scientific reports

Check for updates

OPEN

Massive acquisition of conjugative and mobilizable integrated elements fuels *Faecalibacterium* plasticity and hints at their adaptation to the gut

Gérard Guédon[1], Florence Charron-Bourgoin[1], Thomas Lacroix[2], Toufik Hamadouche[1], Nicolas Soler[1], Badreddine Douzi[1], Hélène Chiapello[2] & Nathalie Leblond-Bourget[1,3]✉

Faecalibacterium **is one of the most abundant bacteria of the human gut microbiota of healthy adults and is recognized to have positive effects on health. Here, we precisely and comprehensively analyzed the conjugative mobilome of four complete *Faecalibacterium* genomes. Despite lacking any plasmid, these bacteria harbor a vast arsenal of 130 elements, including 17 integrative and conjugative elements (ICEs) and 83 integrative and mobilizable elements (IMEs), collectively comprising 14–23% of the genome. Genome comparison of two strains isolated from the same fecal sample (*Faecalibacterium* and *Roseburia* strains) revealed almost identical elements indicating that transfer of ICEs and IMEs shape gut microbiome. ICEs and IMEs from *Faecalibacterium* encode many and diverse predicted functions such as defense and stress response (phages, multidrug, antibiotics, oxidative stress, biliar salts, antimicrobial peptides), nutrient import and metabolisms (Fe^{3+}, carbohydrates) and riboflavin synthesis. This hints at their important role in the survival and adaptation of *Faecalibacterium* strains to the gut ecosystem. A rapid survey of 29 additional *Faecalibacterium* genomes uncovered many putative ICEs and IMEs, reinforcing their role in the rapid and massive evolution of *Faecalibacterium* genomes.**

Keywords ICE, IME, Horizontal gene transfer, Tn*1549*, *Faecalibacterium*, Nutrient import, Phage resistance, Stress response

Faecalibacterium, a member of Oscillospiraceae family (phylum Firmicutes), is a prominent bacterial genus of the human gut microbiota, that plays a crucial role in colon health (for a review 1). By fermentation of polysaccharides, it produces butyrate that is essential for colonocyte energy[1–4]. Reduced *Faecalibacterium* levels are linked to gastrointestinal diseases, obesity, metabolic syndrome, and depression[3,5–8]. Initially defined as a single species in 2002[9], *Faecalibacterium* has recently been subdivided into eight species based on phylogenetic analyses and genome comparison: *F. prausnitzii*, *F. duncaniae*, *F. longum*, *F. butyricigenerans*, *F. hattorii*, *F. gallinarum, F. wellingii and F. taiwanense*[10,11].

Comparative analysis of *Faecalibacterium* genomes shows significant genome plasticity[12] raising questions about the driving forces behind such rapid evolution. The presence of numerous chromosomal genes encoding conjugal transfer proteins[12] suggests that conjugation could be one of these major forces. Since *Faecalibacterium* genomes analyzed by Fitzgerald et al. 2018 lack plasmids[12], we hypothesize that integrative and conjugative elements (ICEs) and integrative and mobilizable elements (IMEs) would play a key role in their genome plasticity.

ICEs, IMEs, and related decayed elements, although less frequently described than conjugative and mobilizable plasmids, are likely the most abundant conjugative elements in bacteria[13–15]. These mobile genetic elements contribute significantly to the genetic variability within and between bacterial species[16,17]. They are widely distributed across various ecological niches and phyla, particularly among Firmicutes[14,15].

ICEs are self-mobile elements that perform excision, conjugative transfer, and integration. The excision of ICEs is catalyzed by integrases belonging to the tyrosine recombinase family, serine recombinase family, or DDE

[1]Université de Lorraine, INRAE, DynAMic, 54000 Nancy, France. [2]Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France. [3]Present address: Université de Lorraine, UMR1128 DynAMic UL-INRAE, 54000 Nancy, France. ✉email: nathalie.leblond@univ-lorraine.fr

transposase family[14]. Transfer of the excised ICE DNA begins when a relaxase cleaves the origin of transfer (*oriT*) and covalently attaches to the 5′ end of the cleaved strand. This one is transferred through a conjugation pore by a coupling protein[14]. Conjugative pores are type IV secretion systems (T4SSs) which rely on ATPases for energy. During conjugative transfer, ICE undergoes rolling-circle replication in both donor and recipient cells. After transfer, the relaxase promotes recircularization of the ICE DNA[18]. Ultimately, the integrase catalyzes the integration of the element in both recipient and donor cells[14].

IMEs are mobilizable elements capable of excision and integration, using their own *oriT* and integrases. However, they cannot transfer autonomously and subvert the conjugation machinery of a co-resident conjugative plasmid or ICE for their transfer. Although most IMEs encode proteins involved in their conjugative transfer, these are generally unrelated or distantly related to those of ICEs[15]. IMEs in Firmicutes typically encode their own relaxase. Many of them also encode a coupling protein[19], but lack other T4SS proteins such as the VirB4 ATPase. Some do not encode any conjugation proteins, which complicates their identification[15].

In Firmicutes, ICEs and IMEs often form complex composite structures[20]: tandems and matryoshkas[14]. Tandems involve elements integrated side by side, resulting from the integration of an incoming element in the *att* site flanking a resident element. Matryoshkas are elements integrated into each other. Composite structures may also enclose highly decayed elements (remnants) deriving from ICEs or IMEs by deletion of all or part of conjugation genes, recombination genes and/or recombination sites. Many remnants (cis mobilizable elements or CIMEs) would be able to transfer by *cis*-mobilization[14].

The vast majority of known ICEs and IMEs harbor adaptive genes that may provide significant benefits to the host bacteria such as symbiosis, pathogenicity and resistance to antibiotics, heavy metals, or phages[14,15].

The aim of this study was to get a comprehensive view of the prevalence and diversity of ICEs and IMEs in *Faecalibacterium* genomes. Our analysis revealed an impressive number of ICEs and IMEs, with up to 42 elements per genome—significantly exceeding the number observed in other Firmicutes genomes. This highlighted the massive contribution of these elements in *Faecalibacterium* genome plasticity. Sequence comparison of elements provides evidence of transfer within gut microbiome strains. This strongly suggests that the gut environment provides highly favorable conditions for conjugative transfer. This study also explored the predicted functions encoded by ICEs and IMEs that probably contribute to *Faecalibacterium* adaptation to the gut environment.

Results

Four different complete and well annotated *Faecalibacterium* genomes were available in RefSeq at the beginning of this work (updated April 2021) (Table 1). Each consists of a single chromosome of approximately 3 Mb and lacks plasmid. All the corresponding strains were isolated from healthy human feces. Recent genomic comparisons, based on the determination of the average nucleotide identity (ANI), led to a complete reclassification of this genus: two of the analyzed strains A2-165^T (JCM 31915^T) and Indica were reclassified as *F. duncanii*, another, APC942/30-2 (JCM 39208), was reclassified as *F. longum*, whereas the last one, APC918/95b, is a genuine *F. prausnitzii*[1].

Most *Faecalibacterium* conjugative elements are included in composite structures

Each ICE or IME was identified in *Faecalibacterium* genomes using ICEscreen[21], and carefully delineated by identifying direct repeats at their ends or through genome comparisons. The analysis revealed a high density of mobile genetic elements including 17 ICEs and 83 IMEs (Table 1). Eighteen other mobile genetic elements (called hereafter MGEs) were also identified using ICEscreen and were delineated. They i) encode a tyrosine- or a serine-recombinase, ii) are devoid of any known genes or pseudogenes involved in conjugation, iii) carry cargo genes similar to those found in ICEs or IMEs, iv) 12 of these 18 MGEs belong to composite structures including ICEs or IMEs. At last, 12 highly decayed ICEs or IMEs (hereafter referred as remnants) were inventoried since being part of a composite structure and delineated. Thus, *Faecalibacterium* genomes contain an average of 33 conjugative elements per genome, making up about 20% of their total genetic content. This underscores the significant impact of conjugation in shaping genomic plasticity of *Faecalibacterium*.

Analysis of the distribution of mobile elements revealed numerous integration sites scattered throughout the chromosomes (Fig. 1). However, ICEs and IMEs are predominantly found in composite structures, with 23

Species	Strain[a]	Chromosome size (bp)	Genome accession number	ICE number[b]	IME number[b]	MGE number	Remnant element number[c]	Total number of elements[c]	Cumulative size of elements (bp)[c]	Percentage of genome corresponding to elements
F. duncaniae	A2-165 ^{T}	3102523	NZ_CP048437.1	8	21	6	1	36	703151	22.7
F. duncaniae	Indica	2868932	NZ_CP023819.1	4	12	3	2	21	419634	14.6
F. longum	APC942/30–2	2833887	NZ_CP026548.1	2	24	2	3	31	508701	18.0
F. prausnitzii	APC918/95b	2970937	NZ_CP030777.1	3	26	7	6	42	637433	21.5
Total				17	83	18	12	130		

Table 1. Prevalence of ICEs, IMEs and MGEs in *Faecalibacterium* genomes. [a]The strains A2-165^T, APC942/30–2 and APC918/95b are also known as JCM 31915^T, JCM 39208 and JCM 39207, respectively. All the four strains were isolated from human feces. [b]The number of ICEs and IMEs includes few slightly decayed elements. [c]Only properly delineated remnant elements enclosed in composite structure were taken into consideration.

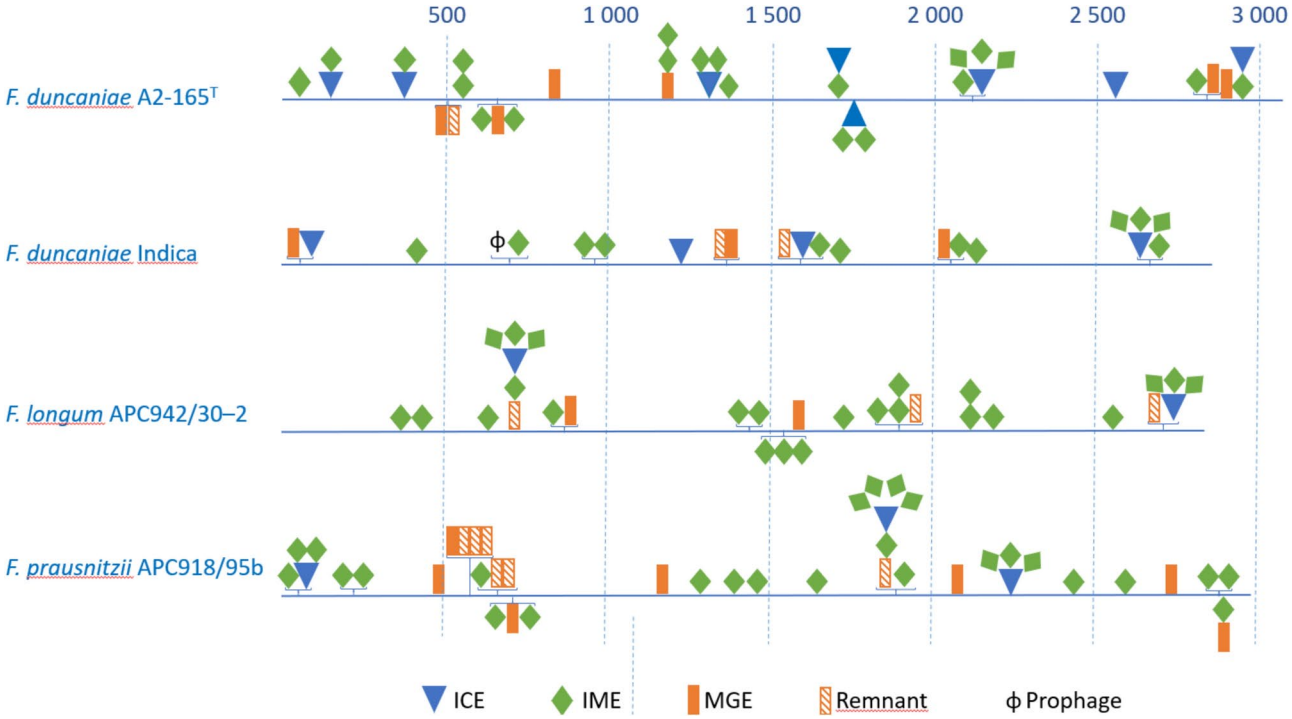


Fig. 1. Mapping of conjugative and mobilizable elements along *Faecalibacterium* chromosomes. The four blue lines represent the chromosomes, and the numbers the position in kb. The origin point of the chromosome corresponds to its origin of replication. Elements in tandem are grouped within braces. The matryoshkas are represented by elements one above the other with the lowest element being the host and the highest the guest. The location of the elements reflects their position along the chromosomes.

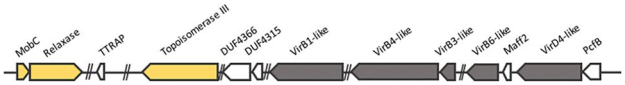
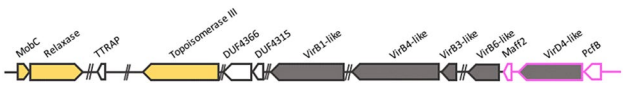
Element type or name	Backbone genes within the conjugation module of *Faecalibacterium* ICEs	Number of elements
A-type Tn*1549*		9
B-type Tn*1549*		7
ICE_*Fpr918_intser*		1

Table 2. Organization of backbone genes within the conjugation module of Tn*1549*. The genes are represented by arrows. Conserved genes encoding proteins with a possible role in DNA processing are shown in yellow. Conserved genes encoding proteins predicted to be part of T4SSs or involved in its assembly are in grey. Conserved genes encoding proteins with unknown function are in white. Slashes indicate the possible presence of additional non-conserved genes. A-type genes are outlined in black, while B-type genes are in pink. ICE_*Fpr918_intser* is a chimeric element.

tandems and 18 matryoshkas identified. Tandems consist of up to four elements, including ICEs, IMEs, MGEs, and remnants. One tandem also contains a prophage in *F. duncaniae* Indica. Fifty-seven elements are included in matryoshkas. Matryoshkas include mainly IMEs integrated within ICEs (29 IMEs in total), with some ICEs containing up to three or four IMEs (6 ICEs in total). Within matryoshkas, up to four levels of nesting are observed. One of the most striking examples of composite element is observed in the *F. prausnitzii* APC918/95b genome. It includes one remnant ICE and one IME in tandem. The remnant carries an IME which itself carries an ICE that holds four additional IMEs.

ICE backbone genes and organization

ICE analysis identified a backbone of 13 conserved protein-encoding genes belonging to the conjugation module (Table 2). Five of them encode components analogous to Gram-negative T4SSs, according to the conserved

protein domains, transmembrane domains, and secondary structure predictions (Table 3). These proteins are: VirB1-like peptidoglycan hydrolase, VirD4-like coupling protein, VirB4-like, VirB3-like, and VirB6-like, which likely form parts of the translocon in the actual proposed model of Gram-positive T4SSs[22]. The remaining eight backbone genes are predicted to encode five proteins with undefined roles in conjugation and relaxosome proteins, including a MOBP relaxase, a MobC auxiliary protein recognizing the origin of transfer (*oriT*), and a topoisomerase III for DNA processing.

The conjugation modules of *Faecalibacterium* ICEs exhibit the typical organization of the Tn*1549* family: they encode a MOB_{P7} relaxase with a PF03432 domain (Fig. S1), a VirD4-like coupling protein with PF02534 and PF12686 domains and a VirB4-like protein with a PF12846 domain[19,23]. Phylogenetic analysis of relaxases, using homologues from *Enterococcus faecalis* Tn*1549* as a reference and two VanG family ICEs as outgroup, confirmed that *Faecalibacterium* ICEs belong to the Tn*1549* family (Fig. 2). Similar results were obtained for other backbone proteins (data not shown), indicating that all *Faecalibacterium* ICEs belong to the Tn*1549* family.

Diversity of ICE conjugation modules

Despite they belong to the same family, ICEs from *Faecalibacterium* exhibit significant diversity in their conjugation modules. One aspect of diversity is the spatial organization of backbone genes, categorized hereafter as A-type and B-type (Table 2). In A-type ICEs, the genes encoding the relaxase and MobC proteins are in convergent orientation relative to the other backbone genes. In B-type ICEs, these genes are in the same orientation as the other backbone genes and are separated by one or more non-conserved genes. Phylogenetic analyses showed that the relaxases of A-type ICEs cluster together and are closely related to that of Tn*1549* from *E. faecalis,* whereas relaxases of B-type ICEs form a distinct cluster (Fig. 2). Similar clustering was observed in the phylogenetic analyses of other backbone proteins (data not shown).

A second level of diversity in conjugation modules of *Faecalibacterium* ICEs is the sequence variability of backbone proteins. The most conserved proteins are VirB3-like, VirB4-like, VirB6-like and VirD4-like proteins (50 to 100% amino-acid identity) whereas proteins with DUF4315 domain are the most divergent (some sharing less than 30% identity).

The third level of diversity in conjugation modules lies in the presence of variable regions, leading to different gene contents for the 17 ICEs. Variable regions of 11 ICEs encode one or two polyvalent proteins. These are large proteins which carry diverse functional domains and are predicted to favor element transfer or success in the recipient cell[24]. Each polyvalent protein contains two to eight domains (Table S1). Nine include a peptidase_M78, which would process autocleavage of the protein into functional units[24]. Variable proteins also include 14 monodomain proteins that carry a domain related to those found in polyvalent proteins (Table S1). Fifteen of these polyvalent or monodomain proteins harbor a YodL domain proposed to play a role in anchoring the invasive element or the polyvalent proteins in the cytoskeleton to facilitate their transport or localization[24].

Gene designation[a] locus-tag protein ID	Length (aa)	Archetypal VirB	Protein groups or conserved domain	E-value	Alphafold prediction/dali correspondence	Transmembrane helices	Suggested function
pcfB CRH10_RS07495 WP_098923918.1	156	None	DUF3801 superfamily	1.25^{e-30}	No relevant hits	0	Not defined
virD4 CRH10_RS07500 WP_098923920.1	601	VirD4	VirD4-superfamily	1.00^{e-126}	Conjugal transfer protein TRWB	2	Coupling protein
maff2 CRH10_RS07505 WP_002574004.1	71	None	Maff2 superfamily	6.37_{e-25}	No relevant hits	2	Not defined
virB6 CRH10_RS07510 WP_098923922.1	287	VirB6	TrbL-2 superfamily	2.48^{e-49}	No relevant hits	5	Translocon component
virB3 CRH10_RS07515 WP_098923924.1	141	VirB3	PrgI superfamily	7.14^{e-29}	No relevant hits	2	Translocon component
virB4 CRH10_RS07520 WP_242958040.1	747	VirB4	VirB4	4.56^{e-147}	VirB4 PDB ID: 4AG6-B	0	ATPase
virB1 CRH10_RS07530 WP_098923928.1	661	VirB1	Cell wall-associated hydrolase, NlpC_P60 family	2.77^{e-39}	No relevant hits	1	Peptidoglycan hydrolase
DUF4315 CRH10_RS07535 WP_097777297.1	83	None	DUF4315	8.45^{e-21}	No relevant hits	0	Not defined
DUF4366 CRH10_RS075540 WP_098923930.1	240	None	DUF4366	1.53^{e-41}	No relevant hits	1	Not defined
topB CRH10_RS07545 WP_098923932.1	707	None	DNA topoisomerase (pfam01131)	4.57^{e-141}	DNA topoisomerase-I	0	DNA topoisomerase III

Table 3. Analogy of backbone proteins from ICEs with Vir proteins of Gram-negative T4SSs. [a]The proteins from ICE_*FprIndica_tRNAarg_2* were used as reference.

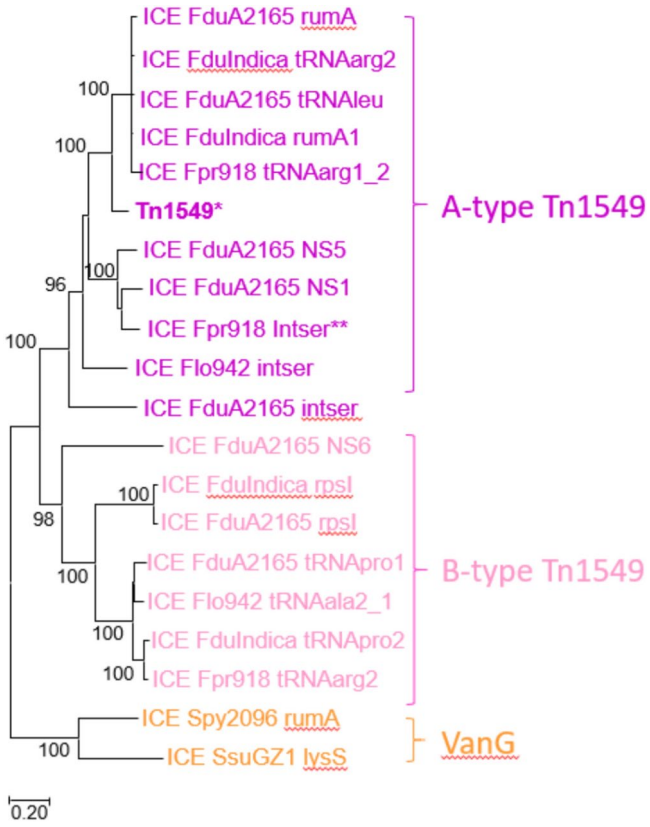


Fig. 2. ICEs from *Faecalibacterium* are all related to Tn*1549*. The phylogenetic tree was constructed using MOB_{P7} relaxase sequences from ICEs. Only bootstrap values > 80% are indicated. In bold, the MOB_{P7} relaxase from ICE Tn*1549* identified in *Enterococcus faecalis*[55]. Relaxase from A-type ICEs are in purple and those from B-type ICEs appear in light pink. ICE_*Fpr918_intser* is a chimeric element resulting from a recombination event between a A-type ICE and a B- type ICE. The relaxases of two ICEs belonging to the VanG family which is closely related to Tn*1549*, appear in orange [23].

	Strains				
IME superfamilies (domain identifier)	***F. duncaniae* A2-165^T**	***F. duncaniae* Indica**	***F. longum* APC942/30–2**	***F. prausnitzii* APC918/95b**	**Total number**
MOB_{Q2} (PF03389)	8	2	5	8	23
MOB_{V1} (cd17242)	6	4	10	9	29
MOB_{P7} (PF03432)	7	5	9	8	29
MOB_{P1} (PRK13878)		1		1	2
	21	12	24	26	

Table 4. IMEs from *Faecalibacterium* belongs to 4 distinct superfamilies.

Fourteen polyvalent or monodomain proteins carry one or two domains involved in the protection of the transferred ICE against restriction-modification (RM) systems harbored by the recipient cell[22]. Six other variable genes encode proteins harboring a DNA methylase-helicase dyad (Table S1). Dyad proteins were proposed to bind and methylate the mobile element DNA to discriminate invasive and host DNA and to protect the former against restriction[24]. Eleven other variable genes encode proteins with DNA methyltransferase domains unrelated to those of dyad proteins and that could also be involved in the protection of transferred DNA.

IMEs belong to four distinct superfamilies

All IME mobilization modules contain an *oriT*. Almost all known IMEs encode proteins necessary for their transfer, typically including a relaxase, often additional relaxosome proteins, and occasionally a coupling protein[15]. In this work, we only searched IMEs that possess a relaxase gene or pseudogene. IMEs were classified into four superfamilies based on their relaxase domains: MOB_{Q2}, MOB_{V1}, MOB_{P7} and MOB_{P1} (Table 4).

The MOB_{Q2} superfamily consists of 23 IMEs that encode a relaxase with a PF03389 domain. Analysis of their conserved motifs allowed their classification in the MOB_{Q2} clade (Fig. S2) with the pTi relaxase from *Agrobacterium tumefaciens* as prototype[25]. Their phylogenetic analysis and 40% amino-acid identity clustering

divided the Mob_{Q2} IMEs in four different families (Fig. 3A1). All MOB_{Q2} IMEs contain a gene encoding a protein with a PF12958 domain (DUF3847) next to the relaxase gene. It appears to coevolve with the relaxase and likely plays an important role in mobilization (Fig. 3A2). Additionally, 16 of the 24 MOB_{Q2} IMEs encode a protein with a PF14198 domain (with unknown function), which may also be involved in their transfer.

The MOB_{V1} superfamily comprises 29 IMEs encoding relaxases with a cd17242 domain, and belonging to the MOB_{V1} clade whose prototype is the MobM relaxase from the pMV158 plasmid of *Streptococcus agalactiae*[26] (Fig. S3). Phylogenetic analysis and 40% amino-acid identity clustering divided the MOB_{V1} IMEs into four families (Fig. 3B).

The MOB_{P7} superfamily comprises 29 IMEs encoding relaxases with a PF03432 domain, and belonging to the MOB_{P7} clade (Fig. S4). This clade is related to the relaxases of *Faecalibacterium* ICEs and conjugative plasmids like pCF10 and pC221/ pC223[25]. Separated alignments of IME and ICE relaxases were constructed since the motif IV identified for ICE relaxases is absent in IME relaxases. MOB_{P7} IMEs belong to a single family, as all their relaxases share over 40% amino acid identity. A conserved gene encoding the auxiliary MobC protein, which is a DNA binding protein, that probably recognize *oriT*, is present upstream of the relaxase gene in all these IMEs. The consistent phylogenetic relationship between MOB_{P7} relaxases (Fig. $S5A_1$) and MobC proteins (Fig. $S5A_2$) indicates their co-evolution.

The MOB_{P1} superfamily includes only two IMEs encoding a relaxase with a PRK13878 domain. Analysis of their conserved motifs (Fig. S6) revealed that they belong to the MOB_{P1} clade[25], whose prototype is the TraI relaxase from RP4 plasmid of *Escherichia coli*. This superfamily of IMEs has never been described before within Firmicutes. These two IMEs also encode a putative VirD4-like coupling protein with a PF12696 domain and five other proteins with unknown functions.

MGEs are probably IMEs devoid of any gene involved in conjugative transfer

In this study, 18 MGEs encoding serine or tyrosine integrase were identified. None carry any typical gene of phages nor its own relaxase gene or pseudogene (annotated by RefSeq). However, they could be mobilizable by other conjugative or mobilizable elements and could function as IMEs. This is supported by the following observations: i) three MGEs contain sequences similar to the putative *oriT* found upstream of *mobC* in IMEs with a MOB_{P7} relaxase (Fig. S7), ii) eight are part of composite structures that include ICEs or IMEs (Fig. 1), and iii) ten carry modules or cargo genes similar to those found in ICEs or IMEs.

Evidence of gene exchanges between conjugative elements and between strains in gut microbiome

Gene or module exchanges between conjugative and mobilizable elements significantly drive their evolution[14]. *Faecalibacterium* elements are no exception to this rule. For example, phylogenetic analysis of the conjugation

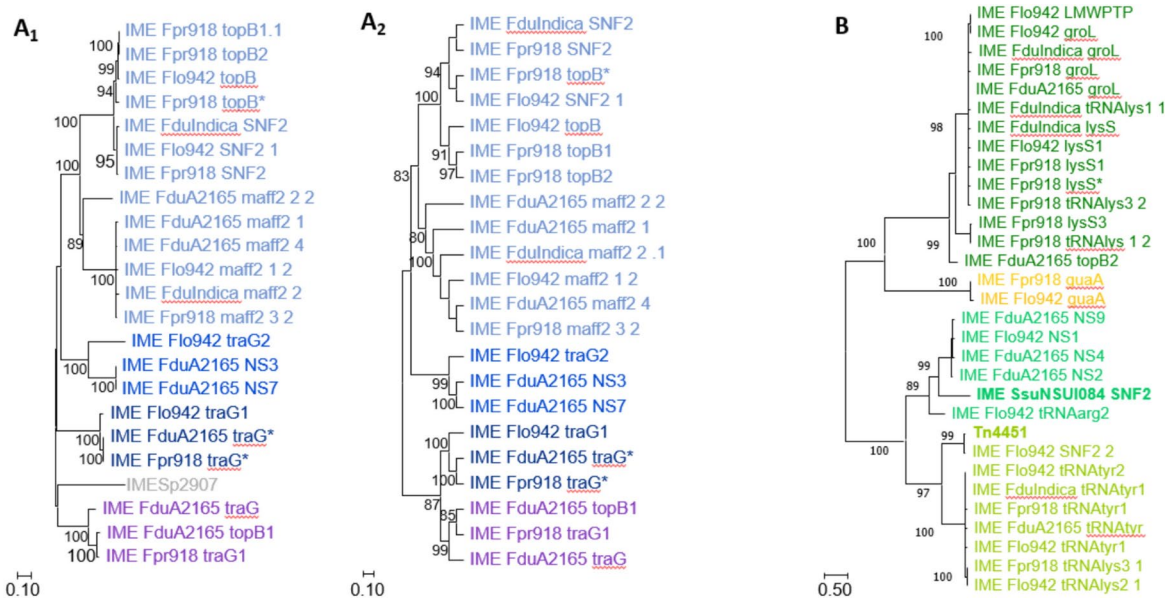


Fig. 3. Phylogenetic analysis of relaxases and other conserved proteins from IMEs. (**A**) Proteins from IMEs belonging to the MOB_{Q2} superfamily. In grey, an IME from *Streptococcus pneumoniae* whose transfer was demonstrated [56]. IME_*Fpr918_traG_2* is missing since it possesses a MOB_{Q2} relaxase pseudogene. **A1**. Phylogenetic tree of MOB_{Q2} relaxases. **A2.** Phylogenetic tree of unknown proteins with a PF12958 domain. (**B**) Proteins from IMEs belonging to the MOB_{V1} superfamily**.** In bold, two IMEs whose transfer was demonstrated: one from *Streptococcus suis* (IME_*SsuNSUI084_SNF2*) and the other from *Clostridium perfringens* (Tn*4451*) [57]. Only bootstrap values > 80% are indicated. IMEs with identical color belong to the same family. At the end of the element names, *indicates their integration in secondary sites and not in their primary site.

module genes in ICE_*Fpr918_intser* revealed that its left part belongs to the A-type phylogroup whereas its right part belongs to the B-type phylogroup. A thorough analysis suggests that this chimeric element likely results from a recombination event between an A-type ICE and an ICE closely related to ICE_*FduA2165_NS6*. This would have occurred between the genes encoding Maff2 and VirB6-like proteins (Table 2). Moreover, two unrelated elements from the same strain *F. prausnitzii* APC918/95b, ICE_*Fpr918_tRNAarg_2* and IME_*Fpr918_lysS_1*, share almost identical riboflavin synthesis module (99.3% nucleotide identity over the entire length of the modules). This suggests that one element acquired this module from the other or that both elements acquired it from an unknown third element.

In this work, we also get evidence of horizontal transfer of conjugative elements across gut bacteria. Indeed, two strains belonging to different species, *F. longum* APC942/30–2 and *F. prausnitzii* APC918/95b, share very closely related IMEs whereas the average nucleotide identity of orthologous genes (ANI) of the two species is 87,9%[1]. First, IME_*Fpr918_maff2_2* of *F. prausnitzii* APC918/95b (excluding an IS*110*-related insertion sequence carried by this IME) shares 98,8% nucleotide identity over its entire length with IME_*Flo42_maff2_1* from *F. longum* APC942/30–2. Second, IME_*Fpr918_lysS_1* of *F. prausnitzii* APC918/95b and IME_*Flo942_lysS_1* _of *F. longum* APC942/30–2 share 95,5% nucleotide identity over their entire length. Sequence comparison also revealed a striking example of horizontal gene transfer involving strains from the same fecal sample[27] but belonging to different bacterial orders, Eubacteriales (*F. duncaniae* A2-165^T, family Oscillospiraceae) and Lachnospirales (*Roseburia hominis* A2-183^T, family Lachnospiraceae)[28]. Indeed, ICE_*FduA2165_NS_6* from *F. duncaniae* A2-165^T and ICE_*RhoA2-183_NS_3* from *R. hominis* A2-183^T exhibit 100% nucleotide identity over their entire length. Furthermore, IME_*FduA2165_NS_2* from *F. duncaniae* exhibits only two different nucleotides with IME_*RhoA2-183_NS_2* from *R. hominis*. Additionally, two other IMEs from *F. duncaniae* A2-165^T share 98,5 and 97,6% identity over their entire length with two IMEs found in *R. hominis* A2-183^T: IME_*FduA2165_NS_7* and IME_*RhoA2-183_hipB*; IME_*FduA2165_traG** and IME_*RhoA2-183_traG_2*, respectively. Similarly, IME_*FduA2165_tRNAtyr* of *F. duncaniae* A2-165^T exhibits only two different nucleotides with IME_*LspYL32_NS_2* of the Lachnospiraceae *E. clostridioformis* YL32 from mouse gut microbiota[28].

Adaptative functions encoded by ICEs, IMEs, MGEs and remnants from *Faecalibacterium* strains

In addition to conjugation or mobilization modules, ICEs and IMEs possess maintenance modules (integration, replication, partition) that primarily benefit the elements themselves rather than the host strains. They also contain numerous genes or modules with unclear biological functions, which are not detailed here. ICEs and IMEs are also known to promote the transfer of functions that can enhance the survival of host strains within the gut or confer a competitive advantage[14,15]. Table S2 lists these putative adaptative functions found in the *Faecalibacterium* conjugative and mobilizable elements. Our analysis did not identify any significant differences in the cargo genes carried by ICEs and IMEs.

First, several elements encode proteins predicted to be involved in the synthesis of compounds that could help bacteria to maintain in gut ecosystems. For example, we found that ICE_*Fpr918_tRNAarg_2* from *F. prausnitzii* APC918/95b and IME_*Flo942_lysS_1* from *F. longum* APC942/30-2 contain nearly identical complete riboflavin synthesis modules, a vitamin important for bacterial survival in the gut[29]. Additionally, *F. duncaniae* Indica has a complete riboflavin synthesis module as part of an undelineated remnant adjacent to an ICE remnant integrated into the 3' end of a tRNAleu gene (Table S2). Although bacteriocin biosynthesis genes are often present in ICEs and IMEs, none were found in this study. Due to the high variability of known bacteriocins[30] and to the few identified in *Faecalibacterium* and other Oscillospiraceae[31], we may not have been able to identify them.

Second, many ICEs and IMEs of *Faecalibacterium* carry highly diverse modules or isolated genes (n = 25) that were predicted to encode the nutrient import across bacterial membranes. (Table S2, Fig. 4). They encode a large variety of carbohydrate importer including (i) four ATP-binding cassette (ABC) transporters with unrelated substrate-binding subunits, (ii) three phosphotransferase systems (PTS) that couple carbohydrate transport with phosphorylation, and (iii) two major facilitator superfamily (MFS) transporters that use electrochemical gradients across the membrane for active substrate transport. These nine carbohydrate import modules or genes are associated with genes that could be involved in the degradation of beta-galactosides such as lactose, mucins produced by enterocytes, plant polysaccharides such as mannans, and fructoselysine, a product of the Amadori reaction resulting from food cooking. ICEs and IMEs also carry modules that could encode transporters for other nutrients i) five ABC transporters specific of Fe^{3+}-siderophores or Fe^{3+}-chelates, ii) three for import and catabolism of di- or tricarboxylates, iii) one high-affinity K^+ transporter and iv) six ABC transporters associated with the import of unidentified nutrient.

Third, many ICEs and IMEs from *Faecalibacterium* also encode proteins predicted to withstand phage predation. In the four *Faecalibacterium* strains analyzed, 47 isolated genes or modules (composed of up to 10 genes) likely be involved in resistance to bacteriophages are carried by ICEs and IMEs (Table S2, Fig. 4). They are predicted to encode 86 putative defense systems belonging to 47 unrelated classes. These highly diverse systems include (i) restriction-modification systems (types I, II, III, IV) that prevent infection by degrading phage DNA, (ii) many abortive infection systems (ten different classes) that trigger cell death to prevent phage replication, (iii) many systems belonging to 13 other classes of phage resistance and many hypothetic defense systems. Only one CRISPR system was identified in an IME from *F. prausnitzii* APC918/95b. However, three other CRISPR systems were found in undelineated or isolated remnants, in *F. longum* APC942/30-2, *F. prausnitzii* APC918/95b and *F. duncaniae* Indica (Table S2). Furthermore, another CRISPR system from *F. longum* APC942/30–2 is located approximately 7.6 kb from the 3' end of a tRNASer gene which is known to be a specific integration site. This suggests that it might have initially been carried by an ICE or an IME, which has subsequently lost its other modules.

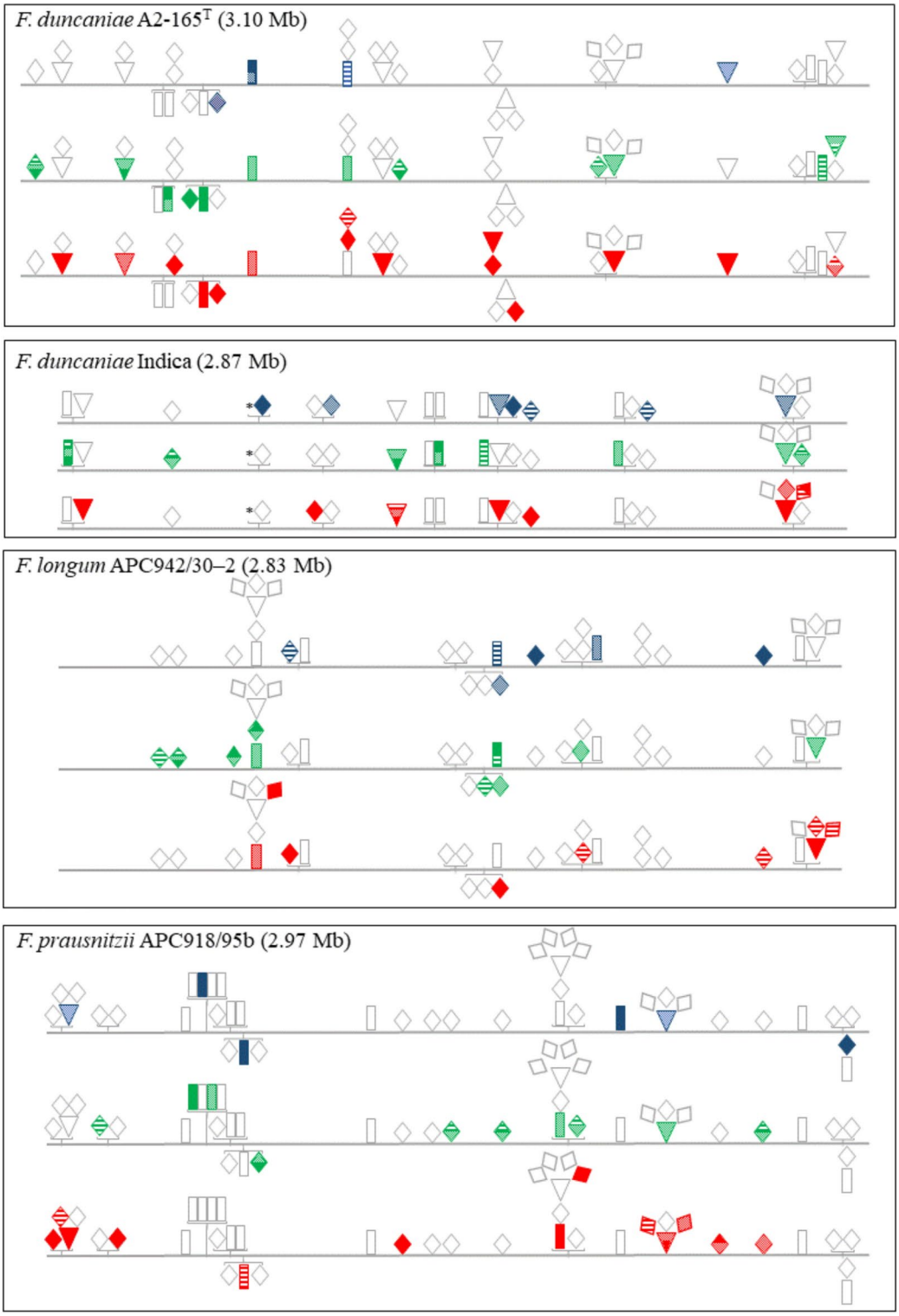


Fig. 4. Repertoire of adaptive functions carried by *Faecalibacterium* conjugative an mobilizable elements. The symbols corresponding to the elements are identical to those shown in (Fig. 1). Solid, striped or dotted colors indicate the functions carried. The blue color lists elements encoding import systems with ● for carbohydrate import, ▤ for iron import and ▩ for other import systems. The green color lists elements encoding phage resistance systems with ● indicating those encoding restriction-modification genes, ▤ those encoding abortive infection systems and ▩ those encoding other systems. The red color lists elements encoding drug or stress resistance systems with ● for efflux systems, ▤ for antibiotic or bacteriocin resistance and ▩ for other systems.

Fourth, many ICEs and IMEs are predicted to encode various other types of resistance (Table S2, Fig. 4) including (i) antibiotics (tetracycline, aminoglycosides, chloramphenicol), (ii) bacteriocins, (iii) DNA damage, (iv) oxidative stress, (v) bile salts, and (vi) arsenate or heavy metals. ICEs and IMEs also carry numerous modules or isolated genes encoding highly diverse transporters potentially involved in the efflux of unidentified toxic compounds, including (i) 12 multidrug efflux ABC transporters and one MFS transporter, (ii) ABC transporters probably involved in the efflux of antimicrobial peptides (n = 6) or toxic compounds (n = 9), and (iii) six multidrug and toxin extrusion (MATE) transporters.

Expanding strain analysis for ICE and IME prevalence and diversity

Data obtained for the four *Faecalibacterium* genomes underscore the remarkable abundance and diversity of ICEs and IMEs. To further validate these findings, we conducted a rapid analysis on 29 additional *Faecalibacterium* genomes available in Refseq since January 2025, including the type strains of two other species *F. wellingii* and *F. taiwanense*. The raw ICEscreen data are presented in Table S3 where the number of ICEs, IMES and partial elements is indicated. This analysis demonstrates that all *Faecalibacterium* genomes are exceptionally rich in ICEs and IMEs. Indeed, a total of 634 elements were identified, comprising 114 ICEs (76 complete and 38 partial), 294 IMEs, and 226 MGEs, which probably correspond to remnants or imperfectly detected IMEs or ICEs. The number of elements varies across strains from 13 to 31 per genome. All complete ICEs belong to the Tn*1549* family and all complete IMEs fall into the four previously identified superfamilies. Three IME superfamilies—MOB_{Q2} (n = 79), MOB_{V1} (n = 115), and MOB_{P1} (n = 98) – are highly prevalent, whereas the fourth MOB_{P7} is relatively rare (n = 2). In addition, each of the 226 "MGEs" encode a relaxase (MOB_{Q2}, MOB_{V1} or MOB_{P1}) but lack both a coupling protein-encoding gene and a VirB4 protein-encoding gene. As a result, many of these elements are also likely IMEs. However, ICEScreen failed to assign an integrase gene in many of them due to their presence in composite structures.

Discussion

This work follows the creation of Firmidata, a publicly accessible reference resource composed of a set of 40 genomes of Firmicutes for which we have identified and annotated all ICEs and IMEs (https://icescreen.migale.inrae.fr/). Mainly focused on streptococci, the database also includes strains from human and animal microbiota (*F. duncaniae* A2-165, *Roseburia hominis* A2-183, *Enterocloster clostridioformis* YL32, *Lachnoclostridium phocaeense* Marseille-P3177)[28]. The current article aims to further analyze ICEs and IMEs in Firmicutes from human intestinal microbiota by examining the three other *Faecalibacterium* well-annotated genomes, available in the Firmidata database at the beginning of the work (April 2021).

The four analyzed *Faecalibacterium* genomes lack plasmids, but carry an impressive number of ICEs, IMEs, MGEs and remnants, constituting about 20% of their genomes. Relative to their genome size, the *Faecalibacterium* strains have the highest number of elements (complete or almost complete) ever described. The rapid survey of the well-annotated genome of 29 other *Faecalibacterium* strains, available in January 2025, confirmed the absence of plasmids and revealed a similar mobile element content. This high prevalence of ICEs and especially of IMEs is a feature shared by *R. hominis* A2-183 (5 ICEs and 17 IMEs) and by the two other Firmicutes from mammalian microbiota—*E. clostridioformis* YL32 (3 ICEs and 15 IMEs) and *L. phocaeense* Marseille-P3177 (6 ICEs and 11 IMEs)[28]. These findings suggest that ICEs and IMEs are widespread in both Oscillospiraceae—to which *Faecalibacterium* belongs—and Lachnospiraceae—to which *Roseburia, Lachnoclostridium* and *Enterocloster* are members.

We revealed in this study that the four *Faecalibacterium* genomes contain approximately five times more IMEs (n = 83) than ICEs (n = 17), a much higher ratio than previously suspected in other bacterial or archaeal genomes[13] or observed in a large set of streptococci[19,23]. The analysis of the additional set of 29 *Faecalibacterium* genomes further confirmed the substantial predominance of IMEs over ICEs. Recent research, including this study, indicate that conjugative and mobilizable plasmids are far from being the only elements that can be transferred by conjugation, nor the most widespread among Firmicutes[15].

All the ICEs identified in the 33 *Faecalibacterium* genomes analyzed belong to the Tn*1549* family. The abundance of this family in the Firmicutes from microbiota could be general (at least in Oscillospiraceae and Lachnospiraceae). Tn*1549*-related ICEs are also abundant in other Firmicutes such as *R. hominis* A2-183 (n = 4), *E. clostridioformis* YL32 (n = 2) and *L. phocaeense* Marseille-P3177 (n = 3). They are also notably common in human gut microbiota metagenomes[32]. Tn*1549*-related ICEs are found in other Firmicutes like *Streptococcus* but are much less prevalent with only 21 instances in 124 genomes[23]. In contrast, the 33 analyzed genomes of *Faecalibacterium* exhibits high diversity of IMEs, which are classified into four distinct superfamilies. Similar diversity was observed in *R. hominis* A2-183^T, *E. clostridioformis* YL32, and *L. phocaeense* Marseille-P3177 (data not shown).

Besides ICEs and IMEs, in-depth analysis of the four *Faecalibacterium* genomes revealed 18 MGEs that are likely IMEs lacking relaxases. Similar MGEs have been already identified in Proteobacteria and Firmicutes[14]. Here, we only considered remnants enclosed in composite structures that we could delineate (n = 12). However, by searching for cargo modules similar to three modules encoding riboflavin synthesis and one encoding a CRISPR system found in ICEs and IMEs, we identified five isolated and/or undelineated remnants, suggesting that many more unidentified remnants are present in the chromosome of the four strains. Therefore, despite the high number of elements identified in this study, these elements represent only a fraction of the chromosomal sequences really acquired through conjugation.

The high proportion of ICEs, IMEs and related elements clearly indicates that *Faecalibacterium* massively acquired chromosomal mobile elements by conjugation. This suggests that the gut microbiota provides a suitable environment for conjugative transfer, probably due to the close contact between bacteria. Evidence for recent horizontal transfer includes the presence of identical or nearly identical IMEs and ICEs in different

Faecalibacterium species, in the Oscillospiraceae *F. duncaniae* A2-165^T and the Lachnospiraceae *R. hominis* A2-183^T, or in the *Oscillospiraceae F. duncaniae* A2-165^T and the *Lachnospiraceae E. clostridioformis* YL32. Altogether, these results support recent horizontal transfer between *Faecalibacterium* strains or between phylogenetically distant species of the gut microbiota.

Our study also revealed the remarkable diversity of cargo genes in both ICEs and IMEs. This suggests that these elements could enhance the ecological success of *Faecalibacterium* strains in the gut by encoding adaptive functions crucial for their growth and survival. First, these elements could contribute to the synthesis of riboflavin and to Fe^{3+} import, which are scarce in the large intestine but vital for the bacteria maintenance in gut ecosystems. Second, the survival of bacteria in gut ecosystems also depends on their ability to import carbohydrate across bacterial membranes and metabolize them (for a review[1]. ICEs and IMEs likely support metabolic flexibility by enabling the adsorption and degradation of a variety of carbohydrates, including dietary plant mannans, fructoselysine produced by cooking food, lactose, and mucins from the host. Third, to persist in the gut, bacteria must resist to various stresses and aggressions, including those caused by the high abundance and diversity of bacteriophages[33]. In this context, the abundance and diversity of phage resistances encoded by ICEs and IMEs could help bacteria withstand such aggressions. Additionally, they carry other genes encoding predicted resistance to oxidative stress that could help *Faecalibacterium* strains to thrive in the intestinal environment, despite O_2 and reactive oxygen species present. Oxygen concentration in the gut is low and variable depending on the location—it decreases from the stomach to the colon and from epithelial cells to the lumen—and health status, with higher levels often associated with many gastrointestinal diseases[1]. Although *Faecalibacterium* is generally described as extremely oxygen-sensitive, some strains, such as *F. duncaniae* A2-165^T, can grow under low-oxygen conditions[1]. Finally, ICEs and IMEs of *Faecalibacterium* also encode putative resistances to other stress factors, such as antimicrobial agents like antibiotics, bile salts and antimicrobial peptides synthetized by other bacteria from the gut or by the host.

Altogether, this study highlights the crucial role of ICEs and IMEs in the evolution of *Faecalibacterium* strains in the gut. The genetic content of these mobile elements suggests that they confer beneficial properties to their bacterial host, increasing its fitness and allowing it to better compete within its particular gut microbial ecosystem. This is consistent with earlier research showing that ICEs are also abundant in Bacteroidetes, another major gut phylum, and are linked to enhance bacterial fitness[34]. Expanding the search for ICEs and IMEs in other intestinal bacteria could provide further insights into the evolutionary dynamics and ecological success shaping the gut microbiota.

Materials and methods

Detection of ICEs and IMEs using ICEscreen

Four complete *Faecalibacterium* chromosomes were retrieved from RefSeq (accession numbers NZ_CP048437.1, NZ_CP023819.1, NZ_CP026548.1 and NZ_CP030777.1) and analyzed using ICEscreen, a bioinformatic pipeline for the detection of ICEs and IMEs in Firmicutes genomes[21]. ICEscreen employs HMM profiles and BLASTP[35] to identify signature proteins of ICEs/IMEs, including integrases, relaxases, coupling proteins, and VirB4 ATPases. ICEscreen analysis of each genome allowed us to identify a large range of ICEs, IMEs, MGEs and remnants. According to ICEscreen nomenclature, elements containing all four signature protein-encoding genes are classified as complete ICEs, while those with a *virB4* gene but lacking one other signature genes or containing pseudogenes are classified as partial ICEs. IMEs are elements that encode both an integrase and a relaxase, with some also encoding a coupling protein. MGEs correspond to elements that encode an integrase but lack any other conjugation-related genes. Remnants are elements that either lack an integrase gene or contain an insufficient number of genes encoding signature proteins to be classified as ICEs or IMEs. All elements identified by ICEscreen, whether isolated or part of composite structures, were manually validated and delineated. Among the highly decayed elements, only those integrated into composite structures were taken into account.

To strengthen our results, we analyzed a new set of 29 *Faecalibacterium* genomes, fully assembled and well-annotated, available in January, 2025. The data presented here are raw outputs from ICEscreen. The identified elements were not delineated and neither did undergo manual validation. For these 29 genomes, manual investigations were performed only for elements (ICEs and IMEs) reported as complete by ICEscreen, and solely to supplement missing data (primarily the family classification of some ICEs and the superfamily classification of some IMEs).

Identification of the ICE, IME, MGE and remnant boundaries

Only elements from the first set of four genomes were delineated. Delimitation of the elements at the gene or nucleotide level was done manually by searching for DRs at both ends of each element. DRs were searched using BLASTN or tBLASTN[35] depending on the nature of the product encoded by the gene targeted by the integrase. When DRs were absent, too short or too degenerated to be detected, three alternative methods were used to define element limits by comparing i) translated disrupted targeted gene sequences with related proteins using tBLASTN, ii) region encoding signature proteins with genomic regions of other *Faecalibacterium* strains devoid of related sequences using MegaBlast or (iii) region encoding signature proteins with closely related ICEs or IMEs from other genera encoding identical or almost identical integrases, using MegaBlast.

Identification of ICE conjugation modules and IME mobilization modules

Conjugation and mobilization modules belonging to the same family share core genes that were identified using Roary[36] with a threshold of 90% presence and 25% minimum amino-acid identity. IME superfamilies and families are based on the catalytic domains and phylogeny of their relaxases, respectively. Relaxase families were aligned using Muscle[37], with manual refinement and edited with BioEdit TA Hall…, Nucleic acids symposium series, 1999—Cited by 51,970", "title":"BioEdit: a user-friendly biological sequence alignment editor and analysis

program for Windows 95/98/NT", "title-short": "BioEdit", "URL": "https://scholar.google.com/citations?view_op=view_citation&hl=en&user=25lPJlMAAAAJ&citation_for_view=25lPJlMAAAAJ:b0M2c_1WBrUC","accessed":{"date-parts":[["2024",7,29]]}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}[38]. ICE transfer proteins were identified through BLASTP. Their topology was analyzed using CCTOP[39] and their homology to VirB proteins from Gram-negative T4SSs was confirmed using Alphafold[40], and DALI[41].

Sequence comparison of elements or modules

Automated nucleotide sequence comparison of elements were performed with Geneious,Prime (2023.0.1) to identify the most closely related elements in different *Faecalibacterium* species. Then, all pairs of very closely related elements over their entire length were compare using Megablast (28-nt word size; match/mismatch scores 1, -2; linear gap costs), and these results were subsequently manually verified since some of these elements are included in complex structures. Almost identical sequences in the complete genomes of *Faecalibacterium duncanii* A2-165^T and *Roseburia hominis* A2-183^T were searched using Megablast (256-nt word size; match/mismatch scores 1, -2; gap costs 0, 2.5). A manual verification of these results was performed to identify identical or almost identical ICEs and IMEs in the two strains.

Phylogenetic analysis

Phylogenetic trees were done with Mega11[42] by using the Maximum Likelihood method. The robustness of the branch support was verified using the bootstrap method with 100 replicates.

Identification of cargo genes

Considering data available on cargo genes in the literature, 31 functional categories related to fitness and 10 category tags were defined. The attribution of cargo genes encoded by ICEs and IMEs to these categories was based on keywords found in the functional annotation of the genes or on alignment with six external resources with significant hits: AMRFinderPlus[43], BACTIBASE[44], VFDB[45], REBASE[46], NORINE[47] and MEROPS[48]. AMRFinderPlus is packaged with its own search engine (AMRfinder) and was used with default parameters. BLASTP was used to query the other resources with the alignment criteria previously described[20]. TCDB[49] was used to attribute more specific annotations to CDS categorized as transmembrane transport (e-value < 1e-50, identities > 30%, positives > 55%). These category assignments were then manually reviewed and completed by i) domain analyses of proteins by CD Search[50], ii) bacteriocin synthesis genes identification, using BAGEL4[51] and antiSMASH[52], iii) analysis of defense mechanisms against bacteriophages, using DefenseFinder[53], and PADLOC[54].

Data availability

The four annotated *Faecalibacterium* genomes, including ICE and IME content, can be accessed freely and openly on Data INRAE (https://doi.org/https://doi.org/10.15454/VI7YRB).

Received: 15 November 2024; Accepted: 24 April 2025
Published online: 16 May 2025

References

1. Martín, R. et al. *Faecalibacterium*: a bacterial genus with promising human health applications. *FEMS Microbiol. Rev.* **47**, 1–18 (2023).
2. Tap, J. et al. Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* **11**, 2574–2584 (2009).
3. Miquel, S. et al. *Faecalibacterium prausnitzii* and human intestinal health. *Curr. Opin. Microbiol.* **16**, 255–261 (2013).
4. Miquel, S. et al. Ecology and metabolism of the beneficial intestinal commensal bacterium *Faecalibacterium prausnitzii*. *Gut Microb.* **5**, 146–151 (2014).
5. Sokol, H. et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. USA* **105**, 16731–16736 (2008).
6. Machiels, K. et al. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* **63**, 1275–1283 (2014).
7. Borges-Canha, M., Portela-Cidade, J. P., Dinis-Ribeiro, M., Leite-Moreira, A. F. & Pimentel-Nunes, P. Role of colonic microbiota in colorectal carcinogenesis: a systematic review. *Rev. Esp. Enferm. Dig.* **107**, 659–671 (2015).
8. Lopez-Siles, M. et al. Alterations in the abundance and co-occurrence of *Akkermansia muciniphila* and *Faecalibacterium prausnitzii* in the colonic mucosa of inflammatory bowel disease subjects. *Front. Cell Infect. Microbiol.* **8**, 281 (2018).
9. Duncan, S. H., Hold, G. L., Harmsen, H. J., Stewart, C. S. & Flint, H. J. Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *Int. J. Syst. Evol. Micro* **52**, 2141–2146 (2002).
10. Sakamoto, M. et al. Genome-based, phenotypic and chemotaxonomic classification of Faecalibacterium strains: proposal of three novel species Faecalibacterium duncaniae sp. nov., Faecalibacterium hattorii sp. nov. and Faecalibacterium gallinarum sp. nov.. *Int. J. Syst. Evol. Micro* **72**, 005379 (2022).
11. Plomp, N. & Harmsen, H. J. M. Description of Faecalibacterium wellingii sp. nov. and two Faecalibacterium taiwanense strains, aiding to the reclassification of Faecalibacterium species. *Anaerobe* **89**, 102881 (2024).
12. Fitzgerald, C. B. et al. Comparative analysis of *Faecalibacterium prausnitzii* genomes shows a high level of genome plasticity and warrants separation into new species-level taxa. *BMC Genom.* **19**, 931–951 (2018).
13. Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F. & Rocha, E. P. C. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet.* **7**, e1002222 (2011).
14. Bellanger, X., Payot, S., Leblond-Bourget, N. & Guédon, G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.* **38**, 720–760 (2014).
15. Guédon, G., Libante, V., Coluzzi, C., Payot, S. & Leblond-Bourget, N. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. *Genes* **8**, 337 (2017).

16. Botelho, J. & Schulenburg, H. The role of Integrative and conjugative elements in antibiotic resistance evolution. *Trends Microbiol.* **29**, 8–18 (2021).
17. Forster, S. C. et al. Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome. *Nat. Commun.* **13**, 1445 (2022).
18. Laroussi, H. et al. Exploration of DNA processing features unravels novel properties of ICE conjugation in gram-positive bacteria. *Nucleic Acids Res.* **50**, 8127–8142 (2022).
19. Coluzzi, C. et al. A glimpse into the world of integrative and mobilizable elements in streptococci reveals an unexpected diversity and novel families of mobilization proteins. *Front. Microbiol.* **8**, 443 (2017).
20. Lao, J. et al. Abundance, diversity and role of ICEs and IMEs in the adaptation of *Streptococcus salivarius* to the Environment. *Genes (Basel)* **11**, 999 (2020).
21. Lao, J. et al. ICEscreen: a tool to detect firmicute ICEs and IMEs, isolated or enclosed in composite structures. *NAR Genom. Bioinform.* **4**, 079 (2022).
22. Bhatty, M., Laverde Gomez, J. A. & Christie, P. J. The expanding bacterial type IV secretion lexicon. *Res. Microbiol.* **164**, 620–639 (2013).
23. Ambroset, C. et al. New insights into the classification and integration specificity of *Streptococcus* integrative conjugative elements through extensive genome exploration. *Front. Microbiol.* **6**, 1483 (2016).
24. Iyer, L. M., Burroughs, A. M., Anand, S., de Souza, R. F. & Aravind, L. Polyvalent Proteins, a pervasive theme in the intergenomic biological conflicts of bacteriophages and conjugative elements. *J. Bacteriol.* https://doi.org/10.1128/jb.00245-17 (2017).
25. Garcillán-Barcia, M. P., Francia, M. V. & Cruz, F. D. L. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* **33**, 657–687 (2009).
26. Pluta, R. et al. Structural basis of a histidine-DNA nicking/joining mechanism for gene transfer and promiscuous spread of antibiotic resistance. *Proc. Natl. Acad. Sci. USA* **114**, E6526–E6535 (2017).
27. Barcenilla, A. et al. Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl. Environ. Microbiol.* **66**, 1654–1661 (2000).
28. Guédon, G. et al. FirmiData: a set of 40 genomes of *Firmicutes* with a curated annotation of ICEs and IMEs. *BMC Res. Notes* **15**, 157 (2022).
29. Soto-Martin, E. C. et al. Vitamin biosynthesis by human gut butyrate-producing bacteria and cross-feeding in synthetic microbial communities. *MBio* **11**, e00886-e920 (2020).
30. Morton, J. T., Freed, S. D., Lee, S. W. & Friedberg, I. A large scale prediction of bacteriocin gene blocks suggests a wide functional spectrum for bacteriocins. *BMC Bioinformat.* **16**, 381–389 (2015).
31. Suryaletha, K. et al. Demystifying bacteriocins of human microbiota by genome guided prospects: An impetus to rekindle the antimicrobial research. *Curr. Protein Pept. Sci.* **23**, 811–822 (2022).
32. Kurokawa, K. et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14**, 169–181 (2007).
33. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098-1109.e9 (2021).
34. Coyne, M. J., Zitomersky, N. L., McGuire, A. M., Earl, A. M. & Comstock, L. E. Evidence of extensive DNA transfer between bacteroidales species within the human gut. *MBio* **5**, e01305-01314 (2014).
35. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
36. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
37. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
38. Hall, T. A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
39. Dobson, L., Reményi, I. & Tusnády, G. E. CCTOP: a consensus constrained TOPology prediction web server. *Nucleic Acids Res.* **43**, W408–W412 (2015).
40. Jumper, J. & Hassabis, D. Protein structure predictions to atomic accuracy with AlphaFold. *Nat. Methods* **19**, 11–12 (2022).
41. Holm, L. & Sander, C. Protein folds and families: sequence and structure alignments. *Nucleic Acids Res.* **27**, 244–247 (1999).
42. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evolut.* **38**, 3022–3027 (2021).
43. Feldgarden, M. et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* https://doi.org/10.1128/AAC.00483-19 (2019).
44. Hammami, R., Zouhir, A., Ben Hamida, J. & Fliss, I. BACTIBASE: a new web-accessible database for bacteriocin characterization. *BMC Microbiol.* **7**, 89 (2007).
45. Liu, B., Zheng, D., Jin, Q., Chen, L. & Yang, J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* **47**, D687–D692 (2019).
46. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—A database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **43**, D298–D299 (2015).
47. Flissi, A. et al. Norine: update of the nonribosomal peptide resource. *Nucleic Acids Res.* **48**, D465–D469 (2020).
48. Rawlings, N. D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632 (2018).
49. Saier, M. H. et al. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res.* **49**, D461–D467 (2021).
50. Wang, J. et al. The conserved domain database in 2023. *Nucleic Acids Res.* **51**, D384–D388 (2022).
51. van Heel, A. et al. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46**, W278–W281 (2018).
52. Blin, K. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
53. Tesson, F. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).
54. Payne, L. J. et al. PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. *Nucleic Acids Res.* **50**, W541–W550 (2022).
55. Tsvetkova, K., Marvaud, J.-C. & Lambert, T. Analysis of the mobilization functions of the vancomycin resistance transposon Tn1549, a member of a new family of conjugative elements. *J. Bacteriol.* **192**, 702–713 (2010).
56. Giovanetti, E., Brenciani, A., Tiberi, E., Bacciaglia, A. & Varaldo, P. E. ICESp2905, the erm(TR)-tet(O) element of *Streptococcus pyogenes*, is formed by two independent integrative and conjugative elements. *Antimicrob. Agents Chemother.* **56**, 591–594 (2012).
57. Crellin, P. K. & Rood, J. I. Tn4451 from *Clostridium perfringens* is a mobilizable transposon that encodes the functional Mob protein, TnpZ. *Mol. Microbiol.* **27**, 631–642 (1998).

Author contributions

Conceived and designed the experiments: G.G, N. L-B. Analyzed the data: G.G., F. C-B., T.H, N.S., B.D., T.L, H.C., N. L-B. Writing original draft: G.G, N. L-B. Reviewing: all authors.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-99981-y.

Correspondence and requests for materials should be addressed to N.L.-B.

Reprints and permissions information is available at www.nature.com/reprints.

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