgggcatgctgttgcg
	21mut_Xba	gttgacTCTAGAgcaggcgcgcggcgtcag
546 bp fragment of *pilA* in pORI19	22mut_H3	aattaaAAGCTTtgaagctgacggtccactgg
	22mut_Xba	catcacTCTAGAgcttccgcctctcctacagg
570 bp fragment of *pilB* in pORI19	23mut_H3	aattaaAAGCTTcgtgtggacggtgttgttcg
	23mut_Xba	catcacTCTAGAggcaacgcgaacatcatgaag
516 bp fragment of *pilC* in pORI19	24mut_H3	aattaaAAGCTTtcttctcggtgctgacgg
	24mut_Xba	catcacTCTAGAtgagccgcacatattcgc
425 bp fragment of *spd2* in pORI19	200mut_H3	aattaaAAGCTTcgaggagaagaccgtcag
	200mut_Xba	gtagacTCTAGAcggcggatgtggaggtga
tet(W) amplification	tetW_F	tcagctgtcgacatgctcatgtacggtaag
	tetW_R	gcgacggtcgaccattaccttctgaaacat
Confirmation of site-specific homologous recombination	20conf	gcaacagcatgcccgaccaag
	21conf	ctcgcctgcctgatcatcg
	22conf	cgtcccgggtcttgtagg
	23conf	gtgtggacggtgttgttcg

	24conf	gtcaacatcggcaaatac
	200conf	catcagcaagaagtggtt
	FinalConf_tet	gttgcggacatacacat
Cloning of *srtC-pilA* in pBbrU for complementation	21_PspX1_R	gagtccCCTCGAGGtcaggacaagtgtttacccg
	22_Cla1_F	acttacATCGATcacacgtcccggcccgactg
Cloning *pilC* in pBbrU for complementation	24_Cla1_F	gagtccATCGATcgctgattccgaagaacgctag
	24_PspX1_R	atctacCCTCGAGGagtcactcggacttcgatgcct
Cloning of the *srtC* promoter in pNZ272	21PP_BgLII	catcacAGATCTgacgcgaaggagcccaagg
	21PP_Pst1	catgagCTGCAGcctggcatgccttccggcc
Cloning of the *pilA* promoter in pNZ272	22PP_BgLII	catcacAGATCTgcgttcgcgttctggctgg
	22PP_Pst1	gtgtatCTGCAGcatcctcatcgtgccgtacg
Cloning of the *pilB* promoter in pNZ272	23PP_BgLII	catcacAGATCTccgagtgacttatggggcgatg
	23PP_Pst1	catgagCTGCAGgtatgggaaattgtctgattg
Cloning of the *pilC* promoter in pNZ272	24PP_BgLII	caagaaAGATCTctgattccgaagaacgctag
	24PP_Pst1	catgagCTGCAGtcgaggttttctcaaacac

Restriction sites are represented in uppercase bold text.

Table 3. Pairwise comparison of the predicted structures of candidate pMP7017-encoded surface proteins and experimentally determined target structure

Reference	Target	C	Description	Organism	Z-score	RMSD (Å)	TM-score	SI (%)	ER	TC (%)	DI	Topology	Family
Spd1	6Z9L	A	PrgA	*Enterococcus faecalis*	16.5	3.05	0.64	15	672	90	N/A		
Spd1 CAP	4D53	A	BB0689 CAP	*Borreliella burgdorferi* B31	12	2.31	0.5	21	113	85	e4d53A1	PR-1-like	CAP
	6Z9K	A	PrgA CAP	*Enterococcus faecalis*	8.9	3.04	0.47	16	139	58	e5jysA1	PR-1-like	CAP_1
	5JYS	A	PRY1 CAP	Saccharomyces cerevisiae S288C	12.8	2.69	0.53	27	122	87	e5jysA1	PR-1-like	CAP_1
PilA CTD	2y5p	D	B-repeat of InlB	*Listeria monocytogenes* EGD	8.3	2	0.2	31	70	97	e2y5pA1	Ubiquitin-like	Flg_new
	2y5p	A	B-repeat of InlB	*Listeria monocytogenes* EGD	8.2	2.03	0.2	23	71	99	e2y5pB1	Ubiquitin-like	Flg_new
	2y5p	B	B-repeat of InlB	*Listeria monocytogenes* EGD	8	2.46	0.2	21	73	100	e2y5pC1	Ubiquitin-like	Flg_new
	2y5p	C	B-repeat of InlB	*Listeria monocytogenes* EGD	8	2.55	0.2	24	72	100	e2y5pD1	Ubiquitin-like	Flg_new
PilA CBM2	6GED	A	PrgB adhesion domain	*Enterococcus faecalis*	12.9	3.03	0.58	11	205	69	e6gedA1	V-region SA I/II	GbpC
	2WD6	A	Variable Domain of SspB	*Streptococcus gordonii*	13	3.07	0.58	7	206	68	e2wd6A1	V-region SA I/II	GbpC
PilA CBM1	6A98	A	aromatic prenyltransferase	*Fischerella ambigua*	10.1	3.06	0.55	12	141	70	e6a98A1	GBD-like	UNK_F_TYPE
	6A9F	A	aromatic prenyltransferase	*Mastigocladus laminosus* UTEX LB 1931	10.3	3.1	0.55	12	142	71	e6a9fA1	GBD-like	UNK_F_TYPE
	2Y6G	A	xylanase	*Rhodothermus marinus*	11.5	3.08	0.52	13	132	79	e2y6gA1	GBD-like	CBM_4_9
	2Y6H	A	xylanase	*Rhodothermus marinus*	11.5	3.08	0.52	13	132	79	e2y6hA1	GBD-like	CBM_4_9
	6A8X	A	aromatic prenyltransferase	*Fischerella ambigua*	10.2	3	0.55	12	141	69	e6a8xA1	GBD-like	UNK_F_TYPE
	6A99	C	aromatic prenyltransferase	*Mastigocladus laminosus* UTEX LB 1931	10.2	3.04	0.55	12	142	71	e6a99C1	GBD-like	UNK_F_TYPE

PilA CBM3	2ZEW	B	Family 16 CBM_1	*Caldanaerobius polysaccharolyticus*	9.3	3.09	0.53	16	121	82	e2zewB1	GBD-like	CBM_4_9
	2ZEY	A	Family 16 CBM	*Caldanaerobius polysaccharolyticus*	9.6	3.12	0.53	16	121	83	e2zeyA1	GBD-like	CBM_4_9
	3OEA	A	CBM16-1	*Caldanaerobius polysaccharolyticus*	9.3	3.1	0.53	16	121	83	e3oeaA1	GBD-like	CBM_4_9
	4XUP	E	CMB22-1/2 domains of Endo-1,4-beta-xylanase C	*Paenibacillus barcinonensis*	10.7	3.12	0.55	9	127	80	e4xupE1	GBD-like	CBM_4_9
	4XUO	A	CBM22-1 xylan-binding domain	*Paenibacillus barcinonensis*	10.7	3.13	0.55	9	127	81	e4xuoA1	GBD-like	CBM_4_9
PilB	6IXY	A-D	major pilin	*Clostridium perfringens* SM101	19.3	5.15	0.43	16	312	70	e6ixyA3	Prealbumin-like	SdrD_B_1
	4UZG	A	SPB1 major pilus	*Streptococcus agalactiae* A909	17.7	3.11	0.37	18	242	87	e4uzgA1	Prealbumin-like	SdrD_B_1
	4P0D	A	T6 major pilin	*Streptococcus pyogenes* MGAS10394	16.1	3.24	0.41	16	272	58	e4p0dA1	Prealbumin-like	EUF07739
Spd2 NTD	6FWV	B	Collagen Adhesion protein BaTIE	*Bacillus anthracis*	9.7	4.58	0.46	8	325	63	e6fwvB4	Prealbumin-like	SdrD_B_1
	6FX6	A	SaTIE-TED	*Staphylococcus aureus*	8.9	3.6	0.28	10	190	80	e6fwvB1	Pilus-presented adhesin	TED_2nd_1

C = Chain, **SI** = Sequence Identity, **ER** = Equivalent Residues, **TC** = Target Coverage, **DI** = Domain Identifier (ECOD Database), **CBM** = Carbohydrate Binding Module, **GBD** = Galactose Binding Domain.

SUPPLEMENTAL MATERIAL

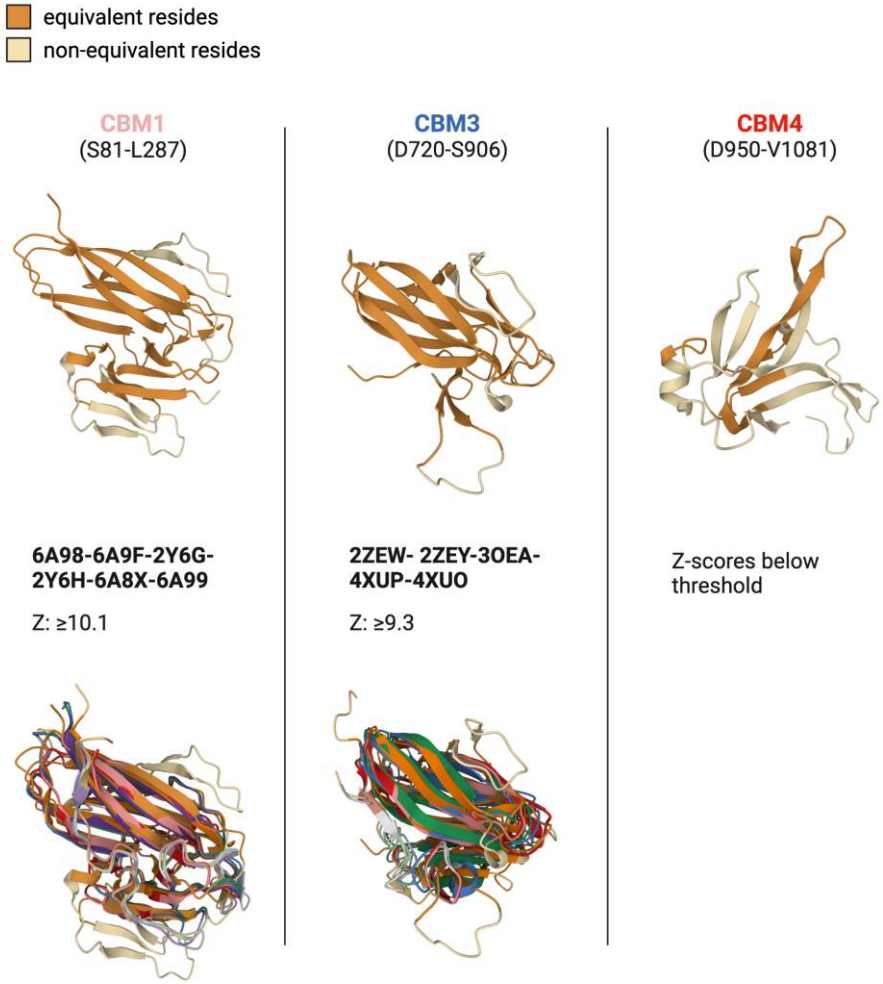


Figure S1. Structural homology of the PilA N-terminal subdomains, CBM1 and CBM3, with the carbohydrate binding modules of various carbohydrate-active enzymes

The predicted tertiary structures of the PilA NTD subdomains, CBM1, CBM3 and CBM4 are displayed (top) and coloured by the equivalency of their modelled resides with experimentally determined the structures (**Table 3**), orange signifies matched residues while off-white denotes non-equivalent residues. The CBM4 subdomain did not produces structural matches above the required thresholds, however the structure is provided for completeness.

Supplementary Table S1. Functional analysis of the alternative pilin proteins encoded within the adhesion *loci* of megaplasmids from *B. longum* subsp. *longum*

Protein	Sorting motif	E-box	Pilin motif	Domain	Accession	E-value
PilD	LPxTG	Y	Y	FctA SF	PF12892	2.45E-22
				FctA SF	PF12892	4.13E-22
				FctA SF	PF12892	1.96E-21
				FctA SF	PF12892	3.42E-18
				FctA SF	PF12892	1.21E-15
				FctA SF	PF12892	1.68E-15
				FctA SF	PF12892	2.62E-13
				FctA SF	PF12892	2.15E-11
				FctA SF	PF12892	3.03E-10
				SdrD_B SF	PF17210	1.55E-04
PilE	LPxTG	Y	N	Flg_new SF	cl09714	3.19E-07
				SGNH_hydrolase SF	cl01053	1.91E-05
				choice_anch_A SF	cl18868	3.17E-03
PilF	LAxTG	N	Y	No domains identified	-	-

REFERENCES

[1] **M. Ventura *et al.***, 'Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum', *Microbiol. Mol. Biol. Rev.*, vol. 71, no. 3, pp. 495–548, Sep. 2007, doi: 10.1128/MMBR.00005-07.

[2] **I. Zanotti *et al.***, 'Evidence for cholesterol-lowering activity by Bifidobacterium bifidum PRL2010 through gut microbiota modulation', *Appl. Microbiol. Biotechnol.*, vol. 99, no. 16, pp. 6813–6829, Aug. 2015, doi: 10.1007/s00253-015-6564-7.

[3] **P. C. Hagen and J. W. Skelley**, 'Efficacy of Bifidobacterium Species in Prevention of Necrotizing Enterocolitis in Very-Low Birth Weight Infants. A Systematic Review', *J. Pediatr. Pharmacol. Ther. JPPT*, vol. 24, no. 1, pp. 10–15, 2019, doi: 10.5863/1551-6776-24.1.10.

[4] **E. Abdolalipour *et al.***, 'Evaluation of the antitumor immune responses of probiotic Bifidobacterium bifidum in human papillomavirus-induced tumor model', *Microb. Pathog.*, vol. 145, p. 104207, Aug. 2020, doi: 10.1016/j.micpath.2020.104207.

[5] **D. Jakubczyk, K. Leszczyńska, and S. Górska**, 'The Effectiveness of Probiotics in the Treatment of Inflammatory Bowel Disease (IBD)—A Critical Review', *Nutrients*, vol. 12, no. 7, p. 1973, Jul. 2020, doi: 10.3390/nu12071973.

[6] **Z. Faghfoori, M. H. Faghfoori, A. Saber, A. Izadi, and A. Yari Khosroushahi**, 'Anticancer effects of bifidobacteria on colon cancer cell lines', *Cancer Cell Int.*, vol. 21, no. 1, Art. no. 1, Dec. 2021, doi: 10.1186/s12935-021-01971-3.

[7] **L. Ruiz, A. Margolles, and B. Sánchez**, 'Bile resistance mechanisms in Lactobacillus and Bifidobacterium', *Front. Microbiol.*, vol. 4, 2013, Accessed: Sep. 26, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2013.00396

[8] **P. Schumann**, '5 - Peptidoglycan Structure', in *Methods in Microbiology*, vol. 38, F. Rainey and A. Oren, Eds. Academic Press, 2011, pp. 101–129. doi: 10.1016/B978-0-12-387730-7.00005-X.

[9] **E. Foroni *et al.***, 'Genetic analysis and morphological identification of pilus-like structures in members of the genus Bifidobacterium', *Microb. Cell Factories*, vol. 10 Suppl 1, p. S16, Aug. 2011, doi: 10.1186/1475-2859-10-S1-S16.

[10] **F. Turroni *et al.***, 'Role of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in modulating bacterium-host interactions', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 27, Art. no. 27, Jul. 2013, doi: 10.1073/pnas.1303897110.

[11] **F. Turroni *et al.***, 'Expression of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in response to environmental gut conditions', *FEMS Microbiol. Lett.*, vol. 357, no. 1, pp. 23–33, Aug. 2014, doi: 10.1111/1574-6968.12509.

[12] **Y. Xiong, Z. Zhai, Y. Lei, B. Xiao, and Y. Hao**, 'A Novel Major Pilin Subunit Protein FimM Is Involved in Adhesion of Bifidobacterium longum BBMN68 to Intestinal Epithelial Cells', *Front. Microbiol.*, vol. 11, 2020, Accessed: Sep. 23, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2020.590435

[13] **T. Shanmugasundarasamy, D. Karaiyagowder Govindarajan, and K. Kandaswamy**, 'A review on pilus assembly mechanisms in Gram-positive and Gram-negative bacteria', *Cell Surf.*, vol. 8, p. 100077, Apr. 2022, doi: 10.1016/j.tcsw.2022.100077.

[14] **H. Ton-That, L. A. Marraffini, and O. Schneewind**, 'Sortases and pilin elements involved in pilus assembly of Corynebacterium diphtheriae', *Mol. Microbiol.*, vol. 53, no. 1, pp. 251–261, 2004, doi: 10.1111/j.1365-2958.2004.04117.x.

[15] **L. El Mortaji, R. Terrasse, A. Dessen, T. Vernet, and A. M. Di Guilmi**, 'Stability and Assembly of Pilus Subunits of Streptococcus pneumoniae*', *J. Biol. Chem.*, vol. 285, no. 16, pp. 12405–12415, Apr. 2010, doi: 10.1074/jbc.M109.082776.

[16] **T. Spirig, E. M. Weiner, and R. T. Clubb**, 'Sortase enzymes in Gram-positive bacteria', *Mol. Microbiol.*, vol. 82, no. 5, pp. 1044–1059, Dec. 2011, doi: 10.1111/j.1365-2958.2011.07887.x.

[17] **A. Susmitha, H. Bajaj, and K. Madhavan Nampoothiri**, 'The divergent roles of sortase in the biology of Gram-positive bacteria', *Cell Surf.*, vol. 7, p. 100055, Dec. 2021, doi: 10.1016/j.tcsw.2021.100055.

[18] **N. Goessweiner-Mohr, K. Arends, W. Keller, and E. Grohmann**, 'Conjugation in Gram-Positive Bacteria', *Microbiol. Spectr.*, vol. 2, no. 4, Aug. 2014, doi: 10.1128/microbiolspec.PLAS-0004-2013.

[19] **A. Schmitt *et al.***, 'PrgB Promotes Aggregation, Biofilm Formation, and Conjugation through DNA Binding and Compaction', *Mol. Microbiol.*, vol. 109, no. 3, pp. 291–305, Aug. 2018, doi: 10.1111/mmi.13980.

[20] **S. A. Revitt-Mills, T. D. Watts, D. Lyras, V. Adams, and J. I. Rood**, 'The ever-expanding tcp conjugation locus of pCW3 from Clostridium perfringens', *Plasmid*, vol. 113, p. 102516, Jan. 2021, doi: 10.1016/j.plasmid.2020.102516.

[21] **C. Gago-Córdoba *et al.***, 'Surface Exclusion Revisited: Function Related to Differential Expression of the Surface Exclusion System of Bacillus subtilis Plasmid pLS20', *Front. Microbiol.*, vol. 10, 2019, doi: 10.3389/fmicb.2019.01502.

[22] **A. Schmitt *et al.***, 'Enterococcal PrgA Extends Far Outside the Cell and Provides Surface Exclusion to Protect against Unwanted Conjugation', *J. Mol. Biol.*, vol. 432, no. 20, pp. 5681–5695, Sep. 2020, doi: 10.1016/j.jmb.2020.08.018.

[23] **F. Bottacini *et al.***, 'Discovery of a Conjugative Megaplasmid in Bifidobacterium breve', *Appl. Environ. Microbiol.*, vol. 81, no. 1, pp. 166–176, Jan. 2015, doi: 10.1128/AEM.02871-14.

[24] **M. Bhatty *et al.***, 'Enterococcus faecalis pCF10-encoded surface proteins PrgA, PrgB (aggregation substance) and PrgC contribute to plasmid transfer, biofilm formation and virulence', *Mol. Microbiol.*, vol. 95, no. 4, pp. 660–677, Feb. 2015, doi: 10.1111/mmi.12893.

[25] **J. J. Almagro Armenteros *et al.***, 'SignalP 5.0 improves signal peptide predictions using deep neural networks', *Nat. Biotechnol.*, vol. 37, no. 4, Art. no. 4, Apr. 2019, doi: 10.1038/s41587-019-0036-z.

[26] **A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer**, 'Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes', *J. Mol. Biol.*, vol. 305, no. 3, pp. 567–580, Jan. 2001, doi: 10.1006/jmbi.2000.4315.

[27] **S. El-Gebali *et al.***, 'The Pfam protein families database in 2019', *Nucleic Acids Res.*, vol. 47, no. D1, pp. D427–D432, Jan. 2019, doi: 10.1093/nar/gky995.

[28] **A. Marchler-Bauer and S. H. Bryant**, 'CD-Search: protein domain annotations on the fly', *Nucleic Acids Res.*, vol. 32, no. Web Server issue, pp. W327–W331, Jul. 2004, doi: 10.1093/nar/gkh454.

[29] **M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger**, 'ColabFold: making protein folding accessible to all', *Nat. Methods*, vol. 19, no. 6, Art. no. 6, Jun. 2022, doi: 10.1038/s41592-022-01488-1.

[30] **L. Holm and C. Sander**, 'Protein Structure Comparison by Alignment of Distance Matrices', *J. Mol. Biol.*, vol. 233, no. 1, pp. 123–138, Sep. 1993, doi: 10.1006/jmbi.1993.1489.

[31] **L. Holm**, 'Using Dali for Protein Structure Comparison', in *Structural Bioinformatics: Methods and Protocols*, Z. Gáspári, Ed. New York, NY: Springer US, 2020, pp. 29–42. doi: 10.1007/978-1-0716-0270-6_3.

[32] **Y. Zhang and J. Skolnick**, 'Scoring function for automated assessment of protein structure template quality', *Proteins Struct. Funct. Bioinforma.*, vol. 57, no. 4, pp. 702–710, 2004, doi: 10.1002/prot.20264.

[33] **M. Egan *et al.***, 'Staying alive: growth and survival of Bifidobacterium animalis subsp. animalis under in vitro and in vivo conditions', *Appl. Microbiol. Biotechnol.*, vol. 102, no. 24, pp. 10645–10663, Dec. 2018, doi: 10.1007/s00253-018-9413-7.

[34] **J. C. D. Man, M. Rogosa, and M. E. Sharpe**, 'A Medium for the Cultivation of Lactobacilli', *J. Appl. Bacteriol.*, vol. 23, no. 1, pp. 130–135, 1960, doi: 10.1111/j.1365-2672.1960.tb00188.x.

[35] **D. G. Anderson and L. L. McKay**, 'Simple and rapid method for isolating large plasmid DNA from lactic streptococci.', *Appl. Environ. Microbiol.*, vol. 46, no. 3, pp. 549–552, Sep. 1983.

[36] **P. Álvarez-Martín, A. B. Flórez, A. Margolles, G. del Solar, and B. Mayo**, 'Improved Cloning Vectors for Bifidobacteria, Based on the Bifidobacterium catenulatum pBC1 Replicon', *Appl Env. Microbiol*, vol. 74, no. 15, Art. no. 15, Aug. 2008, doi: 10.1128/AEM.00074-08.

[37] **D. J. O'sullivan and T. R. Klaenhammer**, 'Rapid Mini-Prep Isolation of High-Quality Plasmid DNA from Lactococcus and Lactobacillus spp', *Appl. Environ. Microbiol.*, vol. 59, no. 8, pp. 2730–2733, Aug. 1993, doi: 10.1128/AEM.59.8.2730-2733.1993.

[38] **J. Sambrook, E. F. Fritsch, and T. Maniatis**, 'Molecular cloning: a laboratory manual.', *Mol. Cloning Lab. Man.*, no. Ed. 2, Art. no. Ed. 2, 1989, Accessed: Mar. 19, 2019. [Online]. Available: https://www.cabdirect.org/cabdirect/abstract/19901616061

[39] **M. O'Connell Motherway, J. O'Driscoll, G. F. Fitzgerald, and D. Van Sinderen**, 'Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in Bifidobacterium breve UCC2003', *Microb. Biotechnol.*, vol. 2, no. 3, Art. no. 3, May 2009, doi: 10.1111/j.1751-7915.2008.00071.x.

[40] **J. Sambrook, E. Fritsch, and T. Maniatis**, *Molecular cloning: A laboratory manual: Vol. 2*, 2. ed. S.l.: Cold Spring Harbor, 1989.

[41] **A. Mazé, M. O'Connell-Motherway, G. F. Fitzgerald, J. Deutscher, and D. van Sinderen**, 'Identification and characterization of a fructose phosphotransferase system in Bifidobacterium breve UCC2003', *Appl. Environ. Microbiol.*, vol. 73, no. 2, pp. 545–553, Jan. 2007, doi: 10.1128/AEM.01496-06.

[42] **J. Law, G. Buist, A. Haandrikman, J. Kok, G. Venema, and K. Leenhouts**, 'A system to generate chromosomal mutations in Lactococcus lactis which allows fast analysis of targeted genes', *J. Bacteriol.*, vol. 177, no. 24, pp. 7011–7018, Dec. 1995, doi: 10.1128/jb.177.24.7011-7018.1995.

[43] **A. Sangrador-Vegas, C. Stanton, D. van Sinderen, G. F. Fitzgerald, and R. P. Ross**, 'Characterization of plasmid pASV479 from Bifidobacterium pseudolongum subsp. globosum and its use for expression vector construction', *Plasmid*, vol. 58, no. 2, pp. 140–147, Sep. 2007, doi: 10.1016/j.plasmid.2007.02.004.

[44] **F. Sievers *et al.***, 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Mol. Syst. Biol.*, vol. 7, p. 539, Oct. 2011, doi: 10.1038/msb.2011.75.

[45] **G. Stecher, K. Tamura, and S. Kumar**, 'Molecular Evolutionary Genetics Analysis (MEGA) for macOS', *Mol. Biol. Evol.*, vol. 37, no. 4, pp. 1237–1239, Apr. 2020, doi: 10.1093/molbev/msz312.

[46] **C. Camacho *et al.***, 'BLAST+: architecture and applications', *BMC Bioinformatics*, vol. 10, p. 421, Dec. 2009, doi: 10.1186/1471-2105-10-421.

[47] **L. Guy, J. Roat Kultima, and S. G. E. Andersson**, 'genoPlotR: comparative gene and genome visualization in R', *Bioinformatics*, vol. 26, no. 18, pp. 2334–2335, Sep. 2010, doi: 10.1093/bioinformatics/btq413.

[48] 'R: The R Project for Statistical Computing'. https://www.r-project.org/ (accessed Nov. 24, 2020).

[49] **G. M. Gibbs, K. Roelants, and M. K. O'Bryan**, 'The CAP Superfamily: Cysteine-Rich Secretory Proteins, Antigen 5, and Pathogenesis-Related 1 Proteins—Roles in Reproduction, Cancer, and Immune Defense', *Endocr. Rev.*, vol. 29, no. 7, pp. 865–897, Dec. 2008, doi: 10.1210/er.2008-0032.

[50] **R. Schneiter and A. D. Pietro**, 'The CAP protein superfamily: function in sterol export and fungal virulence', *Biomol. Concepts*, vol. 4, no. 5, pp. 519–525, Oct. 2013, doi: 10.1515/bmc-2013-0021.

[51] **M. Marino, M. Banerjee, R. Jonquières, P. Cossart, and P. Ghosh**, 'GW domains of the Listeria monocytogenes invasion protein InlB are SH3-like and mediate binding to host ligands', *EMBO J.*, vol. 21, no. 21, pp. 5623–5634, Nov. 2002, doi: 10.1093/emboj/cdf558.

[52] **M. Ebbes, W. M. Bleymüller, M. Cernescu, R. Nölker, B. Brutschy, and H. H. Niemann**, 'Fold and function of the InlB B-repeat', *J. Biol. Chem.*, vol. 286, no. 17, pp. 15496–15506, Apr. 2011, doi: 10.1074/jbc.M110.189951.

[53] **C. Geerds, W. M. Bleymüller, T. Meyer, C. Widmann, and H. H. Niemann**, 'A recurring packing contact in crystals of InlB pinpoints functional binding sites in the internalin domain and the B repeat', *Acta Crystallogr. Sect. Struct. Biol.*, vol. 78, no. Pt 3, pp. 310–320, Mar. 2022, doi: 10.1107/S2059798322000432.

[54] **B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat**, 'The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics', *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D233-238, Jan. 2009, doi: 10.1093/nar/gkn663.

[55] **L. Makowski, W. Olson-Sidford, and J. W. Weisel**, 'Biological and Clinical Consequences of Integrin Binding via a Rogue RGD Motif in the SARS CoV-2 Spike Protein', *Viruses*, vol. 13, no. 2, p. 146, Jan. 2021, doi: 10.3390/v13020146.

[56] **C. Gago-Córdoba *et al.***, 'A Conserved Class II Type Thioester Domain-Containing Adhesin Is Required for Efficient Conjugation in Bacillus subtilis', *mBio*, vol. 12, no. 2, pp. e00104-21, Mar. 2021, doi: 10.1128/mBio.00104-21.

[57] **M. Cronin, M. Knobel, M. O'Connell-Motherway, G. F. Fitzgerald, and D. van Sinderen**, 'Molecular Dissection of a Bifidobacterial Replicon', *Appl. Environ. Microbiol.*, vol. 73, no. 24, Art. no. 24, Dec. 2007, doi: 10.1128/AEM.01630-07.

[58] **C. Platteeuw, G. Simons, and W. M. de Vos**, 'Use of the Escherichia coli beta-glucuronidase (gusA) gene as a reporter gene for analyzing promoters in lactic acid bacteria.', *Appl. Environ. Microbiol.*, vol. 60, no. 2, Art. no. 2, Feb. 1994.

[59] **T. Odamaki *et al.***, 'Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan', *Sci. Rep.*, vol. 8, Jan. 2018, doi: 10.1038/s41598-017-18391-x.

[60] **B. Henderson, S. Nair, J. Pallas, and M. A. Williams**, 'Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins', *FEMS Microbiol. Rev.*, vol. 35, no. 1, pp. 147–200, Jan. 2011, doi: 10.1111/j.1574-6976.2010.00243.x.

[61] **F. Chaffanel *et al.***, 'Surface proteins involved in the adhesion of Streptococcus salivarius to human intestinal epithelial cells', *Appl. Microbiol. Biotechnol.*, vol. 102, no. 6, pp. 2851–2865, 2018, doi: 10.1007/s00253-018-8794-y.

[62] **E. Josefsson, D. O'Connell, T. J. Foster, I. Durussel, and J. A. Cox**, 'The Binding of Calcium to the B-repeat Segment of SdrD, a Cell Surface Protein of Staphylococcus aureus *', *J. Biol. Chem.*, vol. 273, no. 47, pp. 31145–31152, Nov. 1998, doi: 10.1074/jbc.273.47.31145.

[63] **T. Das *et al.***, 'Influence of Calcium in Extracellular DNA Mediated Bacterial Aggregation and Biofilm Formation', *PLOS ONE*, vol. 9, no. 3, p. e91935, Mar. 2014, doi: 10.1371/journal.pone.0091935.

[64] **M. M. King, B. B. Kayastha, M. A. Patrauchan, and M. J. Franklin**, 'Calcium Regulation of Bacterial Virulence', *Adv. Exp. Med. Biol.*, vol. 1131, pp. 827–855, 2020, doi: 10.1007/978-3-030-12457-1_33.

[65] **A. Mølgaard, S. Kauppinen, and S. Larsen**, 'Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases', *Struct. Lond. Engl. 1993*, vol. 8, no. 4, pp. 373–383, Apr. 2000, doi: 10.1016/s0969-2126(00)00118-0.

[66] **A. C. Anderson, S. Stangherlin, K. N. Pimentel, J. T. Weadge, and A. J. Clarke**, 'The SGNH hydrolase family: a template for carbohydrate diversity', *Glycobiology*, vol. 32, no. 10, pp. 826–848, Jul. 2022, doi: 10.1093/glycob/cwac045.

[67] **K. G. Anthony, C. Sherburne, R. Sherburne, and L. S. Frost**, 'The role of the pilus in recipient cell recognition during bacterial conjugation mediated by F-like plasmids', *Mol. Microbiol.*, vol. 13, no. 6, pp. 939–953, Sep. 1994, doi: 10.1111/j.1365-2958.1994.tb00486.x.

[68] **T. J. Foster, J. A. Geoghegan, V. K. Ganesh, and M. Höök**, 'Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus', *Nat. Rev. Microbiol.*, vol. 12, no. 1, pp. 49–62, Jan. 2014, doi: 10.1038/nrmicro3161.

[69] **C. Raynaud, D. Sheppard, J.-L. Berry, I. Gurung, and V. Pelicic**, 'PilB from Streptococcus sanguinis is a bimodular type IV pilin with a direct role in adhesion', *Proc. Natl. Acad. Sci.*, vol. 118, no. 22, p. e2102092118, Jun. 2021, doi: 10.1073/pnas.2102092118.

[70] **P. E. Kolenbrander**, 'Intergeneric Coaggregation Among Human Oral Bacteria and Ecology of Dental Plaque', *Annu. Rev. Microbiol.*, vol. 42, no. 1, pp. 627–656, 1988, doi: 10.1146/annurev.mi.42.100188.003211.

[71] **A. H. Rickard, P. Gilbert, N. J. High, P. E. Kolenbrander, and P. S. Handley**, 'Bacterial coaggregation: an integral process in the development of multi-species biofilms', *Trends Microbiol.*, vol. 11, no. 2, pp. 94–100, Feb. 2003, doi: 10.1016/S0966-842X(02)00034-3.

[72] **R. G. Ledder, A. S. Timperley, M. K. Friswell, S. Macfarlane, and A. J. McBain**, 'Coaggregation between and among human intestinal and oral bacteria', *FEMS Microbiol. Ecol.*, vol. 66, no. 3, pp. 630–636, Dec. 2008, doi: 10.1111/j.1574-6941.2008.00525.x.

[73] **E. Vlková, V. Rada, M. Smehilová, and J. Killer**, 'Auto-aggregation and co-aggregation ability in bifidobacteria and clostridia', *Folia Microbiol. (Praha)*, vol. 53, no. 3, pp. 263–269, 2008, doi: 10.1007/s12223-008-0040-z.

[74] **A. Datta, F. Stapleton, and M. D. P. Willcox**, 'Bacterial Coaggregation and Cohesion Among Isolates From Contact Lens Cases', *Invest. Ophthalmol. Vis. Sci.*, vol. 59, no. 7, pp. 2729–2735, Jun. 2018, doi: 10.1167/iovs.17-23155.

[75] **L. J. Cohen *et al.***, 'Unraveling function and diversity of bacterial lectins in the human microbiome', *Nat. Commun.*, vol. 13, no. 1, Art. no. 1, Jun. 2022, doi: 10.1038/s41467-022-29949-3.

[76] **C. Chang *et al.***, 'Cell-to-cell interaction requires optimal positioning of a pilus tip adhesin modulated by gram-positive transpeptidase enzymes', *Proc. Natl. Acad. Sci.*, vol. 116, no. 36, pp. 18041–18049, Sep. 2019, doi: 10.1073/pnas.1907733116.

[77] **H. Ton-That and O. Schneewind**, 'Assembly of pili on the surface of Corynebacterium diphtheriae', *Mol. Microbiol.*, vol. 50, no. 4, pp. 1429–1438, Nov. 2003, doi: 10.1046/j.1365-2958.2003.03782.x.

[78] **B. E. Uhlin, M. Norgren, M. Båga, and S. Normark**, 'Adhesion to human cells by Escherichia coli lacking the major subunit of a digalactoside-specific pilus-adhesin', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 82, no. 6, pp. 1800–1804, Mar. 1985, doi: 10.1073/pnas.82.6.1800.

[79] **J. L. Telford, M. A. Barocchi, I. Margarit, R. Rappuoli, and G. Grandi**, 'Pili in Gram-positive pathogens', *Nat. Rev. Microbiol.*, vol. 4, no. 7, Art. no. 7, Jul. 2006, doi: 10.1038/nrmicro1443.

[80] **P. G. de Ruyter, O. P. Kuipers, and W. M. de Vos**, 'Controlled gene expression systems for Lactococcus lactis with the food-grade inducer nisin', *Appl. Environ. Microbiol.*, vol. 62, no. 10, pp. 3662–3667, Oct. 1996, doi: 10.1128/AEM.62.10.3662-3667.1996.

[81] **P. Alvarez-Martín, M. O'Connell-Motherway, D. van Sinderen, and B. Mayo**, 'Functional analysis of the pBC1 replicon from Bifidobacterium catenulatum L48', *Appl. Microbiol. Biotechnol.*, vol. 76, no. 6, pp. 1395–1402, Oct. 2007, doi: 10.1007/s00253-007-1115-5.

CHAPTER V

Transcriptional Landscape of the pMP7017 Megaplasmid and its Impact on the *Bifidobacterium breve* UCC2003 Transcriptome

ABSTRACT

The 190 kb megaplasmid pMP7017 of *Bifidobacterium breve* JCM7017 represents the first conjugative and largest plasmid characterised within this genus to date. In the current study, we adopted an integrated approach combining transcriptome data analysis, whole genome comparative analysis and metagenomic data mining to understand the biology of pMP7017 and related megaplasmids, and to assess the impact of plasmid-carriage on the host strain. Data generated revealed variations within basic features of promoter elements which correlate with a high level of transcription on the plasmid and highlights the transcriptional activity of genes encoding both offensive and defensive adaptations, including a Type IIL restriction-modification system, an anti-restriction system and four Type II toxin-antitoxin systems. Furthermore, a highly transcribed tmRNA, which likely provides translational support to the host strain, was identified, making pMP7017 the first plasmid of the *Bifidobacterium* genus and the smallest plasmid known to express a tmRNA. Analyses of synteny and variability among pMP7017 and related plasmids indicates substantial diversity in gene organisation and accessory gene cargo highlighting diverse (co-)evolution and potential host-specific rearrangements and adaptations. Systematic analysis of the codon usage profile of transcriptionally active pMP7017-encoded genes suggests that pMP7017 originated from (sub)species of *Bifidobacterium longum*. Furthermore, mining of metagenomic data suggests the presence of pMP7017-homologs in ~10% of microbiome samples, mostly infant and/or mothers from various geographical locations. Comparative transcriptome analysis of the *B. breve* UCC2003 chromosome in the presence or absence of pMP7017, revealed differential expression of genes representing 8% of the total gene pool. Genes involved in genetic information processing were exclusively upregulated, while altered expression of genes involved in biofilm production and polysaccharide biosynthesis was also observed.

INTRODUCTION

Bifidobacterium is a genus of high GC% gram-positive (G+) anaerobic bacteria, characterised by their Y-shaped or "bifid" morphology from which they derived their name [1]. Bifidobacteria are ubiquitous inhabitants of the gastrointestinal tract of mammals and constitute one of the dominant genera of the intestinal microbiota. They possess diverse saccharolytic capabilities with an estimated 13.7% of their protein coding capacity dedicated to carbohydrate metabolism, which is presumed to be essential for their organismal lifestyle and species-specific adaptions to their respective ecological niches [2]. Probiotics belonging to *Bifidobacterium* spp. have been exploited in the food industry, as prophylactics and therapeutic agents, due to their purported contribution to the maintenance of a balanced microbiota and the plethora of health-promoting attributes with which they have been associated [3]–[8].

While plasmids are common among many intestinal bacterial species, plasmids of bifidobacterial origin were once considered rare as only small cryptic plasmids had been isolated from five species [9]–[17]. However, this assumption was invalidated upon the discovery of a 190 kb conjugative megaplasmid, pMP7017, in host strain *B. breve* JCM7017 and the subsequent identification of pMP7017-homologs in fifteen *B. longum* subsp. *longum* strains [18], [19].

Horizontal gene transfer (HGT) meditated by conjugative plasmids is considered an important driver of genome diversification and ecological adaption in bacteria due to the transfer of accessory genes between different host strains, while also providing a reservoir of genetic variation offering bacteria an excellent platform for rapid functional and evolutionary innovations. More recently, plasmids have been linked to the altered expression of host genes and the manipulation of bacterial phenotypes, with mounting evidence suggesting that plasmid-mediated alterations to host gene expression may affect niche-adaptive fitness [20]–[27].

As such, 'omics' approaches are fundamental in disentangling the plasmid-host relationship and shedding light on the potential regulatory networks, costs of plasmid carriage and potential fitness trade-offs. Furthermore, analysis of transcriptomics data, which constitute the complete set of RNAs (mRNAs and ncRNAs) transcribed from the genome of a given cell at a particular time-point, such

as that derived from high-throughput RNA sequencing (RNAseq), generates indispensable information on gene functionality, level of expression and regulation.

While transcription in bacteria can be modulated in a number of ways, via *cis*-acting elements which alter the affinity of the RNA-polymerase (RNAP) holoenzyme for the promoter or by *trans*-acting regulatory elements which may prevent or enhance promoter binding, variability within the basal features of promoters underpins all transcription regulation [28]. Bacterial promoters are modular in structure, consisting of two RNAP recognition sites, these hexameric sequences are located at the -10 and -35 positions upstream of the first transcribed base (+1), known as the transcriptional start site (TSS), and are separated by a spacer of 15-19 bps whose sequence is generally not conserved. As observed for other bacteria, promoter strength in bifidobacteria has been indicated to correlate with the degree of identity to the consensus -10 and -35 sequences, as well as the length of the spacer [29]. Comparative analysis of promoter regions permits the identification of sequence variations associated with promoter activity and how these variations impact on the corresponding transcription levels.

Further to promoter strength and regulatory mechanisms, the genome-wide effects of codon usage on transcription levels have been demonstrated, essentially this code within a code may not only affect peptide elongation rates but also influence transcription levels by modulating transcription efficiency and mRNA stability [30]. As the coevolution of intracellular tRNA abundance and codon usage is essential in the optimisation and efficient use of cell resources, tRNAs may therefore constitute dynamic rather than static contributors to gene expression and the codon landscape.

In the current study, we used an integrated approach of transcriptomic data analysis and functional characterisation of plasmid pMP7017 to interpret its functional genetic elements and their associated transcription levels, and to elucidate the biological function and importance of these expressed genes utilising a variety of sequence and structural homology-based tools. Further to this, we investigated if the codon usage profile of transcriptionally active genes correlates with their expected expression levels and how this relates to overall tRNA abundance. In addition, transcriptional data sets pertaining to the *B. breve* UCC2003 chromosome in the presence or absence of pMP7017 were compared to uncover the transcriptomic impact of pMP7017 acquisition by the host cell. Finally, we explore the prevalence, synteny

and variability of pMP7017 and related megaplasmids, through comparative analysis of whole genome sequences and mining metagenomic data.

MATERIALS AND METHODS

RNA isolation

B. breve UCC2003 cultures grown in MRS (Difco) supplemented with cysteine HCl and tetracycline until an OD_{600nm} of 0.6 was reached, the cells were then centrifuged at 10,000 x *g* for 2 mins at RT. The obtained cell pellet was suspended in 1 ml of QIAZOL (Qiagen, United Kingdom) and placed in a tube containing 0.8 g of glass beads (diameter, 106 μm; Sigma). The cells were lysed by shaking the mix on a BioSpec homogenizer at 4 °C for 2 min (maximum setting). The mixture was then centrifuged at 12,000 rpm for 15 min, and the RNA-containing upper phase was recovered. RNA was further purified by phenol extraction and ethanol precipitation. Quality and integrity of the RNA was checked by the Tape station 2200 (Agilent Technologies, USA) analysis. RNA concentration and purity were evaluated by Picodrop microlitre Spectrophotometer (Picodrop, UK).

RNA sequencing

2.5 μg of total RNA was treated by the Ribo-Zero Magnetic kit (Illumina) to remove ribosomal RNA, followed by purification of the rRNA-depleted sample by ethanol precipitation. RNA was further processed according to the manufacturer's instructions. The yield of rRNA depletion was checked by Tape station 2200 (Agilent Technologies). Then, 400 ng of the rRNA-depleted RNA sample was fragmented using a Bioruptor NGS ultrasonicator (Diagenode, USA) followed by size evaluation using Tape station 2200 (Agilent Technologies). A whole transcriptome library was constructed using the TruSeq Stranded RNA LT Kit (Illumina). Samples were loaded into a Flow Cell V3 150 cycles (Illumina) according to the manufacturer's instructions. The reads were depleted of adapters, quality filtered (with overall quality, quality window and length filters) and aligned to the *Bifidobacterium* and plasmid reference genomes through the Burrows-Wheeler Aligner [31]. Counting of reads whose sequences correspond to ORFs was performed using HTSeq (http://htseq.readthedocs.io/en/release_0.9.1/) and analysis of the reads per kilobase per million mapped (RPKM) values was performed using the following formula:

$$RPKM = \frac{reads\ mapped\ to\ transcript}{\frac{gene\ Length}{10^3} * total\ reads - \frac{Reads}{10^6}}$$

While the accepted threshold values for expression range in the literature from ≥1 up to ≥2 RPKM, a higher value was chosen in this work to minimise false positives, assigning less weight to false negatives for this analysis. Consequently, a RPKM threshold value for expression of ≥1.8 RPKM was established, as determined by density plot of the normalised RPKM values.

Identification of promoters and transcriptional terminators

For each expressed gene and operon, the start and end coordinates of the transcripts were defined as the first and last base of reads, where an increase or drop in sequence coverage was observed, respectively. 62 bps upstream of each identified TSSs was first searched using Meme Suite (http://meme-suite.org/) to identify the best fit to a canonical −10 and −35 promoter sequence [32]. A promoter list was then generated from the obtained predicted canonical promoter and re-aligned with Meme to ensure the correct positioning of the −10 and −35 hexamers, before being manually annotated in Artemis (http://www.sanger.ac.uk/science/tools/artemis). The WebLogo web server (https://weblogo.berkeley.edu/) was then used to visualise the level of conservation between 25 and 30 aligned promoters of highly and moderately transcribed TUs, respectively [33].

Rho-independent transcriptional terminators were first predicted using ARNold and manually refined in Artemis (http://www.sanger.ac.uk/science/tools/artemis) [34]. Meme Suite (http://meme-suite.org/) was used to identify stem-loop structures and the strength of each GC-rich stem-loop was confirmed using the RNAfold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) [35]. Where appropriate, terminators in addition to the ones predicted above were included following a manual search for the presence of polyT stretches downstream of putative stem-loop structures at the end of transcripts.

Differential expression of *B. breve* UCC2003 chromosomal genes in the presence of pMP7017

Two-way hierarchical clustering was performed using transcriptional data of *B. breve* UCC2003 chromosomal genes in the presence and absence of pMP7017 based on transcripts per million (TPM), using the following formula:

$$TPM = 10^6 x \left(\frac{\left(\frac{reads\ mapped\ to\ gene}{gene\ length}\right)}{Sum\left(\frac{reads\ mapped\ to\ gene}{gene\ length}\right)}\right)$$

From this data a heatmap with K-means clustering was generated. Using the sum of squared error (SSE) the optimum number of clusters was determined to be two, containing a total of 157 differentially expressed genes (DEGs), represented by 30 down-regulated and 127 upregulated genes, after filtering based on mean expression, fold change (fc) ≥ 2 and a P-value of < 0.05 with a false discovery rate (FDR) cut-off of $P < 0.1$ for each compared pair.

Protein families and pathway assignments were achieved utilising the BioCyc (https://biocyc.org/) and KEGG databases (https://www.genome.jp/kegg/) while enrichment analysis was performed using ShinyGO [36]–[38]. Enrichment analysis is a method of functional class scoring which utilises the log2 fold changes for all genes to determine whether gene sets for particular biological pathways are enriched among large positive or negative fold changes in transcription within the differentially expressed gene dataset relative to the reference genome.

Prediction of RNAs

Prediction of tRNAs, sRNAs and RNA-based regulatory elements was performed using the RNAspace web server (http://www.rnaspace.org/) and the structRNAfinder webserver (https://structrnafinder.integrativebioinformatics.me/) [39], [40]. Sequences producing hits with both webtools were then cross-referenced and verified using the Rfam database (http://rfam.xfam.org/) [41]. Identified tRNAs which had conflicting annotated amino acid specificities were annotated as 'tRNA_OTHER'.

Indices of codon usage

Multivariate analysis of codon usage and the corresponding indices of codon usage for all datasets were determined using the program CodonW (J Peden, version 1.4.2 http://codonw.sourceforge.net/). The relative synonymous codon usage (RSCU) is calculated as the observed frequency of a codon divided by the expected frequency. RSCU values range from 0 (codon absent), through 1 (no bias) to 4 (single codon used out of a four-codon family), i.e. values >1 indicate that a codon was used more frequently than expected, where the converse is true for values <1. The Codon Adaptation Index (CAI) and Codon Bias Index (CBI) were used to estimate the extent of bias toward codons preferred in highly expressed genes and those of the host strain genome. CAI values range from 0 to 1.0, an increase in CAI values positively correlate with an increase in codon usage bias and thus expression level or the host strain.

Comparative analysis

Putative pMP7017 proteins were aligned using Diamond with an e-value cut off of 0.0001, >50% sequence coverage and >50% of identity cut-off [42]. The R package, circlize, was then employed to produce a circular heatmap [43], illustrating the percentage of identity of each pMP7017 protein against 43 RepB positive Refseq strains [44]. Reads from metagenomic samples were mapping using Bowtie2 and normalised using RPKM [45]. Reassembly of metagenomes was conducted using metaSPAdes [46].

RESULTS

Transcription-mapping reveals features of DNA regulatory elements which correlate with transcription levels in pMP7017

To identify genes and operons that are actively transcribed in pMP7017 during logarithmic growth and to connect features of promoter elements with levels of transcription, we analysed the pMP7017 transcriptome and assessed the basic features of transcriptionally active promoters and transcription termination sites, employing a combination of high-throughput RNA sequencing (RNAseq) analysis and promoter mapping.

From our analysis highly transcribed genes of pMP7017 possess a promoter that best resembles the accepted canonical bifidobacterial consensus ($TTGACAn_{17}TATAAT$) [29]. The highest level of conservation observed within these motifs are residues A in the 2nd position and T in the 6th position of the -10 box and T in the 2nd position and G in the 3rd position of the -35 box, with an optimum spacer of 17 bps (**Figure 1A**). When comparing promoter sequences from highly transcribed vs those from moderately transcribed TUs we observed a more degenerate promoter consensus in the latter case, in addition to a higher occurrence of a ±1 deviation from the optimal spacer distance of 17 bps (**Figure 1B**). This data is consistent with similar transcriptional profiling and promoter analyses of the *B. breve* UCC2003 genome [29].

Analysis of the approximate 3' end of transcripts based on the obtained RNAseq data suggests that Rho-independent termination represents the more frequently adopted method of transcription termination of highly transcribed genes and operons by pMP7017 (**Figure 1C**). Convergent transcriptional termination is the second most frequent, whereby RNA polymerase conflicts result in efficient bidirectional termination which mark transcript boundaries, a method of transcription termination which is considered pervasive in bacteria [47]. Transcriptional units with moderate levels of expression appear to favour stem-loop structures, mRNA hairpins characterised as GC rich inverted repeat sequences. These structures differ from Rho-independent termination signals due to the absence of a poly U tract (in nascent RNA) and have been suggested as intrinsic termination signals as they can allosterically alter the structure of the RNA polymerase resulting in its release [48]. In addition, examination of the 5' and 3' untranslated regions (UTRs) of highly expressed vs

moderately expressed genes indicate that the 5'-UTR of highly expressed genes tends to be longer on average compared to those of moderately expressed genes which may serve to stabilize the transcript or indeed constitute regulatory functions (**Figure 1D**). We then analysed the genomic regions encompassing the promoters and 5'-UTRs of highly expressed genes to assess for potential regulatory motifs, revealing a conserved 34 bp motif overlapping the -10 promoter element of six genes (5'-AW**A**Y**ATG**Y**TATACTG**AATA**TG**T**C**C**ACAT**AAAAWA-3', 100% conserved residues in bold, -10 underlined). We then used the consensus motif as bait across a 60 bp region upstream of all presumed translation start sites, revealing 13 additional motifs distributed throughout the plasmid (**Figure 2A**). Based on the transcriptional data and promoter mapping in this work, these promoters are predicted to drive the expression of 56 genes, representing ~24.1% of pMP7017's coding capacity. In an effort to establish commonality of function within this presumed regulon, all genes were grouped into functional categories based on the clusters of orthologous groups (COG) classification system (**Figure 2B**). While the role(s) of 64.3% of these genes remain to be determined (S: Function Unknown), the remaining 35.7% representing 20 genes fall within seven functional categories, ordered by relative abundance; (V: Defence Systems), (L: Replication, recombination and repair), (K: Transcription), (T: Signal transduction mechanisms), (D: Cell-cycle control, cell division and partitioning), (F: Nucleotide transport and metabolism) and (M: Cell wall/membrane biogenesis). As this site appears to overlap the -10 promoter element, this motif likely represents a transcription factor binding site, the binding of which may inhibit or enhance transcription. Sequence-homology based analysis revealed the presence of this motif preceding comparable protein coding sequences within presumed pMP7017-homologs, yet this motif is absent among bifidobacterial chromosomes suggesting that it is unique to pMP7017 and related plasmids.

Transcriptional Activity of pMP7017-Genes Encoding Offensive and Defensive Systems likely Ensure its Maintenance and Enhance the Success of its Lateral Transfer

While over half of the predicted gene complement of plasmid pMP7017 is considered transcriptionally active at the mid-logarithmic growth phase of *B. breve* UCC2003, 54.2% of these expressed genes are currently functionally unassigned (**Figure 3**). As the annotation of genomes may become antiquated over time, we endeavoured to elucidate the biological function and importance of these expressed genes using a variety of sequence and structural homology-based tools to infer structural features and assign putative functional roles.

Comprehensive computational analysis of each pMP7017 coding sequence enabled us to assign a putative functional role to an additional 15 genes and further refine the functional annotation of 21 transcriptionally active genes (**Table 1**). In addition, we identified four previously unassigned, transcriptionally active Type II Toxin-Antitoxin systems, which fall within three toxin superfamilies, AtaT/TacT, HipAB and two systems of the ParE/RelE superfamily, all of which are predicted to act through translation elongation inhibition in plasmid-free cells promoting post-segregational killing (**Table 2**).

The highest level of gene transcription was observed for transposable elements, genes specifying replication functions, Type-II TA systems and genes encoding predicted surface-associated proteins. Genes predicted to specify multimer resolution functions, additional recombinases and active partitioning systems were expressed at moderate levels, ensuring the structural and segregationally stability of the plasmid. Moderate levels of transcription were also observed for genes encoding putative defence systems, including a homolog of the so-called Lar anti-restriction protein and a homologue of the Type-IIL restriction modification (RM) enzyme MmeI, corresponding to locus tags B7017_p0162 and B7017_p0174, respectively [49], [50]. While a low level of transcription was observed for locus B7017_p0197, encoding a presumed N6_N4_Mtase.

Although pMP7017 contains Type IE clustered regularly interspaced short palindromic repeats (CRISPR) and encodes their associated (*cas*) genes which together constitute an RNA-based adaptive immunity system, transcriptional activity above the assigned threshold was observed only for the CasE-specifying gene, a protein involved in spacer acquisition when complexed with Cas1, and an internally

located hypothetical gene (locus tag: B7017_p0086). Furthermore, transcription of the associated CRISPR arrays (above our assigned threshold) was not detected, suggesting the pMP7017 CRISPR/*cas* system is non-functional or dormant. Nevertheless, CRISPR spacers that are inserted between direct repeats in CRISPR arrays represent a biological record of past encounters with foreign DNA, as such CRISPR spacers provide a valuable source of information regarding invading elements-host interactions. Characterisation of the pMP7017 spacer sequences revealed two complete matches of spacers 10 and 18 to *Siphoviridae* bacteriophage genomes and partial matches (>45% identity) for 17 of the 25 remaining spacers to various actinobacterial phage sequences (**Table 3**).

Identification of pMP7017-encoded RNAs

In silico characterisation of pMP7017 upon its discovery indicated the presence of 14 putative transfer RNA (tRNA) genes flanking the pMP7017 replicase-encoding gene [18], [51]. Analysis of the transcriptional data of pMP7017 verified the presence of a tRNA array unit at the previously indicated genomic locus, however upon careful inspection this region was found to contain a total of 20 tRNA-specifying genes organised within three tandemly arranged TUs. The data further highlights markedly high transcription for seventeen of the twenty tRNA genes, which collectively correspond to 12 amino acids (**Table 4**). Furthermore, analysis of the pMP7017 transcriptome allowed the identification of two non-coding RNAs (ncRNAs) not previously reported. The first of these is a highly transcribed 397 bp transfer messenger RNA (tmRNA), which is located between the replicase-encoding gene and the tRNA region (nucleotide coordinates 120,876-121,272). tmRNAs are bifunctional chimeric molecule*s* involved in the rescue of stalled ribosomes during translation, while they also function in targeting truncated polypeptides for degradation, a system known as trans-translation [52]. The pMP7017 tmRNA ($tmRNA^{pMP7017}$) single chain 397 bp mature transcript encodes the tag peptide A^{RESUME}NTVSSSRFTLAA- which is recognised by DDE/DHE proteases for proteolytic cleavage of polypeptides produced from truncated mRNAs.

In addition, a 178 bp *manA* motif was identified (nucleotide coordinates 122,656-122,830), this conserved cis-regulatory RNA structure is thought to function as a riboswitch due to its complex architecture [53]. This motif is most commonly

found within the 5'-UTR of genes involved in mannose metabolism or adjacent to tRNAs in some cases, and while its function is not well defined, *manA* is associated with the regulation of genes involved in mannose metabolism, nucleotide synthesis and stress response [54].

pMP7017-Carriage of tRNA Genes Facilitate Configuration of the Cellular tRNA Pool to suit the Codon Usage of Plasmid-Encoded Genes

Empirical patterns of synonymous codon usage, codons that specify the same amino acid, are not used in equal frequencies across species, genomes, or individual genes. This variation has been mainly attributed to mutation biases and adaptation of codon usage to cellular tRNA abundance [55]. Highly expressed genes generally present with an optimal codon usage profile corresponding to enriched tRNA species, as genes which are highly transcribed have been subjected to stronger purifying selection [56]. The translation efficiency of plasmid-encoded genes is thus predominantly determined by its codon bias and the host tRNA pool which impacts the success of both its maintenance and transfer [57]. Horizontally transferred genes with a codon bias deviating from that of the host may sequester ribosomes on plasmid-borne gene transcripts due to reduced ribosome decoding rates, and as a consequence the host strain may incur fitness cost due to decreased ribosome availability [57], [58]. However, plasmids can mitigate these restrictions via the carriage of tRNA genes which have been speculated to facilitate the adaption of the host tRNA pool to the codon usage of plasmid genes, enabling the plasmid to possess a codon bias which in itself would have been mal-adapted to certain host backgrounds [59].

Following the identification of 17 highly transcribed pMP7017-encoded tRNAs, we aimed to establish if the codon usage signature of pMP7017 correlates with the tRNA-species it expresses and to assess how changes in global tRNA abundance due to pMP7017-expressed tRNAs may impact the host. Analysis of the codon usage profile of all predicted coding sequences of pMP7017 based on the Relative Synonymous Codon Usage (RSCU) values, we observed that the nucleotide composition at the third codon position favour Cytosine (C) or Guanine (G) in the third synonymous codon position, with a distinct preference towards C at the wobble position for both data sets, consistent with what has been observed for bifidobacterial genomes (**Figure 4**).

To ascertain if a correlation could be observed between codon usage patterns and observed levels of transcription of individual pMP7017 genes, the RSCU profile of highly transcribed genes and non-transcribed genes (those below the assumed FKPM threshold) at mid-logarithmic growth of host strain *B. breve* UCC2003 were compared. However, two-dimensional clustering of the normalised data did not distinguish a clear pattern between groups (**Figure S1**). We subsequently assessed the overall codon usage biases of highly transcribed genes vs those with a low level of transcription to determine if a correlation exists between the level of transcription and overall nucleotide composition. Highly transcribed genes were determined from this analysis to possess an overall GC content at the wobble position (GC_3) that is 14% higher than those with a low level of transcription (**Figure 5**). This bias towards G+C pairing in the third codon position of highly transcribed genes may have implications for downstream processes, as a high GC_3 content has been suggested to impact translation efficiency through enhanced mRNA stability, improved translation fidelity and increased peptide-elongation rates, ultimately culminating in increased protein production [60]. Furthermore, marginally higher values of adaption based on CAI and CBI indices (See Materials and Methods section) were observed for coding sequences of genes with a lower rate of transcription, while both datasets encode proteins with comparable properties based on *Gravy* and AROMA indices.

Taken together and as expected, these data indicate that highly transcribed genes have been subjected to stronger purifying selection, but are less adapted towards the host strain, *B. breve* JCM7017, and overall tRNA abundance. Further to this, highly transcribed pMP7017 genes possess a codon usage profile that is more like that of the *B. longum* subsp. *longum* NCIMB8809 genome. In addition, this data highlights several highly transcribed pMP7017-encoded tRNAs which correspond to rare or under-utilised codons in both the host and within the pMP7017 proteome, this may enable the codon usage of pMP7017 to be less dependent on the host tRNA pool, and/or indeed, enhance translation speed and fidelity of specific mRNAs enriched for these codons.

We subsequently investigated the relationship between pMP7017-encoded tRNAs, its associated codon usage patterns and the codon usage patterns of the host strain and the *B. longum* subsp. *longum* NCIMB8809 genome. This analysis highlights a subset of pMP7017-expressed tRNAs which specify for rare/low-usage codons (those used at <10% frequency) in the *B. breve* JCM7017 genome, specifically those

corresponding to Arg-AGA, Arg-AGG, Gly-GGG, Gly-GGA, Leu-CTT, Pro-CCA and Ser-TCA (**Supplementary Table S1**). Upon comparing the codon usage signatures based on the RSCU values, of the host strain *B. breve* JCM7017 and pMP7017, it is evident that the codon usage of pMP7017 appears to be optimised towards the pMP7017-encoded tRNA pool (P-value based on χ-squared test: 0.0013) (**Figure 6A**). Furthermore, when comparing the codon usage signatures based on the RSCU values of highly transcribed genes to the overall codon usage pattern of pMP7017, highly transcribed genes contain a higher frequency of pMP7017-encoded tRNA anti-codons than that observed for the total pMP7017 gene pool, however this not of statistical significance (**Figure 6B**).

These data suggest that pMP7017-carriage of tRNA genes may facilitate the adaption of the host tRNA pool to the codon usage of pMP7017 genes, while pMP7017-expressed tRNAs that specify for rare/low-usage codons may alleviate ineffective translation and enhance expression of target genes in the host.

pMP7017-mediated Differential Expression of *B. breve* UCC2003 Chromosomal Genes

While the molecular mechanisms underlying the cost and compensation of plasmid carriage are not well understood, recent analyses of transcriptomics data, experimental evolution and mutational analyses have shed light on the specific genetic interactions between plasmids and associated host chromosomes. To assess if pMP7017-carriage affects host fitness by inducing transcriptional changes, transcriptional data of *B. breve* UCC2003 chromosomal genes in the presence and absence of pMP7017 were compared.

From this data, a heatmap with K-means clustering was generated indicating 157 differentially expressed genes in the presence of pMP7017, represented by 30 down-regulated and 127 upregulated genes (**Figure 7**). Functional hierarchical classification based on KEGG Orthology (KO) groups within *B. breve* UCC2003 specific pathways indicates that down-regulated genes fall within eight protein families, predominantly polysaccharide biosynthesis, membrane transporters and extracellular structures, in addition to the metabolism of amino acids, fatty acids and carbohydrates, as well as genes representing hypothetical functions and mobile elements (**Figure 7**, **Table 6**). Gene enrichment analyses of these down-regulated

genes, using the *B. breve* UCC2003 total gene pool as a background, suggests an enrichment of glycosyltransferases and cell surface proteins within the 30 gene dataset (**Table 7**).

30% of down-regulated genes encode proteins involved in carbohydrate transport and metabolism, over half of which (*agl4*, *malECG* and *apuB*), are involved in malto-oligosaccharide metabolism and share the same MalR1/MalR2/MalR3 regulon [61], [62]. Significant down-regulation was also observed for part of the *eps2* operon (Bbr_0442 – Bbr_0451) representing one half of the bidirectional surface-associated exopolysaccharide (EPS) biosynthesis gene cluster [3]. In addition, gene *accD* (Bbr_1720), putatively encoding an acetyl Co-A β chain involved in the metabolism of fatty acids, which was previously identified as a gene involved in *B. breve* UCC2003 biofilm formation due to the reduction in biofilm biomass exhibited by the *accD* mutant strain compared to the WT [63]. Three down-regulated genes (locus-tags Bbr_0113, Bbr_0114 and Bbr_1888) are predicted to encode extracellular structures, of these, locus-tag Bbr_0114, encoding a SpaH/EbpB family LPXTG-anchored major pilin is the most significantly down-regulated gene with a fold change (fc) of -15.5 (P-value, 2.35^{-117}) compared to the control.

Although a substantial portion (62.2%) of up-regulated genes encode proteins of unknown function, those which could be functionally classified based on KO groups within *B. breve* UCC2003 specific pathways suggest up-regulated genes fall within 18 protein families (**Figure 7**, **Table 8**). Gene enrichment analyses of the 127 up-regulated genes using the *B. breve* UCC2003 total gene pool as a background indicates the specific enrichment of ABC transporters, genes encoding biofilm related proteins, those involved in the diaminopimelate (DAP) pathway and Tad-like proteins (**Table 9**).

DEGs within pathways of genetic information processing are exclusively up regulated, containing 12 genes encoding transcription factors, including two MerR family metal sensing transcriptional activators (Bbr_1519 and Bbr_1523) and a number of transcription repressors including the SOS response regulator, LexA, cold shock protein, CspA, and WhiB, suggested to function in the regulation of cell division. Two genes encoding XerC site-specific recombinases (Bbr_1113 and Bbr_1548) are among the most significantly upregulated genes, these proteins are known to be important in the process of chromosomal dimer resolution and promote heritable stability. While a number of upregulated genes are involved in redox

metabolism, including, Bbr_1717 encoding a Pimeloyl-ACP methyl ester carboxylesterase involved in the biosynthesis of biotin (vitamin B7), and Bbr_1717 encoding a NAD(P)H-dependent FMN reductase involved in Riboflavin (vitamin B2) metabolism, in addition to Bbr_1222 encoding a MutT/nudix family phosphohydrolase predicted to function in the removal of deleterious oxidized nucleotide derivatives and Bbr_0205 encoding a predicted polyketide synthase.

Upregulation was observed for the *pel* operon (Bbr_0044 - Bbr_0048), protein products of orthologous genes of G+ origin have been demonstrated to be responsible for the biosynthesis of Pellicle (Pel) polysaccharides [64]. Furthermore, upregulation was observed for two of the four dihydrodipicolinate synthase-encoding genes, *dapA3* and *dapA4*, indicated to catalyse the rate-limiting step within the DAP pathway, responsible for the biosynthesis of bacterial cell wall [65]. In addition, increased expression was observed of cluster (Bbr_0133 - Bbr_0137) specifying a tight-adherence (Tad) type IVc pili (T4cP) previously identified as an essential colonisation factor promoting colonic epithelial proliferation of *B. breve* UCC2003 *in vivo* [66].

Upregulated Chromosomal Genes in the Presence of pMP7017 possess a Codon Usage Profile Compatible with the Altered Cellular tRNA pool due to pMP7017-expressed tRNAs

Codon usage, an intrinsic feature within open reading frames (ORFs), has recently been documented to influence mRNA levels in fungi and eukaryotes in a translation-independent manner, revealing the emerging role of codon usage in transcription level regulation [30], [67]–[69]. As such, coding sequence composition may also have important implications for gene expression and transcriptional regulation in bacteria. To investigate the potential effect of the pMP7017-encoded tRNAs on gene expression levels in the host, we calculated the average relative synonymous codon usage (RSCU) values for each codon corresponding to pMP7017-expressed tRNA anti-codons for up- and down-regulated *B. breve* UCC2003 chromosomal gene sequences and compared these to the RSCU values for the total *B. breve* UCC2003 and pMP7017 transcriptomes. This analysis revealed that upregulated *B. breve* UCC2003 chromosomal genes contain a more congruent RSCU profile with that of the pMP7017 transcriptome and are likely more compatible with the altered tRNA pool, while down-regulated genes present with a more divergent

RSCU profile (**Figure 8**). Changes in global tRNA abundance due to pMP7017-expressed tRNAs may therefore enhance the stability of transcripts enriched for these cognate codons, or indeed, alter expression of specific transcripts through a mechanism of tRNA modulation as indicated for eukaryotes [30], [67]–[69].

Prevalence, Synteny and Variability of pMP7017 and Related Megaplasmids

Screening of pMP7017 genes against the reads of 1,846 metagenomic samples across different geographical locations revealed 206 pMP7017 positives samples with reads mapping to >40% of megaplasmid genes, however only 10 of these samples, including samples from Estonia, Finland, Japan, Russia and the UK, contained mapped pMP7017 genes above the >70% threshold (**Table 10**). Due to the low abundance and frequency of pMP7017-homologs, it was only possible to recover and reconstruct a single megaplasmid from the final metagenome-assembled genomes.

We then focused on the experimentally verified replicase specifying gene (B7017_p0146) and searched for this gene against all available *B. longum and B. breve* genomes in the Refseq database [51]. 43 of the 375 *B. longum* strains were positive for the replicase, thus were assumed to contain a pMP7017-related megaplasmid (**Figure 9**). A single *B. breve* genome was replicase-positive, *B. breve* MCC 1454 isolated from infant faeces (Biosample: SAMN03978804), however, *B. breve* MCC 1454 shares an average nucleotide identity (ANI) of >99.95% with the pMP7017 host genome *B. breve* JCM7017. As this value is above the accepted threshold for species differentiation (≥95% ANI) these genomes thus constitute the same strain, representing a positive control for the presence of pMP7017 within *B. breve* JCM7017 in the natural environment. Collectively, this data supports the supposition that *B. longum* is likely the natural host of this plasmid family due to the presence of this megaplasmid in ~11.5% of *B. longum* strains compared to ~0.19% of the 537 *B. breve* strains analysed, solely representing *B. breve* JCM7017 [19].

To assess the diversity of gene-carriage among bifidobacterial megaplasmids, fourteen *B. longum* genomes previously demonstrated to contain a pMP7017-related megaplasmid were compared (**Figure 10**) [19]. This comparative analysis revealed the conservation of 88 Markov clustering (MCL) families representing the core genome and includes genes specifying replication and partition functions, in addition to those encoding type four secretion system components. A total of 123 MCL

families, representing 53% of the pMP7017 predicted coding sequences were shared by some but not all strains, thus constituting the dispensable genome. Furthermore, 28 MCL families are unique to pMP7017 and mainly represent functionally unassigned genes.

Further to this analysis, using the established pMP7017 replicase-coding sequence and iteron containing region as bait we were able to identify four additional recently sequenced putative pMP7017 homologs in the NCBI database. When comparing these whole genome sequences to the pMP7017 reference genome, we observed varying degrees of sequence conservation, in addition to a number of frame shifts representing deletions or insertions among sequences and a large inversion (nt coordinates, 84,894-128,826) encompassing genes B7017_p0114 – B7017_p0153 inclusive was observed within the pMP7017 genome (**Figure 11**). These data indicate that the bifidobacterial megaplasmids analysed are structurally related and thus share a common ancestor. However, the diversity among accessory gene cargos as well as divergence between orthologous protein sequences suggest that these megaplasmids have experienced extensive independent evolutionary pressures.

DISCUSSION

In this work the megaplasmid pMP7017 was transcriptionally profiled, while the transcription termination sites of active transcriptional units were mapped. Using this dataset, we were able to identify a consensus promoter typical of bifidobacteria with conserved hexamers TATAAT (−10 region) and TTGACA (−35 region), while a degeneration of the promoter consensus motif is observed for genes with lower levels of transcription. In addition, the most prevalent spacer length among the deduced promoters of highly expressed genes was that of 17 bps, with a ± 1 bp deviation in spacer length observed within moderate strength promoter sequences. Together this data suggests the AT% of the −10 hexamer in addition to the length of the spacer region appear to highly impact gene transcription levels, consistent with what has been observed in the transcriptional profiling and promoter mapping of the *B. breve* UCC2003 genome [29]. As homologs of a large number of our identified genes are present among pMP7017-related plasmids, this dataset constitutes a transcriptomic reference which may be implemented for future investigations of gene expression in members of this plasmid family.

A primary motivation of systems biology research is to understand gene functions and their interactions within an overarching community or global context. Given the rapid increase in next generation sequencing data and more comprehensive databases, updating structural and functional annotations of genes as well as deciphering transcriptional profiles, are essential components in building a foundation toward this understanding. Comprehensive computational analysis of each pMP7017 coding sequence enabled the assignment of putative functional roles to 15 genes (6.6% of the pMP7017 predicted genome) and further refine the functional annotation of 21 transcriptionally active genes (9.2% of the pMP7017 predicted genome). However, 51% of transcriptionally active pMP7017 genes remain functionally unassigned, this value is substantially higher than that of the estimated 26% and 22.5% representing the average percentage of hypothetical genes among all available *B. breve* and *B. longum* genome sequences, respectively [70], [71]. Thus, considering the cost of expression to the host cell it is possible that these genes encode functionally important proteins.

Horizontal transfer of plasmid genes enables the adaptation of bacteria to diverse environmental conditions, facilitating their evolution. However, non-self-genomes impose multi-level pressures on the host chromosome with large mobile elements, such as megaplasmids, depleting cellular resources for their maintenance, horizontal spread, and expression, resulting in a reduction of host fitness. While compensatory evolution can lessen these fitness costs over time, the initial cost(s) associated with plasmid carriage may represent a constraint hindering horizontal and vertical transmission of these genetic elements, thereby reducing their chances of survival in bacterial communities. These biological conflicts are widespread among bacteria, as a result extensive conflict-related adaptations have evolved such as CRISPR/*cas* systems and RM systems in the host, as well as plasmid-encoded TA systems and anti-restriction proteins. In this work, an integrated approach of transcriptome analysis and functional characterisation of pMP7017-encoded proteins revealed transcriptional activity of both offensive and defensive adaptions, including four Type II TA systems, an anti-restriction protein and a presumed Type IIL RM system. While experimental validation of their functionality is required, pMP7017-encoded TA systems likely ensure maintenance and stability of the plasmid in the host, while methylases and/or anti-restriction proteins enhance the success of its horizontal transmission by preventing restriction of plasmid DNA upon entry into a recipient cell.

Furthermore, our analysis enabled the identification of a highly transcribed tmRNA housekeeping sRNA, this unique RNA facilitates the rescue of stalled ribosomes and promotes the degradation of defective mRNAs. Prior to this work, the smallest megaplasmid known to carry a tmRNA was that of the 567.3kb megaplasmid, pCPAA3, isolated from *Curtobacterium pusillum* AA, as such, pMP7017 now represents the smallest plasmid known from literature and the first plasmid of the *Bifidobacterium* genus harbouring a tmRNA. Although the contribution of $tmRNA^{pMP7017}$ to the biology of the megaplasmid and its interaction with the host-encoded trans-translational system remains to be uncovered, $tmRNA^{pMP7017}$ likely aids the host in mitigating toxicity effects resulting from the accumulation of truncated polypeptides and the depletion of translational ribosomes. Equally, pMP7017 may benefit from the unique functional aspects of the specific tmRNA variant it encodes or profit from the increased gene dosage.

While pMP7017 expresses a large number of tRNAs, a characteristic which has been widely suggested to confer benefits to the host, our analyses suggest that

pMP7017-carriage of tRNA genes may in fact facilitate adaption of the cellular tRNA pool to the codon usage of pMP7017-encoded genes. Interestingly, the codon usage profile of highly transcribed pMP7017 genes was established to be more similar to that of the *B. longum* subsp. *longum* genome rather than that of the host bacterium, *B. breve* JCM7017, from which it was originally derived. As both *B. breve* and *B. longum* subsp. *longum* represent common inhabitants of the infant gut, in addition to the co-occurrence of known pMP7017-homologs in both species, these data may support the supposition that an ancestorial plasmid of pMP7017 transferred to *B. breve* JCM7017 from a *B. longum* subsp. *longum* donor. Strain-specific evolution and domestication of pMP7017 may thus have resulted in its diversification. This is reflected in the observed divergence between orthologous protein sequences and a high frequency of genomic reorganization, represented by frameshifts and inversions within the pMP7017 genome relative to pMP7017-related megaplasmids of *B. longum* subsp. *longum* origin. Furthermore, considering the prevalence, albeit low, of pMP7017-related megaplasmids within this specific bacterial lineage, deeper analysis of variations among more distantly related plasmids, would serve in the understanding of the evolution of this plasmid family, revealing the impact(s) of structural variations on gene expression and potentially unveiling strain-specific adaptations.

While no clear discernible pattern was observed between transcriptionally affected chromosomal genes, it is evident that specific gene clusters and genes encoding functionally comparable proteins were similarly affected, together representing 8% of the *B. breve* UCC2003 genome. Although the biological significance of this is not yet apparent, these changes may be consequential to the altered expression of chromosomal regulators, as observed for the 21 kb pathogenicity island (PAI) of *Rhodococcus equi* [23].

It is well established that various polysaccharides have distinct physical properties and functional roles during biofilm formation, affecting both spatial organization and integration. *B. breve* UCC2003 has been demonstrated to contain a large EPS cluster containing two bidirectional EPS operons, *eps1* and *eps2*, specifying two alternative EPSs [3]. In our analysis a down regulation of the *eps2* operon in the presence of pMP7017 was observed with no statistically significant differential expression observed at the *eps1* operon. Furthermore, our data indicates an upregulation of a presumed *pel* polysaccharide biosynthesis cluster [72]. Pel polysaccharide, is a thin polysaccharide that is tightly associated with the

peptidoglycan layer which has been shown to be essential for the formation of mixed integrated biofilm communities in *P. aeruginosa* and *S. aureus* co-cultures, while a reduction of 20 – 40 μm in microcolony height was observed with a *P. aeruginosa* PAO1 *pel* mutant compared to the WT, where the authors propose that the absence of Pel may result in stiffer, more rigid biofilms [73], [74]. Altered expression of host chromosomal genes involved in the formation of surface polysaccharides and adhesion-related functions involved in biofilm production have been reported for a number bacteria upon acquisition of conjugative plasmids, which favour plasmid transfer in these spatially structured populations [20]–[22], [24], [25]. Considering this, it is tempting to speculate that these findings indicate dynamic remodelling of the *B. breve* UCC2003 cell envelope composition which could indirectly enhance horizontal transmission of pMP7017, resulting from the rheological contributions of these EPSs.

In-depth analysis of gene expression profiles which differentiate experimental groups is critical in the discovery and analysis of potential transcriptional networks between chromosomal and plasmid genes and to identify key factors which may contribute to the maintenance of pMP7017 in natural populations. Transcriptome comparison of pMP7017-free and pMP7017-containing hosts with more divergent genotypes may refine the dataset and enable the identification of host-specific and/or general impacts resulting from plasmid carriage, as observed for pCAR1 with three distinct *Pseudomonas* host strains and F-plasmid in two different *Escherichia coli* strains [75], [76]. In addition, investigation of the effects of plasmid carriage on the host cell may elucidate the mechanisms underpinning the acquisition of horizontally acquired genes and expose how pMP7017 and related plasmids adapt to new host cell backgrounds in the natural environment.

FIGURES AND TABLES

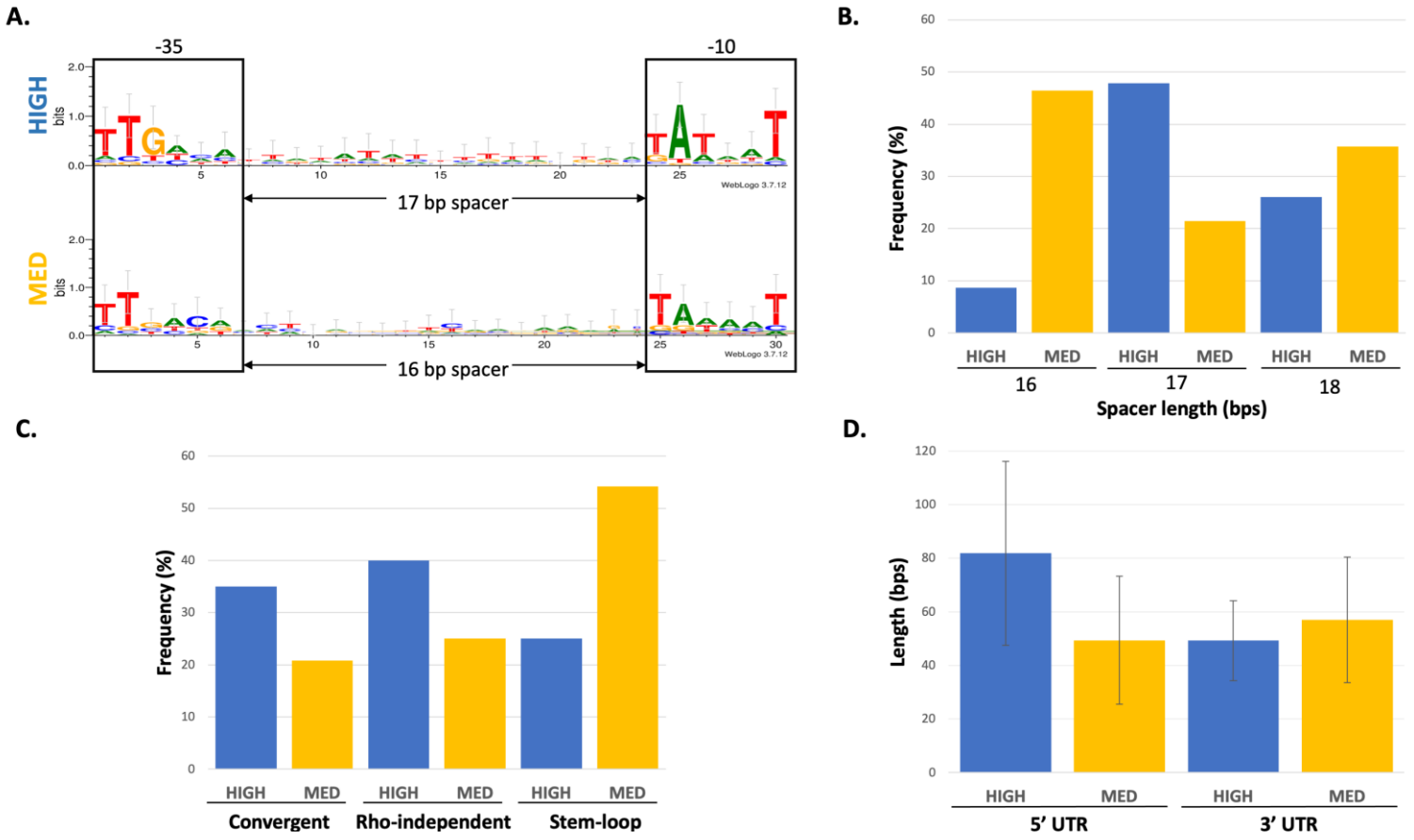


Figure 1. Intrinsic features of pMP7017 promoters, transcriptional unit elements and termination signals, as a function of transcriptional rate

A. Sequence logo showing the consensus motifs identified for the −10 and −35 sequences of the canonical promoter in pMP7017 across the levels of expression high and medium-low. **B.** Bar chart representing the optimal spacer length of highly expressed and medium expressed genes. A spacer of 17 bp is more prevalent in highly expressed genes, while a spacer of 16 bps has the highest occurrence in the promoters of genes with a lower level of expression and 17 bps is less frequently observed. **C.** Bar plot showing the different strategies for transcriptional termination identified in pMP7017, grouped by transcription level (high vs medium). The bar chart suggests a higher prevalence of Rho-independent termination in highly transcribed genes. **D.** Bar plot comparing the length of the 5'-UTR and 3'-UTR regions of genes grouped by transcription level high vs medium. From the chart it is possible to observe a higher average 5'-UTR length in highly transcribed genes as compared to gene with a lower level of transcription.

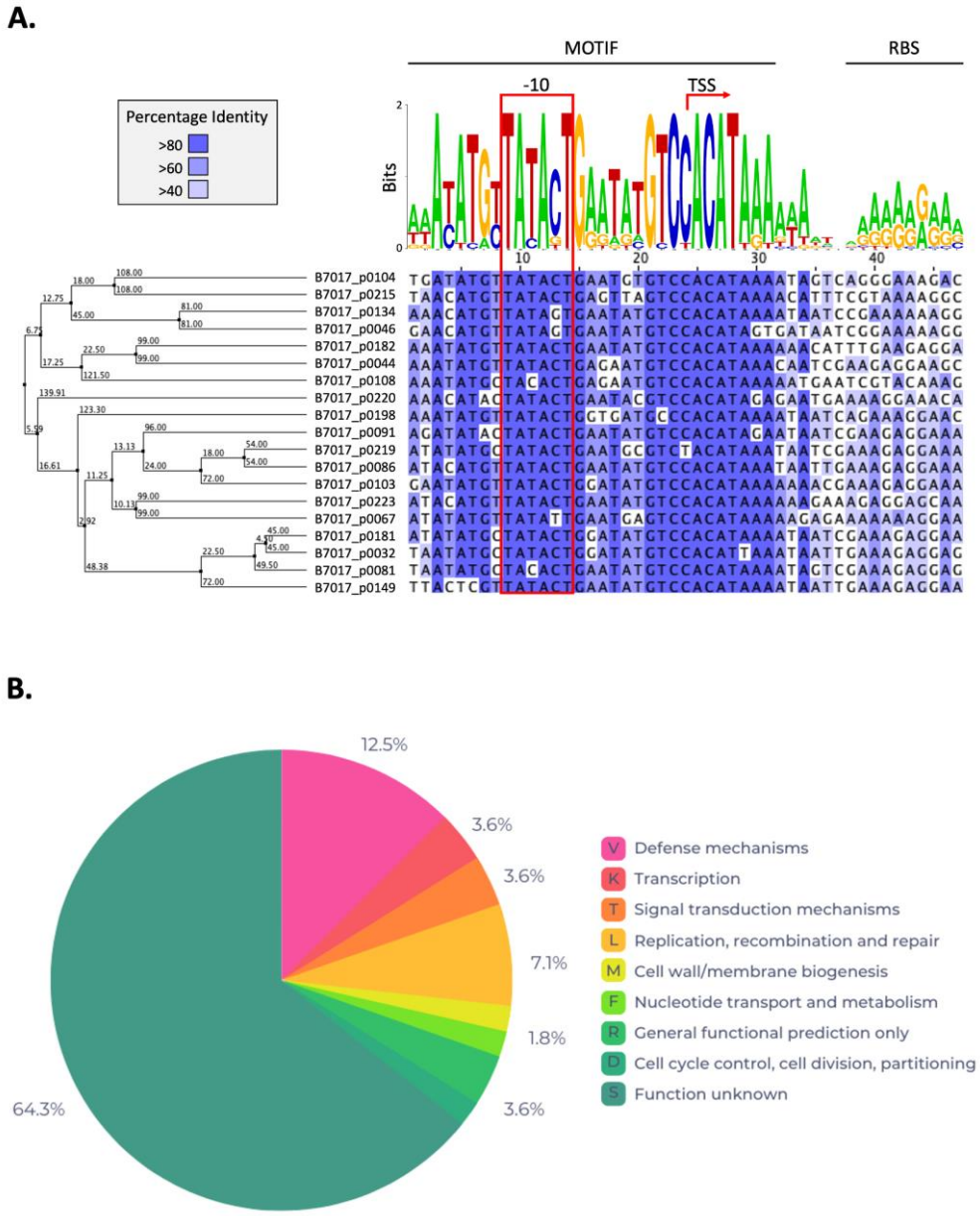


Figure 2. Conserved putative regulatory motif present within the 5' untranslated region of fourteen pMP7017 genes

A. Sequence logo (top) highlighting the conserved residues among the nineteen identified motifs, -10 promoter element (boxed in red), presumed transcriptional start sites (TSS) and downstream ribosome binding sites (RBS). The corresponding sequence alignment (below) is coloured based on the percentage of nucleotide identity (40-100%). The phylogenetic relationship of the given sequences is depicted in the unrooted phylogram based on the average distances of the DNA with bootstrap values (left). **B.** Pie chart representing the relative abundances of Clusters of Orthologous Groups (COG) functional categories within the presumed pMP7017-specific regulon.

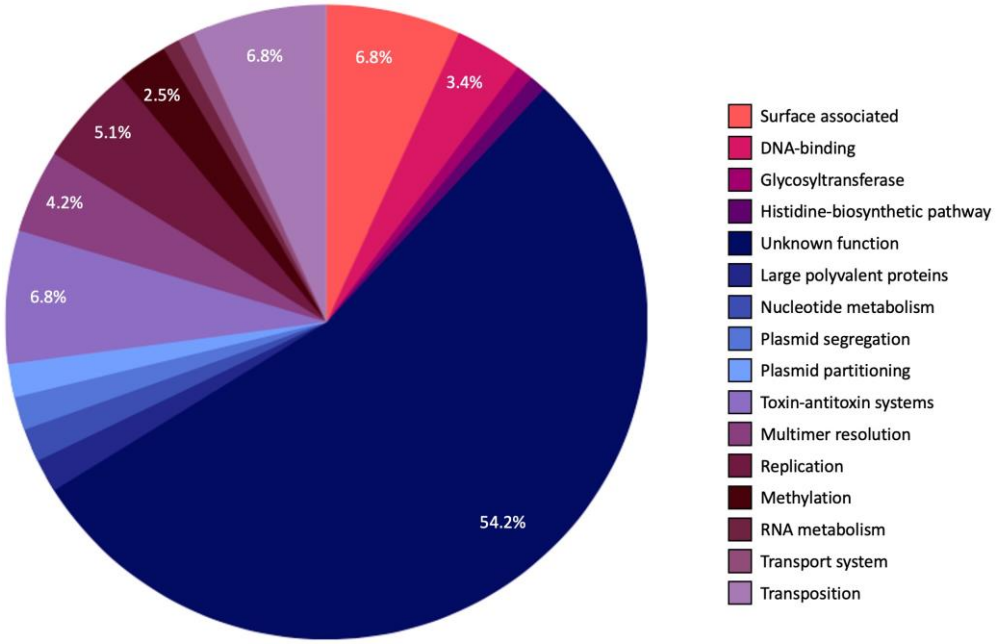


Figure 3. Pie chart representation of transcriptionally active predicted pMP7017-genes, grouped by functional categories

Categories within in the legend are presented as ordered in the pie chart clockwise from the 12 o'clock position.

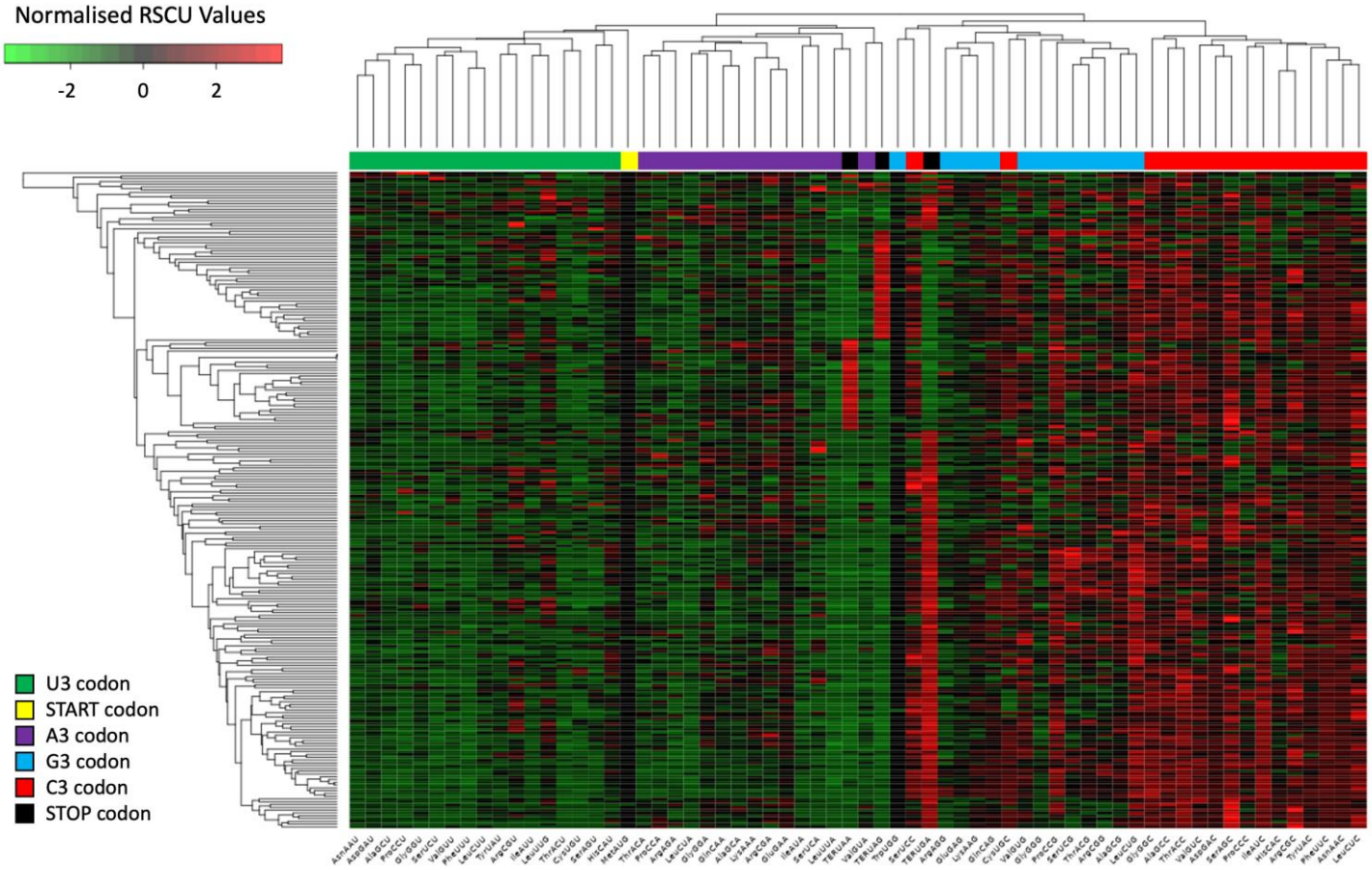
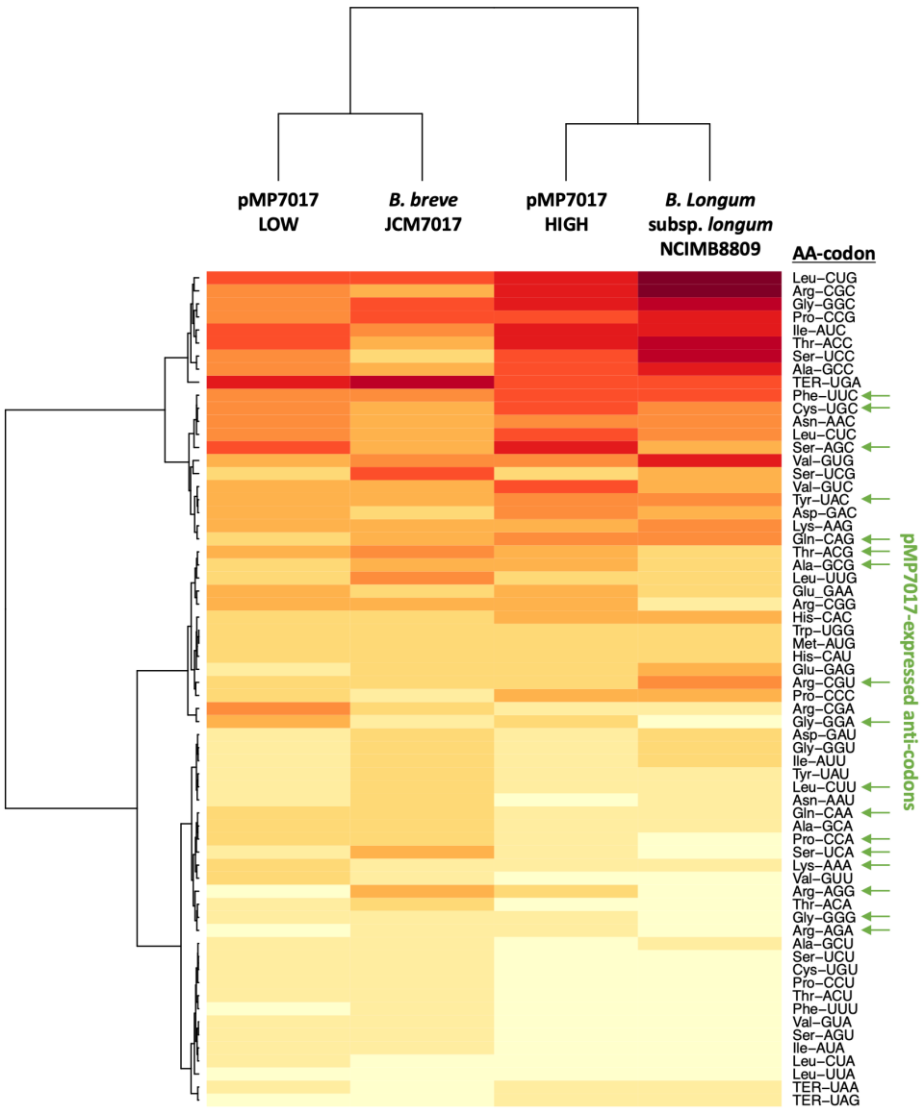


Figure 4. Codon usage bias of individual genes within the pMP7017 genome based on RSCU values

The two dimensional cluster heatmap is based on the frequency of codons for each putative gene within the pMP7017 genome. Genes are in rows and codons are in columns (colour coded by synonymous third codon position - top). Codons corresponding to methionine (yellow), tryptophan and stop codons (black) are included. Colour intensity changes from green to bright red with an increase in frequency of a particular codon.



Parameter	Expression	
	HIGH	LOW
T3s	0.13	**0.20**
C3s	**0.57**	0.43
A3s	0.19	**0.27**
G3s	**0.37**	0.32
GC3s	**0.75**	0.61
GC	0.59	0.59
CAI	0.26	**0.28**
CBI	-0.02	**0.01**
Gravy	**-0.51**	-0.75
Aromo	**0.07**	0.06

Figure 5. Correlation analysis of transcription level with overall nucleotide composition

A3, T3, C3 and G3 = frequency of each individual base A, T, C and G, respectively, at the third position of codons. GC3 = Guanine + Cytosine (G+C) content in the third position of codons, GC = total GC content of the entire gene, CAI = Codon Adaptation Index, CBI = Codon Bias Index, Gravy = grand average of hydropathy index, Aromo = aromaticity index.

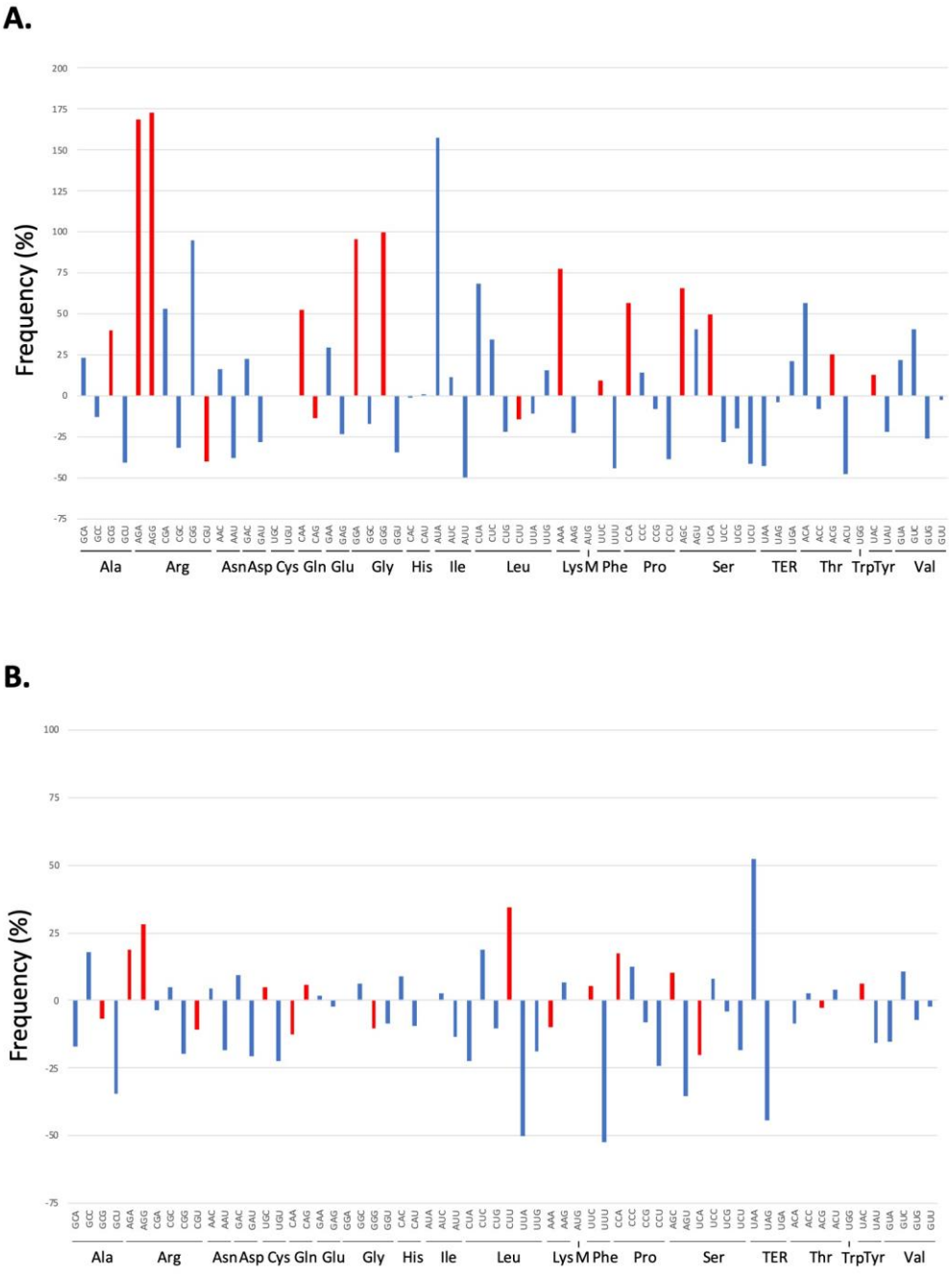


Figure 6. Codon usage frequencies of pMP7017 genes vs JCM7017: tRNAs and codon usage

A. Bar height is based on the codon usage frequency of pMP7017 genes divided by the codon usage frequency of *B. breve* JCM7017. Red bars correspond to pMP7017-expressed tRNA anti-codons, while blue bars represent all other tRNA anti-codons. **B.** Bar height is based on the codon usage frequencies of highly transcribed pMP7017 genes divided by the codon usage frequencies of the total pMP7017 genome.

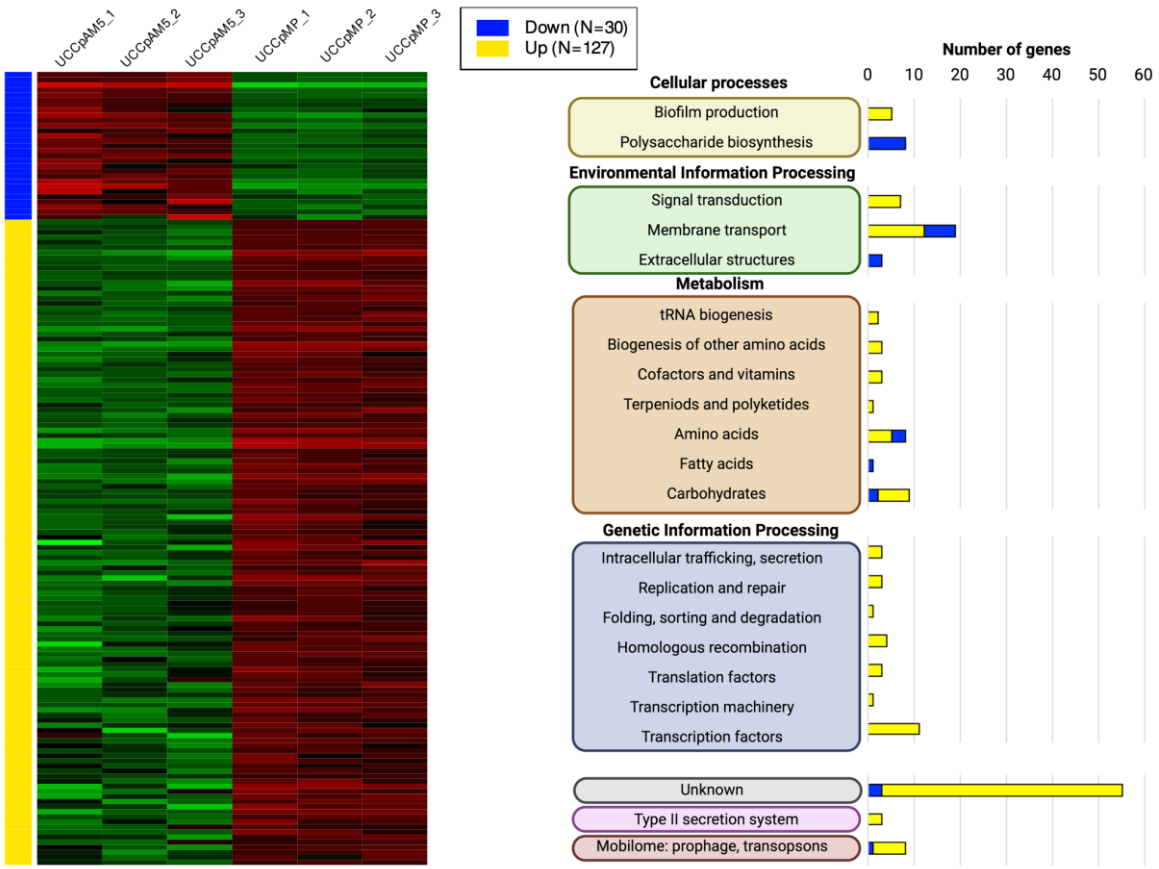


Figure 7. Differential expression of *B. breve* UCC2003 chromosomal genes in the presence of pMP7017

Heat map representing the log2 transformed expression values of *B. breve* UCC2003 chromosomal genes with altered levels of transcription in the presence of pMP7017 at mid-logarithmic growth. Samples UCCpAM5_1-3, represent replicates 1-3 of *B. breve* UCC2003 containing pAM5 (pAM5 is included to control for any transcriptional effects due to the use of antibiotic), while samples UCCpMP_1-3, represent replicates 1-3 of *B. breve* UCC2003 harbouring pMP7017 (left). COG category assignments of differentially expressed genes, grouped by protein families is presented to the left of the chart. The number of DEGs within each protein family is given, down-regulated genes are indicated by blue bars and up-regulated genes are represented by yellow bars. Detailed analysis of each DEG are presented in Table 6 and Table 8.

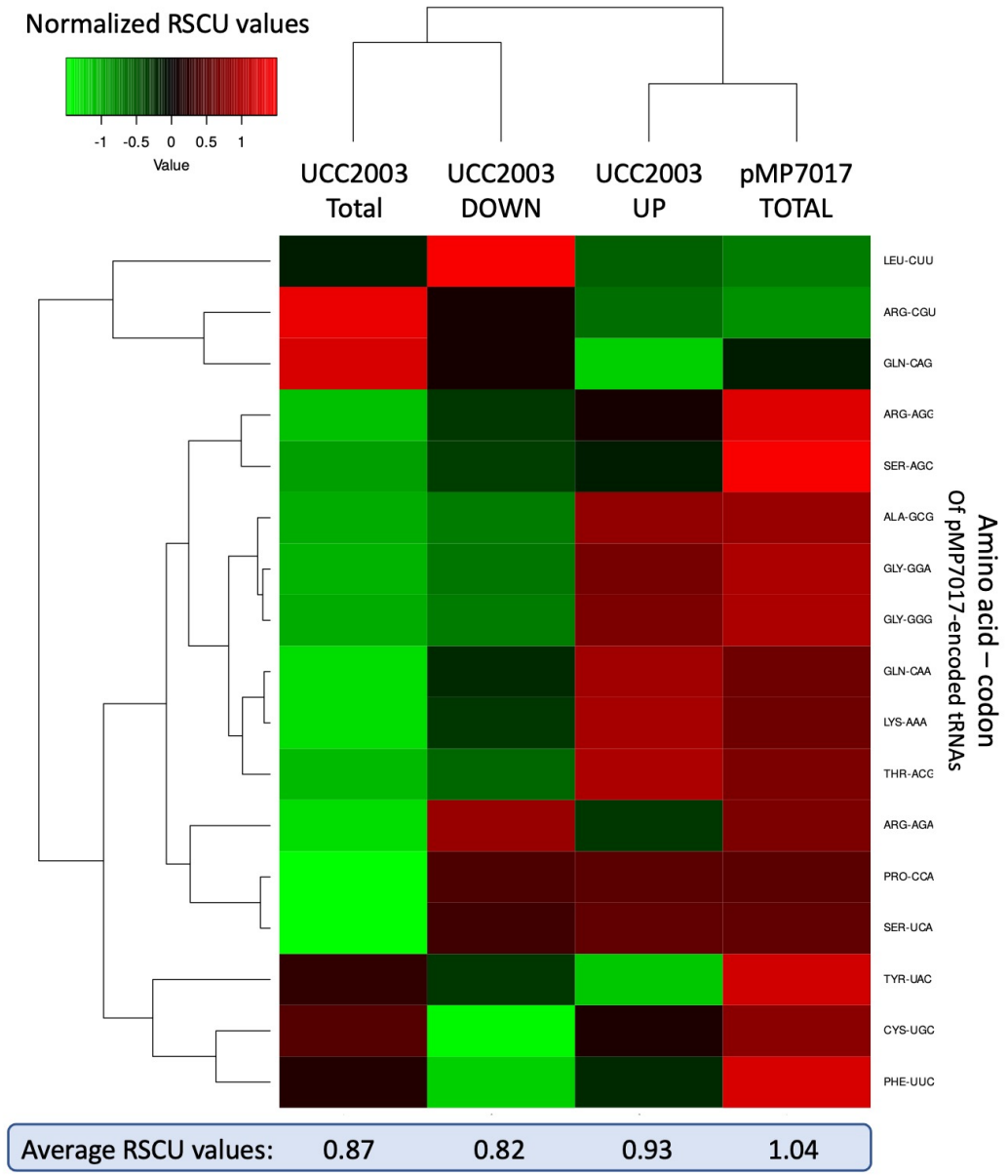


Figure 8. Codon usage comparison of chromosomal DEGs in the presence of pMP7017 and its relationship with pMP7017 encoded tRNA anti-codons

The two-dimensional cluster heatmap comparing the frequencies of codons for all coding sequences within the *B. breve* UCC2003 genome, chromosomal DEGs and the pMP7017 genome. Gene datasets are in columns and codons corresponding to pMP7017-expressed tRNA anti-codons are in rows. Colour intensity changes from green to bright red with an increase in frequency of a particular codon.

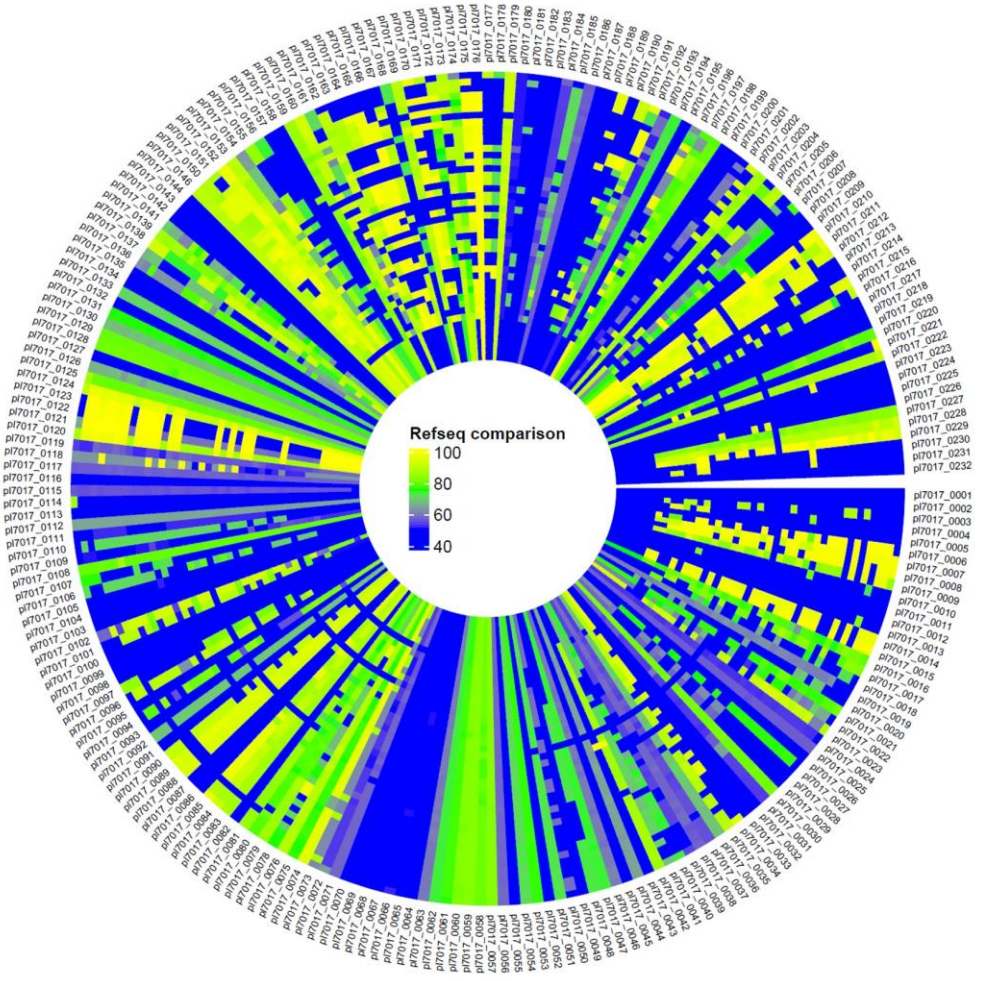


Figure 9. Mapping of genes from 43 *B. longum* pMP7017-replicase positive strains to the pMP7017 genome

The heatmap presents the percentage of sequence identity of each pMP7017 gene (locus tags given in outer ring) against those of the 43 *B. longum* pMP7017-replicase positive Refseq strains presenting with 0.0001 E-value cut off, 50% of sequence coverage and 40% identity cut-off. The colour ranges from blue to yellow with increasing identity.

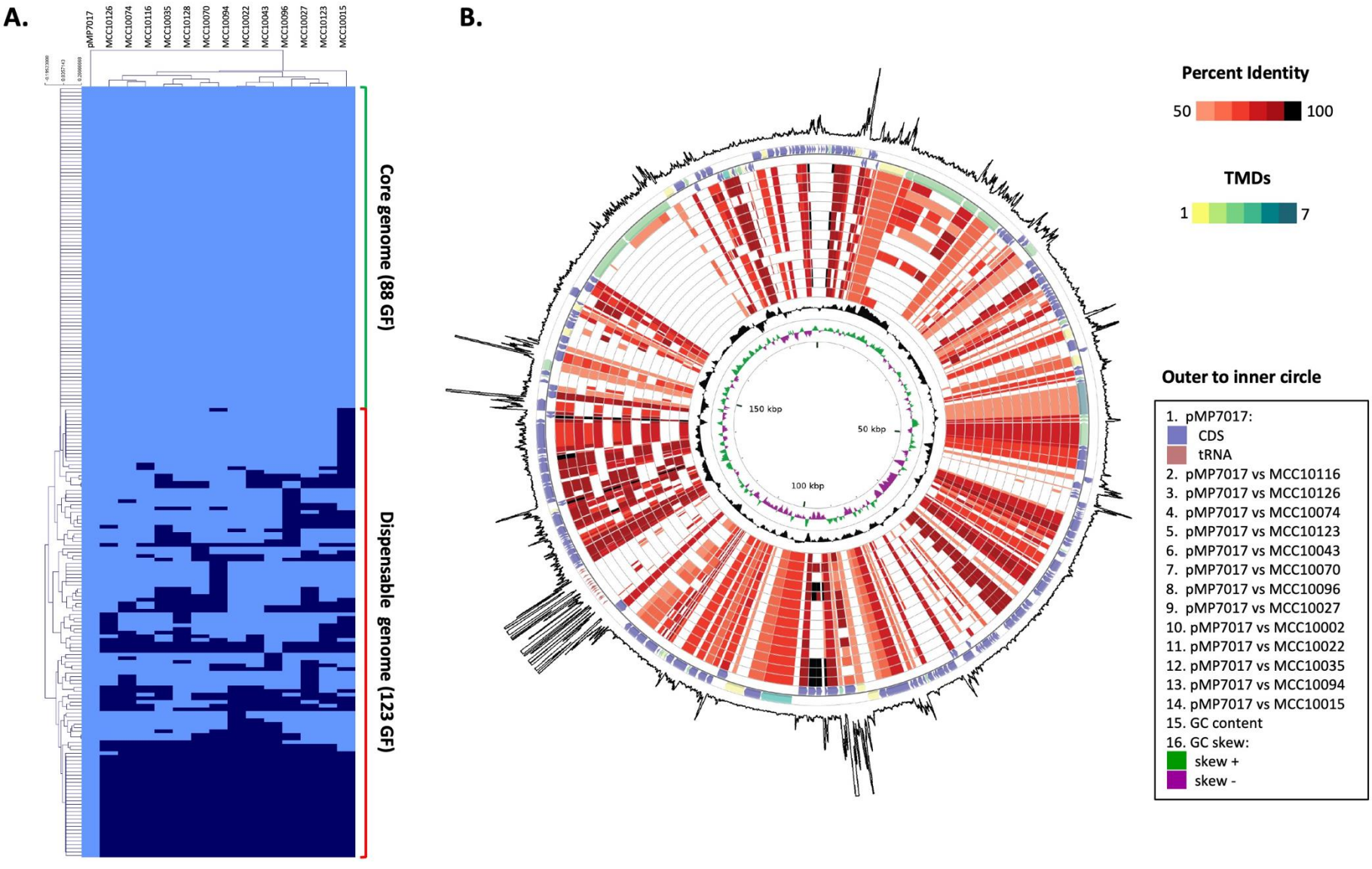

A.
pMP7017
MCC10126
MCC10074
MCC10116
MCC10035
MCC10128
MCC10070
MCC10094
MCC10022
MCC10043
MCC10096
MCC10027
MCC10123
MCC10015
Core genome (88 GF)
Dispensable genome (123 GF)
B.
150 kbp
50 kbp
100 kbp
Percent Identity
50
100
TMDs
1
7
Outer to inner circle
1. pMP7017:
CDS
tRNA
2. pMP7017 vs MCC10116
3. pMP7017 vs MCC10126
4. pMP7017 vs MCC10074
5. pMP7017 vs MCC10123
6. pMP7017 vs MCC10043
7. pMP7017 vs MCC10070
8. pMP7017 vs MCC10096
9. pMP7017 vs MCC10027
10. pMP7017 vs MCC10002
11. pMP7017 vs MCC10022
12. pMP7017 vs MCC10035
13. pMP7017 vs MCC10094
14. pMP7017 vs MCC10015
15. GC content
16. GC skew:
skew +
skew -

Figure 10. Synteny and variability of Bifidobacterial megaplasmid genomes

A. Hierarchical clustering heatmap representing the variability of pMP7017 family plasmids in terms of presence/absence of gene families computed using the all-vs-all blastP algorithm and MCL clustering. The number of core (those present in all genomes) and dispensable (those present in some genomes) gene families are indicated. **B**. Circular genome atlas of pMP7017 with mapped orthologs (defined as reciprocal best BlastP hits with >50% identity over >80% of both protein lengths) in 14 publicly available *B. longum* MCC strain genomes. The outermost circle displays RNAseq mapped reads, while the remaining circles are as outlined in the legend. pMP7017 coding sequence (CDS) are colour coded by the number of predicted transmembrane domains (TMDs).

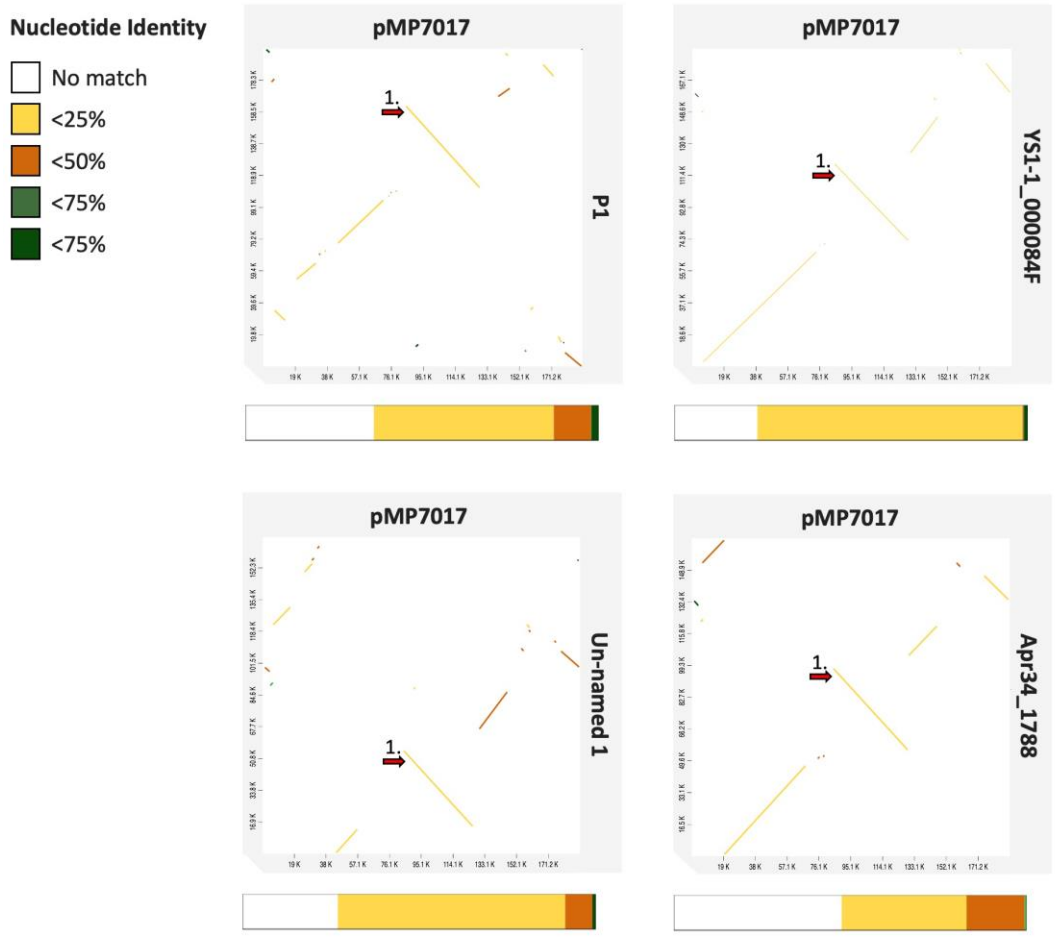


Figure 11. Genome synteny and co-linearity of pMP7017 family plasmids

Dot-plot comparison showing conserved and inverted regions found in pMP7017-homologs relative to the pMP7017 genome. The plots highlight a large inversion (red arrow, numbered (1.)), corresponding to nucleotide coordinates; 84 894-128 826, within the pMP7017 genome.

Table 1. Updated annotation of transcriptionally active genes

Locus_tag	Accession	Current annotation	Proposed annotation	Software	Accession	Description	Confidence (%)	Probability (%)	E-value
B7017_p0014	AIW55067	hypothetical protein with a helix-turn-helix motif	HigA	Phyre2	c6ltyB	Putative antitoxin higa3	99.6	-	-
				HHpred	7EWC-C	Antitoxin HigA2	-	98.62	3.80E-06
				Pfam	HTH_37	Helix-turn-helix domain	-	-	3.30E-12
B7017_p0015	AIW55068	toxin-antitoxin system	HigB	Phyre2	c6af3-C	HigB toxin	99.9	-	-
				HHpred	6AF4_C	HigB toxin	-	99.62	2.90E-14
				Pfam	Gp49	Phage derived protein Gp49-like (DUF891)	-	-	20E-25
B7017_p0019	AIW55072	antidote protein, toxin-antitoxin system	XRE-family transcriptional regulator	Phyre2	c3trb-A	antidote protein HigA	99.9	-	-
				HHpred	6HU8_A	Transcriptional regulator ComR	-	98.69	2.50E-07
				Pfam	HTH_3	Helix-turn-helix 3	-	-	6.00E-08
				CDD	cl22854	HTH_XRE super family			4.43E-23
B7017_p0020	AIW55073	cell surface protein, cysteine-rich extracellular domain	PrgA-like protein	Phyre2	c7kog-B	Myosin heavy chain isoform mhc_x1	91.1	-	-
				HHpred	6Z9L-A	PrgA	-	100	9.30E-45
				Pfam	CAP	Cysteine-rich secretory protein family	-	-	6.80E-07
B7017_p0021	AIW55074	Sortase family protein	Pilin specific class C sortase	Phyre2	c2xwg-	Sortase class C	100	-	-
				HHpred	7CG9_A	Class C sortase	-	99.95	4.20E-25
				Pfam	Sortase	Sortase domain	-	-	1.20E-21
B7017_p0129	AIW55183	DNA-binding helix-turn-helix protein HipB	XRE family transcriptional regulator	Phyre2	c2kpj-A	LexA, sos response regulator	93.5	-	-
				HHpred	3PXP_A	MmyB-like transcription regulator MltR, antibiotic	-	98.29	1.40E-05
				Pfam	HTH_19	Helix-turn-helix domain	-	-	2.60E-05

				CDD	COG3655	YozG; DNA-binding transcriptional regulator, XRE family			2.70E-05
B7017_p0142	AIW55195	Putative restriction enzyme	MmeI-like, N-terminal domain	Phyre2	c5hr4-J	MmeI, DNA-methyltransferase domain	100	-	-
				HHpred	5HR4_J	MmeI	-	99.96	3.70E-27
				Pfam	MmeI_hel	MmeI, helicase spacer domain	-	-	8.10E-37
				Pfam	MmeI_N	MmeI, N-terminal domain	-	-	3.70E-28
B7017_p0143	AIW55196	putative DNA methyl-transferase	MmeI-like, DNA-methyl-transferase domain	Phyre2	c5hr4-J	MmeI	100	-	-
				HHpred	5HR4_J	MmeI	-	100	2.40E-33
				Pfam	MmeI_ Mtase	MmeI, DNA-methyltransferase domain	-	-	4.20E-82
				Pfam	MmeI_ TRD	MmeI, target recognition domain	-	-	7.40E-19
B7017_p0144	AIW55197	Methyl-transferase domain protein	MmeI-like, target recognition domain	Phyre2	c5hr4-J	MmeI	100	-	-
				HHpred	5HR4_J	MmeI	-	99.56	1.00E-12
				Pfam	MmeI_ TRD	MmeI, target recognition domain	-	-	2.10E-28
B7017_p0170	AIW55219	ATPase, AAA superfamily	RecF	Phyre2	c6wg3-B	Structural maintenance of chromosomes 3	99.9	-	-
				HHpred	4I99_B	Chromosome partition protein SMC protein	-	99.75	8.40E-17
				Pfam	AAA_21	AAA_21 super family	-	-	5.89E-13
				CDD	COG1195	Recombinational DNA repair ATPase RecF			1.80E-12
B7017_p0173	AIW55222	hypothetical protein	DUF3841 domain protein	Pfam	DUF3841	Domain of unknown function (DUF3841)	-	-	1.80E-38
B7017_p0174	AIW55223	DNA methylase	MmeI-like RM	Phyre2	c5hr4-J	MmeI	100	-	-
				HHpred	5HR4_J	MmeI	-	100	2.70E-76

				Pfam	MmeI_N	MmeI, N-terminal domain	-	-	1.80E-26
				Pfam	MmeI_ Mtase	MmeI, DNA-methyltransferase domain	-	-	4.50E-83
				Pfam	MmeI_hel	MmeI, helicase spacer domain	-	-	4.30E-41
				Pfam	MmeI_ TRD	MmeI, target recognition domain	-	-	5.40E-63
				Pfam	MmeI_C	MmeI, C-terminal domain	-	-	1.20E-17
B7017_p0197	AIW55246	DNA methylase	N6_N4_Mtase	HHpred	1G60_A	Adenine-specific Methyltransferase MboIIA	-	99.63	1.70E-14
				Pfam	N6_N4_ Mtase	DNA methylase	-	-	5.70E-16
B7017_p0202	AIW55251	AAA ATPase	DnaA	Phyre2	c1w5s-B	Origin recognition complex subunit 2 orc2	99.9	-	-
				HHpred	3PFI_B	Holliday junction ATP-dependent DNA helicase RuvB	-	99.82	1.50E-16
				Pfam	AAA_16	AAA ATPase domain	-	-	3.50E-11
				Pfam	Bac_DnaA	Bacterial DnaA protein			9.20E-11
B7017_p0224	AIW55273	pentapeptide repeat protein	PRP/ubiquitin-protein ligase SopA/YjbI	Phyre2	c4yc5-A	Beta1	99.8	-	-
				HHpred	5JW7_A	E3 ubiquitin-protein ligase SopA	-	99.67	9.50E-17
				Pfam	Penta-peptide	Pentapeptide repeats (8 copies)	-	-	1.90E-13

Table 2. pMP7017-expressed Type II TA systems: activity and process inhibition

Toxin super-family	Toxin	Activity	Inhibition	Gene pair(s)
AtaT/TacT	TacT	Acetylation of elongation tRNAs	Translation elongation	B7017_p0101-B7017_p0102
HipA	HipA	Phosphorylation of Glutamyl-tRNA-synthetase	Translation elongation	B7017_p0120-B7017_p0121
ParE/RelE	RelE	Ribosome-dependent mRNA cleavage	Translation elongation	B7017_p0014-B7017_p0015
				B7017_p0164-B7017_p0165

Table 3. Analysis of CRISPR spacer targets of the pMP7017 Type IE CRIPSR-cas system

Spacer ID	Spacer sequence 5' - 3'	Target	Target	Accession	Coverage (%)
Spacer1	tttccgccgtaatccgattccgcgtgcggaatc	*Microbacterium* phage Morrill	terminase	UQT01735.1	45.5
Spacer2	gaaaattaacccaattaaaccaattgttcctca	*Mycobacterium* phage Superphikiman	Intergenic region	-	48.5
Spacer4	ctattcagaaaatgcgatgtggtcggtagggaa	Mycobacterium phage Anaya	NrdH-like glutaredoxin	AEK08028.1	48.5
Spacer5	cgaatcccagggccatcacggccgtggtccggg	*Mycobacterium* phage KayaCho	hypothetical protein	AGT12995.1	54.5
Spacer6	cgtgggagcagaccaaccccgcctatagggtg	*Streptomyces* phage Ibantik	DNA polymerase I	AWN05242.1	45.5
Spacer7	tggaatcgaactcatcatttccaatccctgctt	*Rhodococcus* phage Grayson	hypothetical protein	AWN04483.1	48.5
Spacer8	ctgggaacacacgcaagacaaacgcgacctgag	*Arthrobacter* phage Persistence	HNH endonuclease	QWY79717	48.5
Spacer9	tttgggtggaatcgtctccgccggtggtgtttt	*Propionibacterium* phage SKKY	minor tail protein	AKO60425	51.5
Spacer10	taggcgacgggctcgacctgaaacagatcgaat	*Siphoviridae* sp. isolate ct0MU3	Integrase	DAJ53678.1	100.0
Spacer13	ccagaggggcaacgtggcggtcgacaactacac	*Microbacterium* phage Alakazam	major tail protein	QIN93839	54.5
Spacer15	agtcggcatgcttgataatctcatcacgctggta	Bixzunavirus I3	hypothetical protein	QWT30384.1	45.5
Spacer16	gctgacgccgaggagaaacagatcaaggaactg	*Mycobacterium* phage Raymond7	type 1 restriction enzyme	QKY79207	57.6
Spacer17	gaacggggaatggtacaccacactgccattgca	*Streptomyces* phage Yaboi	LysM-like	AYB70884.1	48.5
Spacer18	gattctgccgatcgacattccgtcgagcaggac	*Siphoviridae* sp. isolate ctleM7	hypothetical protein	DAP75288.1	100.0
Spacer19	cattcccgtgcccaatatgatcataggggaaaa	*Arthrobacter* phage BenitoAntonio	membrane protein	QXO13000.1	48.5
Spacer21	tcttggcgacgatggcacacgccaccagtgcag	*Mycobacterium* phage Scorpia	head-to-tail adaptor	"QBP29015.1	48.5
Spacer22	cgtgccgtcacggatatgacccgcatcagcgga	*Mycobacterium* phage Sandalphon	tape measure protein	AXC34923.1	45.5
Spacer24	ggggacgggaaagacagtcctgcgaatcatatg	*Streptomyces* phage Mildred21	hypothetical protein	ASR75603.1	51.5
Spacer25	ccggcgttcatacaccatcggcatatcgacatg	*Gordonia* phage Trax	Cro protein	QDM55928.1	48.5

Table 4. pMP7017-expressed RNAs

Family	Locus_tag	Annotation	From	To
tRNA	B7017_ptRNA1	tRNA-Tyr	121,375	121,456
	B7017_ptRNA2	tRNA-OTHER	121,645	121,714
	B7017_ptRNA3	tRNA-Pro	121,962	122,035
	B7017_ptRNA4	tRNA-Gly	122,117	122,190
	B7017_ptRNA5	tRNA-Lys	122,331	122,402
	B7017_ptRNA6	tRNA-Phe	122,566	122,640
	B7017_ptRNA7	tRNA-Arg	122,967	123,041
	B7017_ptRNA8	tRNA-OTHER	123,211	123,286
	B7017_ptRNA9	tRNA-Arg	123,429	123,503
	B7017_ptRNA10	tRNA-Ser	123,508	123,585
	B7017_ptRNA11	tRNA-Ser	123,665	123,752
	B7017_ptRNA12	tRNA-Gln	123,871	123,941
	B7017_ptRNA13	tRNA-Cys	123,949	124,022
	B7017_ptRNA14	tRNA-Gln	124,028	124,098
	B7017_ptRNA15	tRNA-Arg	124,133	124,207
	B7017_ptRNA16	tRNA-Gly	124,463	124,534
	B7017_ptRNA17	tRNA_leu	124,947	125,023
	B7017_ptRNA18	tRNA-OTHER	125,284	125,360
	B7017_ptRNA19	tRNA-Thr	125,493	125,565
	B7017_ptRNA20	tRNA-Ala	125,731	125,803
sRNA	tmRNA	tmRNA	120,876	121,272
	manA	manA	122,711	122,830

Table 5. Correlation analysis of transcription level with overall nucleotide composition

Parameter	Expression	
	HIGH	**LOW**
T3s	0.13	**0.20**
C3s	**0.57**	0.43
A3s	0.19	**0.27**
G3s	**0.37**	0.32
GC3s	**0.75**	0.61
GC	0.59	0.59
CAI	0.26	**0.28**
CBI	-0.02	**0.01**
Gravy	**-0.51**	-0.75
Aromo	**0.07**	0.06

A3, T3, C3 and G3 = frequency of each individual base A, T, C and G at the third position of codons, GC3 = GC content in the third position of codons, GC = total guanine- cytosine content of the entire gene, CAI = Codon Adaptation Index, CBI = Codon Bias Index, Gravy = grand average of hydropathy index, Aromo = aromaticity index.

Table 6. Down-regulated *B. breve* UCC2003 chromosomal genes in the presence of pMP7017

Protein family	Pathway	Locus-tag/ Symbol	KO	description	COG	fc	Adj. P-value
Cellular Processes	Polysaccharide biosynthesis	Bbr_0442	-	Glycosyltransferase polysaccharide	COG0438	3.29	2.24E-07
		Bbr_0443	-	Glycosyltransferase polysaccharide	COG0438	10.95	1.96E-23
		Bbr_0445	-	Glycosyltransferase polysaccharide	COG0463	3.23	1.12E-18
		Bbr_0446	-	Acetyltransferase involved in polysaccharide biosynthesis	COG0110	2.67	8.67E-11
		Bbr_0447	-	Polysaccharide pyruvyl transferase	-	2.96	7.28E-20
		Bbr_0448	-	Glycosyltransferase polysaccharide	COG0463	2.88	1.76E-16
		Bbr_0450	-	Membrane spanning protein, polysaccharide biosynthesis	-	2.20	1.06E-05
		pslL	K21005	polysaccharide biosynthesis protein PslL	COG3594	2.69	8.42E-07
Environmental Information Processing	Extracellular structures	Bbr_0113	K21449	trimeric autotransporter adhesin, SpaA domain	COG4932	2.85	1.44E-21
		Bbr_0114	-	Iso_D2_wall_anc	COG4932	15.34	2.35E-117
		Bbr_1888	-	Uncharacterized surface anchored protein	COG4932	2.01	7.86E-09
	Membrane transport	*malE*	K15770	maltooligosaccharide transport system	COG2182	2.99	3.63E-21
		malC	K15772	maltooligosaccharide transport system permease	COG1175	2.82	8.21E-23
		malG	K15771	maltooligosaccharide transport system permease	COG3833	2.95	1.97E-39
		Bbr_0273	K01990	ABC-2 type transporter	COG1131	7.67	7.99E-34
		Bbr_0274	K01990	ABC-2 type transporter	COG1682	3.77	3.40E-30
		Bbr_1501	K01990	ABC-2 type transporter	COG1131	2.44	1.29E-17
		Bbr_1836	K10439	ribose transport system substrate-binding protein	COG1879	2.61	5.63E-13
Metabolism	Carbohydrate	*agl4*	K01187	alpha glucosidase	COG0366	2.22	1.29E-09
		apuB	K01200	Amylopullulanase	COG1523	2.66	1.44E-21

Protein family	Pathway	Locus-tag/ Symbol	KO	description	COG	fc	Adj. P-value
	Fatty acid	*accD*	K01966	propionyl-CoA carboxylase	COG4799	2.10	1.45E-08
	Amino acid transport and metabolism	NA	-	tRNA Ile	-	3.04	1.45E-06
		NA	-	tRNA-Ala	-	2.96	3.48E-06
		NA	-	tRNA-Ala	-	2.10	1.57E-04
		NA	-	tRNA-Val	-	2.77	4.00E-02
		NA	-	tRNA-Lys	-	2.03	1.37E-05
Mobilome	prophages, transposons	Bbr_0465	-	Bacteriophage abortive infection	-	2.16	3.16E-02
Unknown	Unknown	Bbr_0121	-	TraX protein acetylation of pro-pilin	-	2.16	1.69E-17
		Bbr_0449	-	hypothetical protein	-	2.92	2.30E-07
		Bbr_1748	-	Acetyltransferase	COG0456	4.08	1.01E-20

KO = KEGG Orthology, **COG** = Clusters of Orthologous Genes, **fc** = fold change, **Adj. P-value** = Adjusted P-value.

Table 7. Gene Ontology enrichment analyses of down-regulated *B. breve* UCC2003 chromosomal genes in the presence of pMP7017

Pathway	nGenes	Pathway Genes	Fold Enrichment	Enrichment FDR	Gene/ symbol					
Glycosyl transferases	9	10	187.5	1.88E-19	BBR_0451	Bbr_0442	Bbr_0443	Bbr_0445	Bbr_0446	Bbr_0447
					Bbr_0448	Bbr_0449	Bbr_0450			
Carbohydrate metabolism	11	117	19.6	2.13E-11	*agl4*	*apuB*	Bbr_0442	Bbr_0443	Bbr_0445	Bbr_0446
					Bbr_0447	Bbr_0448	Bbr_0449	Bbr_0450	BBR_0451	
Polysaccharide biosynthesis protein	4	5	166.7	2.24E-08	Bbr_0442	Bbr_0443	Bbr_0445	Bbr_0446		
Cna protein B-type domain	3	5	125.0	9.30E-06	Bbr_0113	Bbr_0114	Bbr_1888			
Transmembrane	11	461	5.0	2.30E-05	Bbr_0113	Bbr_0114	*malC*	*malG*	Bbr_0121	*apuB*
					Bbr_0274	Bbr_0449	Bbr_0450	BBR_0451	Bbr_1888	
Transferase	7	279	5.2	1.08E-03	*apuB*	Bbr_0443	Bbr_0445	Bbr_0446	Bbr_0448	BBR_0451
					Bbr_1748					
Alpha amylase, catalytic domain	2	13	32.1	4.28E-03	*agl4*	*apuB*				
Glycosidase	2	45	9.3	3.62E-02	*agl4*	*apuB*				

Table 8. Up-regulated *B. breve* UCC2003 chromosomal genes in the presence of pMP7017

Protein family	Pathway	Locus-tag/ Symbol	KO	Pathway/description	COG	fc	Adj. P-value
Environmental Information Processing	Signal transduction	Bbr_0062	-	Membrane associated serine protease, rhomboid family	COG0705	2.82	1.60E-05
		Bbr_1060	-	Signal transduction histidine kinase	COG4585	2.65	1.66E-04
Cellular Processes	Biofilm formation	PelF	K21011	biofilm biosynthesis glycosyltransferase	COG0438	3.37	7.88E-05
		PelG	K21012	polysaccharide exporter	COG4267	3.13	6.41E-05
		Bbr_0046	K06330	Spore coat protein CotH	COG5337	2.82	4.67E-09
		Bbr_0047	-	CYTH-like phosphatase	-	2.78	7.60E-05
		Bbr_0048	-	Mg2+ transport	-	2.07	1.82E-11
Environmental Information Processing	Membrane transport	Permease	K02004	ABC-type transport system, involved in lipoprotein release, permease component	COG4591	2.13	2.67E-03
		cldF	K10241	cellobiose transport system permease protein	COG1175	2.41	7.28E-20
		cldG	K10242	cellobiose transport system permease protein	COG0395	2.87	2.98E-16
		Bbr_1341	K06895	L-lysine exporter family protein LysE/ArgO	COG1279	3.37	2.04E-11
		Bbr_1858	K10117	transport system permease protein	COG0395	2.22	8.18E-08
		Bbr_1862	K10117	transport system substrate-binding protein	COG1653	2.13	5.77E-17
		Bbr_0259	K01990	ABC-2 type transporter	COG1134	2.09	9.04E-03
	Signal transduction	Bbr_0859	K08217	Macrolide-efflux protein	COG0477	2.13	6.26E-09
		Bbr_1049	-	VanZ family protein	COG4767	2.49	7.39E-08
		Bbr_1065	-	ABC-type multidrug transport system, permease	COG0577	2.74	1.83E-13
		Bbr_1071	K01992	ABC-type multidrug transport system, permease	COG0842	3.05	8.75E-05
		Bbr_1483	-	Multidrug resistance protein B	COG0842	3.52	1.62E-04
General function prediction only	General function prediction only	Bbr_1270	-	LysM domain protein	COG3170	4.73	1.23E-31
		Bbr_1290	-	Predicted metal-dependent hydrolase, TIM-barrel fold	COG3618	4.14	1.70E-17
		Bbr_1515	-	ATP-binding protein of ABC transporter system	COG0488	2.08	3.46E-07
		Bbr_1741	-	Predicted metal-dependent hydrolase, TIM-barrel fold	COG3618	2.32	2.12E-03
Genetic Information Processing	Transcription factors	Bbr_0032	K05499	LacI family transcriptional regulator	-	2.18	5.06E-08
		Bbr_0208	-	Transcriptional regulator, TetR family	-	2.27	4.94E-15
		Bbr_0849	K15545	NagC/XylR-type transciptional regulator	COG1940	2.76	9.62E-14
		Bbr_1309	-	DNA-binding transcriptional regulator, MarR family	COG1846	2.22	1.29E-09

		Bbr_1340	K18907	GntR family transcriptional regulator, regulator for abcA and norABC	COG1167	2.86	1.24E-13
		cspA	K03704	Cold shock protein	COG1278	2.53	9.04E-14
		Bbr_1389	-	Transcription antiterminator, BglG family	COG3711	2.34	1.71E-02
		Bbr_1509	-	Transcription regulator, HTH_36 domain	-	3.96	1.02E-10
		Bbr_1510	-	WhiB family transcriptional regulator	-	2.98	1.53E-02
		Bbr_1519	-	DNA-binding transcriptional regulator, MerR family	COG0789	2.21	8.34E-05
		Bbr_1523	-	DNA-binding transcriptional regulator, MerR family	COG0789	3.73	1.28E-19
	Transcription machinery	Bbr_1068	-	RNA polymerase sigma-E factor	COG1595	4.18	3.79E-03
	Translation factors	Bbr_1079	-	Ribosome-associated factor Y	COG2197	2.19	1.48E-05
		infC	K02520	Translation initiation factor IF-3	COG0290	2.29	9.80E-24
		Bbr_1516	-	Acetyltransferase (GNAT) family	COG0456	2.24	1.96E-08
	Homologous recombination	RecG	K03655	ATP-dependent DNA helicase RecG	COG1200	3.31	5.23E-04
		Bbr_1113	-	site-specific tyrosine recombinase XerC	COG4973	25.58	5.27E-19
		topB	K03169	DNA topoisomerase III	COG0550	4.40	2.98E-09
		Bbr_1548	-	site-specific tyrosine recombinase XerC	COG0582	14.11	9.14E-21
	Folding, sorting and degradation	Bbr_1222	-	MutT/nudix family phosphohydrolase	COG1051	3.30	3.34E-11
	Replication and repair	Bbr_0585	K03502	DNA polymerase V	COG0389	3.78	8.15E-21
		lexA	K01356	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase)	COG1974	2.25	3.43E-11
		parB1	K03497	ParB family transcriptional regulator	COG1475	3.31	6.15E-04
	Intracellular trafficking, secretion	Bbr_1128	-	Predicted membrane metal-binding protein	COG0658	14.38	3.65E-21
		Relaxase	-	relaxase	COG3843	3.82	1.52E-02
		Bbr_1545	-	MobC	COG3843	3.01	4.23E-02
	terpenoids and polyketides	Bbr_0205	-	type I polyketide synthase	-	2.18	9.55E-09
Metabolism	cofactors and vitamins	Bbr_0209	-	Phosphopantetheinyl transferase	-	2.30	3.01E-10
		Bbr_1717	-	Pimeloyl-ACP methyl ester carboxylesterase	COG0596	3.28	6.74E-04
		Bbr_1718	-	NAD(P)H-dependent FMN reductase	COG0431	4.70	1.70E-04
		Bbr_0342	K02055	Spermidine/putrescine-binding periplasmic protein	COG0687	2.70	4.00E-04

	Amino acid transport and metabolism	BBR_0600	-	Aspartate/methionine/tyrosine aminotransferase	-	4.11	3.29E-03
		Bbr_1061	-	Nucleotide pyrophosphatase	COG2197	2.06	3.51E-02
	other amino acids	oppD3	K13892	glutathione transport system ATP-binding protein	COG1123	2.75	3.56E-09
		oppC2	K13891	glutathione transport system permease protein	COG1173	2.43	6.01E-04
		bgl4	K05349	beta-glucosidase	COG1472	2.30	4.25E-04
	Transfer RNA Biogenesis	truA	K06173	tRNA pseudouridine38-40 synthase	COG0101	2.83	1.38E-12
		dtd	K07560	D-aminoacyl-tRNA deacylase	COG1490	2.92	6.76E-14
	Carbohydrate	lacZ3	-	Beta-galactosidase/beta-glucuronidase	COG2723	2.32	3.44E-04
		BBR_1143	-	Alpha-glucosidase	COG1501	2.13	1.94E-05
		cwlO	K21471	peptidoglycan DL-endopeptidase CwlO	COG0791	3.68	1.27E-03
		acyP	K01512	Acylphosphatase	COG1254	2.45	6.06E-03
		dapA3	K22397	Dihydrodipicolinate synthase	COG0329	3.45	2.71E-19
		dapA4	K22397	Dihydrodipicolinate synthase	COG0329	2.18	6.96E-05
		Bbr_1291	K18333	L-fucose dehydrogenase	COG1028	4.38	5.82E-28
Mobilome	prophages, transposons	BBR_0024	-	Transposase	-	2.15	1.16E-05
		BBR_0159	-	Transposase	-	3.66	1.30E-03
		tnpY	-	Transposase	-	2.41	2.65E-10
		Bbr_0302	-	Transposase	-	2.71	7.26E-03
		Bbr_0457	-	IS21 family transposase	-	2.87	1.60E-05
		Bbr_1401	-	transposase	-	3.73	3.59E-03
		Bbr_1404	-	Transposase	COG4584	2.20	3.67E-02
Secretion system	Type II secretion system	tadA	K02283	Pilus assembly ATPase	COG4962	9.35	2.90E-03
		tadB	K12510	tight adherence pilus assembly protein	COG4965	4.00	4.47E-02
		tadF	-	Pilin precursor	-	5.21	2.57E-02
Unknown	Unknown	Bbr_0055	-	Conserved hypothetical protein	-	2.69	1.45E-08
		Bbr_0145	-	Conserved hypothetical protein	-	2.03	1.49E-02
		Bbr_0202	-	Conserved hypothetical protein	-	3.02	1.23E-05
		Bbr_0218	-	hypothetical protein	-	2.17	1.94E-03
		Bbr_0264	-	hypothetical protein	-	2.47	2.72E-24
		Holin	-	hypothetical phage protein	-	3.84	4.35E-03
		Bbr_0300	-	hypothetical secreted protein	-	2.47	1.60E-18
		Bbr_0315	-	hypothetical phage protein (flanking tRNA10)	-	3.32	5.33E-04

Bbr_0316	-	hypothetical phage protein (flanking tRNA10)	-	3.41	4.52E-02
Bbr_0317	-	hypothetical phage protein	-	4.98	2.58E-08
Bbr_0318	-	hypothetical protein	-	3.60	9.24E-07
Bbr_0319	-	phage transcription regulator protein	-	2.27	7.83E-11
Bbr_0320	-	hypothetical protein	-	2.44	9.87E-12
Bbr_0337	-	hypothetical protein	-	3.86	4.96E-02
Bbr_0362	-	conserved hypothetical protein (flanking tRNA13)	-	2.36	5.46E-05
Bbr_0364	-	hypothetical protein	-	2.15	3.83E-02
Bbr_0366	-	Conserved hypothetical secreted protein	-	2.09	9.42E-06
Bbr_0376	-	Conserved membrane spanning rhomboid family protein	-	3.38	2.49E-08
Bbr_0384	-	Conserved membrane spanning rhomboid family protein	-	4.37	3.41E-37
Bbr_0452	-	hypothetical protein	-	2.72	2.67E-03
Bbr_0454	-	conserved hypothetical protein	-	2.04	4.19E-02
Bbr_0501	-	conserved hypothetical protein	COG1721	3.73	1.56E-05
Bbr_0502	-	MoxR-like ATPase	COG0714	2.61	8.84E-11
Bbr_0572	-	hypothetical protein	-	2.59	3.11E-02
Bbr_0591	-	hypothetical protein	-	2.42	5.35E-03
Bbr_0592	-	hypothetical protein	-	16.44	4.91E-02
Bbr_0593	-	Conserved hypothetical secreted protein	-	2.07	2.65E-03
Bbr_0638	-	conserved hypothetical protein	-	2.41	2.17E-03
Bbr_0729	-	Conserved membrane spanning rhomboid family protein	-	2.51	3.15E-03
Bbr_0836	-	conserved hypothetical protein	-	2.37	1.79E-08
Bbr_1058	-	conserved hypothetical protein	-	3.92	8.18E-08
Bbr_1069	-	Conserved membrane spanning rhomboid family protein	-	4.68	7.45E-05
Bbr_1070	-	Conserved membrane spanning rhomboid family protein	-	2.76	1.35E-02
Bbr_1151	-	hypothetical protein	-	3.26	1.07E-09
Bbr_1306	-	Conserved membrane spanning rhomboid family protein	-	2.21	2.11E-09
Bbr_1308	-	Conserved membrane spanning rhomboid family protein	-	2.41	7.63E-06
Bbr_1509a	-	hypothetical protein	-	4.29	3.61E-03
Bbr_1524	-	PIN_3 RNAse	-	3.13	2.66E-24
Bbr_1528	-	Conserved membrane spanning rhomboid family protein	-	20.90	1.35E-03
Bbr_1537	-	hypothetical protein	-	3.78	4.58E-02
Bbr_1538	-	hypothetical protein	-	5.05	4.16E-02
Bbr_1546a	-	hypothetical protein	-	2.18	4.73E-02

	Bbr_1547	-	hypothetical protein	-	3.98	2.31E-03
	Bbr_1609	-	hypothetical protein	-	2.45	6.43E-10
	Bbr_1728	-	hypothetical protein	-	2.04	1.30E-03
	Bbr_1765	-	Conserved membrane spanning rhomboid family protein	-	4.25	6.85E-15
	Bbr_1772	-	conserved hypothetical protein	-	2.86	3.72E-15
	Bbr_1861	-	conserved hypothetical protein	-	2.32	1.14E-18

Table 9. Gene Ontology enrichment analyses of up-regulated *B. breve* UCC2003 chromosomal genes in the presence of pMP7017

Pathway	nGenes	nGenes total gene pool	Fold Enrichment	Enrichment FDR	Gene/ symbol						
Biofilm biosynthesis	4	6	44.4	2.83E-05	Bbr_0044	Bbr_0045	Bbr_0046	Bbr_0047			
MetI-like superfamily	2	4	33.3	1.30E-02	*cldF*	*cldG*					
Diaminopimelate pathway	2	4	33.3	1.30E-02	*dapA3*	*dapA4*					
Tad-like protein	3	8	25.0	2.37E-03	*tadA*	*tadB*	*tadF*				
Integrase core domain	4	20	13.3	2.46E-03	Bbr_0159	Bbr_0302	Bbr_0457	Bbr_1404			
Fatty acid metabolism, and phospholipase/ carboxylesterase	7	51	9.2	1.98E-04	Bbr_0205	Bbr_0209	dapA3	Bbr_1290	Bbr_1291	*dapA4*	Bbr_1741
SOS response	1	8	8.3	1.85E-01	*lexA*						
Transmembrane, and ABC transporter	13	114	7.6	7.22E-07	Bbr_0044	Bbr_0045	Bbr_0046	Bbr_0047	Bbr_0062	Bbr_0208	*tnpY*
					Bbr_0259	Bbr_1060	Bbr_1061	Bbr_1065	Bbr_1068	Bbr_1071	
Quorum sensing	2	20	6.7	1.14E-01	*oppD3*	*oppC2*					
Transcription regulation	7	75	6.2	1.87E-03	Bbr_0032	Bbr_0208	Bbr_1061	Bbr_1068	Bbr_1079	*lexA*	Bbr_1340
DNA-binding	10	115	5.8	1.82E-04	Bbr_0032	Bbr_0146	Bbr_0208	Bbr_1061	Bbr_1068	Bbr_1079	Bbr_1113
					lexA	Bbr_1340	Bbr_1548				
Transport, and signalling	10	183	3.6	4.38E-03	*cldF*	*cldG*	Bbr_0205	lacZ3	Bbr_0342	Bbr_0849	oppD3
					oppC2	Bbr_1858	Bbr_1862				
Signal	5	93	3.6	5.69E-02	Bbr_0046	Bbr_0342	Bbr_1530	Bbr_1858	Bbr_1862		
Transport	5	104	3.2	7.15E-02	*cldF*	*cldG*	*oppD3*	*oppC2*	Bbr_1858		
Transmembrane	18	461	2.6	1.87E-03	Bbr_0045	Bbr_0062	*cldF*	*cldG*	*tadB*	*tadF*	Bbr_0859

					Bbr_1049 Bbr_1483	Bbr_1060 Bbr_1530	Bbr_1065 oppC2	Bbr_1071 Bbr_1858	Bbr_1081	Bbr_1128	Bbr_1341
Schiff base-forming aldolases	2	7	19.0	2.62E-02	*dapA3*	*dapA4*					
Transcription	7	80	5.8	2.37E-03	Bbr_0032	Bbr_0208	Bbr_1061	Bbr_1068	Bbr_1079	*lexA*	Bbr_1340

Table 10. Prevalence of pMP7017 among metagenomic samples

Study	Country	NGS platform(s)	Total Samples	Mapped Samples
Nishijama 2016 (PMID: 26951067)	Japan	454, Ion PGM and illumina Miseq	400	36
DIABIMMUNE cohort (PMID: 30559407)	Russia	Illumina	89	34
DIABIMMUNE cohort (PMID: 30559407)	Finland	Illumina	102	40
DIABIMMUNE cohort (PMID: 30550407)	Estonia	Illumina	97	29
Backhed 2015	Sweden	Illumina	312	16
Shao 2019	UK	Illumina	588	41
Ferretti 2018, Asnicar 2017	Italy	Illumina	102	6
Wampach 2018	Luxembourg	Illumina	49	3
Chu 2017	USA	Illumina	27	1
		Total samples	**1846**	
		Mapped samples	**206**	
		>70% mapped genes	**10**	

SUPPLEMENTAL MATERIAL

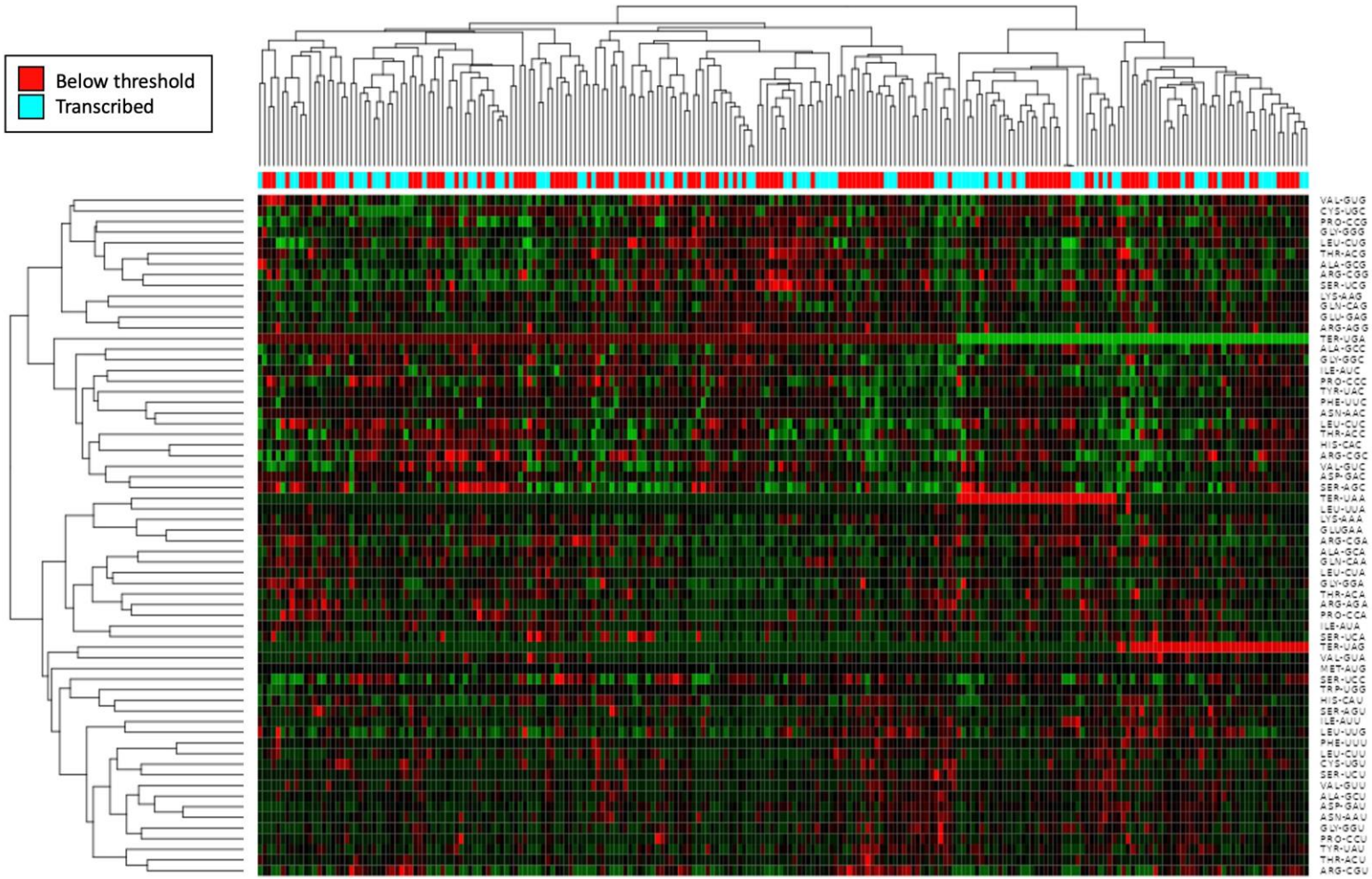


Figure S1. Codon usage comparison of pMP7017-encoded transcribed genes vs non-transcribed genes at mid-log growth of *B. breve* UCC2003

Codon usage based two dimensional cluster heatmap. Heatmap based on the usage (frequency) of codons for the pMP7017 genome. Colour intensity changes from bright red to green with an increase in frequency of a particular codon. Genes transcribed at mid-log growth of *B. breve* UCC2003 are indicated in cyan, while non-transcribed genes (below-threshold) are indicated in red (top).

Supplementary Table S1. Comparison of codon usage frequencies

AA	Codon	*B. breve* JCM7017	pMP7017	*B. longum* subsp. *longum* NCIMB8809
Ala	GCG	0.23	0.33	0.29
Ala	GCA	0.13	0.16	0.21
Ala	GCT	0.14	0.08	0.18
Ala	GCC	0.5	0.43	0.32
Cys	TGT	0.18	0.18	0.32
Cys	TGC	0.82	0.82	0.68
Asp	GAT	0.44	0.32	0.45
Asp	GAC	0.56	0.68	0.55
Glu	GAG	0.55	0.43	0.45
Glu	GAA	0.45	0.57	0.55
Phe	TTT	0.17	0.09	0.23
Phe	TTC	0.83	0.91	0.77
Gly	GGG	0.07	0.15	0.14
Gly	GGA	0.11	0.22	0.21
Gly	GGT	0.22	0.14	0.21
Gly	GGC	0.6	0.5	0.45
His	CAT	0.48	0.48	0.5
His	CAC	0.52	0.52	0.5
Ile	ATA	0.05	0.12	0.15
Ile	ATT	0.29	0.15	0.28
Ile	ATC	0.66	0.73	0.57
Lys	AAG	0.78	0.6	0.69
Lys	AAA	0.22	0.4	0.31
Leu	TTG	0.16	0.18	0.16
Leu	TTA	0.01	0.01	0.05
Leu	CTG	0.47	0.37	0.33
Leu	CTA	0.03	0.04	0.08
Leu	CTT	0.1	0.09	0.14
Leu	CTC	0.22	0.3	0.23
Asn	AAT	0.31	0.19	0.37
Asn	AAC	0.69	0.81	0.63
Pro	CCG	0.54	0.5	0.44
Pro	CCA	0.09	0.15	0.25
Pro	CCT	0.12	0.07	0.15
Pro	CCC	0.25	0.28	0.17
Gln	CAG	0.79	0.68	0.45
Gln	CAA	0.21	0.32	0.55
Arg	AGG	0.04	0.12	0.12
Arg	AGA	0.03	0.07	0.08
Arg	CGG	0.1	0.19	0.2
Arg	CGA	0.08	0.13	0.21
Arg	CGT	0.28	0.17	0.15
Arg	CGC	0.46	0.32	0.25
Ser	AGT	0.05	0.08	0.07
Ser	AGC	0.2	0.34	0.18
Ser	TCG	0.21	0.17	0.29

AA	Codon	*B. breve* JCM7017	pMP7017	*B. longum* subsp. *longum* NCIMB8809
Ser	TCA	0.06	0.09	0.18
Ser	TCT	0.08	0.05	0.11
Ser	TCC	0.4	0.29	0.17
Thr	ACG	0.24	0.3	0.34
Thr	ACA	0.07	0.12	0.18
Thr	ACT	0.12	0.06	0.14
Thr	ACC	0.57	0.52	0.33
Val	GTG	0.51	0.38	0.39
Val	GTA	0.08	0.1	0.13
Val	GTT	0.1	0.1	0.2
Val	GTC	0.3	0.42	0.28
Tyr	TAT	0.37	0.29	0.39
Tyr	TAC	0.63	0.71	0.61

Codons with low frequency (<10%) are highlighted in pink. The most preferred codon for each amino acid is highlighted in blue. Methionine, Tyrosine and stop codons are not included.

REFERENCES

[1] **H. Tissier**, 'Recherches sur la flore intestinale des nourrissons:(état normal et pathologique)', 1990.

[2] **C. Milani *et al.***, 'Genomics of the Genus Bifidobacterium Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment', *Appl Environ Microbiol*, vol. 82, no. 4, pp. 980–991, Feb. 2016, doi: 10.1128/AEM.03500-15.

[3] **S. Fanning *et al.***, 'Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection', *Proceedings of the National Academy of Sciences*, vol. 109, no. 6, pp. 2108–2113, Feb. 2012, doi: 10.1073/pnas.1115621109.

[4] **S. Arboleya, C. Watkins, C. Stanton, and R. P. Ross**, 'Gut Bifidobacteria Populations in Human Health and Aging', *Frontiers in Microbiology*, vol. 7, p. 1204, 2016, doi: 10.3389/fmicb.2016.01204.

[5] **G. Alessandri, M. C. Ossiprandi, J. Macsharry, D. van Sinderen, and M. Ventura**, 'Bifidobacterial Dialogue With Its Human Host and Consequent Modulation of the Immune System', *Front. Immunol.*, 2019, doi: 10.3389/fimmu.2019.02348.

[6] **E. Abdolalipour *et al.***, 'Evaluation of the antitumor immune responses of probiotic Bifidobacterium bifidum in human papillomavirus-induced tumor model', *Microb Pathog*, vol. 145, p. 104207, Aug. 2020, doi: 10.1016/j.micpath.2020.104207.

[7] **M. Esteban-Torres, L. Ruiz, G. A. Lugli, M. Ventura, A. Margolles, and D. van Sinderen**, 'Editorial: Role of Bifidobacteria in Human and Animal Health and Biotechnological Applications', *Frontiers in Microbiology*, vol. 12, 2021, Accessed: Feb. 08, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmicb.2021.785664

[8] **Z. Faghfoori, M. H. Faghfoori, A. Saber, A. Izadi, and A. Yari Khosroushahi**, 'Anticancer effects of bifidobacteria on colon cancer cell lines', *Cancer Cell Int*, vol. 21, no. 1, Art. no. 1, Dec. 2021, doi: 10.1186/s12935-021-01971-3.

[9] **B. Sgorbati, V. Scardovi, and D. J. Leblanc**, 'Plasmids in the genus Bifidobacterium', *J. Gen. Microbiol.*, vol. 128, no. 9, Art. no. 9, Sep. 1982, doi: 10.1099/00221287-128-9-2121.

[10] **B. Sgorbati, V. Scardovi, and D. J. Leblanc**, 'Related structures in the plasmid profiles of Bifidobacterium asteroides, B. indicum and B. globosum', *Microbiologica*, vol. 9, no. 4, pp. 443–454, Oct. 1986.

[11] **M. Iwata and T. Morishita**, 'The presence of plasmids in Bifidobacterium breve', *Letters in Applied Microbiology*, vol. 9, no. 5, pp. 165–168, 1989, doi: 10.1111/j.1472-765X.1989.tb00315.x.

[12] **D. Matteuzzi, P. Brigidi, M. Rossi, and D. Di**, 'Characterization and molecular cloning of Bifidobacterium longum cryptic plasmid pMB1', *Lett Appl Microbiol*, vol. 11, no. 4, pp. 220–223, Oct. 1990, doi: 10.1111/j.1472-765x.1990.tb00165.x.

[13] **M. S. Park, K. H. Lee, and G. E. Ji**, 'Isolation and characterization of two plasmids from Bifidobacterium longum', *Lett Appl Microbiol*, vol. 25, no. 1, pp. 5–7, Jul. 1997, doi: 10.1046/j.1472-765x.1997.00059.x.

[14] **K. O'Riordan and G. F. Fitzgerald**, 'Molecular characterisation of a 5.75-kb cryptic plasmid from Bifidobacterium breve NCFB 2258 and determination of mode of

replication', *FEMS Microbiology Letters*, vol. 174, no. 2, pp. 285–294, May 1999, doi: 10.1111/j.1574-6968.1999.tb13581.x.

[15] **P. Álvarez-Martín, A. B. Flórez, and B. Mayo**, 'Screening for plasmids among human bifidobacteria species: Sequencing and analysis of pBC1 from Bifidobacterium catenulatum L48', *Plasmid*, vol. 57, no. 2, Art. no. 2, Mar. 2007, doi: 10.1016/j.plasmid.2006.07.004.

[16] **Y.-S. Park, K.-H. Kim, J. H. Park, I.-K. Oh, and S.-S. Yoon**, 'Isolation and molecular characterization of a cryptic plasmid from Bifidobacterium longum', *Biotechnol. Lett.*, vol. 30, no. 1, pp. 145–151, Jan. 2008, doi: 10.1007/s10529-007-9504-5.

[17] **G.-S. Moon, U. Wegmann, A. P. Gunning, M. J. Gasson, and A. Narbad**, 'Isolation and characterization of a theta-type cryptic plasmid from Bifidobacterium longum FI10564', *J. Microbiol. Biotechnol.*, vol. 19, no. 4, pp. 403–408, Apr. 2009, doi: 10.4014/jmb.0806.378.

[18] **F. Bottacini *et al.***, 'Discovery of a Conjugative Megaplasmid in Bifidobacterium breve', *Appl. Environ. Microbiol.*, vol. 81, no. 1, pp. 166–176, Jan. 2015, doi: 10.1128/AEM.02871-14.

[19] **T. Odamaki *et al.***, 'Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan', *Sci Rep*, vol. 8, Jan. 2018, doi: 10.1038/s41598-017-18391-x.

[20] **A. Matsumoto, H. Izutsu, N. Miyashita, and M. Ohuchi**, 'Plaque formation by and plaque cloning of Chlamydia trachomatis biovar trachoma', *J Clin Microbiol*, vol. 36, no. 10, pp. 3013–3019, Oct. 1998, doi: 10.1128/JCM.36.10.3013-3019.1998.

[21] **J. M. Ghigo**, 'Natural conjugative plasmids induce bacterial biofilm development', *Nature*, vol. 412, no. 6845, pp. 442–445, Jul. 2001, doi: 10.1038/35086581.

[22] **V. Parashar, M. A. Konkol, D. B. Kearns, and M. B. Neiditch**, 'A plasmid-encoded phosphatase regulates Bacillus subtilis biofilm architecture, sporulation, and genetic competence', *J Bacteriol*, vol. 195, no. 10, pp. 2437–2448, May 2013, doi: 10.1128/JB.02030-12.

[23] **G. B. Coulson *et al.***, 'Transcriptome Reprogramming by Plasmid-Encoded Transcriptional Regulators Is Required for Host Niche Adaption of a Macrophage Pathogen', *Infect Immun*, vol. 83, no. 8, pp. 3137–3145, Aug. 2015, doi: 10.1128/IAI.00230-15.

[24] **K. S. Lang and T. J. Johnson**, 'Transcriptome modulations due to A/C2 plasmid acquisition', *Plasmid*, vol. 80, pp. 83–89, Jul. 2015, doi: 10.1016/j.plasmid.2015.05.005.

[25] **B. S. Weber, P. M. Ly, J. N. Irwin, S. Pukatzki, and M. F. Feldman**, 'A multidrug resistance plasmid contains the molecular switch for type VI secretion in Acinetobacter baumannii', *Proc Natl Acad Sci U S A*, vol. 112, no. 30, pp. 9442–9447, Jul. 2015, doi: 10.1073/pnas.1502966112.

[26] **A. Ranjan *et al.***, 'ESBL-plasmid carriage in E. coli enhances in vitro bacterial competition fitness and serum resistance in some strains of pandemic sequence types without overall fitness cost', *Gut Pathog*, vol. 10, p. 24, 2018, doi: 10.1186/s13099-018-0243-z.

[27] **L. Vial and F. Hommais**, 'Plasmid-chromosome cross-talks', *Environmental Microbiology*, vol. 22, no. 2, pp. 540–556, 2020, doi: 10.1111/1462-2920.14880.

[28] **D. F. Browning and S. J. W. Busby**, 'The regulation of bacterial transcription initiation', *Nat Rev Microbiol*, vol. 2, no. 1, Art. no. 1, Jan. 2004, doi: 10.1038/nrmicro787.

[29] **F. Bottacini *et al.***, 'Global transcriptional landscape and promoter mapping of the gut commensal Bifidobacterium breve UCC2003', *BMC Genomics*, vol. 18, Dec. 2017, doi: 10.1186/s12864-017-4387-x.

[30] **Z. Zhou *et al.***, 'Codon usage is an important determinant of gene expression levels largely through its effects on transcription', *Proceedings of the National Academy of Sciences*, vol. 113, no. 41, pp. E6117–E6125, Oct. 2016, doi: 10.1073/pnas.1606724113.

[31] **H. Li and R. Durbin**, 'Fast and accurate short read alignment with Burrows–Wheeler transform', *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, Jul. 2009, doi: 10.1093/bioinformatics/btp324.

[32] **T. L. Bailey and C. Elkan**, 'Fitting a mixture model by expectation maximization to discover motifs in biopolymers', *Proc Int Conf Intell Syst Mol Biol*, vol. 2, pp. 28–36, 1994.

[33] **G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner**, 'WebLogo: a sequence logo generator', *Genome Res*, vol. 14, no. 6, pp. 1188–1190, Jun. 2004, doi: 10.1101/gr.849004.

[34] **K. Rutherford *et al.***, 'Artemis: sequence visualization and annotation', *Bioinformatics*, vol. 16, no. 10, Art. no. 10, Oct. 2000.

[35] **A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neuböck, and I. L. Hofacker**, 'The Vienna RNA Websuite', *Nucleic Acids Res*, vol. 36, no. Web Server issue, pp. W70–W74, Jul. 2008, doi: 10.1093/nar/gkn188.

[36] **S. X. Ge, D. Jung, and R. Yao**, 'ShinyGO: a graphical gene-set enrichment tool for animals and plants', *Bioinformatics*, vol. 36, no. 8, pp. 2628–2629, Apr. 2020, doi: 10.1093/bioinformatics/btz931.

[37] **M. Kanehisa, S. Goto, S. Kawashima, and A. Nakaya**, 'The KEGG databases at GenomeNet', *Nucleic Acids Res*, vol. 30, no. 1, pp. 42–46, Jan. 2002, doi: 10.1093/nar/30.1.42.

[38] **P. D. Karp *et al.***, 'The BioCyc collection of microbial genomes and metabolic pathways', *Brief Bioinform*, vol. 20, no. 4, pp. 1085–1093, Jul. 2019, doi: 10.1093/bib/bbx085.

[39] **R. Arias-Carrasco, Y. Vásquez-Morán, H. I. Nakaya, and V. Maracaja-Coutinho**, 'StructRNAfinder: an automated pipeline and web server for RNA families prediction', *BMC Bioinformatics*, vol. 19, no. 1, p. 55, Feb. 2018, doi: 10.1186/s12859-018-2052-2.

[40] **M.-J. Cros *et al.***, 'RNAspace.org: An integrated environment for the prediction, annotation, and analysis of ncRNA', *RNA*, vol. 17, no. 11, pp. 1947–1956, Nov. 2011, doi: 10.1261/rna.2844911.

[41] **S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy**, 'Rfam: an RNA family database', *Nucleic Acids Res*, vol. 31, no. 1, pp. 439–441, Jan. 2003.

[42] **B. Buchfink, C. Xie, and D. H. Huson**, 'Fast and sensitive protein alignment using DIAMOND', *Nat Methods*, vol. 12, no. 1, Art. no. 1, Jan. 2015, doi: 10.1038/nmeth.3176.

[43] **Z. Gu, L. Gu, R. Eils, M. Schlesner, and B. Brors**, 'circlize implements and enhances circular visualization in R', *Bioinformatics*, vol. 30, no. 19, pp. 2811–2812, Oct. 2014, doi: 10.1093/bioinformatics/btu393.

[44] **T. Tatusova *et al.***, 'NCBI prokaryotic genome annotation pipeline', *Nucleic Acids Res*, vol. 44, no. 14, pp. 6614–6624, Aug. 2016, doi: 10.1093/nar/gkw569.

[45] **B. Langmead and S. L. Salzberg**, 'Fast gapped-read alignment with Bowtie 2', *Nat Methods*, vol. 9, no. 4, pp. 357–359, Mar. 2012, doi: 10.1038/nmeth.1923.

[46] **S. Nurk, D. Meleshko, A. Korobeynikov, and P. A. Pevzner**, 'metaSPAdes: a new versatile metagenomic assembler', *Genome Res*, vol. 27, no. 5, pp. 824–834, May 2017, doi: 10.1101/gr.213959.116.

[47] **X. Ju, D. Li, and S. Liu**, 'Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria', *Nat Microbiol*, vol. 4, no. 11, pp. 1907–1918, Nov. 2019, doi: 10.1038/s41564-019-0500-z.

[48] **S. Gupta and D. Pal**, 'Clusters of hairpins induce intrinsic transcription termination in bacteria', *Sci Rep*, vol. 11, no. 1, Art. no. 1, Aug. 2021, doi: 10.1038/s41598-021-95435-3.

[49] **G. King and N. E. Murray**, 'Restriction alleviation and modification enhancement by the Rac prophage of Escherichia coli K-12', *Mol Microbiol*, vol. 16, no. 4, pp. 769–777, May 1995, doi: 10.1111/j.1365-2958.1995.tb02438.x.

[50] **R. D. Morgan, T. K. Bhatia, L. Lovasco, and T. B. Davis**, 'MmeI: a minimal Type II restriction-modification system that only modifies one DNA strand for host protection', *Nucleic Acids Res*, vol. 36, no. 20, pp. 6558–6570, Nov. 2008, doi: 10.1093/nar/gkn711.

[51] **R. L. Dineen, C. Penno, P. Kelleher, M. J. B. Bourin, M. O'Connell-Motherway, and D. van Sinderen**, 'Molecular analysis of the replication functions of the bifidobacterial conjugative megaplasmid pMP7017', *Microb Biotechnol*, May 2021, doi: 10.1111/1751-7915.13810.

[52] **B. D. Janssen and C. S. Hayes**, 'The tmRNA ribosome rescue system', *Adv Protein Chem Struct Biol*, vol. 86, pp. 151–191, 2012, doi: 10.1016/B978-0-12-386497-0.00005-0.

[53] **Z. Weinberg, J. Perreault, M. M. Meyer, and R. R. Breaker**, 'Extraordinary Structured Noncoding RNAs Revealed by Bacterial Metagenome Analysis', *Nature*, vol. 462, no. 7273, pp. 656–659, Dec. 2009, doi: 10.1038/nature08586.

[54] **Z. Weinberg *et al.***, 'Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes', *Genome Biol*, vol. 11, no. 3, p. R31, 2010, doi: 10.1186/gb-2010-11-3-r31.

[55] **J. B. Plotkin and G. Kudla**, 'Synonymous but not the same: the causes and consequences of codon bias', *Nat Rev Genet*, vol. 12, no. 1, Art. no. 1, Jan. 2011, doi: 10.1038/nrg2899.

[56] **Y. Liu, Q. Yang, and F. Zhao**, 'Synonymous but not Silent: The Codon Usage Code for Gene Expression and Protein Folding', *Annu Rev Biochem*, vol. 90, pp. 375–401, Jun. 2021, doi: 10.1146/annurev-biochem-071320-112701.

[57] **T. Tuller**, 'Codon bias, tRNA pools and horizontal gene transfer', *Mob Genet Elements*, vol. 1, no. 1, pp. 75–77, 2011, doi: 10.4161/mge.1.1.15400.

[58] **M. Callens, C. Scornavacca, and S. Bedhomme**, 'Evolutionary responses to codon usage of horizontally transferred genes in Pseudomonas aeruginosa: gene retention, amelioration and compensatory evolution', *Microb Genom*, vol. 7, no. 6, p. 000587, Jun. 2021, doi: 10.1099/mgen.0.000587.

[59] **T. Tuller *et al.***, 'Association between translation efficiency and horizontal gene transfer within microbial communities', *Nucleic Acids Res*, vol. 39, no. 11, pp. 4743–4755, Jun. 2011, doi: 10.1093/nar/gkr054.

[60] **M. Gouy and C. Gautier**, 'Codon usage in bacteria: correlation with gene expressivity', *Nucleic Acids Research*, vol. 10, no. 22, pp. 7055–7074, Nov. 1982, doi: 10.1093/nar/10.22.7055.

[61] **M. S. Khoroshkin, S. A. Leyn, D. Van Sinderen, and D. A. Rodionov**, 'Transcriptional Regulation of Carbohydrate Utilization Pathways in the Bifidobacterium Genus', *Frontiers in Microbiology*, vol. 7, 2016, Accessed: Oct. 24, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2016.00120

[62] **N. Lanigan, E. Kelly, A. A. Arzamasov, C. Stanton, D. A. Rodionov, and D. van Sinderen**, 'Transcriptional control of central carbon metabolic flux in Bifidobacteria by two functionally similar, yet distinct LacI-type regulators', *Sci Rep*, vol. 9, no. 1, Art. no. 1, Nov. 2019, doi: 10.1038/s41598-019-54229-4.

[63] **S. M. Kelly *et al.***, 'Bifidobacterial biofilm formation is a multifactorial adaptive phenomenon in response to bile exposure', *Sci Rep*, vol. 10, p. 11598, Jul. 2020, doi: 10.1038/s41598-020-68179-9.

[64] **G. B. Whitfield *et al.***, 'Discovery and characterization of a Gram-positive Pel polysaccharide biosynthetic gene cluster', *PLoS Pathog*, vol. 16, no. 4, p. e1008281, Apr. 2020, doi: 10.1371/journal.ppat.1008281.

[65] **R. E. Impey *et al.***, 'Identification of two dihydrodipicolinate synthase isoforms from Pseudomonas aeruginosa that differ in allosteric regulation', *The FEBS Journal*, vol. 287, no. 2, pp. 386–400, 2020, doi: 10.1111/febs.15014.

[66] **M. O'Connell Motherway *et al.***, 'A Bifidobacterial pilus-associated protein promotes colonic epithelial proliferation', *Mol. Microbiol.*, vol. 111, no. 1, pp. 287–301, 2019, doi: 10.1111/mmi.14155.

[67] **Z. R. Newman, J. M. Young, N. T. Ingolia, and G. M. Barton**, 'Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9', *Proceedings of the National Academy of Sciences*, vol. 113, no. 10, pp. E1362–E1371, Mar. 2016, doi: 10.1073/pnas.1518976113.

[68] **S. Bahiri-Elitzur and T. Tuller**, 'Codon-based indices for modeling gene expression and transcript evolution', *Comput Struct Biotechnol J*, vol. 19, pp. 2646–2663, Apr. 2021, doi: 10.1016/j.csbj.2021.04.042.

[69] **F. Zhao *et al.***, 'Genome-wide role of codon usage on transcription and identification of potential regulators', *Proceedings of the National Academy of Sciences*, vol. 118, no. 6, p. e2022590118, Feb. 2021, doi: 10.1073/pnas.2022590118.

[70] **F. Bottacini *et al.***, 'Comparative genomics of the Bifidobacterium breve taxon', *BMC Genomics*, vol. 15, no. 1, Mar. 2014, doi: 10.1186/1471-2164-15-170.

[71] **K. Albert, A. Rani, and D. A. Sela**, 'Comparative Pangenomics of the Mammalian Gut Commensal Bifidobacterium longum', *Microorganisms*, vol. 8, no. 1, p. 7, Dec. 2019, doi: 10.3390/microorganisms8010007.

[72] **S. Fanning, L. J. Hall, and D. van Sinderen**, 'Bifidobacterium breve UCC2003 surface exopolysaccharide production is a beneficial trait mediating commensal-host interaction through immune modulation and pathogen protection', *Gut Microbes*, vol. 3, no. 5, pp. 420–425, Sep. 2012, doi: 10.4161/gmic.20630.

[73] **S. C. Chew *et al.***, 'Dynamic Remodeling of Microbial Biofilms by Functionally Distinct Exopolysaccharides', *mBio*, vol. 5, no. 4, pp. e01536-14, Aug. 2014, doi: 10.1128/mBio.01536-14.

[74] **S. Periasamy *et al.***, 'Pseudomonas aeruginosa PAO1 exopolysaccharides are important for mixed species biofilm community development and stress tolerance',

Frontiers in Microbiology, vol. 6, 2015, Accessed: Nov. 01, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2015.00851

[75] **B. Harr and C. Schlötterer**, 'Gene expression analysis indicates extensive genotype-specific crosstalk between the conjugative F-plasmid and the E. coli chromosome', *BMC Microbiol*, vol. 6, p. 80, Sep. 2006, doi: 10.1186/1471-2180-6-80.

[76] **M. Shintani *et al.***, 'Response of the Pseudomonas host chromosomal transcriptome to carriage of the IncP-7 plasmid pCAR1', *Environmental Microbiology*, vol. 12, no. 6, pp. 1413–1426, 2010, doi: 10.1111/j.1462-2920.2009.02110.x.

CHAPTER VI

GENERAL DISCUSSION AND FUTURE PERSPECTIVES

Certain members of the genus *Bifidobacterium* hold significant commercial and clinical interest due to their purported probiotic properties, and as such they have been widely utilised in the promotion and maintenance of human health, in addition to their most recent application, health restoration, particularly functioning as adjuvants to enhance the efficacy of the vaccine/treatment by eliciting a stronger immune response [1]–[4]. From an economic perspective, the value of the global microbiome market is expected to advance at a compound annual growth rate of 22.95% during the forecasted period of 2022-2029 [5]. While the market value for engineered bacterial therapeutics is projected to expand at a substantial compound annual growth rate of 35.32% for 2023-2030 [6]. As the scientific, healthcare and public attention continues to escalate key aspects pertaining to *Bifidobacterium* biology and host interactions have been elucidated [7]–[10]. However, despite immense progress numerous gaps in our mechanistic understanding of their benefit(s) to human health remain. Consequently, the European Food Safety Authority concludes that a cause and effect relationship has not yet been established and this regulatory body consequently has not approved a single health claim pertaining to *Bifidobacterium* to date [11].

It is widely accepted that the availability of molecular tools for the study and genetic engineering of bifidobacteria falls short of experimental needs, which subsequently hinders the advancement of research, innovation and potential novel applications of these bacterium [12], [13]. Plasmids are a central asset for these purposes, however due to the lack of characterised residential plasmids of bifidobacterial origin only a small number of suitable vectors are available. Further to this, insert-size constraints, low-transformation efficiencies and plasmid-instability pose major constraints on plasmid-mediated genetic engineering in bifidobacterial research [12], [14]–[16]. This is further exacerbated by the inherent genetic intractability of bifidobacteria, primarily due to the abundance and diversity of chromosomally-encoded restriction modification (RM) systems [17], [18]. While non-cryptic plasmids were once considered an atypical feature of this genus, genome sequencing efforts enabled the identification of a conjugative 190 kb megaplasmid pMP7017 residing in *B. breve* JCM7017 [19]. Subsequent investigations exposed the presence of additional pMP7017-related plasmids within members of the *B. longum* subsp. *longum*, highlighting the prevalence of this emerging family of bifidobacterial megaplasmids [20]. Based on these premises, the replication and conjugative

functions of pMP7017 were investigated in this thesis with a view to exploit these functions for the development of molecular tools for the future genetic manipulation of members of this genus, while aspects pertaining to the plasmid's gene-content, distribution, diversity and persistence were also explored.

To date research efforts concerning the molecular analysis of plasmid replication and its control have focused on plasmids derived from Gram-negative and pathogenic Gram-positive (G+) bacteria. Resultingly, there are major gaps in our knowledge pertaining to plasmid replication systems present in G+ commensals [21]. The scientific investigations described in **Chapter II** concerned the characterisation of pMP7017 replicative functions, which serve to address the shortfall of knowledge concerning bifidobacterial plasmid replication. Based on these data, it was established that pMP7017 is a low copy, iteron-regulated plasmid replicating by a so-called theta-type mechanism. Furthermore, the findings obtained enabled the development of pRD1.3, a high-capacity *E. coli—Bifidobacterium* shuttle vector which can be employed for the cloning and expression of large genes or operons without instability and/or toxicity issues commonly encountered with high-copy plasmid vectors [22]. In addition, a series of pRD1 vectors with varying levels of stability were developed which could be employed in future works exploring stabilisation functions, such as plasmid or chromosomal partitioning cassettes. Further to this, we envisage the development of next-generation genetic tools based on pMP7017-derived vectors for a range of diverse genetic engineering applications by leveraging alternative pMP7017-encoded functions investigated as part of this thesis.

DNA cloning is a routine procedure in standard recombinant DNA technology that relies on plasmid-encoded antibiotic resistance genes to select and maintain recombinant bacteria. This requires extensive screening of colonies to identify recombinant plasmids due to the high propensity of false positives and empty vectors, posing a major drawback which is neither time- nor cost-efficient. A potential solution to overcome these limitations is to develop direct positive selection vectors for use in bifidobacteria by exploiting pMP7017-encoded Type II toxin-antitoxin (TA) systems (**Chapter V**). Toxin disruption- and replacement-based selection vectors contain a toxin gene which is inactivated upon insertion of foreign DNA permitting growth/survival of recombinants harbouring insert-bearing plasmids only [23]–[25]. More complex systems based on TA pairs could also be developed such as the separated-component-stabilization systems developed for *Streptomyces* and *E. coli*,

these toxin neutralization-based strategies involve the use of a recombinant host bacterium with a(n) (anti)toxin gene integrated into the chromosome and a vector expressing the cognate (anti)toxin [26], [27]. The availability of such tools for use in bifidobacteria may also be of significance in the development of genetically enhanced probiotics and/or live biotherapeutics as the host–construct dependency of these systems will ensure the biocontainment of engineered vectors [28]–[32]. Considering this, pMP7017-encoded Type II toxin antitoxin systems described in **Chapter V** could be of considerable value for molecular biology and biotechnological purposes particularly when combined with the high-capacity properties of vectors employing pMP7017 replication functions (**Chapter II**). However, it is necessary to verify the functionality and efficacy of these pMP7017-derived stabilization systems and assess their activity in evolutionarily divergent bacteria to determine their host-range specificities which may dictate the scope of their application [26], [33], [34]. For these purposes, utility of pMP7017-derived unstable vectors, pRD1.9 and pRD1.10 (**Chapter II**), may be advantageous.

In a preceding study, conjugation experiments conclusively demonstrated that pMP7017 has the capacity to spread and persist in various strains of the *B. breve* and *B. longum* (sub)species, however its mechanism of transfer had not been investigated [19]. In **Chapter III** we identify the genetic features of the pMP7017 megaplasmid required for conjugation and based on these findings we propose that pMP7017 employs a type IV secretion system to translocate its DNA. While this work represents the first characterisation of a conjugal system in bifidobacteria and contributes to increased knowledge of G+ conjugal systems, the picture of pMP7017 horizontal spread is far from complete. To better understand conjugation in bifidobacteria it is necessary to identify and characterise the pMP7017 relaxase, a key enzyme of the DNA-processing complex. Relaxases of G+ origin are highly diverse both structurally and mechanistically, a fact reflected in the continuous expansion of relaxase MOB groups due to the identification of new non-canonical relaxases from G+ systems [35]–[40]. Considering this, it is possible that pMP7017 encodes a novel relaxase that mediates the transfer of this conjugative megaplasmid. Comobilization assays (**Chapter III**) indicate a 720 bp region flanking a cluster of transfer genes to contain the presumed enzymatic substrate of the pMP7017 relaxase, therefore this region could be used as bait in pull-down assays in future works to selectively capture proteins that bind to it from cellular lysates, which may allow the identification of the

pMP7017 relaxase. The most unexpected finding from this research is the number of essential ATPases for pMP7017 transfer, which was beyond anticipation, and the genetic organisation of essential components which are contained within two unconnected genomic *loci*. Given that this work was limited to the genetic delineation of conjugation-associated genes, future research efforts may focus on elucidating the biochemical, structural, and/or mechanistic details of this unique bifidobacterial conjugal system and refine or indeed alter the proposed conjugal model accordingly.

Complete elucidation of this bifidobacterial conjugation mechanism may further facilitate the development of genetic tools for a variety of biotechnological applications. By employing an effective bifidobacterial conjugative system as a mode of plasmid/vector introduction, beneficial genes or whole gene clusters without size limitation could be transferred to their *Bifidobacterium* host *in vivo* to impart strain/species-specific advantageous traits, as observed for *Bacillus* spp. [41]. Furthermore, the importance of developing such vector systems could go beyond strain improvement, as they may serve as a promising vehicles for the delivery of other useful gene products such as vaccines or immunostimulants into the human intestinal tract [42]–[45]. However, the viability and scope of these applications hinges in part on the plasmid's host range, which will determine successful bacterium–plasmid combinations and this should be explored in future works. A narrow host range may enable the application of a pMP7017-derived conjugative vector in agriculture or food industry. Generally for these applications high-copy vectors are preferred to ensure adequate expression of these genes of interest. While pMP7017-derived vectors were determined to be low copy (**Chapter II**), gene expression can be fine-tuned, controlled, or significantly amplified through the use of strong and/or inducible promoters and transitional efficiency can be enhanced through codon usage optimisation, for which the analyses conducted in **Chapter V** may be of great use. In **Chapter II** the stability of the pRD1.3 vector was determined to be highly stable in the absence of selective pressure. However, outside of research applications the utility of pMP7017-encoded TA systems to ensure the desired phenotype remains in the population without the use of antibiotics would be more desirable. In contrast, a broader host range would enable its potential use in microbiome editing strategies to eliminate specific bacteria such as antibiotic-resistant bacteria/ pathogens or indeed a precision tool for microbial gene editing.

Controlled manipulation of wild type strains in the natural environment is at the leading-edge of biotechnology. As conjugation is a natural transfer mechanism it is considered the most favourable entry point for microbe engineering within a community context, however, tractable *in situ* interventions requires both efficiency and specificity [46]–[48]. In G+ conjugal systems contact of donor-recipients is mediated by plasmid-encoded surface adhesins, this contact-dependency represents the first rate-limiting step in the conjugal process [49]–[54]. In **Chapter IV** we revealed that conjugal transfer of pMP7017 depends on plasmid-encoded, sortase-dependent cell surface structures which likely support bacterial co-aggregation and facilitate formation of the mating pair. Furthermore, we propose that genetic diversity among adhesion *loci* of pMP7017 and related plasmids may represent a key determinant of plasmid host-range specificity which likely underpins pMP7017s host-range shift. Future research efforts may seek to establish the legitimacy of species-specificity conferred by conjugation-associated adhesions. If this notion can be validated, these functions may be exploited for biotechnological applications to facilitate transfer of pMP7017-derived conjugative vectors to targeted species/strains within (natural) bacterial populations, rendering untargeted bacteria as incompatible recipients [46].

The continued study of bifidobacterial conjugation (**Chapter III**) and plasmid-replication (**Chapter II**) systems may also, in an auxiliary sense, impart knowledge which may contribute to the fight against antimicrobial resistance (AMR). AMR is a progressively increasing burden to health care systems and remains one of the World Health Organizations top 10 biggest threats to global health since its initial declaration in 2014 [55]. Considering the crucial role of conjugative plasmids in the acquisition and transmission of AMR genes, biological approaches to prevent plasmid invasiveness via conjugation-specific inhibition or indeed to eliminate AMR-plasmids via inhibition of plasmid replication systems are envisaged as preeminent strategies to combat microbial resistance [56]–[59]. However, target heterogeneity poses a major concern for the use of these antibacterial strategies seeking to exploit plasmid-encoded traits, as such, they could be effective against AMR plasmids and non-AMR plasmids alike. Considering this, premature utility of such strategies without careful consideration of off-target effects seems imprudent. Better understanding of the differences and commonalities among conjugative- and plasmid-specific replication systems originating from taxonomically diverse bacterial lineages would aid in

improving target-specificity and enable a more comprehensive evaluation of the potential ramifications of target heterogeneity when employing these novel prophylactic antibacterial strategies *in vivo*.

Natural gene circuits predominantly rely on transcriptional and post-transcriptional regulatory factors such as RNAs and proteins to govern cellular activities. These regulatory factors are commonly repurposed to develop, and influence the dynamics of engineered microbial gene circuits [60]–[62]. A major challenge in gene circuit engineering is the burden of synthetic circuits on the host cell, particularity the impact of engineered circuit components on the expression and functionality of endogenous genes and proteins, which can inadvertently reduce host fitness [63]. In **Chapter V** we describe a novel 34 bp pMP7017-specific DNA sequence regulatory motif that overlaps the -10 hexamer of various promoters and speculate that this motif represents a transcription factor binding site. Further to this, we indicate that this motif is unique to pMP7017 and related megaplasmids due to its absence among bifidobacterial chromosomes. Future works may consider the use of DNA pull-down assays to identify the protein(s) to which it binds, as this DNA-protein pair may provide an opportunity to enact precise regulatory control of gene expression in the development of *Bifidobacterium* as an engineering host and overcome the pleiotropic consequences associated with the use of chromosomally-derived regulators [64].

Characterization of plasmid backbone genes i.e. those encoding replication-, conjugation- or stability-functions, as well as copy number control mechanisms and antimicrobial resistance are central to plasmid research due to their significant value for synthetic biology applications; however, plasmid accessory gene cargoes are often overlooked [65], [66]. Comparative analysis of pMP7017 and related plasmids revealed that this megaplasmid family exhibits a highly diverse and dynamic accessory genome which is dominated by genes of unknown function (**Chapter V**). Furthermore, the average percentage of hypothetical genes contained within pMP7017 was determined to be ≥25% higher than those estimated among all available *B. breve* and *B. longum* genome sequences (**Chapter V**). It is likely that this large discrepancy reflects the existence of novel gene functions rather than merely resulting from less comprehensive gene predictions. Considering this, characterisation of hypothetical, or so-called “cryptic” genes harboured by bifidobacterial megaplasmids represents an

interesting prospect for future research, as novel functions or plasmid-encoded adaptive traits may be uncovered such as persistence, phage defence, stress response, or the capacity to exploit new compounds or environments which can facilitate niche migration [67]–[69].

Overall, the work outlined in this thesis provides a wide range of novel findings which markedly advances our understanding of the replicative- and conjugative-functions of pMP7017 and imparts comprehensive insights into the prevalence and diversity of this megaplasmid family. This generated information represents a unique opportunity for the development of advanced genetic tools enabling better mechanistic understanding of *Bifidobacterium* and potentially promote the synthetic biology applications of this health-promoting microorganism. Additionally, a catalogue of pMP7017 mutants were created as part of this thesis which may prove useful for future works, for example the cell surface mutants generated in **Chapter IV** of this thesis may be assessed as potential gut adhesion factors.

REFERENCES

[1] **L. Vitetta, E. T. Saltzman, M. Thomsen, T. Nikov, and S. Hall**, 'Adjuvant Probiotics and the Intestinal Microbiome: Enhancing Vaccines and Immunotherapy Outcomes', *Vaccines (Basel)*, vol. 5, no. 4, p. 50, Dec. 2017, doi: 10.3390/vaccines5040050.

[2] **B. Sun *et al.***, 'Bifidobacterium lactis Probio-M8 Adjuvant Treatment Confers Added Benefits to Patients with Coronary Artery Disease via Target Modulation of the Gut-Heart/-Brain Axes', *mSystems*, vol. 7, no. 2, pp. e00100-22, Mar. 2022, doi: 10.1128/msystems.00100-22.

[3] **C. Hidalgo-Cantabrana, S. Delgado, L. Ruiz, P. Ruas-Madiedo, B. Sánchez, and A. Margolles**, 'Bifidobacteria and Their Health-Promoting Effects', *Microbiology Spectrum*, vol. 5, no. 3, p. 10.1128/microbiolspec.bad-0010–2016, Jun. 2017, doi: 10.1128/microbiolspec.bad-0010-2016.

[4] **M. Esteban-Torres, L. Ruiz, G. A. Lugli, M. Ventura, A. Margolles, and D. van Sinderen**, 'Editorial: Role of Bifidobacteria in Human and Animal Health and Biotechnological Applications', *Frontiers in Microbiology*, vol. 12, 2021, Accessed: Feb. 08, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmicb.2021.785664

[5] 'Human Microbiome Market Size, Scope, Trends, & Industry Analysis'. https://www.databridgemarketresearch.com/reports/global-human-microbiome-market (accessed Jun. 23, 2023).

[6] **R. and M. ltd**, 'Microbiome Therapeutics Market Size, Share & Trends Analysis Report by Type (FMT, Microbiome Drugs), by Application (C. difficile, Crohn's Disease, Inflammatory Bowel Disease, Diabetes), by Region, and Segment Forecasts, 2022-2030'. https://www.researchandmarkets.com/reports/3989902/microbiome-therapeutics-market-size-share-and (accessed Jun. 23, 2023).

[7] **L. Ruiz, S. Delgado, P. Ruas-Madiedo, B. Sánchez, and A. Margolles**, 'Bifidobacteria and Their Molecular Communication with the Immune System', *Front. Microbiol.*, vol. 8, 2017, doi: 10.3389/fmicb.2017.02345.

[8] **L. Ruiz, S. Delgado, P. Ruas-Madiedo, A. Margolles, and B. Sánchez**, 'Proteinaceous Molecules Mediating Bifidobacterium-Host Interactions', *Frontiers in Microbiology*, vol. 7, 2016, Accessed: Jun. 27, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2016.01193

[9] **G. Alessandri, M. C. Ossiprandi, J. MacSharry, D. van Sinderen, and M. Ventura**, 'Bifidobacterial Dialogue With Its Human Host and Consequent Modulation of the Immune System', *Frontiers in Immunology*, vol. 10, p. 2348, 2019, doi: 10.3389/fimmu.2019.02348.

[10] **A. Hickey *et al.***, 'Bifidobacterium breve Exopolysaccharide Blocks Dendritic Cell Maturation and Activation of CD4+ T Cells', *Frontiers in Microbiology*, vol. 12, p. 1441, 2021, doi: 10.3389/fmicb.2021.653587.

[11] 'Food and Feed Information Portal Database | FIP'. https://ec.europa.eu/food/food-feed-portal/screen/health-claims/eu-register (accessed Jul. 10, 2023).

[12] **Z. Sun, A. Baur, D. Zhurina, J. Yuan, and C. U. Riedel**, 'Accessing the inaccessible: molecular tools for bifidobacteria', *Appl Environ Microbiol*, vol. 78, no. 15, pp. 5035–5042, Aug. 2012, doi: 10.1128/AEM.00551-12.

[13] **L. Ruiz, M. Esteban-Torres, and D. van Sinderen**, 'A Resource for Cloning and Expression Vectors Designed for Bifidobacteria: Overview of Available Tools and Biotechnological Applications', *Methods Mol Biol*, vol. 2278, pp. 157–182, 2021, doi: 10.1007/978-1-0716-1274-3_14.

[14] **M. O. Motherway, J. O'Driscoll, G. F. Fitzgerald, and D. V. Sinderen**, 'Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in Bifidobacterium breve UCC2003', *Microbial Biotechnology*, vol. 2, no. 3, pp. 321–332, 2009, doi: 10.1111/j.1751-7915.2008.00071.x.

[15] **S. Fukiya, M. Sakanaka, and A. Yokota**, 'Chapter 15 - Genetic Manipulation and Gene Modification Technologies in Bifidobacteria', in *The Bifidobacteria and Related Organisms*, P. Mattarelli, B. Biavati, W. H. Holzapfel, and B. J. B. Wood, Eds. Academic Press, 2018, pp. 243–259. doi: 10.1016/B978-0-12-805060-6.00015-6.

[16] **F. Zuo, S. Chen, and H. Marcotte**, 'Engineer probiotic bifidobacteria for food and biomedical applications - Current status and future prospective', *Biotechnology Advances*, vol. 45, p. 107654, Dec. 2020, doi: 10.1016/j.biotechadv.2020.107654.

[17] **F. Bottacini *et al.***, 'Comparative genome and methylome analysis reveals restriction/modification system diversity in the gut commensal Bifidobacterium breve', *Nucleic Acids Res*, vol. 46, no. 4, pp. 1860–1877, Feb. 2018, doi: 10.1093/nar/gkx1289.

[18] **M. O' Connell Motherway *et al.***, 'Identification of restriction-modification systems of Bifidobacterium animalis subsp. lactis CNCM I-2494 by SMRT sequencing and associated methylome analysis', *PLoS One*, vol. 9, no. 4, p. e94875, 2014, doi: 10.1371/journal.pone.0094875.

[19] **F. Bottacini *et al.***, 'Discovery of a Conjugative Megaplasmid in Bifidobacterium breve', *Appl. Environ. Microbiol.*, vol. 81, no. 1, pp. 166–176, Jan. 2015, doi: 10.1128/AEM.02871-14.

[20] **T. Odamaki *et al.***, 'Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan', *Sci Rep*, vol. 8, Jan. 2018, doi: 10.1038/s41598-017-18391-x.

[21] **G. del Solar, R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas**, 'Replication and Control of Circular Bacterial Plasmids', *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, pp. 434–464, Jun. 1998.

[22] **R. L. Dineen, C. Penno, P. Kelleher, M. J. B. Bourin, M. O'Connell-Motherway, and D. van Sinderen**, 'Molecular analysis of the replication functions of the bifidobacterial conjugative megaplasmid pMP7017', *Microb Biotechnol*, May 2021, doi: 10.1111/1751-7915.13810.

[23] **P. Bernard and M. Couturier**, 'Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes', *J Mol Biol*, vol. 226, no. 3, pp. 735–745, Aug. 1992, doi: 10.1016/0022-2836(92)90629-x.

[24] **P. Bernard, P. Gabarit, E. M. Bahassi, and M. Couturier**, 'Positive-selection vectors using the F plasmid ccdB killer gene', *Gene*, vol. 148, no. 1, pp. 71–74, Oct. 1994, doi: 10.1016/0378-1119(94)90235-6.

[25] **Y.-J. Choi, T.-T. Wang, and B. H. Lee**, 'Positive selection vectors', *Crit Rev Biotechnol*, vol. 22, no. 3, pp. 225–244, 2002, doi: 10.1080/07388550290789504.

[26] **D. Stieber, P. Gabant, and C. Y. Szpirer**, 'The art of selective killing: plasmid toxin/antitoxin systems and their technological applications', *BioTechniques*, vol. 45, no. 3, pp. 344–346, Sep. 2008, doi: 10.2144/000112955.

[27] **L. Sevillano, M. Díaz, and R. I. Santamaría**, 'Stable expression plasmids for Streptomyces based on a toxin-antitoxin system', *Microb Cell Fact*, vol. 12, no. 1, Art. no. 1, Dec. 2013, doi: 10.1186/1475-2859-12-39.

[28] **L. G. Bermúdez-Humarán and P. Langella**, 'Live bacterial biotherapeutics in the clinic', *Nat Biotechnol*, vol. 36, no. 9, Art. no. 9, Oct. 2018, doi: 10.1038/nbt.4248.

[29] **D. Ağagündüz *et al.***, 'Recent developments in the probiotics as live biotherapeutic products (LBPs) as modulators of gut brain axis related neurological conditions', *J Transl Med*, vol. 20, p. 460, Oct. 2022, doi: 10.1186/s12967-022-03609-y.

[30] **P. W. O'Toole, J. R. Marchesi, and C. Hill**, 'Next-generation probiotics: the spectrum from probiotics to live biotherapeutics', *Nat Microbiol*, vol. 2, no. 5, Art. no. 5, Apr. 2017, doi: 10.1038/nmicrobiol.2017.57.

[31] **C. Mignon, R. Sodoyer, and B. Werle**, 'Antibiotic-Free Selection in Biotherapeutics: Now and Forever', *Pathogens*, vol. 4, no. 2, pp. 157–181, Apr. 2015, doi: 10.3390/pathogens4020157.

[32] **J. Lin, Y. Guo, J. Yao, K. Tang, and X. Wang**, 'Applications of toxin-antitoxin systems in synthetic biology', *Engineering Microbiology*, vol. 3, no. 2, p. 100069, Jun. 2023, doi: 10.1016/j.engmic.2023.100069.

[33] **R. Grady and F. Hayes**, 'Axe–Txe, a broad-spectrum proteic toxin–antitoxin system specified by a multidrug-resistant, clinical isolate of Enterococcus faecium', *Molecular Microbiology*, vol. 47, no. 5, pp. 1419–1432, 2003, doi: 10.1046/j.1365-2958.2003.03387.x.

[34] **R. Leplae, D. Geeraerts, R. Hallez, J. Guglielmini, P. Drèze, and L. Van Melderen**, 'Diversity of bacterial type II toxin–antitoxin systems: a comprehensive search and functional analysis of novel families', *Nucleic Acids Res*, vol. 39, no. 13, pp. 5513–5525, Jul. 2011, doi: 10.1093/nar/gkr131.

[35] **M. P. Garcillán-Barcia, M. V. Francia, and F. de la Cruz**, 'The diversity of conjugative relaxases and its application in plasmid classification', *FEMS Microbiol. Rev.*, vol. 33, no. 3, pp. 657–687, May 2009, doi: 10.1111/j.1574-6976.2009.00168.x.

[36] **S. S. Abby, J. Cury, J. Guglielmini, B. Néron, M. Touchon, and E. P. C. Rocha**, 'Identification of protein secretion systems in bacterial genomes', *Sci Rep*, vol. 6, no. 1, Art. no. 1, Mar. 2016, doi: 10.1038/srep23080.

[37] **M. P. Garcillán-Barcia, S. Redondo-Salvo, L. Vielva, and F. de la Cruz**, 'MOBscan: Automated Annotation of MOB Relaxases', in *Horizontal Gene Transfer: Methods and Protocols*, F. de la Cruz, Ed. New York, NY: Springer US, 2020, pp. 295–308. doi: 10.1007/978-1-4939-9877-7_21.

[38] **D. L. Guzmán-Herrador and M. Llosa**, 'The secret life of conjugative relaxases', *Plasmid*, vol. 104, p. 102415, Jul. 2019, doi: 10.1016/j.plasmid.2019.102415.

[39] **J. A. Wisniewski, D. A. Traore, T. L. Bannam, D. Lyras, J. C. Whisstock, and J. I. Rood**, 'TcpM: a novel relaxase that mediates transfer of large conjugative plasmids from Clostridium perfringens', *Molecular Microbiology*, vol. 99, no. 5, pp. 884–896, 2016, doi: 10.1111/mmi.13270.

[40] **M. V. Francia, D. B. Clewell, F. de la Cruz, and G. Moncalián**, 'Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases', *PNAS*, vol. 110, no. 33, pp. 13606–13611, Aug. 2013, doi: 10.1073/pnas.1310037110.

[41] **E. K. Phillips, J. A. Cannon, Y. Zhou, K. S. Bonifer, and T. B. Reynolds**, 'Conjugation-Mediated Plasmid Transfer Enables Genetic Modification of Diverse

Bacillus Species', *Microbiology Spectrum*, vol. 11, no. 2, pp. e03700-22, Mar. 2023, doi: 10.1128/spectrum.03700-22.

[42] **S.-H. Hong**, 'Influence of Microbiota on Vaccine Effectiveness: "Is the Microbiota the Key to Vaccine-induced Responses?"', *J Microbiol.*, vol. 61, no. 5, pp. 483–494, May 2023, doi: 10.1007/s12275-023-00044-6.

[43] **N. Kazemifard, A. Dehkohneh, and S. Baradaran Ghavami**, 'Probiotics and probiotic-based vaccines: A novel approach for improving vaccine efficacy', *Front Med (Lausanne)*, vol. 9, p. 940454, Oct. 2022, doi: 10.3389/fmed.2022.940454.

[44] **D. J. Lynn, S. C. Benson, M. A. Lynn, and B. Pulendran**, 'Modulation of immune responses to vaccination by the microbiota: implications and potential mechanisms', *Nat Rev Immunol*, vol. 22, no. 1, Art. no. 1, Jan. 2022, doi: 10.1038/s41577-021-00554-7.

[45] **S. E. de Jong, A. Olin, and B. Pulendran**, 'The Impact of the Microbiome on Immunity to Vaccination in Humans', *Cell Host Microbe*, vol. 28, no. 2, pp. 169–179, Aug. 2020, doi: 10.1016/j.chom.2020.06.014.

[46] **M. Robledo *et al.***, 'Targeted bacterial conjugation mediated by synthetic cell-to-cell adhesions', *Nucleic Acids Research*, vol. 50, no. 22, pp. 12938–12950, Dec. 2022, doi: 10.1093/nar/gkac1164.

[47] **C. Ronda, S. P. Chen, V. Cabral, S. J. Yaung, and H. H. Wang**, 'Metagenomic engineering of the mammalian gut microbiome in situ', *Nat Methods*, vol. 16, no. 2, pp. 167–170, Feb. 2019, doi: 10.1038/s41592-018-0301-y.

[48] **F. Zuo and H. Marcotte**, 'Advancing mechanistic understanding and bioengineering of probiotic lactobacilli and bifidobacteria by genome editing', *Current Opinion in Biotechnology*, vol. 70, pp. 75–82, Aug. 2021, doi: 10.1016/j.copbio.2020.12.015.

[49] **M. Bhatty *et al.***, 'Enterococcus faecalis pCF10-encoded surface proteins PrgA, PrgB (aggregation substance) and PrgC contribute to plasmid transfer, biofilm formation and virulence', *Mol Microbiol*, vol. 95, no. 4, pp. 660–677, Feb. 2015, doi: 10.1111/mmi.12893.

[50] **A. Schmitt *et al.***, 'PrgB Promotes Aggregation, Biofilm Formation, and Conjugation through DNA Binding and Compaction', *Mol Microbiol*, vol. 109, no. 3, pp. 291–305, Aug. 2018, doi: 10.1111/mmi.13980.

[51] **G. B. Jensen, L. Andrup, A. Wilcks, L. Smidt, and O. M. Poulsen**, 'The aggregation-mediated conjugation system of Bacillus thuringiensis subsp. israelensis: Host range and kinetics of transfer', *Current Microbiology*, vol. 33, no. 4, pp. 228–236, Oct. 1996, doi: 10.1007/s002849900105.

[52] **S. Leanti La Rosa, M. Camila Montealegre, K. V. Singh, and B. E. Murray**, 'Enterococcus faecalis Ebp pili are important for cell-cell aggregation and intraspecies gene transfer', *Microbiology (Reading)*, vol. 162, no. Pt 5, pp. 798–802, May 2016, doi: 10.1099/mic.0.000276.

[53] **S. A. Revitt-Mills, T. D. Watts, D. Lyras, V. Adams, and J. I. Rood**, 'The ever-expanding tcp conjugation locus of pCW3 from Clostridium perfringens', *Plasmid*, vol. 113, p. 102516, Jan. 2021, doi: 10.1016/j.plasmid.2020.102516.

[54] **A. Susmitha, H. Bajaj, and K. Madhavan Nampoothiri**, 'The divergent roles of sortase in the biology of Gram-positive bacteria', *The Cell Surface*, vol. 7, p. 100055, Dec. 2021, doi: 10.1016/j.tcsw.2021.100055.

[55] 'World Health Organization (WHO)'. https://www.who.int (accessed Jun. 27, 2023).

[56] **J. J. Williams and P. J. Hergenrother**, 'Exposing Plasmids as the Achilles' Heel of Drug-Resistant Bacteria', *Curr Opin Chem Biol*, vol. 12, no. 4, pp. 389–399, Aug. 2008, doi: 10.1016/j.cbpa.2008.06.015.

[57] **C. O. Vrancianu, L. I. Popa, C. Bleotu, and M. C. Chifiriuc**, 'Targeting Plasmids to Limit Acquisition and Transmission of Antimicrobial Resistance', *Front Microbiol*, vol. 11, p. 761, May 2020, doi: 10.3389/fmicb.2020.00761.

[58] **E. Cabezón, F. de la Cruz, and I. Arechaga**, 'Conjugation Inhibitors and Their Potential Use to Prevent Dissemination of Antibiotic Resistance Genes in Bacteria', *Front. Microbiol.*, vol. 8, 2017, doi: 10.3389/fmicb.2017.02329.

[59] **Z. Shen, C. M. Tang, and G.-Y. Liu**, 'Towards a better understanding of antimicrobial resistance dissemination: what can be learnt from studying model conjugative plasmids?', *Military Med Res*, vol. 9, no. 1, Art. no. 1, Dec. 2022, doi: 10.1186/s40779-021-00362-z.

[60] **R. W. Bradley, M. Buck, and B. Wang**, 'Tools and Principles for Microbial Gene Circuit Engineering', *Journal of Molecular Biology*, vol. 428, no. 5, Part B, pp. 862–888, Feb. 2016, doi: 10.1016/j.jmb.2015.10.004.

[61] **J. A. N. Brophy and C. A. Voigt**, 'Principles of genetic circuit design', *Nat Methods*, vol. 11, no. 5, pp. 508–520, May 2014, doi: 10.1038/nmeth.2926.

[62] **V. K. Mutalik, L. Qi, J. C. Guimaraes, J. B. Lucks, and A. P. Arkin**, 'Rationally designed families of orthogonal RNA regulators of translation', *Nat Chem Biol*, vol. 8, no. 5, pp. 447–454, Mar. 2012, doi: 10.1038/nchembio.919.

[63] **A. Costello and A. H. Badran**, 'Synthetic Biological Circuits Within an Orthogonal Central Dogma', *Trends Biotechnol*, vol. 39, no. 1, pp. 59–71, Jan. 2021, doi: 10.1016/j.tibtech.2020.05.013.

[64] **J. Chen, X. Chen, and C. L. Ho**, 'Recent Development of Probiotic Bifidobacteria for Treating Human Diseases', *Frontiers in Bioengineering and Biotechnology*, vol. 9, 2021, Accessed: Jun. 29, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fbioe.2021.770248

[65] **W. W. L. Hsiao, K. Ung, D. Aeschliman, J. Bryan, B. B. Finlay, and F. S. L. Brinkman**, 'Evidence of a Large Novel Gene Pool Associated with Prokaryotic Genomic Islands', *PLOS Genetics*, vol. 1, no. 5, p. e62, Nov. 2005, doi: 10.1371/journal.pgen.0010062.

[66] **N. S. Mitić, S. N. Malkov, J. J. Kovačević, G. M. Pavlović-Lažetić, and M. V. Beljanski**, 'Structural disorder of plasmid-encoded proteins in Bacteria and Archaea', *BMC Bioinformatics*, vol. 19, no. 1, p. 158, Apr. 2018, doi: 10.1186/s12859-018-2158-6.

[67] **O. van Mastrigt, E. Di Stefano, S. Hartono, T. Abee, and E. J. Smid**, 'Large plasmidome of dairy Lactococcus lactis subsp. lactis biovar diacetylactis FM03P encodes technological functions and appears highly unstable', *BMC Genomics*, vol. 19, no. 1, Art. no. 1, Dec. 2018, doi: 10.1186/s12864-018-5005-2.

[68] **P. Kelleher, J. Mahony, F. Bottacini, G. A. Lugli, M. Ventura, and D. van Sinderen**, 'The Lactococcus lactis Pan-Plasmidome', *Front. Microbiol.*, vol. 10, 2019, doi: 10.3389/fmicb.2019.00707.

[69] **D. Davray, D. Deo, and R. Kulkarni**, 'Plasmids encode niche-specific traits in Lactobacillaceae', *Microb Genom*, vol. 7, no. 3, p. mgen000472, Nov. 2020, doi: 10.1099/mgen.0.000472.

DEDICATION

In loving memory

Annabelle

This is for you.

She has the soul of a storm and the freedom of the wind.

She has eyes like the stars with moonshine within.

With the strength of the sea and lightning in her toes,

She dances on the breeze of change wherever she goes.

Hotter than a volcano and deeper than the oceans

A whirling, swirling myriad of feelings & emotions.

She is daughter of the earth, primordially inspired.

She is water, she is wood.

She is earth.

She is fire!

A POEM BY MS MOEM

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my supervisor Prof. Douwe van Sinderen for his mentorship, guidance, and patience through the years. Douwe, this endeavour would not have been possible without your continued support and for that I am extremely grateful. I must also extend my gratitude to my co-supervisor Dr. Mary O'Connell-Motherway who taught me many of the wet lab skills which I relied upon throughout my PhD.

I would like also to sincerely thank all staff within the school of Microbiology, in particular Dr. David Clarke, Dr. Gerry Reen, Prof. Paul O'Toole, Prof. Jennifer O'Mahony and Dr. Brian O'Donnell for their help and support. To those I had the pleasure of collaborating with, Dr. Francesca Bottacini and Dr. Christophe Penno, thank you for your insightful contributions. I also acknowledge Science Foundation Ireland for their financial contribution which enabled this project and the APC Microbiome Institute.

To my friends and colleagues within UCC, from Lab 4.25 to 3.37, past and present, thank you for the many coffees and chats, the company during late nights in the lab and for all the weird and wonderful times outside the lab, you know who you are.

My family of course have been a huge support, they keep my feet on the ground, remind me what's important in life and will always do whatever they can to be there for me and each other. I would not be in this position today if it wasn't for everything my parents did for my sisters and me. I would particularly like to thank my mom for her "it's all good girl" attitude and for generally being hilarious, my dad for never missing a slagging opportunity even when I'm in the depths of despair and for all the lifts, my big sis Mandy for always being proud of me and for her sage advice and Hails for being simultaneously the best craic and the grounding force of our mad house... and for listening to me rambling on about 'interesting' things every morning in the car. Dusty Bin thanks you.

To my wonderful partner Max, I thank you for all your love, encouragement and for kicking my ass when I needed it. You are the best!

To my amazing, independent soul sisters Trish and Gemma, thank you both for your friendship and all the shared adventures which got me through the years...

from the Pink Floyd ballet in Italy, to hanging on for dear life to a cistern on the back of a truck while driving through a storm in Thailand with a broken shoe.

Lolo, je te remercie pour tous les bons moments, la bonne cuisine et le pastis. Ta maison m'a donné de nombreux moments de répit dont j'avais bien besoin. "*C'est bon* ?"

Tata et Toto, je vous remercie tous les deux pour toutes les conversations stimulantes et pour m'avoir aidé lorsque mon ordinateur portable est tombé en panne pendant la rédaction. Vous êtes vraiment des sauveurs !

À toute la famille Bourin-Rapine, vous n'entendrez plus jamais les mots "C'est presque fini", enfin !!! Comme on dit dans cette famille, PAtriAaaarchalll !

Of course, I couldn't leave out my furry friends. Honestly if I didn't have my little Fifi spoon to cuddle at night, my boisterous Harri Beau to keep me laughing/running/jumping/shouting, the boys; Louis, Leo, Bo, Reggie and Milo, the girls; Snoops The Cat, Dilly and Diva, then it would have been a much more arduous few years.

Honourable mentions: Jambons from Daybreak, quiz nights at Brick Lane, the camembert from Bradleys, the Irish tag-rugby federation, Elbow Lane's sirloin, BDSM and Vicarstown, Element's flat white, 'Break Stuff' by Limp Bizkit, the Blarney stone, Lidl, the 222 bus and Bob.