AMERICAN SOCIETY FOR MICROBIOLOGY | mBio

Check for updates

| Bacteriology | Research Article

Linear dicentric chromosomes in bacterial natural isolates reveal common constraints for replicon fusion

Ram Sanath-Kumar,[1] Arafat Rahman,[2] Zhongqing Ren,[1] Ian P. Reynolds,[1] Lauren Augusta,[1] Clay Fuqua,[1] Alexandra J. Weisberg,[2] Xindan Wang[1]

AUTHOR AFFILIATIONS See affiliation list on p. 10.

ABSTRACT Multipartite bacterial genome organization can confer advantages, including coordinated gene regulation and faster genome replication, but is challenging to maintain. *Agrobacterium tumefaciens* lineages often contain a circular chromosome (Ch1), a linear chromosome (Ch2), and multiple plasmids. We previously observed that in some stocks of the C58 lab model, Ch1 and Ch2 were fused into a linear dicentric chromosome. Here we analyzed *Agrobacterium* natural isolates from the French Collection for Plant-Associated Bacteria and identified two strains distinct from C58 with fused chromosomes. Chromosome conformation capture identified integration junctions that were different from the C58 fusion strain. Genome-wide DNA replication profiling showed that both replication origins remained active. Transposon sequencing revealed that partitioning systems of both chromosome centromeres were essential. Importantly, the site-specific recombinase XerCD is required for the survival of the strains containing the fusion chromosome. Our findings show that replicon fusion occurs in natural environments and that balanced replication arm sizes and proper resolution systems enable the survival of such strains.

IMPORTANCE Most bacterial genomes are monopartite with a single, circular chromosome. However, some species, like *Agrobacterium tumefaciens*, carry multiple chromosomes. Emergence of multipartite genomes is often related to adaptation to specific niches, including pathogenesis or symbiosis. Multipartite genomes confer certain advantages; however, maintaining this complex structure can present significant challenges. We previously reported a laboratory-propagated lineage of *A. tumefaciens* strain C58 in which the circular and linear chromosomes fused to form a single dicentric chromosome. Here we discovered two geographically separated environmental isolates of *A. tumefaciens* containing fused chromosomes with integration junctions different from the C58 fusion chromosome, revealing the constraints and diversification of this process. We found that balanced replication arm sizes and the repurposing of multimer resolution systems enable the survival and stable maintenance of dicentric chromosomes. These findings reveal how multipartite genomes function across different bacterial species and the role of genomic plasticity in bacterial genetic diversification.

KEYWORDS *Agrobacterium tumefaciens*, chromosome fusion, multipartite genome, Hi-C, natural isolates

Most bacteria have a single chromosome. However, approximately 10% of bacteria contain other chromosome-type replicons in addition to the main chromosome (1). These multipartite genomes are common in bacteria involved in pathogenesis and symbiosis (2–9). Evolutionary drivers for the emergence of these complex genomes include adaptation to different niches, as seen in the nitrogen-fixing symbiont *Sinorhi-*

Editor Arash Komeili, University of California Berkeley, Berkeley, California, USA

Address correspondence to Alexandra J. Weisberg, alexandra.weisberg@oregonstate.edu, or Xindan Wang, xindan@iu.edu.

Ram Sanath-Kumar and Arafat Rahman contributed equally to this article. These authors are listed in reverse alphabetical order.

The authors declare no conflict of interest.

See the funding table on p. 11.

Received 9 April 2025
Accepted 22 April 2025
Published 20 May 2025

zobium meliloti, human pathogens including *Vibrio cholerae* and *Brucella melitensis*, and plant pathogens like *Agrobacterium tumefaciens* (4–9).

Most multipartite genomes are stably maintained, although there are examples of bacterial taxa that undergo frequent replicon fusion and splitting. For instance, *Rhizobium* sp. strain NGR234, which has three replicons, was reported to have spontaneous replicon fusion and splitting at ~10^{-3} frequency (10). Similarly, researchers have generated *S. meliloti* strains in which the tripartite genome was successively fused into two replicons, then into a single replicon (11). The frequency of each fusion step was about 6×10^{-3}. Although the fusion strains can be maintained under laboratory conditions, the fusion replicons split at a frequency of ~6×10^{-3}. Interestingly, the preferential site for splitting was the same as the site at which fusion occurred (11). Thus, the tripartite genome of *S. meliloti* is maintained even under a highly dynamic state (11). These studies signify both the stability and the flexibility of multipartite genomes.

A specific subset of *A. tumefaciens* lineages has a complex genome consisting of a circular chromosome (Ch1), a linear chromosome (Ch2), and two or more large plasmids (pAt, pTi, and others) (8, 12). We previously found that, surprisingly, particular stocks of the commonly used lab strain *A. tumefaciens* C58 have undergone fusion of Ch1 and Ch2 into a stable linear dicentric chromosome over the course of laboratory cultivation (13). Here, we have identified two environmental *A. tumefaciens* isolates containing a fused linear chromosome. We analyzed their integration sites and examined genetic requirements for their survival. Comparing these natural isolates with our previously identified C58 fusion strain provides insights into the plasticity of and requirements for maintaining fusion chromosomes.

RESULTS

Two natural isolates of *A. tumefaciens* contain a fused chromosome

To explore whether chromosome fusion naturally occurs in environmental isolates, we performed long-read nanopore whole-genome sequencing for over 400 *Agrobacterium* strains, including several from the International Centre for Microbial Resources-French Collection for Plant-Associated Bacteria (CFBP). After assembling the sequencing reads into complete genomes, we searched for strains with potential fusion chromosomes. We identified two strains that were isolated from geographically separated grape-growing localities, each having a single large linear chromosome in place of binary replicons: CFBP_2407 was isolated in Hérault, France, in 1982; and CFBP_2642 was isolated in Gironde, France, in 1985 (14). The two locations were separated by about 450 km (~280 miles). This genomic configuration was reminiscent of the chromosomal fusion in C58 (13). To find the closest relative of these strains with a binary chromosome structure, we constructed a core genome phylogeny made from 4,192 core genes extracted from publicly available genomes. In this phylogeny, both CFBP_2407 and CFBP_2642 are most closely related to each other, with the closest sister clade containing strains 47-2, IL15, and 1D1418 (Fig. 1A). The three strains of this sister clade are all predicted to have typical bipartite genomes based on their long-read assemblies. We chose strain 47-2, isolated from a grapevine gall in Israel, as a comparator for CFBP_2407 and CFBP_2642. Contig synteny analysis identified conserved regions and revealed a high degree of similarity (>99.3% average nucleotide identity [ANI]) between 47-2 and the fusion chromosome strains (Fig. 1B).

Identifying the sites of chromosome fusion

To further investigate the genome architecture of these three strains and rule out genome misassembly, we performed genome-wide chromosome conformation capture (Hi-C) analysis. The 47-2 strain has a binary chromosome structure, as indicated by clear breakpoints between Ch1 (2,890 kb) and Ch2 (2,410 kb) (Fig. 2A). In contrast, both CFBP_2407 and CFBP_2642 exhibited a continuous genomic map of Hi-C data without any breakpoint, indicating the presence of a single, fused chromosome of ~5,361 kb (Fig.

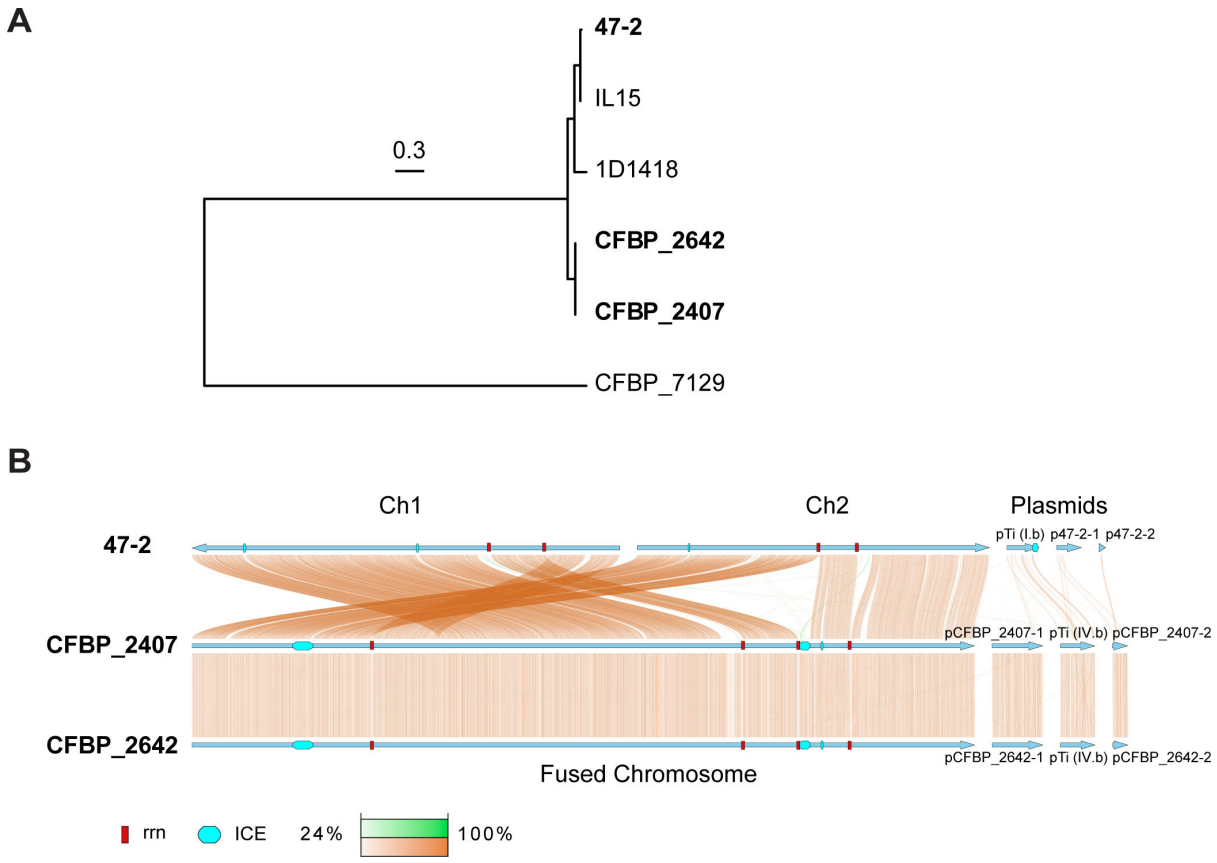

FIG 1 Phylogenetic analysis of natural *A. tumefaciens* isolates. (A) Maximum likelihood phylogeny based on 4,192 genes conserved in the six analyzed strains. The tree is midpoint rooted. Branches with ultrafast bootstrap >95% and SH-aLRT >80% are black; branches with other support are gray. Branch scale is indicated by a bar. (B) Genome synteny between 47-2, CFBP_2407, and CFBP_2642. Blue arrows indicate individual replicons and their direction in the assembly. Colored bars linking replicons of different strains indicate regions of similarity. Copper links indicate the same orientation; green links indicate inversions. Darker colors indicate greater sequence similarity. Red boxes indicate the position of *rrn* regions. Cyan rounded rectangles indicate the position of integrative and conjugative elements (ICEs).

2B and C). Consistent with our nanopore long-read assemblies, strains CFBP_2407 and CFBP_2642 lacked interactions at the termini of the chromosome, indicating a linear chromosome as opposed to the circular Ch1 in strain 47-2 (red circles on the maps of Fig. 2B and C).

To identify the predicted chromosome fusion sites, we aligned the genome assemblies of the two fusion strains to that of the binary strain 47-2 and examined breaks in the alignment (Fig. 1B). We identified putative breakpoints at 2,156 kb on Ch1 and 1,161 kb on Ch2, where the fusion likely occurred. To further corroborate these predicted chromosome fusion sites, we mapped the Hi-C reads of CFBP_2407 and CFBP_2642 to the binary control strain 47-2. As expected, this comparison revealed a shuffled Hi-C map, indicating substantial genome reorganization in the fusion isolates compared with the binary control (Fig. 3). The shuffled Hi-C maps revealed two breakpoints consistent with the synteny analysis of the genome assemblies. Notably, both breakpoints were located within two ~6 kb regions containing identical ribosomal RNA (*rrs*, *rrl*, and *rrf*) and tRNA genes (*trnI*, *trnA*, and *trnfM*) (Fig. 4A), indicating that the fusion event occurred by recombination of homologous regions on different chromosomes. Importantly, both CFBP_2407 and CFBP_2642 contain a 77 kb region that resembles an integrative and conjugative element (ICE) inserted immediately adjacent to the fusion junction (Fig. 4A). We screened all 3,222 new and publicly available genome sequences of members of the agrobacteria/rhizobia complex for other strains carrying the predicted ICE. However, no additional strains were found to carry a similar ICE sequence. Notably, the fusion

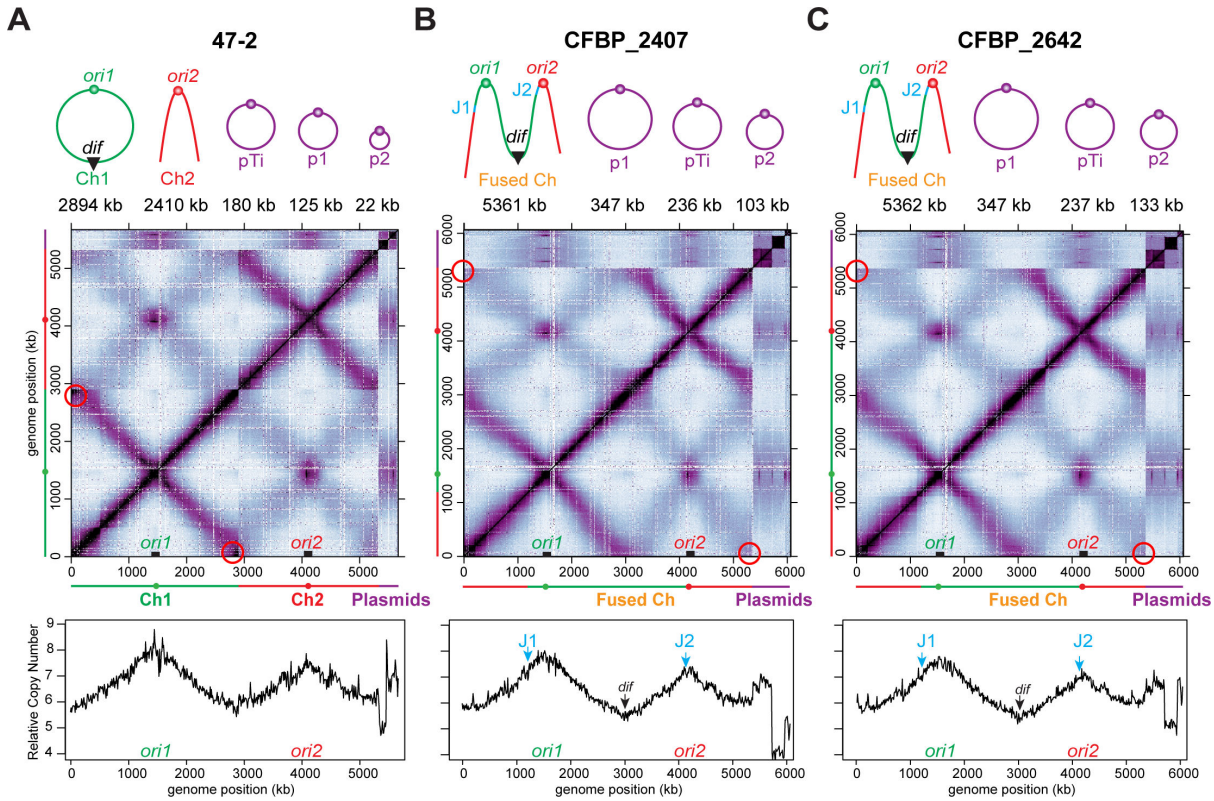

FIG 2 Natural *A. tumefaciens* isolates CFBP_2407 and CFBP_2642 exhibit linear dicentric chromosomes. (A–C) CFBP_2407 and CFBP_2642 contain a fused chromosome, while 47-2 serves as a binary control. (Top) Schematics of the genome composition of the indicated strains. J1 and J2 mark the junctions of chromosome fusion. (Middle) Normalized Hi-C interaction maps of the respective genomes. Red circles indicate the terminus regions of the circular Ch1 (A) or the linear fusion chromosomes (B and C). (Bottom) Marker frequency analysis of exponentially growing cells. The *y*-axis depicts the number of reads.

junctions in these natural isolates were different from those identified in the previously characterized lab strain C58 in terms of both location and gene content (13) (Fig. 4B).

The two replication origins on the fusion chromosomes remain active and are essential

A surprising finding from our previous analysis of the C58 fusion strain was that both origins of the fused chromosome independently initiated replication (13). To understand the replication profiles of these two fusion strains, we performed marker frequency analysis (MFA) using whole-genome sequencing. Both origins of replication (*ori1* and *ori2*) in the fusion chromosomes of CFBP_2407 and CFBP_2642 were firing at a similar frequency to those in the binary 47-2 strain (Fig. 2A through C, bottom), as signified by comparable origin copy numbers. All the strains had similar replication rates as denoted by the slope from each origin peak (Fig. 2A through C, bottom). To examine gene essentiality, we performed transposon sequencing (Tn-seq) of the two fusion strains and binary control strain 47-2. We found that both centromeric partitioning systems (*parAB* for Ch1 and *repABC* for Ch2) were essential in all three strains, regardless of chromosome configuration (Fig. 5A and B). Thus, the fusion chromosomes in the natural isolates retained the same replication and segregation programs as their binary counterparts.

XerCD-*dif* recombination system is required for the viability of the fusion strains

Our Tn-seq results showed that two genes, *xerC* and *xerD*, were essential in the fusion strains but not in the binary 47-2 strain (Fig. 5C and D). XerC and XerD are site-specific recombinases that act at the *dif* site to resolve chromosome dimers that block cell

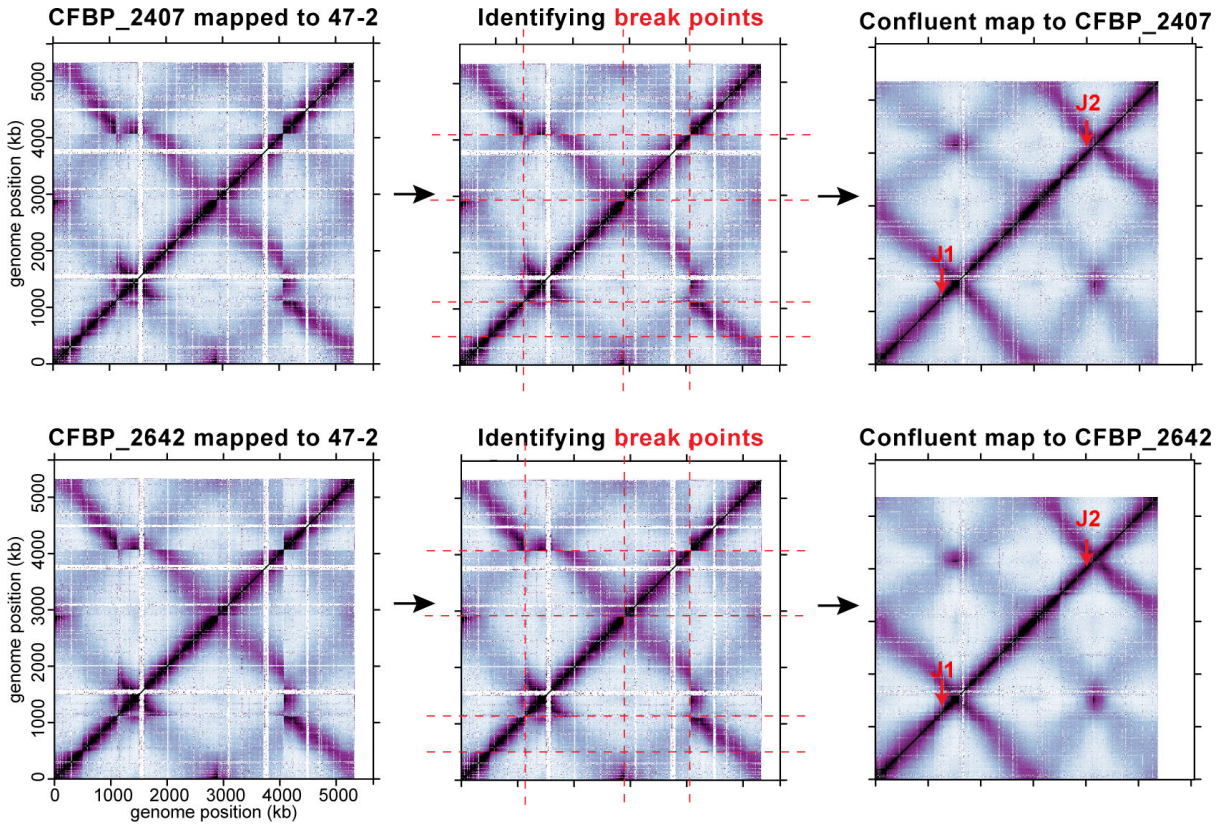

FIG 3 Identifying the junction of chromosome fusion. (Left) Hi-C reads from fusion isolates (CFBP_2407 and CFBP_2642) were mapped to the genome of the binary 47-2 strain. White spaces indicate unmapped regions on the 47-2 genome. (Middle) The breakpoints of the Hi-C maps were highlighted with red dashed lines. The maps were cut and reassembled to best match the confluent maps generated using sequenced genomes (right). Based on this approach, we identified genetic loci of the probable fusion junctions (red arrows).

division (15–18). Effective XerCD recombination at *dif* requires activation by the DNA translocase, FtsK, which resides at the closing septum (19, 20). For this system to work, FtsK must be directed by the FtsK orienting polar sequences (KOPS) (GGGNAGGG) and translocate toward the *dif* sites, and eventually bring *dif* to the closing septum (21–23). Thus, after recombination, the chromosome dimer is resolved, and the septum is clear of DNA to allow unimpeded cell division. On circular bacterial chromosomes, *dif* sites are found at the convergence point of KOPS (21). Through sequence analysis, we found the *dif* sequence is located roughly midway between *ori1* and *ori2* at the convergence of KOPS (Fig. 5E). These results are consistent with the idea that the functional FtsK-XerCD-*dif* system is required for the survival of these naturally occurring fusion strains.

The observation that *ori1* and *ori2* are active and the essentiality of the two partitioning systems indicate that the two origins of the fusion chromosomes remain independent for replication and segregation. The independent *ori* function results in a situation where for every cell division, ~50% of cells have the two replication origins of the fusion chromosome in opposite daughter cells (Fig. S1A). If this situation is unresolved, the affected cells and all their progeny will be impaired for chromosome segregation and cell division, which causes DNA breakage and cell death; in a few generations, the vast majority of cells will be affected (Fig. S1B). The presence of the *dif* site at the convergence of KOPS sites near the closing septum enables FtsK to efficiently activate XerCD recombination at the *dif* site to resolve the partitioning issue, ensuring that each daughter cell receives a complete and properly partitioned genome (Fig. S1A).

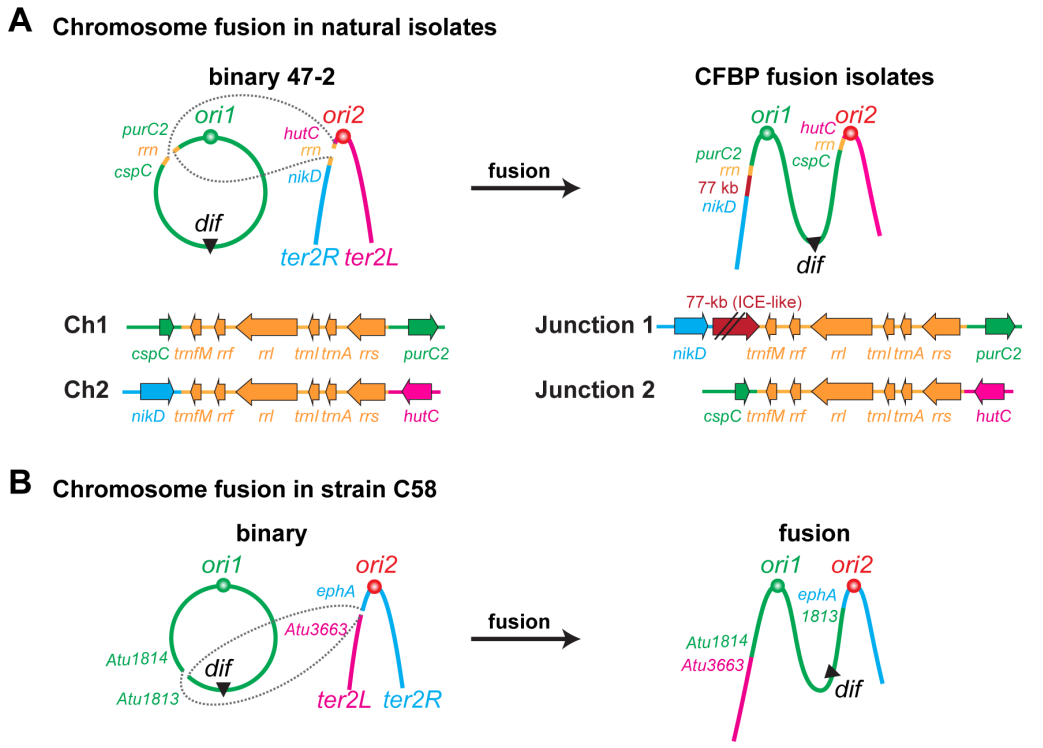

FIG 4 Sites of chromosome fusion in natural isolates and the lab C58 strain. (A) (Upper panel) Schematic illustration of chromosome fusion events in the natural isolates of CFBP_2407 and CFBP_2642 compared with the binary control 47-2. CFBP fusion isolates have a 77 kb insertion (maroon) resembling an integrative conjugative element. (Lower panel) Genetic context of the fusion junctions in the binary and fusion strains. The 77 kb ICE region is not drawn to scale. (B) Schematic illustration of chromosome fusion events in C58 (13).

DISCUSSION

In this study, we identified two natural *Agrobacterium* isolates with fused chromosomes analogous to our previous finding of a dicentric chromosome in specific stocks of the lab strain C58. The fact that two geographically distinct but closely related isolates have precisely the same change suggests that chromosome fusion occurred prior to their geographic separation. The distribution of these strains may have been facilitated by agricultural practices, such as transfer of infected grapevine material. Comparisons between the three fusion strains reveal several similarities of these fused chromosomes (Fig. 4A and B): (i) the recombination that resulted in the fusions occurring between highly similar sequences on different chromosomes; (ii) the fusion strains retain the same replication and segregation programs with individual active replication origins and separate, essential partitioning systems; (iii) the replication arms on either side of each origin have very similar sizes; and (iv) the XerCD-*dif* system is required for the survival of these fusion strains.

The chromosomal fusion sites of the two natural isolates are distinct from those in the laboratory C58 fusion. The C58 fusion sites were two ~1.5 kb sequences that differed by only 11 nucleotides (13) (Fig. 4B). In contrast, the predicted fusion sequence of CFBP_2407 and CFBP_2642 was identical ribosomal RNA (*rrn*) loci (Fig. 4A). There are four nearly identical copies of the *rrn* locus in the genomes of most agrobacteria. In most strains, including 47-2, these are present in two copies on each chromosome, *rrn*1 and *rrn*2 on Ch1 and *rrn3* and *rrn4* on Ch2 (Fig. S2). In the two fusion strains we are reporting here, recombination occurred between *rrn*2 on Ch1 and *rrn*3 on Ch2, which resulted in relatively balanced replichores (Fig. 4A). Similarly, the C58 fusion chromosome we reported earlier also has roughly balanced replichores (Fig. 4B). Since through evolution, replichores are typically balanced in size to maintain correct replication timing (24) and

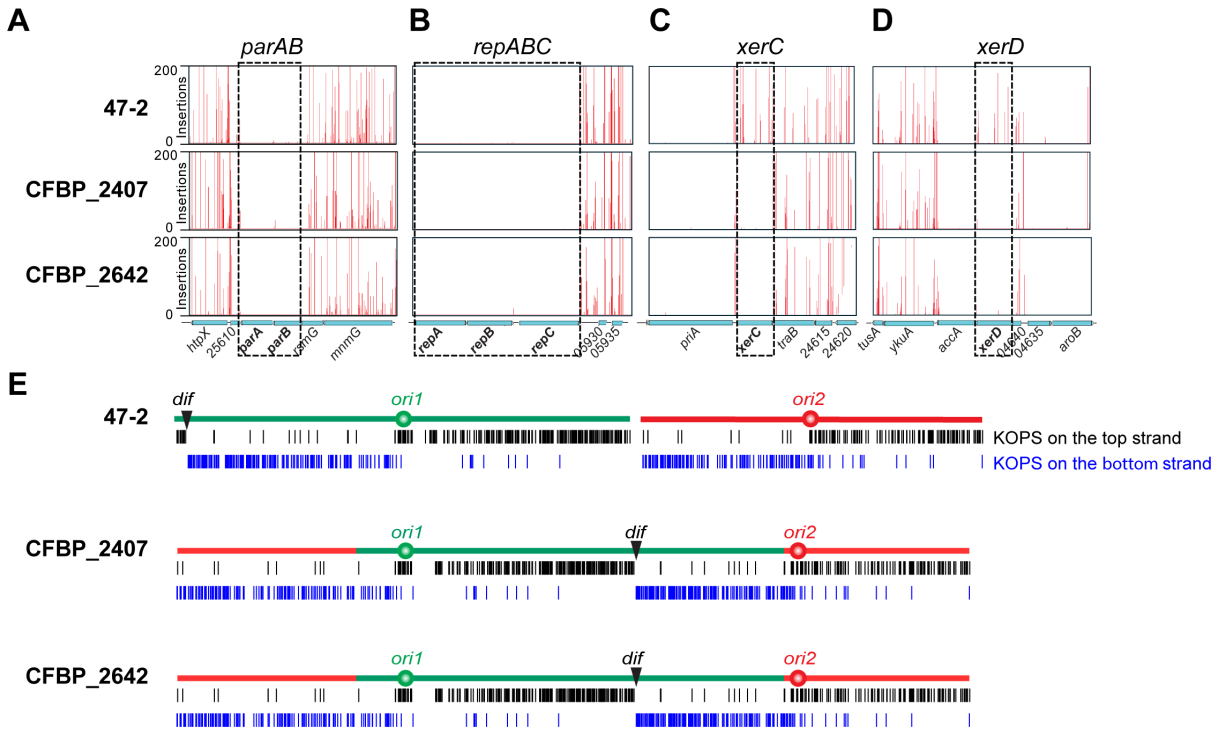

FIG 5 Gene essentiality in the three natural isolates. Transposon sequencing profiles of Ch1 partitioning system *parAB* (A) and Ch2 partitioning system *repABC* (B) and genes encoding for tyrosine recombinases *xerC* (C) and *xerD* (D). The *x*-axis depicts genome position, and the *y*-axis represents the number of transposon insertions. (E) Distribution of the FtsK orienting polar sequences (KOPS) (GGGNAGGG) on the chromosomes of the indicated strains. Black and blue bars indicate KOPS sequences on the top and bottom DNA strands, respectively. The *dif* site is illustrated by a black triangle.

imbalanced replication forks can result in severe fitness effects (25), we speculate that balanced replichores are one important reason for the viability of chromosomal fusion strains.

The presence of an ICE immediately adjacent to the fusion site raises questions about its potential role in the fusion process. We hypothesize that the integration of the ICE may have been a trigger for this fusion event. Mobile genetic elements can mobilize genes important for many phenotypes; however, the act of integration can also play a role in microbial evolution (26). Mobile genetic elements such as ICEs have been implicated in facilitating large conformational changes in bacterial chromosomes (27, 28). ICEs target specific sequences, called attachment (*att*) sites, in the chromosome for integration. The predicted *att* site targeted by this ICE is located within the identical repeat region and in multiple copies in the fusion strain and 47-2 genomes. It is plausible that the ICE recombinase may have induced the fusion event by promoting recombination between a Ch1 and Ch2 copy of the *att* site during integration. While no other sequenced agrobacteria/rhizobia complex strains were found to carry this exact ICE, we identified other diverse ICEs that target tRNA loci and other conserved sequences for integration. Future work will investigate whether integration or excision of this ICE from the chromosome can induce fusion or splitting of chromosomes.

Our findings raise important questions about the maintenance and genomic malleability of multipartite genomes. The fusion and splitting of replicons are drivers for genome evolution and speciation. The products of these events can be detrimental to chromosome segregation and can result in extinction of certain lineages from the population. However, it can also be beneficial in some cases. In *V. cholerae*, chromosome fusion was observed when the origin or centromere of one chromosome is mutated; chromosome fusion allowed Ch2 to piggyback on Ch1 so that essential genes were inherited (29). In humans, Robertsonian chromosome translocations are the fusions of two chromosomes which result in viable individuals, but the fused chromosome also

has only one functional centromere (30). Here in *A. tumefaciens*, chromosome fusion has been observed in both lab strains and environmental isolates. Importantly, these strains require two origins and two centromeres. The balanced replication arms and a sequence-specific resolution system are the key to the survival of these fusion strains. Our results support the notion that multipartite genomes, and particularly replicon fusions, represent a form of genomic plasticity that can facilitate adaptation but elevates the essentiality of the otherwise non-essential systems such as the XerCD recombinases to maintain stability. The examination of natural populations of bacteria has now revealed additional examples of chromosome fusions, expanding our understanding of what facilitates and constrains this process. Further exploration for natural examples in more bacterial taxa with multipartite genomes will provide greater insights into the mechanisms controlling genetic diversification and genome evolution.

MATERIALS AND METHODS

Bacterial growth

A. tumefaciens strains were grown at 30°C with aeration. For Hi-C and marker frequency analysis, cells were grown in defined AT minimal medium with glucose and $(NH4)_2SO_4$ (ATGN) (31). Single colonies were isolated on ATGN plates and inoculated into 5 mL of ATGN medium and incubated overnight in a tube roller. The next morning, the cultures were diluted into 30 mL ATGN liquid with a starting OD_{600} of 0.15. Cultures were grown in a shaking water bath for 6 h to reach an OD_{600} of 0.5–0.6 before harvest.

Nanopore sequencing and analysis

The Wizard Genomic DNA kit (Promega, Madison, WI) was used to extract DNA from overnight cultures of strains 47-2, CFBP_2407, and CFBP_2642. The Rapid Barcoding Kit 96 v.14 (SQK-RBK114.96; Oxford Nanopore Technologies, Oxford, UK) was used to prepare multiplexed long-read sequencing libraries for these strains. Libraries were sequenced on PromethION flow cells on an Oxford Nanopore P2 solo sequencer with super accuracy (SUP) basecalling. Unicycler v.0.5.0 with the parameter "--mode bold" was used to generate hybrid assemblies for strains CFBP_2407 and CFBP_2642 (32). For strain 47-2, Canu v.2.2 with the parameter "useGrid = false genomesize = 5.6 m" was used to assemble the genome from long-read sequences (33). Polypolish v.0.5.0 with the default parameters was used to polish the long-read assembly with Illumina short reads (34). Publicly available *Agrobacterium* genomes were downloaded from the National Center for Biotechnology Information on 19 July 2023. Beav v.1.0.0 with parameter "--agrobacterium --tiger_blast_database" and a blast database of the public genome data set was used to annotate genomes (35). FastANI v.1.34 with the default parameters was used to calculate pairwise ANI between all strains (36).

AutoMLSA2 v.0.9.0 with the default parameters was used to generate a multilocus sequence analysis (MLSA) phylogeny of all publicly available *Agrobacterium* strains and the two fusion strains (37). Twenty-three translated reference gene sequences from strain C58 were used as queries for the MLSA analysis. Genes used were *acnA*, *aroB*, Atu0781, Atu1564, Atu2640, *cgtA*, *coxC*, *dnaK*, *glyS*, *ham1*, *hemF*, *hemN*, *hom*, *leuS*, *lysC*, *murC*, *plsC*, *prfC*, *rplB*, *rpoB*, *rpoC*, *secA*, and *truA* (38). PIRATE v.1.0.5 with the default parameters was used to cluster genes from six strains belonging to the MLSA tree clade containing CFBP_2407 and CFBP_2642 into orthologous groups (39). To prepare input files for PIRATE, the agat v.1.4.0 script agat_convert_sp_gxf2gxf.pl with the "--gff_version_output 3" parameter was used to repair GFF3 files prior to ortholog clustering (40). MAFFT v.7.525 with the default parameters was used to align sequences of 4,192 conserved genes from all genomes (41). The catfasta2phyml v.1.2.0 tool with the parameter "-f --concatenate" was used to concatenate gene alignments (42). IQ-TREE2 v.2.3.1 with the parameter "-bb 1000 -alrt 1000" was used to infer a maximum likelihood phylogeny from the core genome alignment (43). A custom Python script was used to generate synteny plots (https://github.com/acarafat/PySyntenyViz).

The publicly available genome sequences of all members of the agrobacteria/rhizobia complex were downloaded on 19 August 2023. The Beav pipeline was used to annotate genes and mobile genetic elements in each genome (35). Sourmash v.4.8.11 sketch with the parameter "dna -p scaled = 1000" was used to create *k*-mer signatures for the 3,222 new and previously sequenced genomes and the 77 kb ICE sequence (44). Sourmash index was used to index and build a database of *k*-mer signatures for all genomes. Finally, sourmash search with the parameter "--containment" was used to search for instances of the 77 kb ICE sequence in each genome.

Chromosome conformation capture (Hi-C)

The Hi-C procedure performed here was as previously described (13, 45). In brief, we grew *A. tumefaciens* strains to exponential phase (~0.6 OD_{600}) in ATGN broth at 30°C (31) before cells were cross-linked with 3% formaldehyde at room temperature for 30 min. Cross-linking was quenched with 125 mM glycine for 5 min and spun down for pellets normalized to 1.2 OD_{600} units. Pellets were lysed using Ready-lyse lysozyme (Epicentre, R1802M) then treated with 0.5% SDS. Solubilized chromatin was digested with HindIII for 2 h at 37°C. Digested DNA ends were filled in with Klenow DNA polymerase and biotin-14-deoxyadenosine triphosphate, deoxyguanosine triphosphate, deoxycytidine triphosphate, and deoxythymidine triphosphate. Biotinylated products were ligated with T4 DNA ligase at 16°C for about 20 h. The DNA products were subsequently reverse cross-linked at 65°C for about 20 h in the presence of EDTA, proteinase K, and 0.5% SDS. The reactions were extracted twice using phenol:chloroform:isoamyl alcohol (25:24:1) (PCI), precipitated with ethanol, and resuspended in 20 µL of 0.1× Tris-EDTA (TE) buffer. Biotin from non-ligated ends was removed using T4 DNA polymerase (4 h at 20°C) followed by extraction with PCI. The DNA was then sonicated for 12 min with 20% amplitude using a Qsonica Q800R2 water bath sonicator. Sheared DNA was library prepped using the NEBNext UltraII kit (E7645), and biotinylated DNA fragments were purified using 10 µL streptavidin beads. DNA-bound beads were used for PCR in a 50 µL reaction for 14 cycles. PCR products were purified using Ampure beads (Beckman, A63881) and sequenced at the Indiana University Center for Genomics and Bioinformatics using NextSeq2000. Paired-end sequencing reads were mapped to the combined hybrid assemblies for strains CFBP_2407, CFBP_2642, and 47-2 described in "Nanopore sequencing and analysis." Reads in which both reads uniquely aligned to the genome were processed and sorted by HindIII restriction fragments and binned into 10 kb bins using the HiC-Pro pipeline (46). Analyses and visualizations were done using R. The 47-2 genome was rearranged for *ori1* at the center of Ch1 with the following rearrangement: Ch1 1,040.763–2,893.626 kb and 0.001–1,040.762 kb.

Whole-genome sequencing (WGS) and marker frequency analysis (MFA)

Cells were grown to exponential phase (~0.6 OD_{600}, 5×10^9 CFU/mL) in ATGN broth at 30°C (31), in which 5 mL was collected and cells pelleted by centrifugation at 5,000 × *g* for 10 min. Genomic DNA (gDNA) was extracted using the QIAGEN DNeasy Kit (69504), sonicated using a Qsonica Q800R2 water bath sonicator at 20% continuous amplitude for 6 min. Sheared gDNA was library prepared using the NEBNext UltraII kit (E7645) and sequenced using Illumina NextSeq2000. The reads were mapped to the combined hybrid assemblies for strains CFBP_2407, CFBP_2642, and 47-2 using CLC Genomics Workbench (CLC Bio, QIAGEN). The mapped reads were normalized by the total number of reads. Relative copy numbers were calculated by dividing normalized reads with the averaged total number of reads at the terminus of Ch1 or at the left terminus of the fused chromosome. Plotting and analysis were performed using R.

Transposon insertion sequencing (Tn-seq)

Tn-seq was performed as previously described (13). In brief, the *Mariner* transposon-based plasmid pTND2823 (gift from Triana Dalia and Ankur Dalia at Indiana University)

was transformed into diaminopimelic acid auxotroph *MFDpir Escherichia coli* (47) used to conjugate with *A. tumefaciens* strains. *A. tumefaciens* conjugants were plated on 10 large plates (150 mm diameter, VWR 25384–326) with Luria broth (LB) supplemented with 300 mg/mL kanamycin, aiming for 1×10^6 kanamycin-resistant colony forming units. The plates were incubated at 30°C for 2 days before each plate was scraped for all colonies and combined into a single pool. Five OD_{600} units of cells from the pool were used for gDNA isolation using QIAGEN DNeasy Blood and Tissue kit (69504). Three micrograms of gDNA was digested with MmeI (NEB R0637S) for 90 min and treated with calf intestinal alkaline phosphatase (Quick CIP, NEB M0525L) for 60 min at 37°C. The DNA was extracted using phenol-chloroform, precipitated using ethanol, and resuspended in 15 µL ddH_2O. The digested end was ligated to an annealed adapter (48) using T4 DNA ligase and incubated at 16°C for about 16 h. Adapter-ligated DNA was amplified with the primers complementary to the adapter and the transposon inverted repeat sequence. The PCR product was gel purified using a Monarch DNA Gel Extraction Kit (NEB T1020S) and sequenced at the IU Center for Genomics and Bioinformatics using NextSeq2000. Sequencing reads were mapped to the combined hybrid assemblies for strains CFBP_2407, CFBP_2642, and 47-2 and analyzed using a previously described approach (48, 49). The results were visualized using Artemis (https://www.sanger.ac.uk/tool/artemis/).

ACKNOWLEDGMENTS

We thank the IU CGB for next-generation sequencing, OSU Department of Botany and Plant Pathology for computing infrastructure, Ankur Dalia for the transposon plasmid, Doron Teper and Shulamit Manulis for sharing strain 47-2, and the International Centre of Microbial Resource-French Collection for Plant-associated Bacteria (CIRM-CFBP) for providing strains CFBP_2407 and CFBP_2642. We thank the reviewers for valuable suggestions to improve our manuscript.

This research is a contribution of the GEMS Biology Integration Institute, funded by the National Science Foundation DBI Biology Integration Institutes Program, Award #2022049 (to C.F. and X.W.). A.R. and A.J.W. were supported by startup funds from Oregon State University. X.W. was partially supported by the National Institutes of Health (R01GM141242, R01GM143182, and R01AI172822).

AUTHOR AFFILIATIONS

[1]Department of Biology, Indiana University, Bloomington, Indiana, USA
[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, USA

PRESENT ADDRESS

Zhongqing Ren, Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA

AUTHOR ORCIDs

Ram Sanath-Kumar http://orcid.org/0009-0004-5793-2498
Arafat Rahman http://orcid.org/0000-0003-4222-9615
Zhongqing Ren http://orcid.org/0000-0003-1788-5362
Clay Fuqua http://orcid.org/0000-0001-7051-1760
Alexandra J. Weisberg http://orcid.org/0000-0002-0045-1368
Xindan Wang http://orcid.org/0000-0001-6458-180X

FUNDING

Funder	Grant(s)	Author(s)
National Institute of General Medical Sciences	R01GM141242, R01GM143182	Xindan Wang
National Institute of Allergy and Infectious Diseases	R01AI172822	Xindan Wang
National Science Foundation	2022049	Clay Fuqua Xindan Wang

AUTHOR CONTRIBUTIONS

Ram Sanath-Kumar, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Arafat Rahman, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing | Zhongqing Ren, Formal analysis, Investigation, Methodology | Ian P. Reynolds, Formal analysis, Investigation, Methodology | Lauren Augusta, Formal analysis, Methodology | Clay Fuqua, Conceptualization, Project administration, Supervision, Writing – review and editing | Alexandra J. Weisberg, Conceptualization, Funding acquisition, Resources, Supervision, Visualization, Writing – original draft, Writing – review and editing | Xindan Wang, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review and editing

DIRECT CONTRIBUTION

This article is a direct contribution from Clay Fuqua, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Turlough Finan, McMaster University, and Vaughn Cooper, University of Pittsburgh School of Medicine.

DATA AVAILABILITY

Hi-C and whole-genome sequencing data were deposited to GEO (GSE285100). The Nanopore sequencing data were deposited to the Sequence Read Archive (PRJNA1201236). Any additional information required to analyze the data reported in this paper is available from the corresponding authors upon request. Strains reported in this study are available from the corresponding authors with a completed materials transfer agreement.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental Material (mBio01046-25-s0001.pdf). Supplemental figures and tables.

REFERENCES

1. Harrison PW, Lower RPJ, Kim NKD, Young JPW. 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends Microbiol 18:141–148. https://doi.org/10.1016/j.tim.2009.12.010
2. Almalki F, Choudhary M, Azad RK. 2022. Analysis of multipartite bacterial genomes using alignment free and alignment-based pipelines. Arch Microbiol 205:25. https://doi.org/10.1007/s00203-022-03354-2
3. diCenzo GC, Finan TM. 2017. The divided bacterial genome: structure, function, and evolution. Microbiol Mol Biol Rev 81:e00019-17. https://doi.org/10.1128/MMBR.00019-17
4. Rosenberg C, Boistard P, Dénarié J, Casse-Delbart F. 1981. Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. Mol Gen Genet 184:326–333. https://doi.org/10.1007/BF00272926
5. Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, et al. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293:668–672. https://doi.org/10.1126/science.1060966
6. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, et al. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406:477–483. https://doi.org/10.1038/35020000
7. DelVecchio VG, Kapatral V, Redkar RJ, Patra G, Mujer C, Los T, Ivanova N, Anderson I, Bhattacharyya A, Lykidis A, Reznik G, Jablonski L, Larsen N, D'Souza M, Bernal A, Mazur M, Goltsman E, Selkov E, Elzer PH, Hagius S, O'Callaghan D, Letesson J-J, Haselkorn R, Kyrpides N, Overbeek R. 2002. The genome sequence of the facultative intracellular pathogen *Brucella*

melitensis. Proc Natl Acad Sci USA 99:443–448. https://doi.org/10.1073/pnas.221575398
8. Ren Z, Liao Q, Karaboja X, Barton IS, Schantz EG, Mejia-Santana A, Fuqua C, Wang X. 2022. Conformation and dynamic interactions of the multipartite genome in *Agrobacterium tumefaciens*. Proc Natl Acad Sci USA 119:e2115854119. https://doi.org/10.1073/pnas.2115854119
9. Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Qurollo B, Goldman BS, Cao Y, Askenazi M, Halling C, et al. 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science 294:2323–2328. https://doi.org/10.1126/science.1066803
10. Mavingui P, Flores M, Guo X, Dávila G, Perret X, Broughton WJ, Palacios R. 2002. Dynamics of genome architecture in *Rhizobium* sp. strain NGR234. J Bacteriol 184:171–176. https://doi.org/10.1128/JB.184.1.171-176.2002
11. Guo X, Flores M, Mavingui P, Fuentes SI, Hernández G, Dávila G, Palacios R. 2003. Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures. Genome Res 13:1810–1817. https://doi.org/10.1101/gr.1260903
12. Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF Jr, et al. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science 294:2317–2323. https://doi.org/10.1126/science.1066804
13. Liao Q, Ren Z, Wiesler EE, Fuqua C, Wang X. 2022. A dicentric bacterial chromosome requires XerC/D site-specific recombinases for resolution. Curr Biol 32:3609–3618. https://doi.org/10.1016/j.cub.2022.06.050
14. Ridé M, Ridé S, Petit A, Bollet C, Dessaux Y, Gardan L. 2000. Characterization of plasmid-borne and chromosome-encoded traits of *Agrobacterium* biovar 1, 2, and 3 strains from France. Appl Environ Microbiol 66:1818–1825. https://doi.org/10.1128/AEM.66.5.1818-1825.2000
15. Blakely G, May G, McCulloch R, Arciszewska LK, Burke M, Lovett ST, Sherratt DJ. 1993. Two related recombinases are required for site-specific recombination at dif and cer in *E. coli* K12. Cell 75:351–361. https://doi.org/10.1016/0092-8674(93)80076-q
16. Cornet F, Hallet B, Sherratt DJ. 1997. Xer recombination in *Escherichia coli*. Site-specific DNA topoisomerase activity of the XerC and XerD recombinases. J Biol Chem 272:21927–21931. https://doi.org/10.1074/jbc.272.35.21927
17. Sciochetti SA, Piggot PJ, Sherratt DJ, Blakely G. 1999. The *ripX* locus of *Bacillus subtilis* encodes a site-specific recombinase involved in proper chromosome partitioning. J Bacteriol 181:6053–6062. https://doi.org/10.1128/JB.181.19.6053-6062.1999
18. Sciochetti SA, Piggot PJ, Blakely GW. 2001. Identification and characterization of the *dif* site from *Bacillus subtilis*. J Bacteriol 183:1058–1068. https://doi.org/10.1128/JB.183.3.1058-1068.2001
19. Grainge I, Lesterlin C, Sherratt DJ. 2011. Activation of XerCD-dif recombination by the FtsK DNA translocase. Nucleic Acids Res 39:5140–5148. https://doi.org/10.1093/nar/gkr078
20. Aussel L, Barre FX, Aroyo M, Stasiak A, Stasiak AZ, Sherratt D. 2002. FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. Cell 108:195–205. https://doi.org/10.1016/s0092-8674(02)00624-4
21. Pérals K, Cornet F, Merlet Y, Delon I, Louarn JM. 2000. Functional polarization of the *Escherichia coli* chromosome terminus: the *dif* site acts in chromosome dimer resolution only when located between long stretches of opposite polarity. Mol Microbiol 36:33–43. https://doi.org/10.1046/j.1365-2958.2000.01847.x
22. Bigot S, Saleh OA, Lesterlin C, Pages C, El Karoui M, Dennis C, Grigoriev M, Allemand J-F, Barre F-X, Cornet F. 2005. KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. EMBO J 24:3770–3780. https://doi.org/10.1038/sj.emboj.7600835
23. Sivanathan V, Allen MD, de Bekker C, Baker R, Arciszewska LK, Freund SM, Bycroft M, Löwe J, Sherratt DJ. 2006. The FtsK γ domain directs oriented DNA translocation by interacting with KOPS. Nat Struct Mol Biol 13:965–972. https://doi.org/10.1038/nsmb1158
24. Rocha EPC. 2004. Order and disorder in bacterial genomes. Curr Opin Microbiol 7:519–527. https://doi.org/10.1016/j.mib.2004.08.006
25. Darling AE, Miklós I, Ragan MA. 2008. Dynamics of genome rearrangement in bacterial populations. PLoS Genet 4:e1000128. https://doi.org/10.1371/journal.pgen.1000128
26. Baltrus DA. 2013. Exploring the costs of horizontal gene transfer. Trends Ecol Evol 28:489–495. https://doi.org/10.1016/j.tree.2013.04.002
27. Haskett TL, Terpolilli JJ, Bekuma A, O'Hara GW, Sullivan JT, Wang P, Ronson CW, Ramsay JP. 2016. Assembly and transfer of tripartite integrative and conjugative genetic elements. Proc Natl Acad Sci USA 113:12268–12273. https://doi.org/10.1073/pnas.1613358113
28. Weisberg AJ, Sachs JL, Chang JH. 2022. Dynamic interactions between mega symbiosis ICEs and bacterial chromosomes maintain genome architecture. Genome Biol Evol 14:evac078. https://doi.org/10.1093/gbe/evac078
29. Val M-E, Kennedy SP, Soler-Bistué AJ, Barbe V, Bouchier C, Ducos-Galand M, Skovgaard O, Mazel D. 2014. Fuse or die: how to survive the loss of Dam in *Vibrio cholerae*. Mol Microbiol 91:665–678. https://doi.org/10.1111/mmi.12483
30. Gerton JL. 2024. A working model for the formation of Robertsonian chromosomes. J Cell Sci 137:jcs261912. https://doi.org/10.1242/jcs.261912
31. Morton ER, Fuqua C. 2012. Laboratory maintenance of *Agrobacterium*. Curr Protoc Microbiol Chapter 1:Unit3D.1. https://doi.org/10.1002/9780471729259.mc03d01s24
32. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595
33. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116
34. Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol 18:e1009802. https://doi.org/10.1371/journal.pcbi.1009802
35. Jung JM, Rahman A, Schiffer AM, Weisberg AJ. 2024. Beav: a bacterial genome and mobile element annotation pipeline. mSphere 9:e00209-24. https://doi.org/10.1128/msphere.00209-24
36. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9
37. Davis Ii EW, Weisberg AJ, Tabima JF, Grunwald NJ, Chang JH. 2016. Gall-ID: tools for genotyping gall-causing phytopathogenic bacteria. PeerJ 4:e2222. https://doi.org/10.7717/peerj.2222
38. Zhang YM, Tian CF, Sui XH, Chen WF, Chen WX. 2012. Robust markers reflecting phylogeny and taxonomy of rhizobia. PLoS One 7:e44936. https://doi.org/10.1371/journal.pone.0044936
39. Bayliss SC, Thorpe HA, Coyle NM, Sheppard SK, Feil EJ. 2019. PIRATE: a fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. Gigascience 8:giz119. https://doi.org/10.1093/gigascience/giz119
40. Ruano DG. 2024. AGAT: another GFF analysis toolkit to handle annotations in any GTF/GFF format
41. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436
42. Nylander J. 2024. catfasta2phyml
43. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
44. Irber L, Pierce-Ward NT, Abuelanin M, Alexander H, Anant A, Barve K, Baumler C, Botvinnik O, Brooks P, Dsouza D, et al. 2024. Sourmash v4: a multitool to quickly search, compare, and analyze genomic and metagenomic data sets. J Open Source Software 9:6830. https://doi.org/10.21105/joss.06830
45. Akgol Oksuz B, Yang L, Abraham S, Venev SV, Krietenstein N, Parsi KM, Ozadam H, Oomen ME, Nand A, Mao H, Genga RMJ, Maehr R, Rando OJ, Mirny LA, Gibcus JH, Dekker J. 2021. Systematic evaluation of chromosome conformation capture assays. Nat Methods 18:1046–1055. https://doi.org/10.1038/s41592-021-01248-7
46. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 16:259. https://doi.org/10.1186/s13059-015-0831-x
47. Ferrières L, Hémery G, Nham T, Guérout A-M, Mazel D, Beloin C, Ghigo J-M. 2010. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. J Bacteriol 192:6418–6427. https://doi.org/10.1128/JB.00621-10

48. Meeske AJ, Sham L-T, Kimsey H, Koo B-M, Gross CA, Bernhardt TG, Rudner DZ. 2015. MurJ and a novel lipid II flippase are required for cell wall biogenesis in *Bacillus subtilis*. Proc Natl Acad Sci USA 112:6437–6442. https://doi.org/10.1073/pnas.1504967112

49. van Opijnen T, Lazinski DW, Camilli A. 2014. Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. Curr Protoc Mol Biol 106:7. https://doi.org/10.1002/0471142727.mb0716s106