Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# A conserved shufflon as an intra genus regulatory strategy in gut microbiota

**Nathalie G. Gruber**
Institute of Science and Technology Austria, Klosterneuburg, Austria
**Bernhard Hekele**
Centre for Microbiology and Environmental Systems Science, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria; BOKU University, Vienna, Austria
**Sanaz Khadem**
Centre for Microbiology and Environmental Systems Science, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria
**Natalia O. Dranenko**
Center for Bio- and Medical Technologies, Moscow, Russia
**Ekaterina S. Kolodyaznaya**
Center for Bio- and Medical Technologies, Moscow, Russia
**Michael Hennessey-Wesen**
Institute of Science and Technology Austria, Klosterneuburg, Austria; Evolutionary and Synthetic Biology Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan
**Roderich Roemhild**
Institute of Science and Technology Austria, Klosterneuburg, Austria https://orcid.org/0000-0001-9480-5261
**Fyodor A. Kondrashov**
Evolutionary and Synthetic Biology Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan
**David Berry**
Centre for Microbiology and Environmental Systems Science, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria https://orcid.org/0000-0002-8997-608X
**Calin C. Guet**
Institute of Science and Technology Austria, Klosterneuburg, Austria
**Olga O. Bochkareva**
olga.bochkareva@univie.ac.at

Centre for Microbiology and Environmental Systems Science, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria https://orcid.org/0000-0003-1006-6639

**Biological Sciences - Article**

**Keywords:**

**Posted Date:** March 5th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-8987032/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** There is **NO** Competing Interest.

# A conserved shufflon as an intra genus regulatory strategy in gut microbiota

Nathalie G. Gruber[1], Bernhard Hekele[2,3], Sanaz Khadem[2], Natalia O. Dranenko[4], Ekaterina S. Kolodyaznaya[4], Michael Hennessey-Wesen[1,5], Roderich Roemhild[1], Fyodor A. Kondrashov[5], David Berry[2], Calin C. Guet[1], Olga O. Bochkareva[2]

[1]Institute of Science and Technology Austria, Klosterneuburg, Austria
[2]Centre for Microbiology and Environmental Systems Science, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria
[3]BOKU University, Vienna, Austria
[4]Center for Bio- and Medical Technologies, Moscow, Russia
[5]Evolutionary and Synthetic Biology Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

**Gut microorganisms inhabit a highly dynamic environment characterized by fluctuating nutrients, phage predation, and host immune pressure. To cope with such unpredictability, bacteria often rely on phase variation – the rapid generation of phenotypic heterogeneity via reversible DNA inversions. While the presence of multiple inversions, so called shufflons, in polysaccharide utilization loci was reported in several bacterial gut species, their role in gene expression regulation is poorly understood. Here we report and characterize an intra genus conserved phase variation system in Bacteroidaceae, a dominant group of the mammalian gut microbiota. This system consists of a cassette of silent *susCD* transmembrane transporter genes and an invertase that drives the rapid positioning of individual inactive genes into a single active expression site. Consequently, individual bacteria within a colony express different transporters, while collectively maintaining expression of the full gene repertoire. The composition of the transporter cassette is dynamic and evolves through horizontal gene transfer across the genus while preserving a conserved architecture. This shufflon system generates persistent genetic and phenotypic diversity, enabling Bacteroidaceae populations to adapt rapidly to the fluctuating conditions in the mammalian gut.**

## Main

How microbes adapt to dynamic and heterogeneous environments is a central question of microbiology. Many pathogens and commensals survive in a constantly changing environment by rapidly displaying reversible phenotypic variability, a phenomenon known as phase variation[1–5]. The defining feature of phase variation is its extraordinarily rapid rate of genetic change that enables individual bacterial cells within a single colony to display different phenotypes. This intra population heterogeneity is thought to be a bet-hedging strategy that allows microorganisms to maximize the chances that at least some cells will survive any environmental challenge, such as immune system targeting and phage attacks[3,4,6]. For instance, switching of surface adhesins allows pathogens to colonize and infect the host while evading the host's immune responses, leading to persistent infections[7–10]. Similarly, variation in capsular polysaccharides leads to a dynamic equilibrium between phage predation and bacterial growth, enabling the bacterial population to persist despite selective pressures[11–13].

Phase variation can be generated by several molecular mechanisms that produce reversible DNA changes[3,4,6]. One of them is through the inversion of DNA segments by recombination of flanking inverted repeats (IRs)[14–16]. Such inversions can modify genes and/or promoters, which lead to functional differences (**Box a**)[14,16]. When multiple IR copies are present, they enable site-specific DNA inversions between different IR pairs. This creates several possible DNA arrangements and diverse genotypes through DNA shuffling[17,18]. This generates alternative variants of a specific molecular function. The first such system, termed a shufflon, was identified on a plasmid of *Enterobacteriaceae,* where it produces variants of the PilV pilus protein, and thereby alters conjugative specificity (**Box b**)[17]. Shufflons also create variants in phage defense of type 1 restriction-modification systems of *Streptococcus* (**Box c**)[19], and in polysaccharide utilization loci (PUL) of *Bacteroides,* an abundant genus of the mammalian gut microbiome[18].

Human hosts break down complex carbohydrates (glycans) incompletely, leaving a highly diverse set of glycans for degradation by the gut microbiota[20–24]. Bacteroides consume this diversity of glycans, dedicating up to 20% of their genome to a glycan-specific molecular machinery encoded in PULs[25–27]. Many genes in *Bacteroides* species are regulated by serine- and tyrosine-specific recombinases (invertases) that create site-specific DNA inversions that cause promoter flips, domain swapping, and inter operon gene reorientation[28–33]. *Bacteroides* genomes encode numerous invertases, including master invertases that regulate multiple genes[28,30,34]. The

master invertases Mpi, Tsr19, and Tsr0667 toggle expression of genes involved in critical *Bacteroides* survival programs, including capsular and extracellular polysaccharide synthesis, anti-phage defense, and polysaccharide utilization[12,16,29]. Specifically, switching expression of PUL-associated transporters encoded by *susCD* genes enables *Bacteroides* species to utilize different glycan niches, responding in real time to variable nutrient availability in the highly competitive mammalian gut environment[18,35–39].

In *B. fragilis,* the Tsr0667 invertase controls DNA shuffling in three different *susCD* gene clusters, including its own locus, and DNA inversions at nine other *susCD*-containing loci[18]. However, the functional role of DNA shuffling in regulating these genes and the resulting protein composition remains poorly understood. Recently, similarly organized shufflons have been reported independently by different teams in several *Bacteroides* species[11,29,40]. We noticed that these shufflons encode closely related orthologs of the Tsr0667 invertase, which targets the *susCD* genes flanked by the same IRs, suggesting their common evolutionary origin[40]. Given the crucial role of *susCD* genes for *Bacteroides* survival, we hypothesized that these shufflons represent a key widespread shufflon system in members of the gut microbiome. Thus, we performed a genome-wide analysis of the Tsr0667-mediated shufflons in Bacteroidaceae to functionally characterize this phase variation system and uncover the mechanisms driving its evolution. By combining computational analysis with experiments, we achieved the first deep investigation of a shufflon at the level of an entire bacterial family and revealed an ancient core phase variation system in the mammalian gut microbiota.

## The Tsr-mediated *susCD* shufflon is an ancient core phase variation system in Bacteroidaceae

First, we searched for shufflon-encoded invertase orthologs (Tsr0667 from *B. fragilis* was used as a query) across Bacteroidaceae genomes. In diverse members of the taxon, we found *tsr0667* orthologs flanked by genes encoding TonB-dependent transporters from the SusCD protein family. Thus, we preliminarily identified the subset of *tsr0667*-associated *susCD* transporter genes, as those genes that are in the vicinity of the *tsr0667* orthologs and which are flanked by the same pair of core genes (**Figure 1a**). In all of them, we found that genomic motifs (**Figure 1b)** similar to the 22-nucleotide inverted repeat sequence previously reported in *B. fragilis* and *B. thetaiotaomicron tsr0667*-forming shufflons[40], indicating that these *susCD* genes are part of the invertase-transporter

shufflon system. Hereinafter, we will simply use *tsr* as a common name for all identified *tsr0667* orthologs and Tsr BS to refer to the pool of invertase binding sites in Tsr0667-mediated shufflons.

The identification of Tsr BS repeats in *susCD* genes allowed us to complete the annotation of the *tsr*-mediated shufflons. In each analyzed genome, only one *susCD* transporter gene in the shufflon contained a full-length gene sequence, and this functional gene was located immediately upstream of the *tsr* invertase gene itself (**Figure 2a**). The other *susCD* transporter genes were shorter, lacking the 5' end sequence **(SFigure S1)**. Notably, many *susCD* genes close to the *tsr* invertase gene were annotated as pseudogenes or as having incomplete coding DNA sequences (CDSs), indicating that they may not be independently functional. The invertase mechanism of action provides a simple model for the function of all *susCD* genes as part of the shufflon system **(Figure 2b)**. The full-length *susCD* transporter gene pair is located upstream of the invertase where it is fused with an active promoter and a ribosomal binding site (RBS). The Tsr invertase can replace most of the coding sequence of the expressed *susCD* transporter with any of the partial *susCD* transporter genes that have the Tsr BS. In other words, the invertase can select any of the available deactivated transporter genes from the cassette of pseudogenized genes and swap it with the currently active transporter gene in the upstream genomic location, thereby activating the former and deactivating the latter. This genetic system architecture allows for long-term evolutionary flexibility, as different transporter cassette genes can be gained or lost without disrupting the core shuffling mechanism.

The domain architecture of the SusCD complex itself explains how the recombination mechanism creates intra population functional diversity. The N-terminal region that contains the Tsr BS corresponds to the periplasmic functional domain of SusC, while the invertase swapped region encodes the larger outer membrane domain (**Figure 2c**). The Tsr BS motif is part of the periplasmic domain, with the variable positions of the motif matching synonymous sites of the protein sequence.

**Tsr-mediated *susCD* shuffling produces multiple genotypes across Bacteroidaceae**

The similarity of the genetic locus architecture across the gut Bacteroidaceae and the evolutionary conservation of the Tsr BS that we uncovered above suggested that this mechanism is active across the entire taxon. To test this, we selected three distantly related species of Bacteroidaceae with highly diverged sets of *susCD* transporters and dissimilarities in the Tsr BS sequences: *Bacteroides*

*thetaiotomicron (B. theta)* with three transporters in a 20 kb genomic region, *Bacteroides uniformis* with six transporters spread across 42 kb, and *Phocaeicola vulgatus* with six transporter genes present in 35 kb of the genome**.** All cultures were obtained from the Gut Bacteria Collection, University of Vienna. To experimentally test whether all *susCD* transporter genes with predicted Tsr BS recombine, we designed a set of primer pairs for PCR as follows. Each strain had one primer on the 5'-end of the active transporter gene aligned with the gene reading frame (*forward*) and a set of primers binding different predicted Tsr-mediated transporter genes on a complementary strand (*reverse)*. Thus, the presence of PCR products produced by different forward-reverse primer pairs indicates inversion activity (**Figure 3a**). This leads to the replacement of the 5′-fragment of the expressed *susCD* transporter with all other partial genes from the cassette. Remarkably, PCR amplification of genes from a single colony produced products for each *forward-reverse* primer pair combination in each of the three species **(Figure 3b)**. This means that within a single colony, there is a diversity of individual cells that collectively express all of the possible transporter gene combinations of the cassette, likely leading to maximum levels (in terms of possible number of different combinations) of phenotypic variability. The conserved genomic organization suggests this shuffling mechanism operates across all species where we identified the presence of the shufflon system.

To obtain an assessment of activated transporter distributions within a clonal population, we sequenced genomic DNA of *B. uniformis* and *P. vulgatus* with long-reads. We chose these species as they have large transporter gene repertoires. Consistent with the PCR results, we observed the reads corresponding to multiple genotypic variants **(Figure 3c)**. The presence of a major variant in the sequencing reads of the samples explains the success of standard genome assembly procedure, resulting in one complete assembly even when a highly heterogeneous colony is sequenced. This observation starkly exemplifies the limitations of the classic 'one bacterial colony - one genome' model as a basis for bacterial genome assemblers for Bacteroidaceae samples. Taken together, the bioinformatic analysis and the experimental results suggest that all truncated *susCD* cassette genes with a Tsr BS*,* including those labeled as pseudogenes, are in fact active genes that are concurrently expressed in Bacteroidaceae populations through the phase variation mechanism.

**Reconstitution of Tsr-mediated shuffling in *Escherichia coli* produces phenotypically heterogeneous colonies**

To experimentally verify that the different Tsr invertases work in concert with the predicted inverted repeats, we used *E. coli* as a genetic model system. We inserted the 22-nucleotide repeats from *B. fragilis* (Tsr BS) between the promoter regions of *mhpR* and *lacZ*, thereby creating a reporter system in which inversion disrupts *lacZ* transcription and lactose metabolism **(Figure 4a)**. Taking inspiration from the *Bacteroides* genomes, we spaced the genomic inverted repeats by inserting them two genes apart from each other**.** To estimate the rate at which the *B. fragilis* invertase (Tsr) recognizes these motifs and produces chromosomal inversions, we cloned a Tsr invertase under a TetR inducible promoter on a very low-copy plasmid. This design enabled us to examine the kinetics of invertase expression. The *E. coli* strain with the inverted repeats is referred to as the 'IR' strain, while the strain with the additional invertase on a plasmid is referred to as the 'IRX' strain.

To test whether Tsr BS repeats alone could mediate spontaneous recombination, we examined the presence of $Lac^+$ and $Lac^-$ genotypes in mutant cultures. In IR strains and IRX strains without induction, PCR analysis revealed that only the original $Lac^+$ genotype was present, which indicates that the inverted repeats alone do not result in spontaneous recombination in *E. coli* **(Figure 4b)**. After induction with aTc, IRX cultures displayed bands in both, the $Lac^+$ and $Lac^-$ genotypes, showing that active inversions were produced by the *Bacteroides* Tsr invertase even outside its original genomic context.

To study the invertase activity, we isolated the genomic DNA from cultures induced with aTc for 3 min, 5 min, 1 h, and 3 h, to investigate the ratio of $Lac^+$ to $Lac^-$ genotypes in these samples. Strikingly, we found that both, $Lac^+$ and $Lac^-$ genotypes were present after just five minutes of invertase induction **(Figure 4c, S2)**. By analyzing the DNA electrophoresis gel and PCR products, we estimated that approximately two-fold more $Lac^+$ to $Lac^-$ genotypes were present in the IRX culture in which the invertase was induced for just three minutes.

We confirmed these genotypic results phenotypically using a blue-white screening assay. On selection plates with X-gal, a colorless compound turns blue when cleaved by β-galactose, and thus one can distinguish blue colonies when *lacZ* is transcribed ($Lac^+$) and white colonies when transcription of *lacZ* is disrupted due to inversion ($Lac^-$). Using this method, we observed the kinetics of the $Lac^+$/$Lac^-$ ratios in IRX, while IR continuously remains in $Lac^+$ **(Figure 4d)**.

Individual colonies induced for more than three hours displayed mixed phenotypes, indicating that invertase activity continues within clonal populations and generates intra population heterogeneity **(Figure 4e)**. The temporal dynamics of the $Lac^+/Lac^-$ ratio in *E. coli* mutants expressing Tsr support the hypothesis that a large degree of stochastic variation is intrinsic to the invertase-mediated shuffling process in Bacteroidaceae.

**Horizontal gene transfer increases the diversity of the Tsr-mediated transporters repertoire**

The maintenance of the *tsr-susCD* system allows us to study its evolution and gain insights into the underlying processes. We first focused on the sequence evolution of its components: the Tsr invertases, the Tsr BS repeats, and the SusCD transporters. To better resolve the evolutionary patterns in this complex genetic system, we restricted this analysis to the genus *Bacteroides*, in which a total of 134 genome assemblies from 21 species were publicly available. The intergenic region between *tsr* and the activated *susCD* gene pair showed remarkable length conservation across the entire genus, spanning a narrow window of 452-457 nucleotides between the starts of these genes **(Figure S3)**. Phylogenetic analysis **(Figure S4)** revealed that vertical transmission of the DNA fragment with the core machinery - the *tsr* invertase gene, the *tsr-susCD* intergenic region with the *susCD* promoter, and the shuffled part of *susC* genes - was common throughout the clade. Such conservation of the distance between bi-directed *tsr* and the activated *susCD* gene pair may indicate that their transcription is co-regulated by the shared genetic control elements, which is a hypothesis that can be tested in future studies.

By contrast, *susCD* transporters were highly variable even among strains of one species, in both number and composition (**Figure 5a**). For instance, we found a total of eight distinct transporter genes across *B. ovatus*. However, each genome has only a subset of them (**Figure 5b**). Phylogenetic analysis of Tsr-mediated transporters and the phyletic patterns of their representatives (**Figure 5c, d**) indicates high rates of their horizontal transfer between members of the gut microbiome. This reveals an important feature of the *tsr-susCD* shufflon system, namely that the large diversity of its cassette repertoire is driven by horizontal exchange among different members of the gut microbiota. The presence of mobile elements adjacent to silent *susCD* genes **(Figure 1a)** may suggest their role in facilitating horizontal gene transfer (HGT)[41].

Evolution of the *tsr-susCD* shufflon system via horizontal exchange of Tsr-mediated transporters explains the maintenance of shufflon machinery across Bacteroidaceae and completes

the mechanistic picture. Each population maintains its own collection of cassette genes, which defines the spectrum of accessible phenotypes. This collection can be updated through horizontal acquisition from other populations harboring the same system. The genetic linkage of the invertase and the promoter of target genes ensures tight, localized control. Silent cassette genes, in turn can be distributed throughout the genome, with the only prerequisite being to maintain strand orientation. As an intuitive analogy, shufflon evolution resembles a jukebox: a mechanical arm selects a single record from a pre-loaded library and activates it by placing it on the turntable. At any moment, each jukebox operates with its available records, while the library itself is constantly curated through record additions and deletions to optimize its composition.

**At large taxonomic scale the *tsr-susCD* shufflon system is subject to divergent evolution**

Horizontal exchange of silent *susCD* transporters between different *Bacteroides* members raises the question about the specific binding affinity of Tsr and Tsr BS. The majority of Tsr BS sequences are identical across the transporter cassette genes encoded in one genome (**Supplementary Table S1**), including the transporter cassette genes that were likely acquired by HGT**.** This suggests that newly acquired transporter cassette genes must either carry compatible Tsr BS sequences or rapidly acquire them through recombination or mutation to function within the existing system.

Phylogenetic analysis of Tsr BS evolution in *Bacteroides* revealed sites with clade-specific mutations (**Figure 5e**), suggesting that these positions are important for Tsr affinity. To test whether Tsr-Tsr BS pairs exhibit species-specific compatibility, we designed an *E. coli* mutant 'IRY' that contained genetic components from two evolutionary distant *Bacteroides*: Tsr BS from *B. fragilis* and the *tsr* gene from *B. uniformis*. The *B. uniformis* Tsr BS differs by three mutations from the *B. fragilis* motif. Even after 3h induction of the invertase, the IRY cultures displayed only the original $Lac^+$ genotype, indicating an undetectably low rate of inversions produced by this Tsr-Tsr BS pair (**Figure 5f**). These data suggest that while many *Bacteroides* species can exchange the cassette genes by HGT, the phase variation system itself is also undergoing divergent evolution coupled to the change in Tsr-Tsr BS specificity and therefore compatibility.

## Discussion

We uncovered that Tsr-mediated shuffling replaces the expressed *susCD* genes with the partially silent ones from the alterable gene cassette repertoire. These shufflon systems are functional across all Bacteroidaceae and act as an intra genus regulatory strategy in human gut microbiota. The system operates at two distinct levels: intra genomically, it diversifies phenotypes within a population, while inter genomically, horizontal gene transfer enables exchange of a vast phenotypic repertoire. Together, these two levels of recombination couple short-term population diversification with long-term evolutionary flexibility. Our results indicate the extent to which the molecular regulation strategy in the gut microbiome relies on different temporal and organizational levels.

The *tsr-susCD* intra genus shufflon system represents a fundamentally different adaptive alternative to the global transcription factor-based (TF) regulation common to most bacteria that we know, such as *E. coli*[28,42]. While the hierarchical TF networks provide optimal responses to environmental changes, such as nutrient composition, they demand resource-intensive regulatory infrastructure. The non-motile gut commensal Bacteroidaceae evolved a decentralized, modular and highly recombinogenic regulatory strategy based on random diversification of metabolic heterogeneity within populations[28]. While this strategy is reminiscent of bet hedging, our finding of a highly conserved intra genus shufflon structure suggests an alternative interpretation. One may also consider this as a *de facto* genus level gene regulatory strategy, where hierarchical TFs have been replaced by decentralized invertases. From a population dynamics point of view, this strategy has the advantage that at any given time, all possible gene expression states are present, while the beneficial one can simply replicate at a faster speed. When conditions change, the gene expression states are nevertheless maintained in the population and a different gene expression state can in turn replicate faster. It is remarkable that *Bacteroides* as a genus dispensed fully with the universal -10 and -35 promoter architecture of most known bacterial species, and instead evolved a unique -7 and -33 architecture of RNAP recognition, which may reflect the lack of global transcription factors[43].

Conservation of the *tsr-susCD* shufflon on such a large taxonomic scale is unique for the *tsr0667* invertase gene. In contrast, the regulatory networks and genomic arrangement of other well-studied *B. fragilis* master invertases (*mpi* and *tsr19*) vary even across *Bacteroides* species[32]. In *B. fragilis* the *mpi* gene is paired with the *tsr19* gene in a head-to-head orientation[30]. Tsr19

invertase mediates the on-off switching of downstream genes involved in extracellular polysaccharide biosynthesis. Additionally, it is supposed to control the expression of *mpi*, which regulates the expression of fourteen different loci, including seven polysaccharide synthesis loci[44]. This global coordination allows rapid alteration of a large array of surface structures contributing to *B. fragilis'* survival under conditions of inflammation[12]. In contrast, *B. theta,* a prominent symbiont and metabolic generalist, shows co-localization of most invertible regions with their cognate invertases, suggesting local control[34]. While exhibiting less resistance to environmental stressors*, B. theta* possesses much broader dietary polysaccharide utilization capacities, relying on a significantly higher number of PULs encoded in its genome[36]. Given this intra genus diversity in regulatory strategies of invertible regions, maintenance of the *tsr-susCD* shufflon system across Bacteroidaceae indicates a central role for this genetic system in adaptation of glycan-degrading bacteria in the gut environment, suggesting it is essential for Bacteroidaceae fitness.

*SusCD* genes encode protein complexes, which are the essential components of outer membrane transport systems in Bacteroidaceae[36,37,39]. While they are primarily associated with glycan uptake, emerging evidence suggests a broader substrate specificity and a multifaceted role for *Bacteroides* survival in the gut environment. For instance, they have been shown to transport antimicrobial ubiquitin homologues to drive intra species bacterial competition[45] and may be involved in phage infection[46]. A single *Bacteroides* genome can encode over one hundred different SusCD proteins. However, the functions and substrate specificities of most of them remain uncharacterized[35], including those that are subject to Tsr-mediated regulation. We showed that Tsr-mediated shufflons provide the genomic basis for diversification of *susCD* gene expression, allowing individual populations to exploit multiple, likely distinct metabolic niches. However, identification of the targets of individual SusCD proteins is needed in order to understand the connection between cassette composition and the spectrum of phenotypes within a population.

Horizontal gene transfer plays a fundamental role in adaptive evolution for individual bacterial populations and microbial communities[47,48]. The gut of a healthy individual typically contains up to twenty *Bacteroides* species, the most abundant being *B. uniformis, B. vulgatus, B. caccae,* and *B. theta*[49]. Gut Bacteroidaceae exhibit tremendous intra species diversity shaped by HGT, suggesting horizontal exchange of diverse genes occur actively within the gut microbiome of individual hosts[50–52]. We demonstrate that the transporter gene cassette repertoire in a single Bacteroidaceae population evolves horizontally, presumably through exchange within the

microbiome community, allowing for rapid reorganization of the cassette composition. To our knowledge, this is the first indication that bacterial shufflons combine rapid *intra* genomic recombination with the slower *inter* genomic recombination in order to diversify the functional repertoire of individual populations. Taken together, our findings contribute to the understanding of gene regulation in the gut microbiome on the metagenomic level, revealing a new layer of bacterial genome complexity that operates at the genus level.

**References**


1. Van der Woude, M. W. Phase variation: How to create and coordinate population diversity. *Curr. Opin. Microbiol.* **14**, 205–211 (2011).
2. Ruiz, L. M. R., Williams, C. L. & Tamayo, R. Enhancing bacterial survival through phenotypic heterogeneity. *PLoS Pathog.* **16**, e1008439 (2020).
3. Bayliss, C. D., Clark, J. L. & van der Woude, M. W. 100+ years of phase variation: the premier bacterial bet-hedging phenomenon. *Microbiology* **171**, (2025).
4. Fraser, A. J., McMahon, F. E. & Atack, J. M. Microbial Primer: Phase variation-survival and adaptability by generation of a diverse population. *Microbiology* **170**, 001492 (2024).
5. Byrne, A. S., Bissonnette, N. & Tahlan Kapil. Mechanisms and implications of phenotypic switching in bacterial pathogens. *Can. J. Microbiol.* **71**, 1-19 (2024).
6. Phillips, Z. N., Tram, G., Seib, K. L. & Atack, J. M. Phase-variable bacterial loci: How bacteria gamble to maximise fitness in changing environments. *Biochem. Soc. Trans.* **47**, 1131–1141 (2019).
7. Cayrou, C., Barratt, N. A., Ketley, J. M. & Bayliss, C. D. Phase Variation During Host Colonization and Invasion by Campylobacter jejuni and Other Campylobacter Species. *Front. Microbiol.* **12**, (2021).
8. Vargas, R. *et al.* Phase variation as a major mechanism of adaptation in Mycobacterium tuberculosis complex. *Proc. Natl. Acad. Sci. U. S. A.* **120**, (2023).
9. Li, J. & Zhang, J.-R. Phase Variation of Streptococcus pneumoniae. *Microbiol. Spectr.* **7**, 10-1128 (2019).
10. Anjuwon-Foster, B. R. & Tamayo, R. Phase variation of Clostridium difficile virulence factors. *Gut Microbes* **9**, 76–83 (2018).
11. Shkoporov, A. N. *et al.* Long-term persistence of crAss-like phage crAss001 is associated with phase variation in Bacteroides intestinalis. *BMC Biol.* **19**, 163 (2021).
12. Carasso, S. *et al.* Inflammation and bacteriophages affect DNA inversion states and functionality of the gut microbiota. *Cell Host Microbe* **32**, 322-334 (2024).
13. Cortés-Martín, A. *et al.* Adaptations in gut Bacteroidales facilitate stable co-existence with their lytic bacteriophages. *Gut Microbes* **17**, 2507775 (2025).
14. Clegg, S., Hancox, L. S. & Yeh, K.-S. Salmonella typhimurium Fimbrial Phase Variation and FimA Expression. *J. Bacteriol.* **178**, 542–545 (1996).

15. Trzilova, D. & Tamayo, R. Site-Specific Recombination – How Simple DNA Inversions Produce Complex Phenotypic Heterogeneity in Bacterial Populations. *Trends in Genetics* **37**, 59–72 (2021).
16. Jiang, X. *et al.* Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. *Science* **363**, 181–187 (2019).
17. Komano, T., Kubo, A., Kayanuma, T., Furuichi, T. & Nisioka, T. Highly Mobile DNA Segment of Incla Plasmid R64: A Clustered Inversion Region. *J. Bacteriol.* **165**, 94–100 (1986).
18. Nakayama-Imaohji, H. *et al.* Identification of the site-specific DNA invertase responsible for the phase variation of SusC/SusD family outer membrane proteins in Bacteroides fragilis. *J. Bacteriol.* **191**, 6003–6011 (2009).
19. De Ste Croix, M. *et al.* Phase-variable methylation and epigenetic regulation by type I restriction–modification systems. *FEMS Microbiol. Rev.* **41**, S3–S15 (2017).
20. Rowland, I. *et al.* Gut Microbiota Functions: Metabolism of Nutrients and Other Food Components. *European Journal of Nutrition* vol. **57,** 1-25 (2018).
21. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335 (2012).
22. Makki, K., Deehan, E. C., Walter, J. & Bäckhed, F. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* **23**, 705–715 (2018).
23. Patrick, S. A tale of two habitats: Bacteroides fragilis, a lethal pathogen and resident in the human gastrointestinal microbiome. *Microbiology* **168**, 001156 (2022).
24. Cronin, P., Joyce, S. A., O'toole, P. W. & O'connor, E. M. Dietary fibre modulates the gut microbiota. *Nutrients* **13**, 1655 (2021).
25. Feng, J. *et al.* Polysaccharide utilization loci in Bacteroides determine population fitness and community-level interactions. *Cell Host Microbe* **30**, 200-215.e12 (2022).
26. Coyne, M. J. & Comstock, L. E. Niche-specific features of the intestinal Bacteroidales. *J. Bacteriol.* **190**, 736–742 (2008).
27. Joglekar, P. *et al.* Genetic Variation of the SusC/SusD Homologs from a Polysaccharide Utilization Locus Underlies Divergent Fructan Specificities and Functional Adaptation in Bacteroides thetaiotaomicron Strains. *mSphere* **3**, 10-1128 (2018).
28. Ryan, D. Gene regulation in Bacteroides fragilis: adaptive control in a dynamic host environment. *Microbiology and Molecular Biology Reviews, **89***, e00225-25 (2025)
29. Porter, N. T. *et al.* Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in Bacteroides thetaiotaomicron. *Nat. Microbiol.* **5**, 1170–1181 (2020).
30. Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. Mpi Recombinase Globally Modulates the Surface Architecture of a Human Commensal Bacterium. *PNAS* **100**, 10446-10451 (2003).

31. Roche-Hakansson, H., Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. Bacteroides fragilis synthesizes a DNA invertase affecting both a local and a distant region. *J. Bacteriol.* **189**, 2119–2124 (2007).
32. Comstock, L. E. Lessons from the model gut Bacteroidales Bacteroides fragilis and Bacteroides thetaiotaomicron and future opportunities. *J. Bacteriol.* **207**, e00346-25 (2025).
33. Chanin, R. B. *et al.* Intragenic DNA inversions expand bacterial coding capacity. *Nature* **634**, 234–242 (2024).
34. Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**, 2074–2076 (2003).
35. Bolam, D. N. & van den Berg, B. TonB-dependent transport by the gut microbiota: novel aspects of an old problem. *Curr. Opin. Struct. Biol.* **51**, 35–43 (2018).
36. Pollet, R. M., Martin, L. M. & Koropatkin, N. M. TonB-dependent transporters in the Bacteroidetes: Unique domain structures and potential functions. *Mol. Microbiol.* **115**, 490–501 (2021).
37. Gray, D. A. *et al.* Insights into SusCD-mediated glycan import by a prominent gut symbiont. *Nat. Commun.* **12**, p. 12 (2021).
38. White, J. B. R. *et al.* Outer membrane utilisomes mediate glycan uptake in gut Bacteroidetes. *Nature* **618**, 583–589 (2023).
39. Braun, V. Substrate Uptake by TonB-Dependent Outer Membrane Transporters. *Mol. Microbiol.* **122**, 929–947 (2024).
40. Zioutis, C. *et al.* Rapid genetic diversification of Bacteroides thetaiotaomicron in mono-associated mice revealed through deep population-level sequencing. Preprint at *biorXiv* https://doi.org/10.1101/2025.06.24.661302 (2025).
41. Munshi, I. D. *et al.* Emerging concept of genomic islands in bacterial adaptation and pathogenicity. *Res. Microbiol.* **176**, 104303 (2025).
42. Balleza, E. *et al.* Regulation by transcription factors in bacteria: Beyond description. *FEMS Microbiol. Rev.* **33**, 133–151 (2009).
43. Bayley, D. P., Rocha, E. R. & Smith, C. J. Analysis of cepA and other Bacteroides fragilis genes reveals a unique promoter structure. *FEMS Microbiol. Lett.* **193**, 149–154 (2000).
44. Kuwahara, T. *et al.* Genomic analysis of Bacteroides fragilis reveals extensive DNA inversions regulating cell surface adaptation. *PNAS*, **44**, 14919-14924 (2004).
45. Tong, M. *et al.* A highly conserved SusCD transporter determines the import and species-specific antagonism of Bacteroides ubiquitin homologues. *Nature Communications* **15**, 8794 (2024).
46. Noinaj, N., Guillier, M., Barnard, T. J. & Buchanan, S. K. TonB-dependent transporters: Regulation, structure, and function. *Annu. Rev. Microbiol.* **64**, 43–60 (2010).
47. Darmon, E. & Leach, D. R. F. Bacterial Genome Instability. *Microbiology and Molecular Biology Reviews* **78**, 1–39 (2014).

48. Peng, H. *et al.* Longitudinal gut microbiota tracking reveals the dynamics of horizontal gene transfer. *Nat. Commun.* (2025).
49. Rinninella, E. *et al.* What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* **7**, 14 (2019).
50. Groussin, M. *et al.* Elevated rates of horizontal gene transfer in the industrialized human microbiome. *Cell* **184**, 2053-2067.e18 (2021).
51. Yin, Z. *et al.* Pan-genome insights into adaptive evolution of bacterial symbionts in mixed host-microbe symbioses represented by human gut microbiota Bacteroides cellulosilyticus. *Science of the Total Environment* **927**, (2024).
52. Da Silva Morais, E., Grimaud, G. M., Warda, A., Stanton, C. & Ross, P. Genome plasticity shapes the ecology and evolution of Phocaeicola dorei and Phocaeicola vulgatus. *Sci. Rep.* **14**, 10109 (2024).
53. Goldfarb, T. *et al.* NCBI RefSeq: Reference sequence standards through 25 years of curation and annotation. *Nucleic Acids Res.* **53**, D243–D257 (2025).
54. Larsson, A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
55. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
56. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
57. Li, W. & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
58. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
59. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **32**, e4792 (2023).
60. Näsvall, J., Knöppel, A. & Andersson, D. I. Duplication-Insertion Recombineering: A fast and scar-free method for efficient transfer of multiple mutations in bacteria. *Nucleic Acids Res.* **45**, (2017).
61. Lutz, R. & Bujard, H. Independent and Tight Regulation of Transcriptional Units in Escherichia Coli via the LacR/O, the TetR/O and AraC/I 1-I 2 Regulatory Elements. **25**, 1203-1210 (1997).
62. Padmanabhan, S., Banerjee, S. & Mandi, N. Screening of Bacterial Recombinants: Strategies and Preventing False Positives. *Molecular Cloning - Selected Applications in Medicine and Biology* p.1-19 (2011).

## Methods

### Computational genomics

#### Genome collection and loci annotation

For identification of *tsr-susCD* homologous regions across Bacteroidaceae, we downloaded complete genomes available in the RefSeq database[53] as of March 2025. Homologs of the *trs0667* invertase were found using BLAST search with the WP_011108198.145 protein from *Bacteroides fragilis* YCH46 as a query (Supplementary File Table S1). Prediction of SusC-like TonB-dependent transporters was done using a Hidden Markov Model (HMM) constructed for orthogroup #4NDXS available in eggnog database v.52 (Supplementary File Table S2). A cassette of the transporters affected by Tsr-moderated phase variation was preliminarily identified as the ones encoded in a locus with *tsr* invertase homologs between the *hflX* GTPase and *yjjG* dUMP-nucleotidase genes and then revised by the identification of Tsr BS motifs (Supplementary File Table S3).

#### Comparative genomics and phylogenetics

Sequence alignments were performed with the muscle algorithm implemented in AliView software[54]. The logos were constructed using WebLogo v.3.[55] Phylogenetic trees were constructed using iqtree v.2[56]. To identify recent horizontal gene transfer events among Tsr-mediated transporters, we clustered them using Cd-hit v.4.8.1[57] with 70% identity as a threshold.

#### Protein structure

The sequences of *susCD* genes under locus tags BT_RS11460 and BT_RS11465 in *B. theta* VPI 5482 (assembly id GCF_000011065.1) were used for illustration of the SusCD protein complex structure. AlphaFold[58] was used to predict the structure (pTM = 0.9), and ChimeraX v. 1.10.1[59] was used for its visualization.

### Bacterial strains and growth conditions

#### *Bacteroides* spp

*Bacteroides thetaiotaomicron* VPI 5482, *Bacteroides uniformis* ATCC 8492, and *Phocaeicola vulgatus* ATCC 8482 cultures were obtained from the Gut Bacteria Collection, University of Vienna. Brain Heart Infusion (BHI) medium supplemented with yeast extract (5 g/L), hemin (5 mg/L), vitamin K1 (0.20 mg/L), L-cysteine (1 g/L), and sodium bicarbonate (1 M) was used for bacterial cultivation. Cultures were incubated overnight at 37°C in an anaerobic chamber with an atmosphere of 10% $CO_2$, 5% $H_2$, and 85% $N_2$. Bacterial genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, Cat. No. 51304) following the manufacturer's instructions.

The target *susCD* genes from each bacterial species were amplified by PCR using Dream-Taq PCR MM 2x (K1082). From each PCR product, 5 µL was loaded onto a 1% agarose gel and visualized by electrophoresis with GelRed staining, using a 1 kb DNA ladder (Thermo Fisher; SM0313) as a size reference. Following confirmation of the desired bands, PCR products were

purified using the innuPREP PCRpure Kit (IST Innuscreen, Analytik Jena; Cat. No. 845-KS-5010250) for downstream analysis.

**Long-read sequencing**

High molecular weight DNA was extracted using the Monarch HMW DNA extraction kit for tissue (NEB). Pellets were initially treated with an enzymatic cocktail (innuPREP Bacteria Lysis Booster, Innuscreen GmbH) to facilitate chemical cell wall lysis. Extracts were left at room temperature for three days before quantification and fragment estimation. Fragments below 10 kb were removed with the Short Read Eliminator kit (PacBio). Each sample was diluted and equimolarly barcoded five times using the SQK-NBD114.24 (Oxford Nanopore Technologies). About 100 fmol of a 53 kb library was loaded on a R10.4.1 flowcell (FLO-MIN114, Oxford Nanopore Technologies) and sequenced for 72 h on a MinION MK1B (Oxford Nanopore Technologies) using Minknow (v. 25.05.14, Oxford Nanopore Technologies). Reads were basecalled using dorado basecaller integrated in Minknow using super accuracy mode (model: r1041_e82_400bps_sup_v5.0.0). Reads are provided in the NCBI repository under SRA ID PRJNA1414423.

***Escherichia coli* experiments**

The experimental phase variable IRX system was constructed in *E. coli* K-12 MG1655 (DA24100, obtained from Dan I. Andersson, Uppsala). To generate IR, we used AcatsacA DIRex recombineering[60] to insert two inverted repeats (IR) on either side of the *lacZ* promoter region. The utilized oligo sequences to generate the DIRex half-cassettes can be found in Supplementary Table S4. The inserted inverted repeat sequence *'AGTTCTGCAAAGACTTTGCAGA'* is identical to that of *Bacteroides fragilis* (locus BF0388 – BF0389). IR left (IRL) was introduced between *lacZ* and its promoter; IR right (IRR) was introduced between *mhpR* and the *mhpR* promoter. Therefore, the inversion produced by the natively associated *Tsr0667* invertase shuffles the orientation of the *lac* promoter, allowing transitions between Lac+ and Lac– phenotypes. The finalized construct was back-transduced into the wild type using P1*vir* phage in order to remove potential background mutations. Spontaneous excision of the AcatsacA DIRex cassette was selected on 5% sucrose. Scarless insertion of IRL and IRR was confirmed in pure-streaked clones using Sanger sequencing. IRX is the *E. coli* strain IR transformed with the plasmid pInvertase_X. IRY is the *E. coli* strain IR transformed with the plasmid pInvertase_Y.

Cells were grown overnight, shaking (250 rpm) in lysogeny broth (LB) (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl; supplemented with 1.5% agar for plates, pH 7,0) at 37 °C. Growth media were supplemented with appropriate antibiotics when required, as follows: kanamycin (50 μg/mL), and ampicillin (100 μg/mL).

**Plasmid construction**

To construct the plasmid pInvertase, plasmid pZS*21 was used as a backbone (*kanR* marker, very low-copy plasmid)[61]. The *tsr* invertase gene was amplified by PCR and introduced into the plasmid under a TetR inducible promoter using Gibson Assembly. Using this method, two plasmids were

constructed: pInvertase_X (*Bacteroides fragilis* NCTC 9343, WP_011108198) and pInvertase_Y (*B. uniformis* ATCC 8492, WP_005824324.1). Primers used are listed in Supplementary Table S4.

**Verification of DNA inversion**

Single colonies of strains IR and IRX were inoculated. The resulting ON cultures were diluted to an OD of 0.04 in 3 ml of LB medium. The invertase was induced for the indicated time by induction using 20 ng/ml anhydrotetracycline (aTc). The culture was spun down, and the DNA was isolated immediately using the Wizard Genomic DNA Purification Kit by Promega.

gDNA was measured using Qbit, and diluted to an initial concentration of 220 ng/µL. PCR was performed on serial 10-fold dilutions of each of the 220ng gDNAs. PCRs were performed for both $Lac^+$ and $Lac^-$ genotypes and the *acrR* gene. *AcrR* was chosen as a control for the PCR products as a constitutively expressed gene in relatively close genomic proximity to the *lac* operon. For the PCRs, OneTaq (Quick-Load 2xMM; M0486) and Phusion (HS Flex 2x MM; M0536) were used. 5 µL of each PCR product was loaded onto the gel. The individual resulting PCR products were visualized by gel electrophoresis using SYBR Safe and 1 kb Plus DNA ladder (Neb; N0550). For PCR cleanup, the Monarch PCR&DNA Cleanup Kit was used (T1030L). Primers can be found in Supplementary Table S4. PCR products were measured using Qbit to double-check which dilution still had a PCR product. PCR products were sequenced to verify the specificity of the products.

**Blue-White screening**

Blue-white screening was used as described in Padmanabhan *et al* (2011)[62]. In brief, overnight cultures of IR and IRX were diluted to OD of 0.04. Induction of the invertase with a concentration of 20 ng/ml aTc was stopped at the indicated time points. Cultures were spun down at 1000 rpm for 1.5 min in a tabletop centrifuge, resuspended in 1x PBS and further diluted in PBS to achieve about 20-200 colonies per plate. Selection plates were LB agar supplemented with 50 µg/ml kanamycin and 40 µg/mL Xgal, pH 7.0. Experiments were done in three biological replicates on three different days and plated in four technical replicates for each time point of induction (min: 0, 3, 5, 30, 60, 180, 1080).

**Acknowledgements**

The study was funded by Fonds zur Förderung der Wissenschaftlichen Forschung (FWF), Grant ESP 253-B (OOB) and FWF Grant I5127 (CCG). We thank the staff of the Joint Microbiome Facility, CeMESS, and especially Sara Malinowski and Joana Seneca, for ONT sequencing; Dmitry Solovey for help with 3D structure modelling tools; Mikhail Gelfand, Aleksandra Galitsyna, and Sofya Garushyants for proofreading the manuscript.

**Author contributions**

OOB conceived the study; DB and CCG provided the resources; OOB, NGG, CCG provided supervision, NOD, ESK, and OOB analyzed genomic data; BH analyzed the sequencing data; MHW, RR, NGG and OOB designed experiments; RR and NGG created the mutants; NGG, RR, BH, and SK conducted experiments; OBB, FAK, NGG, DB, and CCG interpreted the results; FAK, CCG, OBB and NGG wrote the draft; OBB, NOD, and NGG designed the figures. All authors proofread and approved the manuscript.

**Ethics declarations/Competing interests**

The authors declare no competing interests.

**Supplementary tables**

Table S1. Tsr0667 invertase orthologs and their predicted binding sites in Bacteroidaceae complete genomes.

Table S2. Annotation of *susC* family genes in Bacteroidaceae

Table S3. Composition of *tsr-susCD* shufflon system loci in Bacteroidaceae

Table S4. Primers

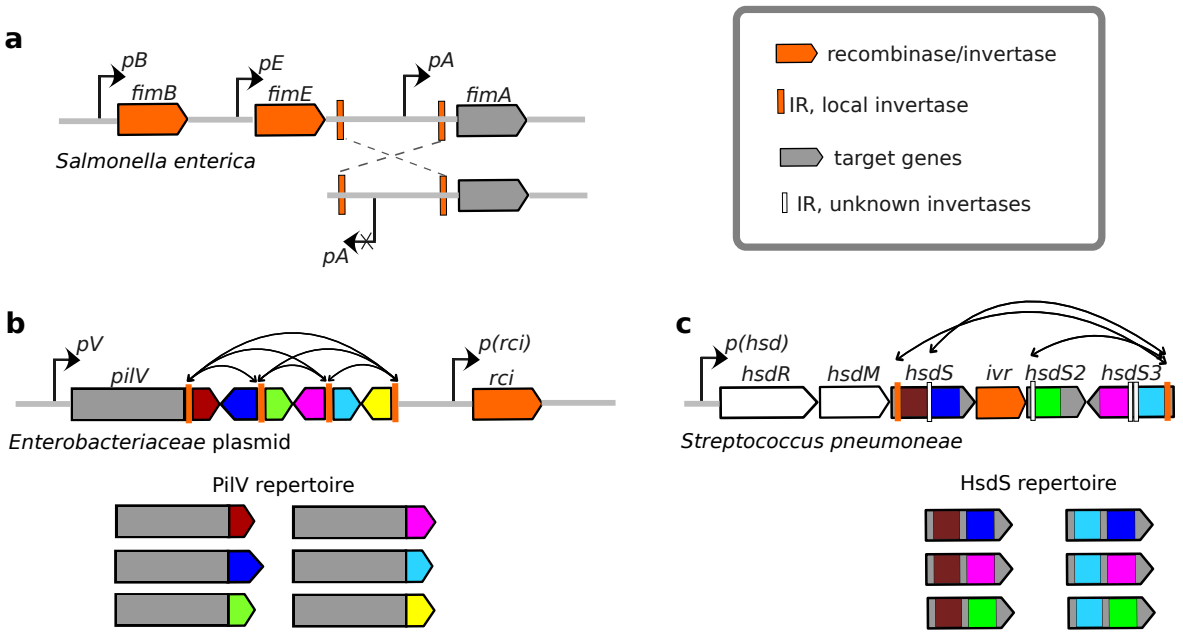
**Extended data figures**

Figure S1. An illustrative example of the alignment of the *susC* proteins encoded in the *tsr-susCD* shufflons.

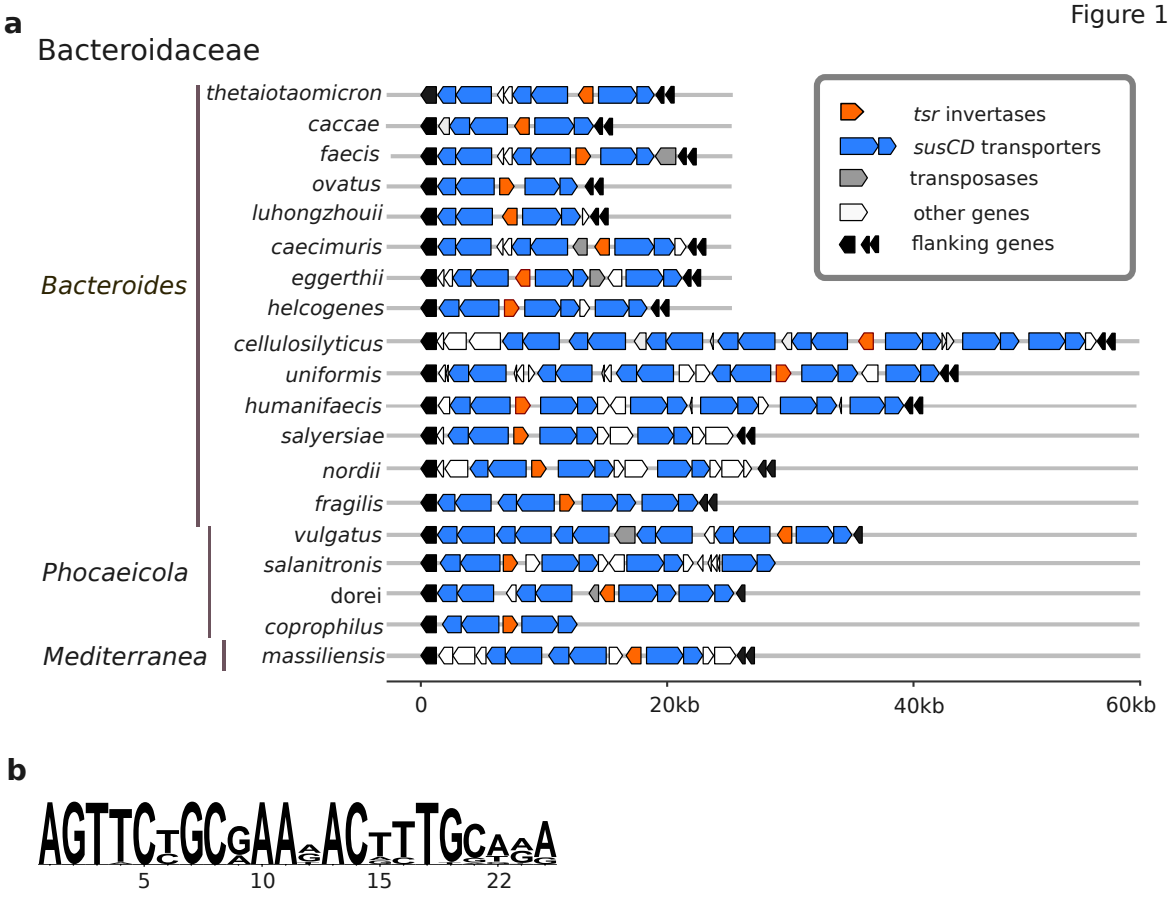
Figure S2. Invertase activity in *E. coli.*

Figure S3. Logo of the intergenic regions between the *tsr* gene and the expressed *susCD* transporter gene with functional annotation of regulatory elements.

Figure S4. Comparison of the *Bacteroides* phylogenetic trees.




**Box. Examples of phase variable loci mediated by recombinases. (a)** Site-specific inversion of the DNA segment containing the promoter regulates the expression of pili in *Salmonella*. In the "on" orientation, the promoter allows *fimA* expression; inversion to the "off" orientation, *fimA* is not expressed. Inversions are produced by FimB and FimE recombinases (invertases). **(b)** Shufflons in the enterobacterial incompatibility plasmid R64 generate diversity in *pilV* gene expression. This system produces PilV protein variants with a constant N-terminal region, but variable C-terminal segments. These variants alter a key component of the type IV pilus, determining the range of recipient bacterial cells to which the plasmid can successfully transfer genetic material during conjugation. **(c)** Shufflons in Type 1 restriction-modification systems alter methylation specificity. These systems consist of three host specificity determinants genes: *hsdR, hsdM,* and *hsdS*. Each of these genes produces a subunit of the complete enzyme, known as the restriction, modification, and specificity subunits, respectively. The HsdS protein functions as a DNA-binding specificity subunit containing two target recognition domains, each recognizing one half-site of the bipartite recognition sequence. This locus in *Streptococcus pneumoniae* includes one expressed *hsdS* gene and two silent *hsdS* genes truncated by an invertase. Three series of IRs (only one of them recognized by the invertase in the locus) allow the formation of six different HsdS alleles.




620

**Figure 1. Computational prediction of the shufflon system in Bacteroidaceae. (a)** Composition of the Tsr0667 homologs loci in a subset of representative genomes. For full annotation, see Supplementary Table S1. The shufflon system consists of a *tsr* invertase gene and a variable number of Tsr-mediated *susCD* transporter genes flanked by *hflX* and *yjjG/marC* genes. The conservation of genetic location, along with the consistency of species phylogeny (Supplementary Figure 4), indicates vertical inheritance of the locus. **(b)** Logo of Tsr binding sites found in Tsr-mediated *susCD* transporter genes in Bacteroidaceae.

Figure 2

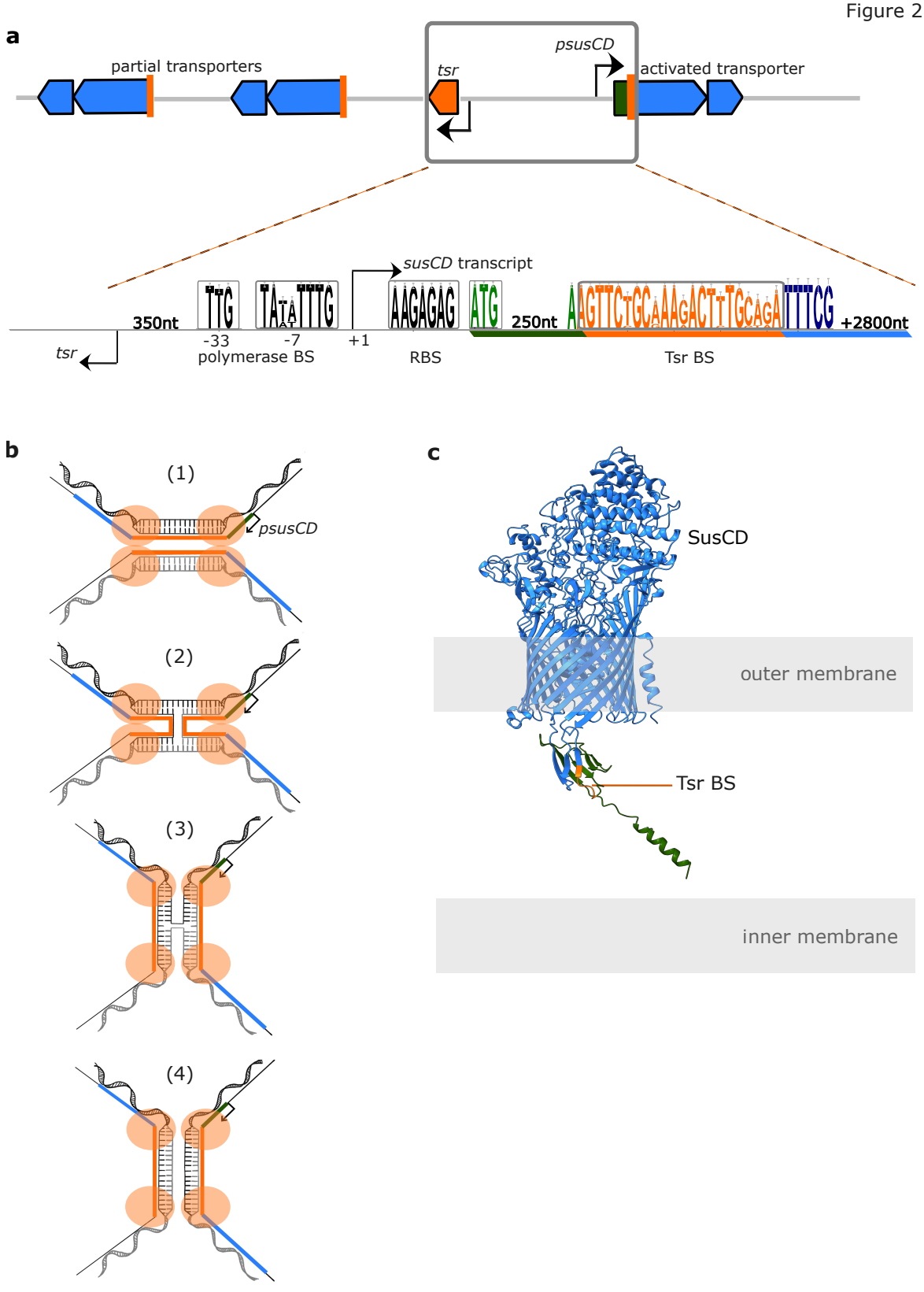

a
partial transporters
tsr
psusCD
activated transporter
susCD transcript
350nt
-33
-7
+1
250nt
+2800nt
tsr
polymerase BS
RBS
Tsr BS
b
(1)
psusCD
(2)
(3)
(4)
c
SusCD
outer membrane
Tsr BS
inner membrane

**Figure 2. Alteration of *susCD* transporter expression and its product composition via DNA shuffling. (a)** The role of the Tsr invertase in the regulation of transporter gene expression. The *tsr* invertase gene is inherited in genetic linkage with the DNA fragment comprising the promoter and the 5'-end of the *susC* gene fragment flanked by Tsr BS. Within the locus, only one Tsr-mediated transporter gene, which is located immediately upstream of the *tsr* invertase, is a full-length gene sequence. All other Tsr-mediated transporter genes are incomplete sequences, starting from the Tsr BS. **(b)** The molecular mechanism of site-specific recombination. Through DNA cleavage and subsequent ligation catalyzed by the Tsr invertase, diverse partial genes may substitute a fragment of the full-length transport gene. **(c)** 3-D structure of SusCD transporter complex (*B. theta* VPI 5482, assembly id GCF_000011065.1, gene locus_tags BT_RS11460 and BT_RS11465) and its localization in the bacterial envelope. Colors indicate the protein domains, which are encoded by the unique 5' end fragment (green), Tsr BS motif (orange), and multiple recombined fragments (blue).

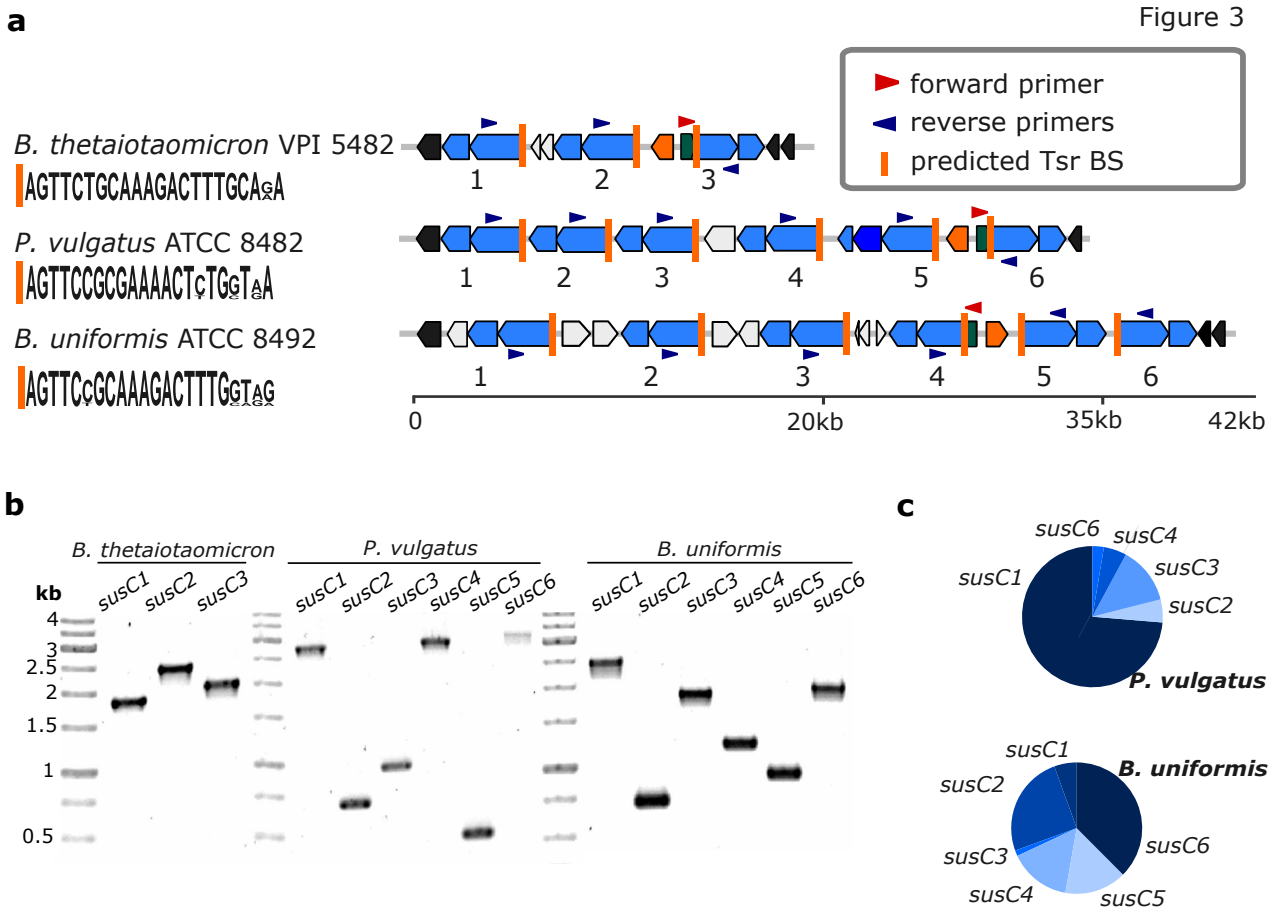



**Figure 3. Tsr-mediated DNA shuffling causes the presence of multiple genotypes in Bacteroidaceae populations. (a)** Composition of the *tsr* locus in the strains selected for experimental verification; color scheme as in Figure 1a. For each strain, a unique forward primer was designed to bind the 5' end of the *susC* gene encoded downstream of the invertase gene (red arrow) in combination with a set of reverse primers corresponding to the array of transporter genes in a locus (blue arrows). **(b)** All strains display the bands corresponding to all possible combinations of forward and reverse primers, confirming the computational prediction that shuffling activity is maintained across members of the Bacteroidaceae. **(c)** Number of reads corresponding to the junction of a unique 5'end fragment with different Tsr-mediated incomplete transporters.

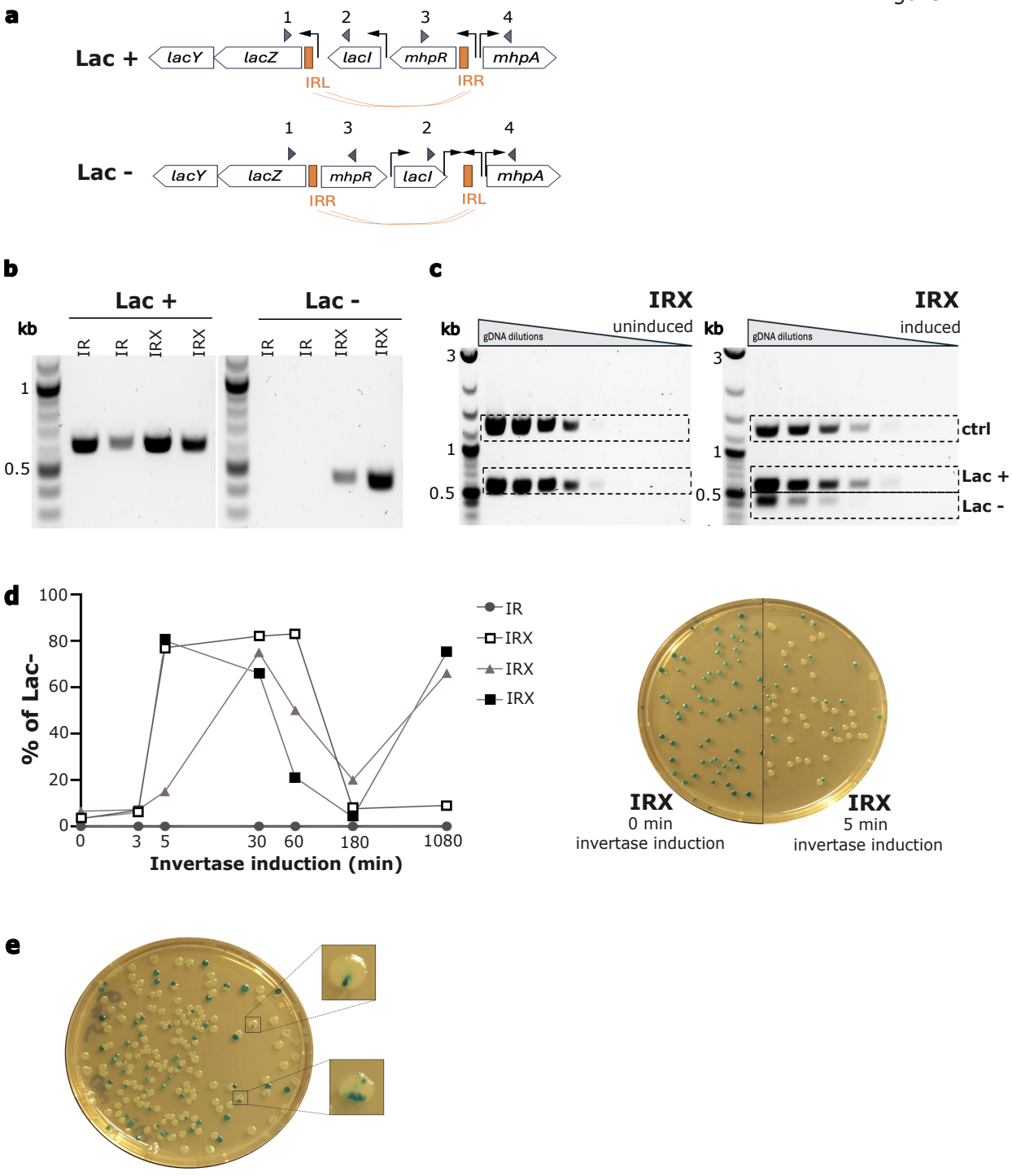
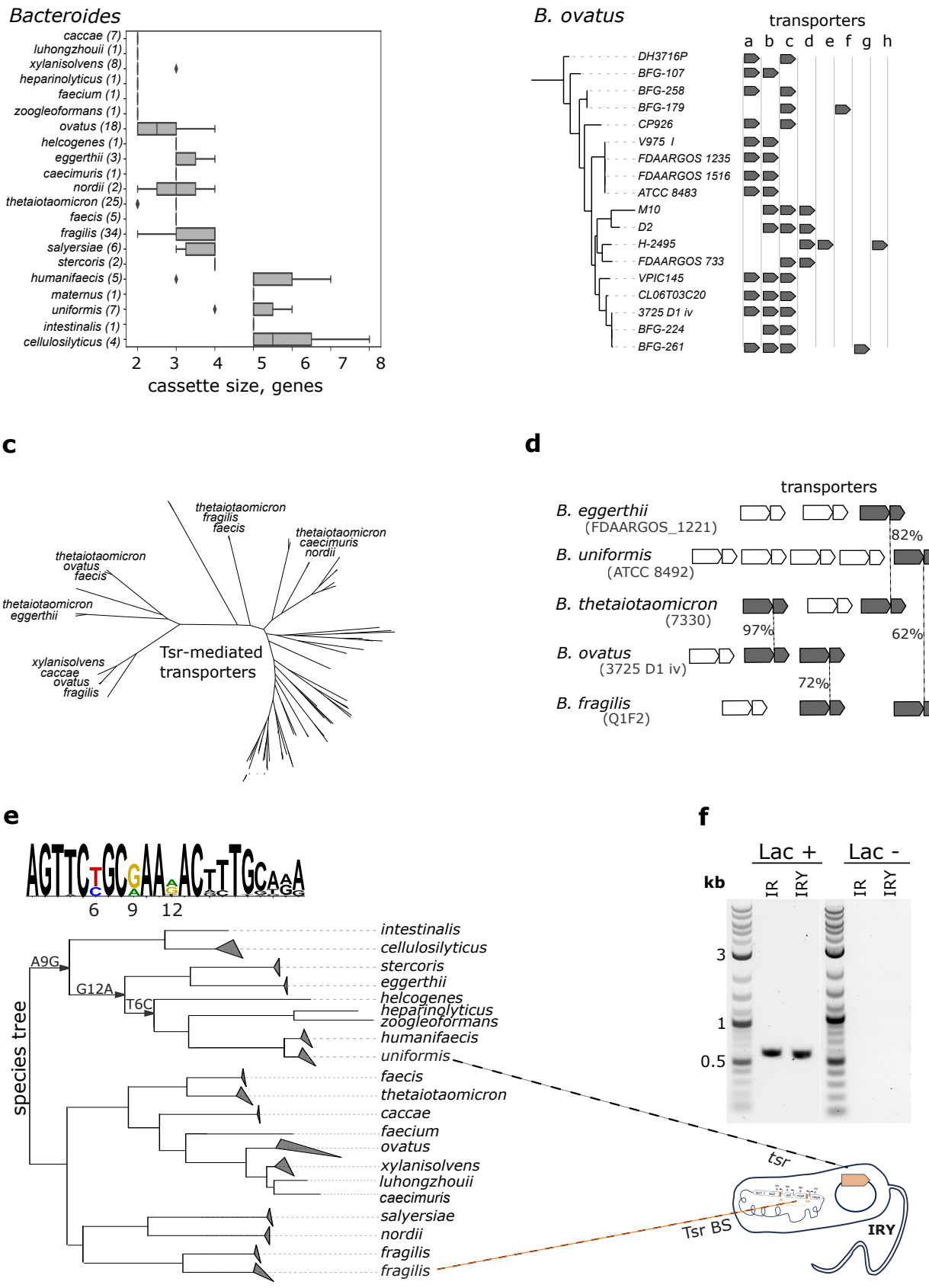



**Figure 4. Reconstitution of the Tsr-mediated shuffling in *Escherichia coli* leads to heterogeneous genotypes. (a)** Schematic of chromosomal integration within the *lac* operon of *E. coli*. Tsr BS are depicted in orange: inverted repeat left (IRL) and right (IRR). In the original orientation, the *lacZ* promoter is active (Lac$^+$). Recombination between IRs inverts the *lacZ* promoter, yielding a Lac$^-$ genotype. Genotypes can be resolved by PCR using primers depicted as grey triangles. Promoter orientations are indicated as black arrows. The *E. coli* strain, IR, carries chromosomally

integrated inverted repeats; IRX additionally expresses an inducible Tsr invertase on a low-copy plasmid. **(b)** PCR of IR strains yields only $Lac^+$ bands, while IRX displays both $Lac^+$ and $Lac^-$ after invertase induction, indicating a mixed genotype. (Two biological replicates are shown for each strain) **(c)** PCR on serial dilutions of genomic DNA from uninduced and five-minute-induced IRX. Only $Lac^+$ is detected without induction; both $Lac^+$ and $Lac^-$ appear post-induction, with $Lac^+$ DNA roughly two-fold as abundant. *AcrR* serves as a transcriptional control. **(d)** Time course of $Lac^-$ emergence during invertase induction quantified via blue-white screening ($Lac^+$ = blue, $Lac^-$ = white). Plate images show IRX before and after five minutes of induction. The experiment was done in triplicate. **(e)** IRX colonies display mixed (blue and white) phenotypes within single colonies after induction for more than three hours.

Figure 5


a
Bacteroides
caccae (7)
luhongzhouii (1)
xylanisolvens (8)
heparinolyticus (1)
faecium (1)
zoogleoformans (1)
ovatus (18)
helcogenes (1)
eggerthii (3)
caecimuris (1)
nordii (2)
thetaiotaomicron (25)
faecis (5)
fragilis (34)
salyersiae (6)
stercoris (2)
humanifaecis (5)
maternus (1)
uniformis (7)
intestinalis (1)
cellulosilyticus (4)
2 3 4 5 6 7 8
cassette size, genes
b
B. ovatus
transporters
a b c d e f g h
DH3716P
BFG-107
BFG-258
BFG-179
CP926
V975 I
FDAARGOS 1235
FDAARGOS 1516
ATCC 8483
M10
D2
H-2495
FDAARGOS 733
VPIC145
CL06T03C20
3725 D1 iv
BFG-224
BFG-261
c
thetaiotaomicron fragilis faecis
thetaiotaomicron caecimuris nordii
thetaiotaomicron ovatus faecis
thetaiotaomicron eggerthii
xylanisolvens caccae ovatus fragilis
Tsr-mediated transporters
d
transporters
B. eggerthii (FDAARGOS_1221)
B. uniformis (ATCC 8492)
B. thetaiotaomicron (7330)
B. ovatus (3725 D1 iv)
B. fragilis (Q1F2)
82%
97%
62%
72%
e
6 9 12
A9G
G12A
T6C
species tree
intestinalis
cellulosilyticus
stercoris
eggerthii
helcogenes
heparinolyticus
zoogleoformans
humanifaecis
uniformis
faecis
thetaiotaomicron
caccae
faecium
ovatus
xylanisolvens
luhongzhouii
caecimuris
salyersiae
nordii
fragilis
fragilis
tsr
Tsr BS
IRY
f
Lac +
Lac -
kb
IR
IRY
IR
IRY
3
1
0.5

662

**Figure 5. Evolution and diversification of the Tsr-mediated *susCD* shufflon system. (a)** The number of Tsr-mediated transporters in jukebox cassettes in *Bacteroides*. The number of available complete genome assemblies is indicated in brackets, the bars show the variation and the mean value for each species. **(b)** Composition (phyletic pattern) of Tsr-mediated transporter genes in *B. ovatus* strains. **(c)** Phylogenetic tree of Tsr-mediated transporters. Tree structure indicates high evolutionary divergence of Tsr-mediated *susCD* transporters and their transmission across *Bacteroides* via HGT. **(d)** Pairwise comparison of the transporters in the Tsr-mediated cassettes of five selected strains of different species. Pairs with >60% amino acid identity are indicated by dotted lines. **(e)** Positions in Tsr BS with clade-specific mutations and their corresponding phylogenetic nodes on the *Bacteroides* species tree. **(f)** Elucidation of Tsr-Tsr BS compatibility using *E. coli* mutants (IRY) that contained Tsr BS from *B. fragilis* cassette genes and the *tsr* invertase gene from *B. uniformis*. IRY cultures maintained only the original $Lac^+$ genotype, indicating that inversions occur at an undetectably low rate.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TS1recombinasesandIRs.xlsx
- ST2annotationsusCgenes.xlsx
- ST3PVlocicomposition.xlsx
- ST4Primers.xlsx
- Extendeddatafiguresfeb272026.pdf