Published online 18 January 2018 *Nucleic Acids Research, 2018, Vol. 46, No. 5* **2521–2536**
doi: 10.1093/nar/gky005

A novel family of tyrosine integrases encoded by the temperate pleolipovirus SNJ2

Jiao Wang[1], Yingchun Liu[1], Ying Liu[1,2], Kaixin Du[1], Shuqi Xu[1], Yuchen Wang[1], Mart Krupovic[2] and Xiangdong Chen[1,*]

[1]State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan 430072, China and [2]Unit of Molecular Biology of the Gene in Extremophiles, Department of Microbiology, Institut Pasteur, Paris 75015, France

Received October 13, 2017; Revised December 26, 2017; Editorial Decision December 29, 2017; Accepted January 08, 2018

ABSTRACT

Genomes of halophilic archaea typically contain multiple loci of integrated mobile genetic elements (MGEs). Despite the abundance of these elements, however, mechanisms underlying their site-specific integration and excision have not been investigated. Here, we identified and characterized a novel recombination system encoded by the temperate pleolipovirus SNJ2, which infects haloarchaeon *Natrinema* sp. J7-1. SNJ2 genome is inserted into the *tRNA^{Met}* gene and flanked by 14 bp direct repeats corresponding to attachment core sites. We showed that SNJ2 encodes an integrase (Int^{SNJ2}) that excises the proviral genome from its host cell chromosome, but requires two small accessory proteins, Orf2 and Orf3, for integration. These proteins were co-transcribed with Int^{SNJ2} to form an operon. Homology searches showed that Int^{SNJ2}-type integrases are widespread in haloarchaeal genomes and are associated with various integrated MGEs. Importantly, we confirmed that SNJ2-like recombination systems are encoded by haloarchaea from three different genera and are critical for integration and excision. Finally, phylogenetic analysis suggested that Int^{SNJ2}-type recombinases belong to a novel family of archaeal integrases distinct from previously characterized recombinases, including those from the archaeal SSV- and pNOB8-type families.

INTRODUCTION

Integrases are site-specific recombinases, which catalyze recombination between two specific DNA molecules, resulting in integration and excision of mobile genetic elements (MGEs) into and from host chromosomes (1). This process occurs without an external energy source and involves several, well-orchestrated steps comprised of strand breakage, exchange and rejoining (2). Recombinases are widespread in both cellular organisms and various MGEs, including plasmids (3,4), viruses (5), pathogenicity islands (6), genomic islands (7), integrons and gene cassettes (8), and integrative and conjugative elements (9). Correspondingly, they play multiple fundamental biological roles in the three domains of life, including integration/excision of viral genomes (10), elimination of dimers from replicated chromosomes during chromosome segregation (Xer/dif system) (11) and modulation of cell-surface components for immune escape (Hin/hix system) (12). They also maintain proper copy number, partition plasmid (Flp/frt system) and bacteriophage (Cre/lox system) genomes (13,14), and acquire, reorder and stockpile various genes (IntI system) (15).

Most integrases can be subdivided into two recombinase superfamilies, the tyrosine and serine superfamilies, respectively. This classification is based on their corresponding nucleophilic amino acid residues, which attack DNA phosphodiester bonds to form a covalent protein–DNA linkage (16,17). Tyrosine recombinases can be further subdivided into two groups based on the directionality of the catalyzed site-specific recombination: complex unidirectional and simple bidirectional tyrosine-type integrases, respectively (2). This directionality is modulated by a class of small accessory proteins known as recombination directionality factors (RDFs). RDFs are typically small proteins lacking clear sequence conservation (18) that play architectural roles in the reactions catalyzed by their cognate recombinases. Some RDFs also function as transcriptional regulators (19–21). Importantly, there is no apparent genomic coupling between *integrase* and *rdf* genes, and the two genes are often separated in the genome. In contrast, simple bidirectional tyrosine-type integrases catalyze DNA recombination in a reversible fashion without the assistance of other factors (22).

Most tyrosine recombinases possess seven conserved residues: Arg_I, Glu/Asp_I, Lys, His_{II}, Arg_{II}, His/Trp_{III} (where Roman numerals correspond to the three catalytic signature motifs I, II and III) and the essential catalytic Tyr residue (23). These residues play catalytic roles in several en-

*To whom correspondence should be addressed. Tel: +86 27 68754533 82; Fax: +86 27 68752560; Email: xdchen@whu.edu.cn

Table 1. Analysis of the conserved residues of SNJ2 and other integrases based on sequence alignment of representative tyrosine recombinase superfamily proteins

Consensus	Arg_I	Glu/Asp_I*	Lys[a]	His_{II}	Arg_{II}	His/Trp_{III}*	Tyr
Int SNJ2	R165	**G168**	K199	H295	R298	**A321**	Y330
Int Lambda	R212	D215	K235	H308	R311	H333	Y342
Flp *S. cerevisiae*	R191	D194	K223	H305	R308	W330	Y400
XerC *E. coli*	R148	E151	K172	H240	R243	H266	Y275
XerD *E. coli*	R148	E151	K172	H244	R247	H270	Y279
Cre P1	R173	E176	K201	H289	R292	W315	Y324
Int BJ1	R170	E173	K197	**T264**	R267	W290	Y300
Int phiCh1	R48	E51	**R79**	H172	R175	W198	Y207
Int SSV1	R211	E214	**R240**	**K278**	R281	**R304**	Y314
Int pNOB8	R277	E280	K305	**Y379**	R382	**R405**	Y417
Int pTN3	R286	E289	K336	**A382**	R385	**R417**	Y427

The traditional consensus tetrad that differentiates between the bacterial and archaeal sources of *int* is shaded in light gray. Substitutions in the conserved sites are bolded. *Two conserved residues (Glu/Asp_I and His/Trp_{III}) are present as one of two alternative amino acid types in tyrosine recombinases. a, the catalytic lysine (arginine) was determined to be present if it was located within four residues of Int^{SNJ2} K199 in the alignment.

zymes. Specifically, the Tyr and Lys residues serve as nucleophile and general acid catalysts, respectively (24). The Arg_I and Arg_{II} residues neutralize the negative charge during the transition state and activate the scissile phosphate by the catalytic tyrosine residue. Finally, the His/Trp_{III}, His_{II} and Glu/Asp_I residues stabilize the transition state (23–25). Importantly, bacterial tyrosine-type integrases contain a signature $R_I...H_{II}XXR_{II}...Y$ tetrad (where X is any residue).

In archaea, there are reports of two unique families of MGE-encoded integrases: SSV-type ($R_I...K_{II}XXR_{II}...Y$) and pNOB8-type ($R_I...Y_{II}XXR_{II}...Y$) integrases, respectively. Integrases of both families harbor non-canonical substitutions within the active site tetrad (Table 1). SSV-type integrases are encoded by the spindle-shaped fuselloviruses, SSV1 and SSV2, that infect hyperthermophilic archaea of the genus *Sulfolobus* (phylum *Crenarchaeota*) (26–30). Uniquely, their integrase (*int*) gene is interrupted upon integration by an *att* site located within this gene (31). By contrast, pNOB8-type integrases are exclusively found in the conjugative pNOB8-like plasmids of *Sulfolobus* (31,32). Unlike SSV-type integrases, pNOB8-type integrases retain an intact *int* gene upon integration because their *att* site is located upstream of the *int* gene (31). Many archaeal head-tailed viruses (order *Caudovirales*) have been predicted to carry integrase genes (33); however, none of these have been studied experimentally.

The accumulation of archaeal genome sequences in public databases has fuelled the discovery of novel integrating MGEs that encode diverse tyrosine recombinases (34). However, very little is known about these integrases due to the lack of efficient genetic systems and relatively poor annotation of the archaeal genomes. Previously, we identified a novel provirus in a haloarchaeon *Natrinema* sp. J7-1. This virus, SNJ2, is the first isolated temperate member of the family *Pleolipoviridae* (35). We also found 17 SNJ2-like proviruses in the genomes of other haloarchaea belonging to 10 different genera. These proviruses all share a unique conserved gene cluster and appear to utilize a similar integration strategy (36). Haloarchaea isolated from widespread geographical locations all contain SNJ2-like proviral regions, suggesting that temperate pleolipoviruses represent a ubiquitous group of viruses 'hiding' in archaeal genome databases. Identification of the site-specific recombination machinery encoded by these viruses is necessary to understand their life cycle. In this study, we characterized a novel archaeal recombination system for SNJ2 that was important for the excision, integration and maintenance of this virus in host cells. Furthermore, this recombination system was comprised of genes for *int* and occasionally, auxiliary factors and the cognate *att* site. These were encoded by various proviral regions or MGEs in haloarchaeal genomes from 22 different genera and 5 families. Collectively, our results provide the first insight into the mechanisms controlling a site-specific recombination system used by a prevalent group of haloarchaeal MGEs, which may promote host genome plasticity and horizontal gene transfer (HGT).

MATERIALS AND METHODS

Bioinformatic analysis of site-specific recombination elements

Alignments of multiple protein sequences were performed using MUSCLE (37) or ClustalW (38). For homology detection, we used a combination of the NCBI Conserved Domain Search (39), PSI-BLAST (40) and HHpred (41). For phylogenetic analyses, protein sequences were aligned with Promals3D (42). Poorly aligned (low information content) positions were removed using the *gappyout* function of Trimal (43). The maximum likelihood phylogenetic tree of multiple tyrosine recombinase families was constructed using the latest version of the PhyML program (44). This program automatically selects the best-fit substitution model for a given alignment. The best model identified by PhyML was LG +G4 +I +F (LG, Le-Gascuel matrix; G4, Gamma shape parameter: fixed, number of categories: 4; I, proportion of invariable sites: fixed; F, equilibrium frequencies: empirical). The phylogenetic tree of the 38 Int^{SNJ2}-type integrases was inferred using the neighbour-joining method in the MEGA software package (version 7.0) (45). Potential open reading frames (ORFs) larger than 50 codons were predicted using ORF Finder (NCBI) or the Vector NTI Advance® 11.5 software. Start codons were determined by selecting for the largest possible ORFs with one of the three start codons (ATG, GTG and TTG). Pairwise identity percentages between amino acid sequences were computed using the EMBOSS Needle tool (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). Promoters were predicted using the Neural Network Promoter Prediction tool (http://www.fruitfly.org/seq_tools/promoter.html).

Strains, culture conditions and transformation procedures

All strains used in this study are listed in Supplementary Table S1. J7 and other haloarchaeal strains were cultured on Halo-2 or 18% modified growth medium (MGM) as previously described (46). Agar plates contained 15 g Bacto Agar (BD) per litre. Casamino Acids medium (Hv-Ca) was prepared according to the online protocol (http://www.haloarchaea.com/resources/halohandbook/Halohandbook_2009_v7.2mds.pdf), except Casamino Acids (Sigma-Aldrich) replaced the peptone and yeast extract in 18% MGM to provide selection pressure. When needed, 5-fluoroorotic acid (5-FOA) was added at a

final concentration of 0.04 mg/ml, while mevinolin (Mev) antibiotic was added at 5 μg/mL in 18% MGM for haloarchaeal cultures. *Escherichia coli* strains, DH5α and JM110, were cultured in Luria-Bertani medium at 37°C. When needed, ampicillin (0.1 mg/ml) and chloramphenicol (0.05 mg/ml) were added to the media. The modified polyethylene glycol (PEG) method was used to transform halobacteria as previously described (46). DH5α and JM110 were transformed using the $CaCl_2$ method (47).

Plasmid construction

Plasmid information is provided in Supplementary Table S2, and the oligonucleotides used for plasmid construction are summarized in Supplementary Table S3. Five classes of plasmids were constructed and classified according to functionality, as follows: (i) pNBF (kindly provided by Prof. Yuping Huang, Wuhan University) used for construction of *pyrF* deletion mutants carried upstream and downstream flanking sequences of *pyrF* (whose product was homologous to the enzyme orotidine-5'-phosphate decarboxylase), with an Mev resistance gene. The pNBK-ORFx and pUC-mev-pyrF-ORFx used to generate the *orfx* deletion and *orfx::pyrF* deletion-insertion mutants were constructed based on the basic vectors, pNBK-F (gift from Prof. Yuping Huang) and pUC-mev (46). pNBK-ORFx contained target gene flanking sequences with an intrinsic *pyrF* cassette as the selective marker, while pUC-mev-pyrF-ORFx contained the target gene flanking sequences separated by the *pyrF* cassette with Mev resistance. (ii) pYCJ-1416 used for complementation of *int*-null mutants. The promoter and coding regions of int^{SNJ2} were amplified by polymerase chain reaction (PCR) using DNA from J7-1-F *Δorf2–3* as a template, and then cloned into the vector pYC-J (46). (iii) Int^{SNJ2} and Int^{SNJ2}-type integration vectors, pCF-X and pCF-Z, were used in the integration assay. To construct these plasmids, fragments X and Z containing different parts of the Int operon were amplified using the genomic DNA of strains J7-1-F *Δorfx* and three haloarchaea, *Haloterrigena thermotolerans* DSM 11522, *Natronorubrum bangense* JCM 10635 and *Natrinema versiforme* JCM 10478, respectively. These were cut and ligated into the vector pUC-M-pyrF (48) (pCF), and were selectively marked by the *pyrF* cassette. We created different point mutations in pCF-2207 to generate Orf2/Orf3 frame-shift fragments and conserved site substitution of Int^{SNJ2} fragments using PCR-based overlap extension mutagenesis (49). (iv) Plasmids containing truncated Int^{SNJ2} and Int^{SNJ2}-type excision recombination vectors, pSU-Y and pFJ6-W, were created using fragments Y and W with the pCF-X and Z as templates, respectively. These were cut with HindIII–XbaI and MfeI–SphI, then cloned into the corresponding basic vectors, pSU19 (50) and pFJ6 (48). (v) The mini-chimeric vector, pUC-mev-oriN-MH, for the excision assay was constructed using four steps. First, an SnaB–MfeI oriC10 (48) replicon fragment digested from pFJ6 was ligated into the vector pUC-mev to generate pUC-mev-oriN. Second, two fragments, *PropyrF (promoter of pyrF)-attBOP* and *pyrF-attB'OP'*, were amplified using overlapping PCR. Third, the two fragments were fused together by overlapping PCR containing the restriction sites of MfeI and HindIII at the 5' end, respectively. Finally, the digested fragment was ligated into the MfeI–HindIII-digested pUC-mev-oriN vector.

Construction of Δ*pyrF* and $\Delta orfx^{SNJ2}$ strains

PyrF deletion in *Natrinema* sp. J7-1, J7-3 was performed as previously described in *Haloferax volcanii* (51). Briefly, the non-replicative vector pNBF transformants were cultured in 18% MGM with Mev antibiotic to promote the first homologous recombination. Then, second recombination was promoted by culturing in the rich medium, Halo-2. Cultures were then plated on Halo-2 plates with 5-FOA to select for *pyrF* deletion mutants, designated as J7-1-F and J7-3-F, respectively. The *orfx* and *orfx::pyrF* strains were constructed based on J7-1-F in a similar manner. The knockout vectors, pNBK-ORFx and pUC-mev-pyrF-ORFx, containing the 5' and 3' flanking sequences of the *orfx* gene were transformed into J7-1-F. The resultant strains underwent twice homologous recombination for allelic exchange by culturing in different medium. The cultures were then plated on Halo-2 plates with 5-FOA and Hv-Ca to select for deletion and deletion-insertion recombinants, respectively. At least three independent clones of each strain were confirmed by PCR with primer pairs (listed in Supplementary Table S3) located at the outer edge and inside the deletion regions.

Induction, verification of virus and infection procedures

Natrinema sp. J7 cultures at the stationary phase were treated with 1 μg/ml mitomycin C (MMC; Roche) for 30 min at 37°C. Cells were then collected by centrifugation and resuspended in fresh Halo-2 medium, followed by incubation for about 48 h to obtain culture lysates containing virus particles (36). Cells and debris were removed by centrifugation (10 000 × *g*, 10 min), followed by filtration through a Millipore Millex filter (0.45 μm). The filtrates were collected and stored. For infection, the filtrate was incubated with J7 culture for 1 h at 37°C. The culture was then plated on the Halo-2 or Hv-Ca medium to select for infected strains. To verify the presence of the marked virus particles in culture supernatants, the supernatant was first treated with DNase I (TaKaRa) to remove chromosomal and proviral DNA and then used as template for PCR to detect the circularized viral genome with specific primer pairs (Supplementary Table S3).

Co-transcription analyses and determination of transcript 5′/3′-ends

For analysis of *orf*1–4 co-transcription, total RNA was first extracted from exponentially growing cultures J7-1 with/without MMC treatment using TRIzol (Invitrogen). RNA quantity and quality were quantified by Eppendorf BioSpectrometer fluorescence (at 260, 280 and 230 nm) and visualized using agarose gel electrophoresis stained with ethidium bromide. Purified RNA was pre-treated with gDNA Eraser Reagent to remove genomic DNA before it was used as a template for Reverse Transcription-PCR (RT-PCR). RT-PCR was performed according to the manufacturer's protocol (PrimeScript™ RT reagent Kit with

gDNA Eraser; TaKaRa) using reverse transcriptase and random primers to amplify the cDNA. To quantify genomic DNA contamination, negative controls consisting of RT-PCR without reverse transcriptase were included for each reaction mixture. The resulting cDNA was used as a template for transcription analysis, with specific primers (Supplementary Table S3) designed to amplify the intragenic and intergenic regions of *orf1-orf4*. To determine the transcript 5'- and 3'-ends, RNA circularization and extraction were performed as previously described (52). Briefly, purified self-ligated RNA was treated with gDNA eraser to remove genomic DNA and then used for reverse transcription. The cDNA of 5'-3' ligated RNA was then amplified with gene-specific primer pairs orf1-edge/orf3-edge, followed by a second PCR with nested primer pairs Nest-F/Nest-R (Supplementary Table S3 and Figure S5). Nested PCR considerably enhanced amplification specificity and eliminated false-positive fragments from the first PCR reaction. Nested PCR products of the 5'-3' ligated RNA were subsequently analyzed by sequencing.

Quantitative and qualitative detection of site-specific recombination *in vivo*

For the excision assay (schematic shown in Supplementary Figure S6A), a series of Int-expressing plasmids, pSU-Y (10 μl), were constructed and transformed into J7-3-F/pUC-mev-oriN-MH competent cells (2 ml, OD_{600} at 0.8–1.0) containing the recombination sites, *attBOP* (257 bp) and *attB'OP'* (162 bp; hybrid of virus attachment site *attP'OP* and host chromosome attachment site *attBOB'*; Supplementary Figure S6A). For the integration assay (Supplementary Figure S6B), a series of Int^{SNJ2} and Int^{SNJ2}-type recombination vectors, pCF-X and pCF-Z, were transformed into J7-3-F competent cells. Transformed cells were recovered with 1 ml of Halo-2 medium at 37°C for 12 h and then plated on solid Hv-Ca medium and incubated at 45°C for 7–8 days. Integration and excision frequencies were calculated as the number of transformants on solid Hv-Ca medium obtained with 1 μg of plasmid DNA. Qualitative tests of the recombination junctions were performed using PCR and enzyme digestion. The PCR-based recombination assay was previously described (36). MMC-treated/untreated cultures were centrifuged to collect cells, and cell pellets were lyzed by suspension in the same volume of distilled water. Lysates were used as PCR amplification templates. The primer pairs (Supplementary Table S3) used for detection of the *att* sites were designed to target the borders between the MGE (proviral) genome and the host chromosome.

RESULTS

Temperate virus SNJ2 encodes a novel tyrosine integrase

The pleolipovirus SNJ2 was previously isolated from *Natrinema* sp. J7-1 cell culture supernatants (36,53). It carries a 14 bp attachment core site identical to the 3'-distal region of the host *tRNA-Met* gene. This enables the virus to site-specifically integrate into the host chromosome by recombining with the *tRNA-Met* gene, and subsequently be excised as a circular DNA molecule (36). Here, SNJ2 genome annotation revealed that ORF1 encodes a putative integrase, Int^{SNJ2} (accession no: AFO55992). Among profile HMMs available on the HHpred website, Orf1 best matched (HHpred probability of 100) the site-specific recombinases XerC (COG4973) and XerD (COG4974) (54). Alignment of Int^{SNJ2} with representative tyrosine recombinases revealed that this enzyme contains five invariant active site residues of the $R_I...K...H_{II}XXR_{II}...Y$ pentad (Table 1 and Supplementary Figure S1), which are conserved in canonical tyrosine recombinases (23). However, in Int^{SNJ2}, the typical Glu/Asp_I site and His/Trp_{III} sites were substituted with Gly and Ala residues, respectively (Table 1). Consistently, Int^{SNJ2} had low pairwise sequence identity (1.7–21.3%) with tyrosine recombinases from all other characterized families (Supplementary Table S4).

To further characterize the relationship between SNJ2-like integrases and other tyrosine recombinases, we created a maximum likelihood phylogenetic tree of representative sequences from diverse tyrosine recombinase families. The global phylogeny generally complied with the predefined tyrosine recombinase classification and available biochemical data (Figure 1 and Supplementary Figure S2). Notably, archaeal recombinases involved in chromosome dimer resolution, named XerA (55), were at the base of a bacterial clade, including the paralogous XerC and XerD recombinases. Putative integrases of archaeal BJ1 (56)-like and phiCh1 (57)-like viruses formed a common clade with Cre-like recombinases of P1-like phages and yeast Flp-like recombinases, distinguishing them from other bona fide viral integrases (Figure 1). For phiCh1-like viruses, this placement is consistent with the experimental evidence that phiCh1 recombinase catalyzes inversion (rather than integration) of a gene cassette encoding tail fibre proteins (58). Importantly, SNJ2-like integrases formed a large, well-supported clade with the MGE-encoded integrases from hyperthermophilic archaea. In this clade, SSV-type integrases encoded by crenarchaeal MGE formed a sister group to pTN3-like integrases encoded by plasmids and viruses infecting members of the order *Thermococcales* (phylum *Euryarchaeota*) (59). Such topology supports the possibility that the two integrase families evolved from a common ancestor, possibly prior to the divergence of the *Crenarchaeota* and *Euryarchaeota* phyla. Furthermore, SNJ2-like and pNOB8-like family integrases were at the base of this archaeal clade, suggesting that the unorthodox SSV-type and pTN3-like integrases evolved from the more typical archaeal integrases, which act on *att* sites located outside of the *int*. Finally, the SNJ2-like integrases were close to the root of the major clade of archaeal MGE-encoded integrases (Figure 1), suggesting that SNJ2-like integrases diverged from the remaining archaeal homologs early in their evolution. These results strongly suggest that the Int^{SNJ2} belongs to a novel family of tyrosine integrases.

Int^{SNJ2} excises and integrates SNJ2 provirus into the host chromosome

To determine the biological function of the putative Int^{SNJ2} protein, we constructed a *Natrinema* sp. J7-1-F strain in which ORF1 (Int^{SNJ2}) was deleted from the SNJ2 proviral region (Supplementary Table S1 and Figures S3A and

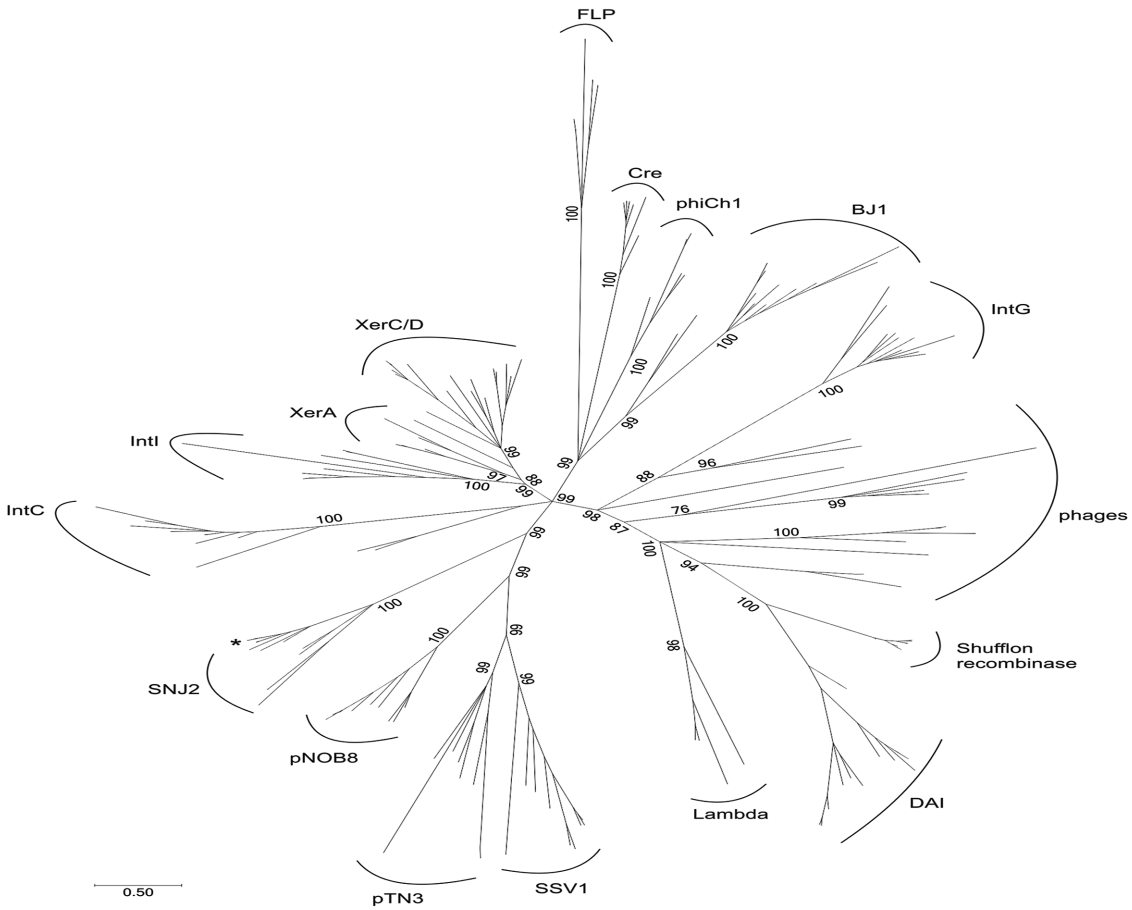
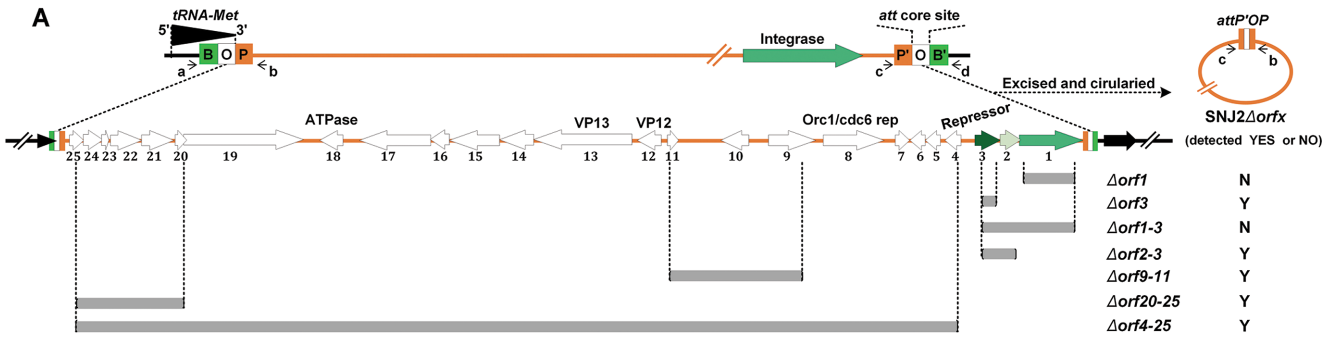


Figure 1. Maximum likelihood phylogenetic analysis of tyrosine recombinases. The tree contains representative members from different tyrosine recombinase families, including shufflon-specific DNA recombinases (4); dusA-associated integrases (DAI); phage integrases; integrases from integrative and conjugative elements (IntC); integron integrases (IntI); genomic island integrases (IntG); site-specific recombinases involved in chromosome dimer resolution in archaea (XerA) and bacteria (XerC/D); yeast Flp-like flippases (FLP); phage P1-like recombinases (Cre); archaeal SSV-type, pNOB8-type and pTN3-type integrases; and integrases encoded by haloarchaeal BJ1-like and phiCh1-like viruses. The SNJ2 clade contains putative integrases encoded by proviruses that are integrated in various haloarchaeal genomes. The SNJ2 integrase is indicated with an asterisk. Numbers at the main branch points represent the Bayesian-like transformation of aLRT (aBayes) local support values. The scale bar represents the number of substitutions per site. Branches with support values below 70% were collapsed. The tree in which all nodes are labeled is shown in Supplementary Figure S2.

B), referred to as J7-1-F *Δorf1*. To determine which SNJ2 provirus products affected excision, a series of J7-1-F *Δorfx* strains were constructed in which different regions of the integrated SNJ2 provirus were deleted (Supplementary Figure S3B and Table S1). The in-frame deleted regions spanned the entire proviral genome, except the hybrid *attBOP* and *attP'OB'* sites (B/B' and P/P' are the host and virus DNA flanking the 14 bp core site (O) respectively; Figure 2A). Deleted ORFs were not essential for SNJ2 excision if we were able to detect the expected circularized form of the provirus with a concomitant reconstitution of the *attP'OP* site (Supplementary Figure S3B). For essential ORFs, however, we would not detect the *attP'OP* site. The *attP'OP* site was not detected in culture supernatants of strains J7-1-F *Δorf1* and J7-1-F *Δorf1–3*, indicating that Orf1 (Int^{SNJ2}) was critical for SNJ2 excision (Figure 2A and Supplementary Figure S3B). Alternatively, the ORFs corresponding to proviruses SNJ2*Δorf3*, SNJ2*Δorf2–3*, SNJ2*Δorf9–11*, SNJ2*Δorf4–25* and SNJ2*Δorf20–25* were not necessary for SNJ2 excision (Figure 2A and Supplementary Figure S3B). To confirm these results, we complemented strains J7-1-F *Δorf1* and J7-1-F *Δorf1–3* with plasmid pYCJ-1416 (Supplementary Table S2) carrying Int^{SNJ2} and tested whether excision was restored upon MMC induction. As shown in Supplementary Figure S3C, we confirmed that Int^{SNJ2} was the only product required for SNJ2 provirus excision.

To test if Int^{SNJ2} was also necessary for provirus integration, we first generated defective SNJ2 viruses lacking ORF1 (encoding Int^{SNJ2}). We introduced the *orfx::pyrF* (x = 1, 1–3) mutation into J7-1-F cells complemented with pYCJ-1416. Upon induction, Int^{SNJ2} excision produced defective SNJ2 viruses marked with *pyrF* (60) (SNJ2*Δorfx::pyrF*, Supplementary Figure S4A). Control mutants were also constructed (*orfx::pyrF)* (x = 2–3, 20–24, 4–24) within the J7-1-F strain to generate their corresponding SNJ2 variants. PCR amplification of *pyrF* and the circularized viral genome containing the *attP'OP* site confirmed virus particle production. PCR of *radA* expression in J7 cells confirmed that there was no contamination by



B

SNJ2Δorfx::pyrF	PCR targets			Existence of virus particles in culture supernatants (Yes or NO)
	radA	*pyrF*	*attP'OP*	
Δorf1::pyrF	-	+	+	Y
Δorf1-3::pyrF	-	+	+	Y
Δorf2-3::pyrF	-	+	+	Y
Δorf20-24::pyrF	-	+	+	Y
Δorf4-25::pyrF	-	-	-	N

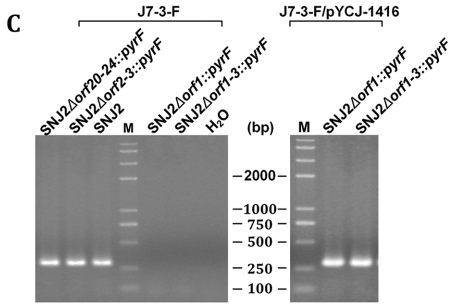


Figure 2. Excision and integration of SNJ2 variants lacking one or more ORFs. (**A**) Schematic containing the different *orfx* deletions in the SNJ2 provirus genome and results of SNJ2*Δorfx* excision. SNJ2*Δorfx* excision from J7-1-F *Δorfx* cells was detected by PCR of the reconstituted *attP'OP* site using primer set b/c (short arrow). 'Y' or 'N' indicated that the *attP'OP* fragment could (Y) or could not (N) be detected from the corresponding knockout strains. (**B**) Detection of SNJ2*Δorfx::pyrF* virus particles in the supernatants of corresponding J7 cultures. The *radA* gene located in the J7 chromosome was used as a control to detect chromosomal DNA contamination. +/−, with/without correct PCR products. (**C**) Integration of the SNJ2*Δorfx::pyrF* viruses was detected by successful amplification of the integration site *attBOP* (292 bp) using primer pairs a/b when J7-3-F and J7-3-F/pYCJ-1416 were infected by the corresponding viruses. Lane M, Trans 2K PlusII marker. Wild-type SNJ2 and H_2O were used in this system as positive and negative controls, respectively.

chromosomal DNA. All marked and defective SNJ2 particles were induced and detected in the J7 culture except for the SNJ2*Δorf4–24::pyrF,* which was not packaged because all genes encoding virus structural proteins were lost (Figure 2B and Supplementary Figure S4B).

Finally, to determine the role of Int^{SNJ2} in integration, SNJ2*Δorf1::pyrF*, SNJ2*Δorf1–3::pyrF,* SNJ2*Δorf2–3::pyrF* and SNJ2*Δorf20–24::pyrF* viruses were incubated with the *Natrinema* sp. J7-3-F cells (Supplementary Table S1) lacking SNJ2 and containing an empty *attBOB'* site. The *attBOP* site would only be detected by PCR from the infected strains using the host-specific primer-a and the virus-specific primer-b (Figure 2A) if the SNJ2 variants could integrate into the host chromosome. Cells infected with SNJ2*Δorf2–3::pyrF,* SNJ2*Δorf20–24::pyrF,* or the wild-type SNJ2 viral infection, but not SNJ2*Δorf1::pyrF* or SNJ2*Δorf1–3::pyrF*, yielded detectable amplification products (Figure 2C), suggesting that Int^{SNJ2} was critical for integration. Accordingly, SNJ2*Δorf1::pyrF*-infected cells were unstable when passaged onto Halo-2 rich culture medium, and could not be selected on Hv-Ca medium. In contrast, incubating J7-3-F/pYCJ-1416 (Supplementary Table S1) cells with SNJ2*Δorf1::pyrF* and SNJ2*Δorf1–3::pyrF* restored integration (Figure 2C), and allowed cells to be stably selected on Hv-Ca medium. These results indicate that SNJ2 lacking *int* could adsorb to and infect J7-3-F cells, but lost the ability to integrate into the host chromosome (Figure 2B and C). Taken together, these findings demonstrated that int^{SNJ2} (*orf1*) was required to maintain the SNJ2 genome in host cells and to site-specifically recombine with the host chromosome.

Int^{SNJ2} is co-transcribed with *orf2* and *orf3*

Previous research indicated that int^{SNJ2} and the adjacent ORF2 and ORF3 are transcribed in the same direction while the nearby ORF4 is transcribed in the opposite direction (Figure 2A) (36). These three genes overlap with one another by 8 (*orf3-orf2*) and 20 (*orf2-orf1*) bp, respectively, indicating that they likely form an operon and are co-transcribed from a promoter upstream of *orf3*. To test this hypothesis, we analyzed the transcription products of int^{SNJ2} and its adjacent ORFs using RT-PCR of the cDNA extracted from J7-1 cells. Int^{SNJ2}, *orf2, orf3* and *orf4* were all transcribed (Figure 3A), and co-transcription of int^{SNJ2} and *orf2*, *orf2* and *orf3*, and int^{SNJ2} and *orf3* was readily detected. However, this was not detected for *orf3* and *orf4*, or int^{SNJ2} and *orf4* (Figure 3A, right panel). These results confirmed that int^{SNJ2} forms an operon with *orf2* and *orf3*, and that the three genes are co-transcribed.

To next determine the *orf1–3* operon transcription start and termination sites, the 5'- and 3'-ends of the transcripts were identified by sequencing the PCR products yielded from two consecutive nested PCR reactions using the circularized cDNA as the template (as illustrated in Supplementary Figure S5). All transcripts initiated at two sites,

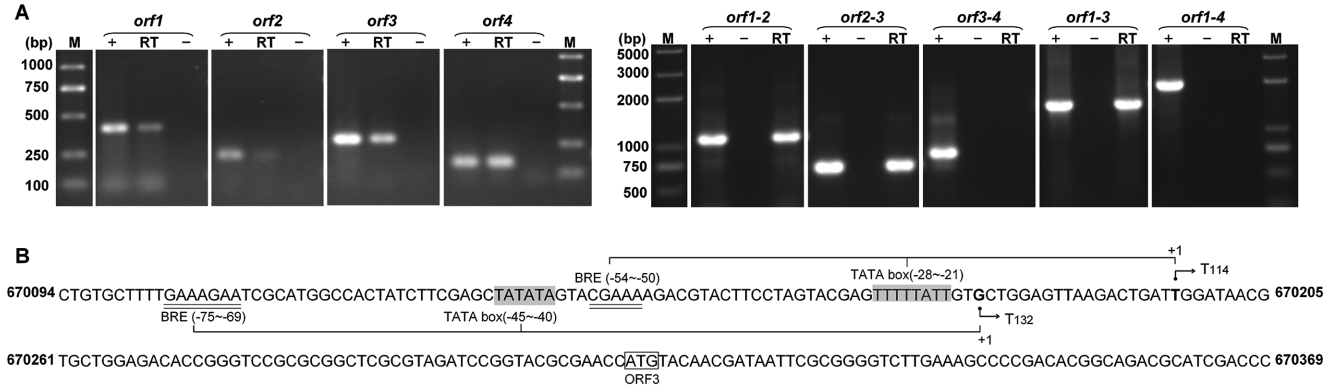


Figure 3. *Orf1*, *orf2* and *orf3* form an operon with the transcription start site located upstream of *orf3*. (**A**) Co-transcription analysis of *orf1* to *orf2* (1034 bp), *orf2* to *orf3* (635 bp) and *orf1* to *orf3* (1474 bp). *Orf4* was not co-transcribed with *orfs1–3*. Lanes: +, positive control with J7–1 genomic DNA as the template; –, negative control with total RNA as the template; RT, RT (Reverse Transcription)-PCR with cDNA as the template; M, DNA marker. (**B**) Promoter and transcription start sites of the operon formed by *orf1-orf3*. Two transcription start sites (+1) are bolded and indicated by a bent arrow. The ORF3 start codon ATG is highlighted in an open box. The predicted TATA boxes are shown in dark gray and the BRE (transcription factor B recognition element) CGAAA motif is underlined.

either 132 or 114 bp upstream of the start codon of *orf3* (ATG^{ORF3}), respectively (Figure 3B). However, transcripts terminated at different positions downstream of int^{SNJ2}, yielding transcripts of variable length. These results suggested that this operon contained two transcription start sites. Indeed, two promoters were identified upstream of *orf3*, including one typical promoter containing the transcription factor B recognition element (BRE, -54 CGAAA -50) (52) upstream of the TATA box (-28 TTTTTATT -21), and another containing a putative purine-rich BRE element (61) (-75 GAAAGAA -69) and the TATA box (-45 TATATA -40) (Figure 3B). The existence of two initiation sites implied that Int^{SNJ2} operon expression is under certain regulation, similar to that of phosphotransferase system in *Haloferax mediterranei* (62).

Orf2 and Orf3 have limited effects on SNJ2 excision but play important roles in virus integration

Genes within an operon often function in the same pathway. Thus, we tested the possibility that *orf2* and *orf3* act together with Int^{SNJ2} to integrate or excise SNJ2 provirus (Figure 2). Products of these two ORFs were detected by tandem mass spectrometry (LC-ESI MS/MS) (36), indicating that both ORFs were translated into proteins. *Orf2* (accession no. AFO55991) encodes a 111-amino acid protein containing a strongly-predicted coiled-coil domain in the N-terminal region, which could mediate homotypic or heterotypic protein–protein interactions. The 141-aa protein product of *orf3* (accession no. AFO55990) contains a MarR-like winged helix-turn-helix DNA-binding motif. Notably, this motif is also found in a group of RDFs that control the recombination direction of the complex site-specific recombination system (63–65).

To investigate whether Orf2 and Orf3 affect site-specific recombination, we first attempted to establish an excision system in J7–3-F by replacing its *attBOB'* site with a fragment containing the *attBOP-pyrF-attB'OP'*; however, this approach was unsuccessful. Instead, we used a chimeric plasmid, pUC-mev-oriN-MH (Supplementary Table S2) containing the replication origin of *Natrinema* sp. J7 (oriN, oriC10 of J7 with an autonomously replicating sequence (48) and the *attBOP-pyrF-attB'OP'* fragment (Supplementary Figure S6A). In the presence of Int^{SNJ2} and accessory factors, recombination between the *attBOP* and *attB'OP'* (Supplementary Figure S6A) on pUC-mev-oriN-MH would invert rather than excise the fragment and concomitantly form the new junction sites, *attBOB'* and *attP'OP,* respectively. The reoriented *pyrF* could complement the *pyrF*-defective strain, J7-3-F, allowing for selection on Hv-Ca plates (Supplementary Figure S6A). Then, the recombination efficiency was evaluated by counting the number of transformants obtained per μg of plasmid DNA. PCR amplification of the recombinant *attBOB'* and *attP'OP* sites and SacI restriction digestion could further verify the recombinants by producing a different restriction pattern of the newly formed *attBOB'-pyrF* fragment compared to the one obtained using the original plasmid (Supplementary Figure S6C). Int^{SNJ2} or its related factors were provided from a series of non-replicating plasmids, pSU19-Y, that carried the int^{SNJ2} operon with different gene deletions: single-gene-deletion of *Δorf1* ($Y = 1048$), *Δorf2* ($Y = 1936$), *Δorf3* ($Y = 1844$), double-gene-deletion of *Δorf2-orf3* ($Y = 1416$) and triple-gene-deletion: *Δorf1–3* ($Y = 367$) (Figure 4A and Supplementary Table S2). Expression of Int^{SNJ2} alone from plasmid pSU19-1416 yielded a recombination efficiency of ∼80% compared to that induced by expression of Int^{SNJ2} with Orf2 and Orf3 from pSU19-2017 (Figure 4A). Deletion of either *orf2* (pSU19-1936) or *orf3* (pSU19-1844) did not affect the recombination efficiency. However, deletion of Int^{SNJ2} or the operon promoter (pSU19-1048 or -367) completely abolished recombination as shown by a lack of transformants (Figure 4A). These results indicated that Int^{SNJ2} mediated excision independent of Orf2 and Orf3, although Orf2 and Orf3 could partially promote this process.

We also studied how these two proteins affected integration. A series of non-replicating plasmids (pCF-X marked by *pyrF*) carrying the *attP'OP* region and different int^{SNJ2} operon gene deletions (Supplementary Table S2) were con-

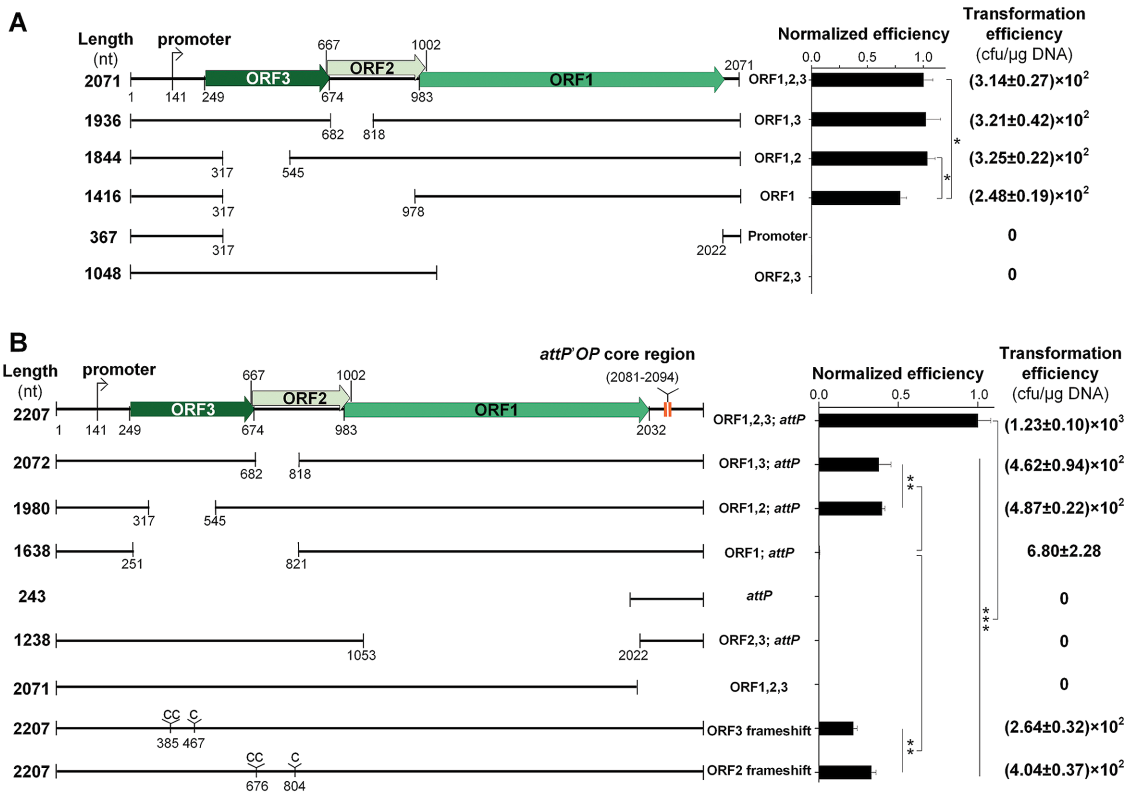


Figure 4. The effects of Orf2 and Orf3 on Int^{SNJ2}-mediated recombination efficiency. (**A**) Scheme of truncated recombination elements and their corresponding excision efficiency. Excision efficiencies are shown as black bars normalized to the fragment of 2071 (whole *orf1-orf3* operon) represented as averages and standard deviations. This assay was performed in triplicate for each construct. Significance testing was performed using a one-sample *t*-test (*$P < 0.05$). (**B**) Scheme of truncated recombination elements and their corresponding integration efficiency. The recombination efficiencies (average) of different plasmids are normalized to pCF-2207 (whole *orf1-orf3* operon with *attP'OP* site), with the efficiency (average ± standard deviation) of integration indicated ($n \geq 3$). Significance testing was performed using a one-sample *t*-test (***$P < 0.001$, **$P < 0.01$).

structed and transformed into J7-3-F cells (Figure 4B and Supplementary Figure S6B). Because J7-3-F is defective in *pyrF*, only transformants that successfully integrated into the host chromosome expressed *pyrF* and grew on Hv-Ca plates (Supplementary Figure S6B). Integrants were further verified by PCR amplification of the recombined *attBOP* and *attP'OB'* (Supplementary Figure S6D). The integration efficiency was evaluated by counting the number of transformants per μg of plasmid DNA. Transformation of the plasmid, pCF-2207, containing the whole operon yielded ~1.23 × 10^3 transformants per μg of DNA on selection plates (Figure 4B). As expected, transformations of plasmids lacking the *int^{SNJ2}* gene, pCF-1238 (*Δint^{SNJ2}*) or pCF-243 (*Δorf1–3*), did not produce transformants (Figure 4B). Interestingly, plasmid pCF-1638 (*Δorf2–3*) yielded less than ten transformants per μg of plasmid DNA, suggesting that Orf2 and Orf3 significantly enhanced integration. Deletion of either Orf3 (pCF-1980) or Orf2 (pCF-2072) decreased the integration efficiency to approximately 40% of the efficiency conferred by the whole operon (Figure 4B). These results indicated that Orf2 and Orf3 partially redundantly promoted Int^{SNJ2} integration. To confirm this, frame-shift mutations were introduced into *orf2* and *orf3* in the plasmid carrying the whole operon. These frame-shift mutations dramatically reduced recombination efficiency (Figure 4B), confirming that both Orf2 and Orf3 were critical to efficient integration.

SNJ2-type site-specific recombination elements are widespread in haloarchaea and are associated with various pleolipoproviral regions and other MGEs

The above results indicated that Int^{SNJ2}-mediated site-specific recombination require the virus-encoded accessory factors for integration but not excision. This distinguished SNJ2 from the typical lambda-type tyrosine Int (66) or phiC31-type serine Int families (67,68), which both require RDFs for excision. Int^{SNJ2} was also distinct from the archaeal SSV-type integrases that catalyze recombination in a simple and reversible fashion (30). To next determine the distribution of these novel recombination systems, we searched the non-redundant protein database at NCBI for the presence of the Int^{SNJ2} homologs using the Int^{SNJ2} amino acid sequence as a query using a Position-Specific Iterated (PSI-) BLAST search. Over 500 integrases encoded on haloarchaeal chromosomes showed an identity higher than 30% with Int^{SNJ2} and the sequence coverage over 90%. From these hits, we selected 46 (denoted as No. 2 to 47) Int^{SNJ2}-like recombinases with the pairwise identities ranging from 34.6 to 88.8% (Supplementary Table S5). These integrases were found in chromosomes of haloarchaea be-

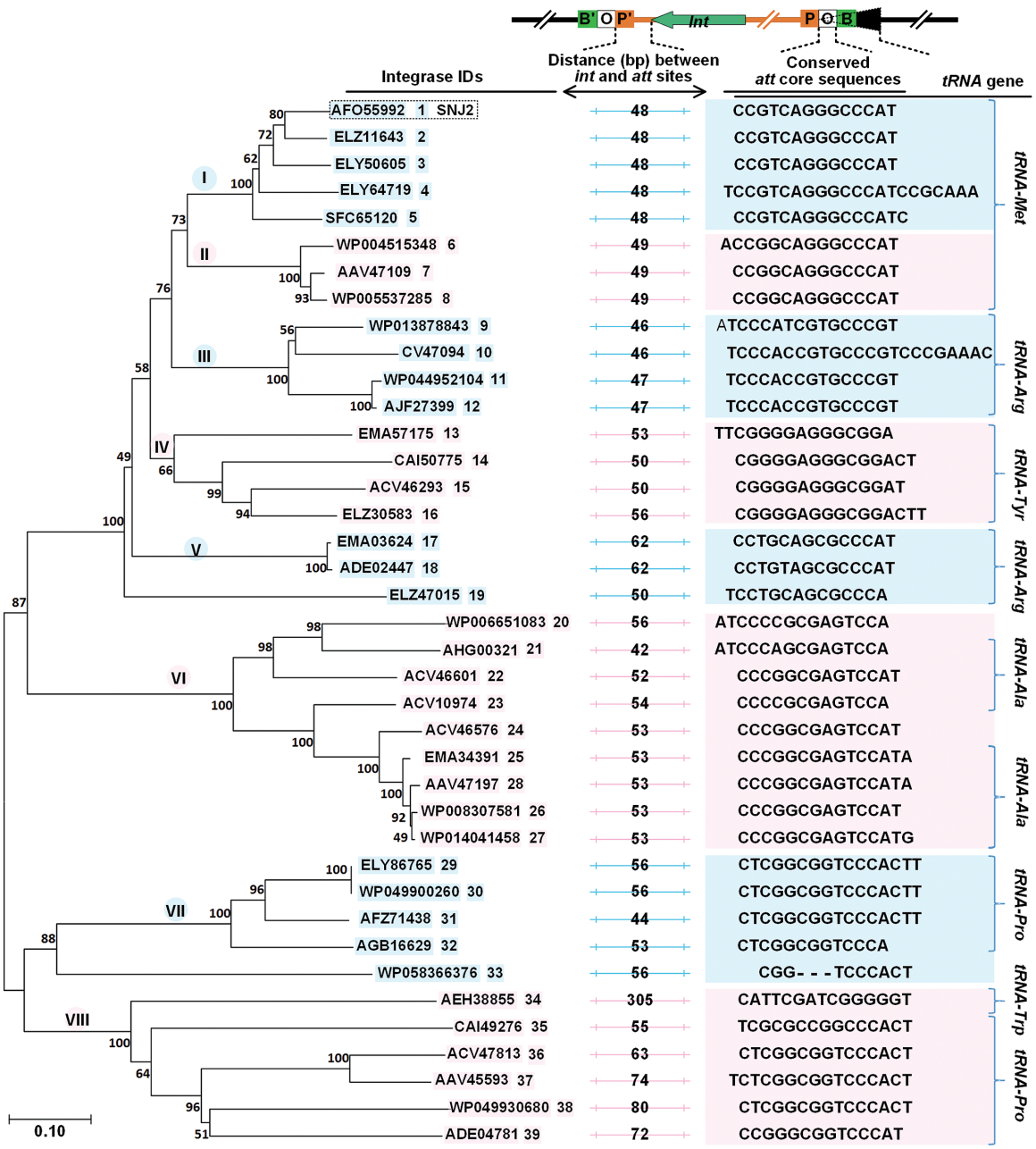


Figure 5. Relatedness of putative Int^{SNJ2}-type integrases and their corresponding conserved core target sequences. The phylogenetic tree of 39 putative haloarchaeal tyrosine integrase was inferred using MEGA7. Numbers at nodes represent percentages of bootstrap support based on a neighbor-joining analysis of 1000 resampled datasets. GenBank identifiers of protein sequences used to generate this phylogenetic tree are presented following the ordinal numbers (1–39, Supplementary Table S5). The distance (bp) between the 3' end of *int* genes and 5' end of *att* core sites are shown in the middle panel. The conserved *att* core sequences targeted by integrases and overlapping with the same *tRNA* genes belonging to eight groups are shown.

longing to 22 different genera and 5 families (Supplementary Table S5). Alignment of the corresponding amino acid sequences revealed that they were all tyrosine recombinases containing the $R_I...K...H_{II}XXR_{II}...Y$ pentad, but with the Glu/Asp_I site replaced by Gly/Ala_I (proportion: 15/31) and His/Trp_{III} replaced by Ala/Val_{III} (proportion: 35/11) (Supplementary Figure S7).

To identify features conserved in the SNJ2-like recombination systems, we attempted to delineate the exact borders of the MGEs encoding these integrases by searching for direct repeats corresponding to the *att* sites and flanking the integrated MGEs. Putative *att* sites were predicted for all MGEs except for elements residing in the incompletely sequenced genomes of *H. thermotolerans* DSM 11522 (No. 20) and *Natrialba chahannaoensis* JCM 10990 (No. 44 and 45) (Supplementary Table S5). In almost all cases, one border of the MGE overlapped with a *tRNA* gene, like the Int^{SNJ2} recombination system. Three exceptions were the *att* sites of MGEs of *Halomicrobium mukohataei* DSM 12286 (No. 24), *Halopenitus* sp. DYS4 (No. 33), and *Natronococ-*

cus jeotgali DSM 18795 (No. 47), which were all located in the intergenic regions (Supplementary Table S5).

Phylogenetic analysis of these Int^{SNJ2}-like integrases revealed that they could be further subdivided into eight subgroups (Supplementary Table S5). In each of the subgroups, integrases and the corresponding *att* sites were coupled. Specifically, the sequences of the *att* sites in each subgroup tended to be identical, using the same *tRNA* target gene, with the notable exception of subgroup VIII (Figure 5 and Supplementary Table S5). In this subgroup, all members targeted *tRNA-Pro* except for the MGE of *Haloplger xanaduensis* SH-6 (No. 34, accession no. AEH38855), which instead used *tRNA-Trp*. Furthermore, the distances between the 3' end of the *int* genes and the 5' end of the *att* sites were often similar or identical within each subgroup (Figure 5). For instance, in subgroup I, all four *att* sites were identical (14 bp-long) and located 48 bp downstream of the 3' end of the *int* genes (Figure 5). The uniform arrangement of the recombination elements in each subgroup suggested that they originated from a common ancestor and spread in different chromosomes by HGT.

Because integration reaction with the SNJ2 recombination system was highly dependent on Orf2 and Orf3, we searched for potential homologs of these proteins in proximity of the *int* genes carried by other haloarchaeal MGEs and containing a coiled-coil structure and DNA-binding domain. In the MGEs, we identified 11 proteins containing coiled-coil domains (ORFs indicated with bright green arrows, Figure 6) and 25 proteins with DNA-binding domains (ORFs indicated in dark green, Figure 6). However, further analysis revealed that only five MGEs, No. 4, 20, 29, 30 and 38, had SNJ2 Orf2 and Orf3 homologs. In 14 MGEs, no candidates for either domain-containing protein were identified, presumably because they had defective recombination systems or evolved to function without accessory factors.

Analysis of the 38 MGEs containing SNJ2-like recombination systems revealed that 33 were pleolipoproviruses (or remnants thereof), with 4 belonging to the genus *Alphapleolipovirus* and 29 to *Betapleolipovirus* (Figure 6). The remaining five were MGEs with the majority of their ORFs encoding functionally uncharacterized proteins with no known counterparts (Figure 6). As expected, all putative pleolipoprovirus regions encoded conserved core genes (VP3-like membrane protein, VP4-like spike protein and VP8-like ATPase shown with purple arrows) (35,69). Moreover, 11 of these putative pleolipoproviral regions contained SNJ2 Orf8-like protein (36), which is homologous to archaeal Orc1/Cdc6 proteins (origin recognition complex 1/cell division control protein 6) (70). This suggested that these proviral regions may be excised from the host to form circular genomes and replicate using host machinery. Notably, two elements, No. 2 (assembled, Supplementary Figure S8) and 19, lacked all the core pleolipoviral genes encoding proteins involved in virion formation and, as a result, were reduced to gene cassettes encoding only the putative replication-associated protein and the integration/excision module (Figure 6).

SNJ2-type site-specific recombinases from three haloarchaea are functional

We demonstrated that Int^{SNJ2}-type site-specific recombination systems are widespread in haloarchaea and are associated with various MGEs. To determine if these recombinases were functional, we further analyzed SNJ2-type integrases from subgroup I (Figure 5). These integrases were encoded within haloarchaea belonging to different genera of the order *Natrialbales* and showed high sequence identity (≥81.4%) to Int^{SNJ2} (Supplementary Table S5). Like SNJ2, their associated MGEs were excised from the host chromosome as closed circular molecules, subsequently reforming the native *tRNA* sequences (Figure 7A). Nucleotide sequence alignments of the putative DNA recombination junctions (*attBOP*, *attP'OB'*, *attP'OP*, *attBOB'*) from *Natrinema* sp. J7-1, *H. thermotolerans* DSM 11522, *N. bangense* JCM 10635, *N. versiforme* JCM 10478 and *Halobiforma haloterrestris* DSM 13078 (indicated as No. 1, 2, 3, 4, 5) revealed not only the conserved common core *att* site (O), but also highly conserved flanking sequences (B arm on the chromosome and P/P' arms on MGEs) (Figure 7A). Additionally, all five haloarchaea harbored an identical 23 bp continuous sequence in the B arm. The B arms of JCM 10635 (No. 3) and DSM 13078 (No. 5) were more similar to that of J7-1 (No. 1) than to DSM 11522 (No. 2) and JCM 10478 (No. 4), as they shared more consensus nucleotide positions and longer *tRNA* genes. Thus, we chose the three representative haloarchaea (No. 2, 3, 4) to test whether their MGEs were active. PCR and sequencing data obtained from the amplicons of the circularized *attP'OP* sites (Supplementary Figure S9A) and reformed *tRNA* genes confirmed that the precise MGE termini and the putative attachment sites were involved in recombination, indicating that the proviral regions were active (Figure 7A).

To test the ability of the three MGEs to site-specifically integrate into the SNJ2 free strain, J7-3-F (*ΔpyrF*), that carried only the *attBOB'* site, we amplified and ligated fragments containing *int-attP'OP* cassette (including the native predicted promoter, integrase and *att* site, Figure 7B) of each of the three proviral regions into the non-replicating plasmid, pCF, carrying *pyrF*. Following plasmid transformation into J7-3-F (*ΔpyrF*) cells, only transformants from plasmids carrying the *int-attP'OP* operons integrated into the chromosome would grow on Hv-Ca plates (as illustrated in Supplementary Figure S6B). As expected, integrants were obtained from cells transformed with plasmids carrying any one of the three tested *int-attP'OP* operons, but not with plasmids lacking the *int* or the *attP'OP* site (Figure 7B). Furthermore, fragments corresponding to the *attBOP* and *attP'OB'* sites were amplified from cultures of these integrants (Supplementary Figure S9B), suggesting that all three recombinases mediated site-specific integration (Figure 7B).

Together, these results showed that the three tested Int^{SNJ2}-type recombination systems catalyzed cross-reactive integration in SNJ2 host J7-3-F cells, suggesting that they may be also able to substitute for Int^{SNJ2} during excision. To test this, we used strains, J7-1-F *Δorf1–3* and J7-1-F *Δorf1*, with an intact *att* site but defective for SNJ2

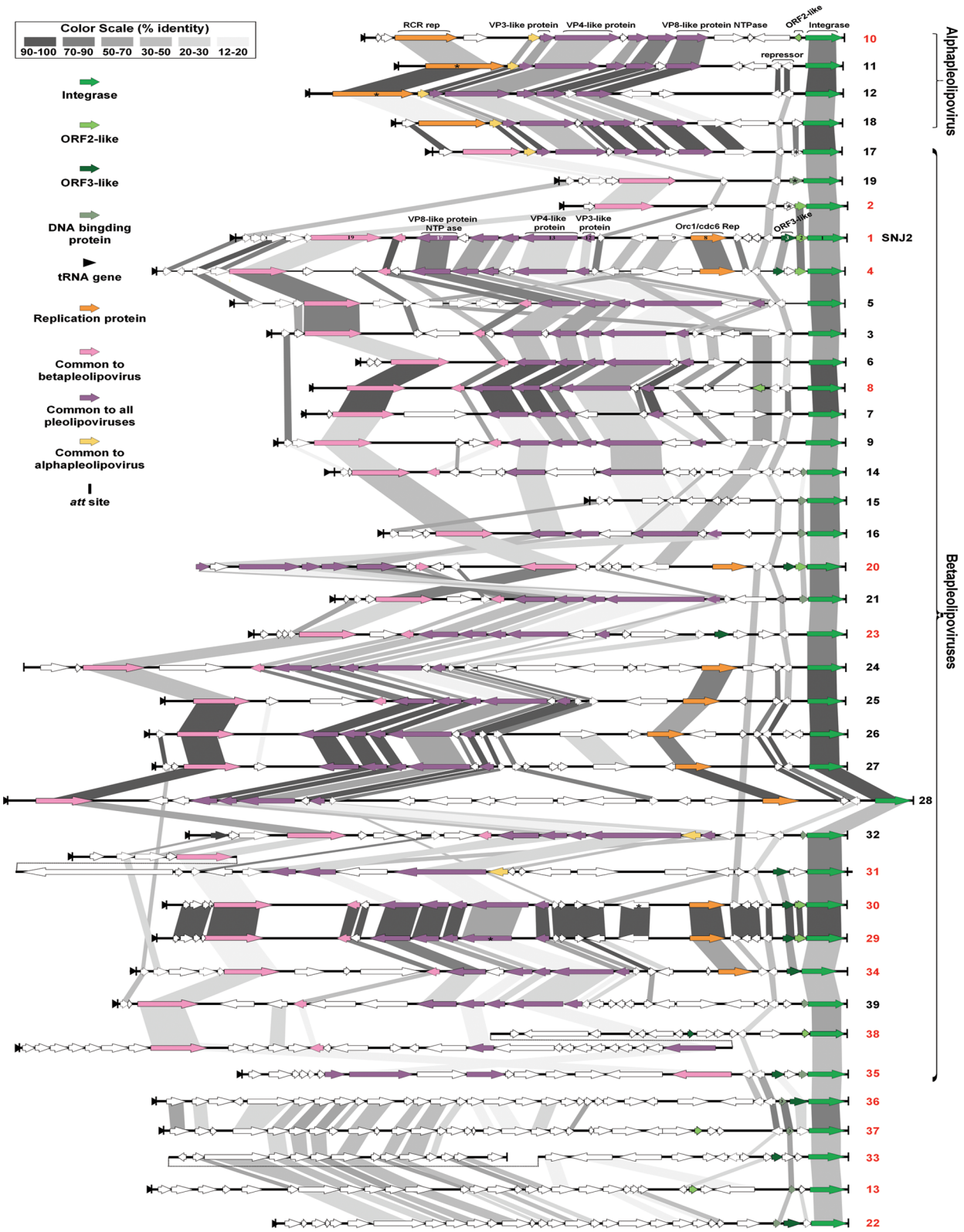


Figure 6. Int^{SNJ2}-type integrases are associated with various MGEs. Genes belonging to the same functional categories are indicated by the use of the same colors (top left legend) in the genome comparison between SNJ2 (1) and the other MGEs (2–39, ordinal numbers consistent with Figure 5). The level of identity between different homologs is indicated in shades of gray (top left legend). The proviral regions encoding the typical Int^{SNJ2}-type recombination elements are numbered in red, including integrases (green arrow) and Orf2^{SNJ2}-like (bright green arrow)/Orf3^{SNJ2}-like (dark green solid arrow) proteins. Newly predicted putative ORFs are marked by an asterisk.

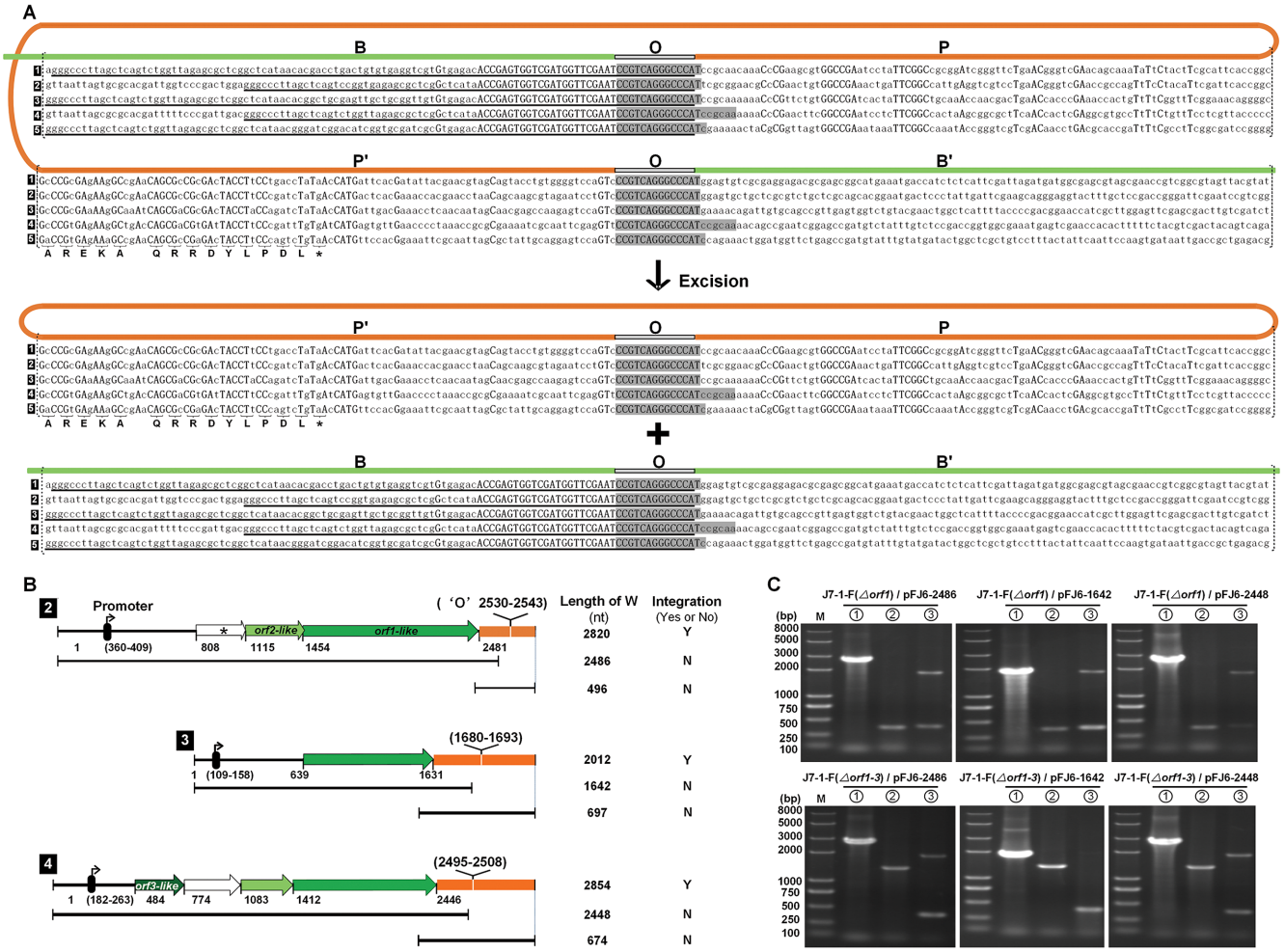


Figure 7. Int^{SNJ2}-type recombination elements mediated excision and integration in *Natrinema* sp. J7. (**A**) Diagram of MGE excision in J7-1 (1), *Haloterrigena thermotolerans* DSM 11522 (2), *Natronorubrum bangense* JCM 10635 (3), *Natrinema versiforme* JCM 10478 (4) and *Halobiforma haloterrestris* DSM 13078 (5). The *att* core (O) site (sequences shaded in gray) overlapped with the 3' end of the *tRNA* gene (underlined sequences) of the corresponding organism. The archaeal chromosome and proviral (MGEs) DNA flanking the core sites are designated as B/B' (green line) and P/P' (orange line) arms, respectively. Identical nucleotides are bolded (uppercase) in the alignments of the sites (B, B', O, P, P') from five different haloarchaea. Consensus amino acid sequences of Int (C-terminal) are shown below the nucleotide sequences. (**B**) Cross integration reactivity of three Int^{SNJ2}-type recombination elements (origin from strains 2, 3, 4 in [**A**]). Int^{SNJ2}-type elements are shown as operons with putative promoter (predicted by the promoter predictor NNPP version 2.2). 'Y' or 'N' indicated the fragments with different operon length that could (Y) or could not (N) integrate into J7–3-F. (**C**) Cross excision reactivity of Int^{SNJ2}-type recombination elements in the excision defective strains J7–1-F *Δorf1* and J7–1-F *Δorf1–3*. Excision was confirmed by PCR amplification of the fragments W (2486, 1642, 2448) in defective strains, which were complemented by plasmids pFJ6-W (lane ①, circularized *attPOP'* (lane ②, 363 bp for SNJ2*Δorf1*; 1355 bp for SNJ2*Δorf1–3*) and restored chromosome junction *attBOB'* (lane ③, 387 bp).

excision (as illustrated in Figure 2A). Plasmids pFJ6-W (W = 2486, 1642 and 2448; Figure 7B and Supplementary Table S2) containing the integrase genes, putative SNJ2 *orf2* or *orf3*-like accessory protein-encoding genes, and the predicted promoters of the three proviral regions were constructed and transformed into Int^{SNJ2}-defective host cells. The transformed cultures of J7-1-F *Δorf1–3* and J7-1-F *Δorf1* were induced by MMC, and cell-free supernatants were analyzed for viral presence (SNJ2*Δorf1–3* and SNJ2*Δorf1*). PCR revealed *attP'OP* site amplification from culture supernatants (Figure 7C), indicating that all three integrases recognized the *att* site on the J7 chromosome and excised the defective SNJ2 provirus. Collectively, these results suggest that SNJ2-type recombinases from subgroup I mediate both integration and excision in J7 cells.

DISCUSSION

Int^{SNJ2} belongs to a novel family of tyrosine recombinases

In this study, we identified over 500 SNJ2-type integrases (including Int^{SNJ2}) encoded within haloarchaeal genomes. Comparison of these integrases with tyrosine recombinases from other known families revealed that SNJ2-type integrases belong to a distinct family. Three main features distinguish SNJ2-type integrases from the known hyperthermophilic archaeal SSV-type and pNOB8-type integrases. First, Int^{SNJ2}-type integrases have a catalytic

tetrad of $R_I...H_{II}XXR_{II}...Y$ that is conserved in tyrosine recombinases from Bacteria and Eukarya (Table 1), while SSV- and pNOB8-type integrases usually contain the $R_I...K_{II}XXR_{II}...Y$ and $R_I...Y_{II}XXR_{II}...Y$ tetrad sequences, respectively. Second, in Int^{SNJ2}-like integrases, the two conserved sequence motifs Glu/Asp_I and His/Trp_{III}, which play important structural roles and stabilize the transition state of the integrase–DNA complex, are replaced with Gly/Ala and Ala/Val, respectively (Supplementary Figure S7). Further mutagenesis of these conserved residues in Int^{SNJ2} confirmed their essentiality for recombination activity (Supplementary Table S6). Third, Int^{SNJ2}-type recombinases are deeply rooted within the major clade of archaeal tyrosine recombinases. Specifically, SNJ2-type integrases diverged from the remaining archaeal homologs early in evolution to catalyze recombination in the intracellular environment specific to halophilic archaea (Figure 1).

SNJ2-type integrases could be classified into eight subgroups based on the similarity of their amino acid sequences. Interestingly, each subgroup had similar or sometimes identical core sequences of the *att* sites (i.e. the *tRNA* genes used as insertion targets) and distances between the 3' end of *int* genes and the 5' end of the core sequences of the *att* sites (Figure 5). However, the integrated MGEs in each Int subgroup did not necessarily belong to the same MGE family. For instance, subgroups I and V included both alphapleolipoviruses and betapleolipoviruses, whereas the novel uncharacterized MGEs, despite their relatively close genomic relationship (Figure 6), were distributed across subgroups IV, VI, VII and VIII, and were accordingly integrated into diverse *tRNA* genes (Figure 5). This indicated that the integrase genes, auxiliary factors and cognate *att* site form a recombination module that is horizontally exchanged between diverse haloarchaeal MGEs. Furthermore, phylogenetic analysis indicated that integrases do not cluster according to cellular taxonomy, but rather according to the specificity of the target *att* site, indicating that their evolution is coupled. A similar observation has been made for SSV-type integrases encoded by hyperthermophilic spindle-shaped viruses (71), suggesting that this may be a general feature of archaeal site-specific recombination systems. Further investigation of the relationship between the structure of integrases and their recognized sites may reveal the underlying basis of this phenomenon (72,73).

The role of Int^{SNJ2} in the maintenance of SNJ2 virus in host cells

All previously isolated pleolipoviruses are not lytic, and are instead capable of continuously releasing virions without rupturing the cell envelope. This only slightly retards host cell growth. Such viruses are produced by infected cells even after several culture passages (74,75). Similar to other members of *Pleolipoviridae*, SNJ2 is non-lytic, although it was difficult to measure host cell growth during viral release because SNJ2 is co-produced with another virus, the sphaerolipovirus SNJ1 that is lytic upon induction (36). Notably, even though pleolipovirus-derived proviruses are found in numerous haloarchaeal genomes (36,74,76), to date, SNJ2 is the only isolated temperate pleolipovirus. In this study, Int^{SNJ2} was critical not only for the excision and integration of the SNJ2 genome, but also improved its maintenance in host cells. As expected, integrase-deficient virus, SNJ2*Δint::pyrF*, still infected J7-3-F host cells, but lost the ability to integrate into the host chromosome. Unexpectedly, the integrase-deficient virus could not persist in infected cells, as revealed by the lack of colonies when infected cells (J7-3-F SNJ2*Δint::pyrF*) were grown on selective Hv-Ca plates. This contrasts with other studied pleolipoviruses, which remain in host cells even without integration into the host genome (74,75). It is possible that Int^{SNJ2} evolved to function in virus replication and/or viral genome partitioning during cell division. Indeed, such roles have been previously reported for tyrosine recombinases encoded by the coliphage P1 and the two-micron plasmid of *Saccharomyces cerevisiae*, which are maintained in host cells due to the presence of Cre/*lox* and Flp/*frt* site-specific recombination systems, respectively (14,77,78). Thus, while most non-integrating betapleolipoviruses and alphapleolipoviruses persist in the carrier state for several host passages, SNJ2 relies on the Int^{SNJ2} not only for integration/excision, but possibly also for episomal viral genome partitioning into daughter cells.

The roles of $Orf2^{SNJ2}$ and $Orf3^{SNJ2}$ during Int^{SNJ2}-mediated recombination

Although Int^{SNJ2} alone catalyzed SNJ2 provirus excision in J7 cells, $Orf2^{SNJ2}$ and $Orf3^{SNJ2}$ significantly increased integration efficiency (Figure 4). $Orf2^{SNJ2}$ and $Orf3^{SNJ2}$ formed an operon with Int^{SNJ2}, but not all identified Int^{SNJ2}-like recombination systems harbored the two integration factors. This suggested that Int^{SNJ2} was distinct from the simple bidirectional recombinases (e.g. Cre, Flp and SSV-type integrases), which catalyze recombination alone in a simple and reversible fashion, or the complex unidirectional recombinases (e.g. λ, HP1, P2), which need accessory proteins (e.g. IHF and Xis or Cox) to catalyze excision (18,20,79). $Orf2^{SNJ2}$ contains a coiled-coil domain and might belong to the Prefoldin superfamily of molecular chaperones that modulate protein folding and stabilization, suggesting that Orf2 may interact with Int^{SNJ2} to promote its folding or stabilize certain intermediate conformations during recombination. $Orf3^{SNJ2}$ contains the MarR-like winged helix-turn-helix domain and is likely to act as a DNA-binding protein that interacts with the conserved flanking sequences (B/B' and P'/P arms in Figure 7A) of the core *att* site (O) during integration. We hypothesize that $Orf2^{SNJ2}$ and $Orf3^{SNJ2}$ act as molecular companions of Int^{SNJ2} to influence the structure of the Int^{SNJ2}–DNA complex and increase recombination efficiency thereby regulating the balance between integration and excision activities of Int^{SNJ2} *in vivo*.

Int^{SNJ2}-type site-specific recombination systems promote fluidity of the haloarchaeal genomes

Various MGEs, including viruses, plasmids, and transposons, play key roles in the evolution of their hosts by promoting genome plasticity and HGT through site-specific and homologous recombination (59,80,81). We showed that SNJ2-type site-specific recombination system is widespread and associated not only with various pleolipoproviral regions but also with novel MGEs of unknown provenance

(Supplementary Table S5). Notably, the SNJ2-like recombination module provided the only connection between these novel MGEs and pleolipoviruses (Figure 6). Such loose connectivity between different families of MGEs is consistent with the previous observation that the global archaeal virus network is sparsely interconnected through a small number of connector genes that function at the MGE–host interface (82). The Int^{SNJ2}-type-mediated integration of MGEs promotes gene flow into the host chromosomes. For instance, *H. mukohataei* DSM 12286 and *Haloarcula marismortui* ATCC 43049 harbors five (No. 10, 15, 22, 24 and 36) and three (No. 7, 28 and 37) MGEs (Figure 6), respectively. Moreover, integrases encoded by proviruses residing within one haloarchaeal genus can catalyze recombination in haloarchaea belonging to at least two other genera, suggesting that Int^{SNJ2}-type recombination systems can promote inter-generic gene flow. This was previously observed in haloarchaea residing in an isolated Antarctic lake (83). Finally, recombination plays a key role in the evolution of MGEs and even drives transitions between different MGE classes (82,84–86). In this context, the two MGEs evidently derived from betapleolipoviruses through loss of the virion morphogenetic module (indicated No. 2 and 19 in Figure 6) are particularly interesting. Likely, these two elements propagate as integrative plasmids devoid of the extracellular stage typical of bona fide viruses, and epitomize the dynamics of evolutionary transition between viruses and plasmids.

In summary, a novel archaeal tyrosine recombinase encoded by the SNJ2 virus mediates site-specific recombination, which substantially differs from the previously characterized simple or complex recombination systems. Int^{SNJ2}-like integrases are widespread in genomes of halophilic archaea associated with various pleolipoproviruses and novel MGEs of unknown provenance. Future studies involving the recombination systems of haloarchaeal MGEs are necessary to further classify genome evolution and gene flow in haloarchaea, and to discover new viruses hidden in the abundant 'dark matter' (34) islands in archaeal genomes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Prof. Yuping Huang (College of Life Sciences, Wuhan University) for kindly providing the pNBF plasmid. We also thank Dr Shisheng Du (Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center) for helpful revising the manuscript.

FUNDING

National Natural Science Foundation of China [31570174]; National Found for Fostering Talents of Basic Sciences [J1103513]; Research (Innovative) Fund of Laboratory Wuhan University. Funding for open access charge: National Natural Science Foundation of China [31570174].
Conflict of interest statement. None declared.

REFERENCES

1. Stark,W.M. (2017) Making serine integrases work for us. *Curr. Opin. Microbiol.*, **38**, 130–136.
2. Hirano,N., Muroi,T., Takahashi,H. and Haruki,M. (2011) Site-specific recombinases as tools for heterologous gene integration. *Appl. Microbiol. Biotechnol.*, **92**, 227–239.
3. Brown,D.P., Idler,K.B., Backer,D.M., Donadio,S. and Katz,L. (1994) Characterization of the genes and attachment sites for site-specific integration of plasmid pSE101 in Saccharopolyspora erythraea and Streptomyces lividans. *Mol. Gen. Genet.*, **242**, 185–193.
4. Kubo,A., Kusukawa,A. and Komano,T. (1988) Nucleotide sequence of the rci gene encoding shufflon-specific DNA recombinase in the IncI1 plasmid R64: homology to the site-specific recombinases of intergrase family. *Mol. Genet. Genomics*, **213**, 30–35.
5. Wang,Z., Xiong,G. and Lutz,F. (1995) Site-specific integration of the phage phi CTX genome into the Pseudomonas aeruginosa chromosome: characterization of the functional integrase gene located close to and upstream of attP. *Mol. Gen. Genet.*, **246**, 72–79.
6. Lesic,B., Bach,S., Ghigo,J.-M., Dobrindt,U., Hacker,J. and Carniel,E. (2004) Excision of the high-pathogenicity island of Yersinia pseudotuberculosis requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol. Microbiol.*, **52**, 1337–1348.
7. Farrugia,D.N., Elbourne,L.D., Mabbutt,B.C. and Paulsen,I.T. (2015) A novel family of integrases associated with prophages and genomic islands integrated within the tRNA-dihydrouridine synthase A (dusA) gene. *Nucleic Acids Res.*, **43**, 4547–4557.
8. Hall,R.M. and Collis,C.M. (1995) Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.*, **15**, 593–600.
9. Johnson,C.M. and Grossman,A.D. (2015) Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.*, **49**, 577–601.
10. Landy,A. (1989) Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.*, **58**, 913–949.
11. Barre,F.X., Soballe,B., Michel,B., Aroyo,M., Robertson,M. and Sherratt,D. (2001) Circles: the replication-recombination-chromosome segregation connection. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8189–8195.
12. Hughes,O.Z.N. and K,T. (1998) In vivo identification of intermediate stages of the DNA inversion reaction catalyzed by the Salmonella Hin recombinase. *Genetics*, **149**, 1649–1663.
13. Liu,Y.-T., Sau,S., Ma,C.H., Kachroo,A.H., Rowley,P.A., Chang,K.M., Fan,H.F. and Jayaram,M. (2014) The partitioning and copy number control systems of the selfish yeast plasmid: an optimized molecular design for stable persistence in host cells. *Microbiol. Spectr.*, **2**, doi:10.1128/microbiolspec.PLAS-0003-2013.
14. Austin,S., Ziese,M. and Sternberg,N. (1981) A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell*, **25**, 729–736.
15. Mazel,D. (2006) Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.*, **4**, 608–620.
16. Groth,A.C. and Calos,M.P. (2004) Phage Integrases: Biology and Applications. *J. Mol. Biol.*, **335**, 667–678.
17. Fogg,P.C., Colloms,S., Rosser,S., Stark,M. and Smith,M.C. (2014) New applications for phage integrases. *J. Mol. Biol.*, **426**, 2703–2716.
18. Lewis,J.A. and Hatfull,G.F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.*, **29**, 2205–2216.
19. Saha,S., Haggard-Ljungquist,E. and Nordstrom,K. (1987) The cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. *EMBO J.*, **6**, 3191–3199.
20. Frumerie,C., Sylwan,L., Ahlgren-Berg,A. and Haggard-Ljungquist,E. (2005) Cooperative interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox. *Virology*, **332**, 284–294.
21. Cali,S., Spoldi,E., Piazzolla,D., Dodd,I.B., Forti,F., Deho,G. and Ghisotti,D. (2004) Bacteriophage P4 Vis protein is needed for prophage excision. *Virology*, **322**, 82–92.

22. Meinke,G., Bohm,A., Hauber,J., Pisabarro,M.T. and Buchholz,F. (2016) Cre recombinase and other tyrosine recombinases. *Chem. Rev.*, **116**, 12785–12820.
23. Gibb,B., Gupta,K., Ghosh,K., Sharp,R., Chen,J. and Van Duyne,G.D. (2010) Requirements for catalysis in the Cre recombinase active site. *Nucleic Acids Res.*, **38**, 5817–5832.
24. Krogh,B.O. and Shuman,S. (2000) Catalytic mechanism of DNA topoisomerase IB. *Mol. Cell*, **5**, 1035–1041.
25. Nunes-Duby,S.E., Kwon,H.J., Tirumalai,R.S., Ellenberger,T. and Landy,A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.*, **26**, 391–406.
26. Clore,A.J. and Stedman,K.M. (2007) The SSV1 viral integrase is not essential. *Virology*, **361**, 103–111.
27. Letzelter,C., Duguet,M. and Serre,M.C. (2004) Mutational analysis of the archaeal tyrosine recombinase SSV1 integrase suggests a mechanism of DNA cleavage in trans. *J. Biol. Chem.*, **279**, 28936–28944.
28. Serre,M.C., Letzelter,C., Garel,J.R. and Duguet,M. (2002) Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. *J. Biol. Chem.*, **277**, 16758–16767.
29. Zhan,Z., Ouyang,S., Liang,W., Zhang,Z., Liu,Z.J. and Huang,L. (2012) Structural and functional characterization of the C-terminal catalytic domain of SSV1 integrase. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 659–670.
30. Zhan,Z., Zhou,J. and Huang,L. (2015) Site-specific recombination by SSV2 Integrase: substrate requirement and domain functions. *J. Virol.*, **89**, 10934–10944.
31. She,Q., Shen,B. and Chen,L. (2004) Archaeal integrases and mechanisms of gene capture. *Biochem. Soc. Trans.*, **32**, 222–226.
32. She,Q., Phan,H., Garrett,R.A., Albers,S.V., Stedman,K.M. and Zillig,W. (1998) Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from an archaeon. *Extremophiles*, **2**, 417–425.
33. Krupovic,M., Forterre,P. and Bamford,D.H. (2010) Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. *J. Mol. Biol.*, **397**, 144–160.
34. Makarova,K.S., Wolf,Y.I., Forterre,P., Prangishvili,D., Krupovic,M. and Koonin,E.V. (2014) Dark matter in archaeal genomes: a rich source of novel mobile elements, defense systems and secretory complexes. *Extremophiles*, **18**, 877–893.
35. Pietilä,M.K., Roine,E., Sencilo,A., Bamford,D.H. and Oksanen,H.M. (2016) Pleolipoviridae, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. *Arch. Virol.*, **161**, 249–256.
36. Liu,Y., Wang,J., Liu,Y., Wang,Y., Zhang,Z., Oksanen,H.M., Bamford,D.H. and Chen,X. (2015) Identification and characterization of SNJ2, the first temperate pleolipovirus integrating into the genome of the SNJ1-lysogenic archaeal strain. *Mol. Microbiol.*, **98**, 1002–1020.
37. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
38. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
39. Marchler-Bauer,A., Bo,Y., Han,L., He,J., Lanczycki,C.J., Lu,S., Chitsaz,F., Derbyshire,M.K., Geer,R.C., Gonzales,N.R. *et al.* (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.*, **45**, D200–D203.
40. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
41. Soding,J., Biegert,A. and Lupas,A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244–W248.
42. Pei,J. and Grishin,N.V. (2014) PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol. Biol.*, **1079**, 263–271.
43. Capella-Gutierrez,S., Silla-Martinez,J.M. and Gabaldon,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
44. Guindon,S., Dufayard,J.F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.
45. Kumar,S., Stecher,G. and Tamura,K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, **33**, 1870–1874.
46. Wang,Y., Sima,L., Lv,J., Huang,S., Liu,Y., Wang,J., Krupovic,M. and Chen,X. (2016) Identification, characterization, and application of the replicon region of the halophilic temperate sphaerolipovirus SNJ1. *J. Bacteriol.*, **198**, 1952–1964.
47. Huff,J.P., Grant,B.J., Penning,C.A. and Sullivan,K.F. (1990) Optimization of routine transformation of Escherichia coli with plasmid DNA. *Biotechniques*, **9**, 570–577.
48. Wang,Y., Chen,B., Sima,L., Cao,M. and Chen,X. (2017) Construction of expression shuttle vectors for the Haloarchaeon natrinema sp. J7 based on its chromosomal origins of replication. *Archaea*, **2017**, 4237079.
49. Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.
50. Bartolomé,B., Jubete,Y., Martínez,E. and de la Cruz,F. (1991) Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. *Gene*, **102**, 75–78.
51. Bitan-Banin,G., Ortenberg,R. and Mevarech,M. (2003) Development of a gene knockout system for the halophilic archaeon Haloferax volcanii by use of the pyrE gene. *J. Bacteriol.*, **185**, 772–778.
52. Brenneis,M., Hering,O., Lange,C. and Soppa,J. (2007) Experimental characterization of cis-acting elements important for translation and transcription in halophilic archaea. *PLoS Genet.*, **3**, e229.
53. Zhang,Z., Liu,Y., Wang,S., Yang,D., Cheng,Y., Hu,J., Chen,J., Mei,Y., Shen,P., Bamford,D.H. *et al.* (2012) Temperate membrane-containing halophilic archaeal virus SNJ1 has a circular dsDNA genome identical to that of plasmid pHH205. *Virology*, **434**, 233–241.
54. Castillo,F., Benmohamed,A. and Szatmari,G. (2017) Xer site specific recombination: double and single recombinase systems. *Front. Microbiol.*, **8**, 453.
55. Cortez,D., Quevillon-Cheruel,S., Gribaldo,S., Desnoues,N., Sezonov,G., Forterre,P. and Serre,M.C. (2010) Evidence for a Xer/dif system for chromosome resolution in archaea. *PLoS Genet.*, **6**, e1001166.
56. Pagaling,E., Haigh,R.D., Grant,W.D., Cowan,D.A., Jones,B.E., Ma,Y., Ventosa,A. and Heaphy,S. (2007) Sequenceanalysis of an Archaeal virus isolated from a hypersaline lake in InnerMongolia, China. *BMC Genomics*, **8**, 410.
57. Klein,R., Baranyi,U., Rössler,N., Greineder,B., Scholz,H. and Witte,A. (2002) Natrialbamagadii virus phiCh1: first complete nucleotide sequence and functionalorganization of a virus infecting a haloalkaliphilic archaeon. *Mol. Microbiol.*, **45**, 851–863.
58. Klein,R., Rossler,N., Iro,M., Scholz,H. and Witte,A. (2012) Haloarchaeal myovirus phiCh1 harbours a phase variation system for the production of protein variants with distinct cell surface adhesion specificities. *Mol. Microbiol.*, **83**, 137–150.
59. Cossu,M., Badel,C., Catchpole,R., Gadelle,D., Marguet,E., Barbe,V., Forterre,P. and Oberto,J. (2017) Flipping chromosomes in deep-sea archaea. *PLoS Genet.*, **13**, e1006847.
60. Peck,R.F., Dassarma,S. and Krebs,M.P. (2000) Homologous gene knockout in the archaeon Halobacterium salinarum with ura3 as a counterselectable marker. *Mol. Microbiol.*, **35**, 667–676.
61. Soppa,J. (1999) Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. *Mol. Microbiol.*, **31**, 1589–1592.
62. Cai,L., Cai,S., Zhao,D., Wu,J., Wang,L., Liu,X., Li,M., Hou,J., Zhou,J., Liu,J. *et al.* (2014) Analysis of the transcriptional regulator GlpR, promoter elements, and posttranscriptional processing involved in fructose-induced activation of the phosphoenolpyruvate-dependent sugar phosphotransferase system in Haloferax mediterranei. *Appl. Environ. Microbiol.*, **80**, 1430–1440.
63. Lewis,J.A. and Hatfull,G.F. (2000) Identification and characterization of mycobacteriophage L5 excisionase. *Mol. Microbiol.*, **35**, 350–360.

64. Bibb,L.A. and Hatfull,G.F. (2002) Integration and excision of the Mycobacterium tuberculosis prophage-like element, phiRv1. *Mol. Microbiol.*, **45**, 1515–1526.
65. Salmi,D., Magrini,V., Hartzell,P.L. and Youderian,P. (1998) Genetic determinants of immunity and integration of temperate Myxococcus xanthus phage Mx8. *J. Bacteriol.*, **180**, 614–621.
66. Abremski,K. and Gottesman,S. (1981) Site-specific recombination Xis-independent excisive recombination of bacteriophage lambda. *J. Mol. Biol.*, **153**, 67–78.
67. Khaleel,T., Younger,E., McEwan,A.R., Varghese,A.S. and Smith,M.C. (2011) A phage protein that binds phiC31 integrase to switch its directionality. *Mol. Microbiol.*, **80**, 1450–1463.
68. Pokhilko,A., Zhao,J., Ebenhoh,O., Smith,M.C., Stark,W.M. and Colloms,S.D. (2016) The mechanism of varphiC31 integrase directionality: experimental analysis and computational modelling. *Nucleic Acids Res.*, **44**, 7360–7372.
69. Sencilo,A., Paulin,L., Kellner,S., Helm,M. and Roine,E. (2012) Related haloarchaeal pleomorphic viruses contain different genome types. *Nucleic Acids Res.*, **40**, 5523–5534.
70. Hori,Y., Shirahige,K., Obuse,C., Tsurimoto,T. and Yoshikawa,H. (1996) Characterization of a novel CDC gene (ORC1) partly homologous to CDC6 of Saccharomyces cerevisiae. *Mol. Biol. Cell*, **7**, 409–418.
71. Held,N.L. and Whitaker,R.J. (2009) Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. *Environ. Microbiol.*, **11**, 457–466.
72. Farruggio,A.P. and Calos,M.P. (2014) Serine integrase chimeras with activity in E. coli and HeLa cells. *Biol. Open*, **3**, 895–903.
73. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
74. Pietila,M.K., Roine,E., Paulin,L., Kalkkinen,N. and Bamford,D.H. (2009) An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. *Mol. Microbiol.*, **72**, 307–319.
75. Pietila,M.K., Atanasova,N.S., Manole,V., Liljeroos,L., Butcher,S.J., Oksanen,H.M. and Bamford,D.H. (2012) Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. *J. Virol.*, **86**, 5067–5079.
76. Dyall-Smith,M. and Porter,K. (2015) A strange family, or how a new pleolipovirus reveals its friends and relatives. *Mol. Microbiol.*, **98**, 995–997.
77. Futcher,A.B. (1986) Copy number amplification of the 2 micron circle plasmid of Saccharomyces cerevisiae. *J. Theor. Biol.*, **119**, 197–204.
78. Petes,T.D. and Williamson,D.H. (1994) A novel structural form of the 2 micron plasmid of the yeast Saccharomyces cerevisiae. *Yeast*, **10**, 1341–1345.
79. Abremski,K. and Gottesman,S. (1982) Purification of the bacteriophage lambda xis gene product required for lambda excisive recombination. *J. Biol. Chem.*, **257**, 9658–9662.
80. Brügger,K., Redder,P., She,Q., Confalonieri,F., Zivanovic,Y. and Garrett,R.A. (2002) Mobile elements in archaeal genomes. *FEMS Microbiol. Lett.*, **206**, 131–141.
81. Wagner,A., Whitaker,R.J., Krause,D.J., Heilers,J.H., van Wolferen,M., van der Does,C. and Albers,S.V. (2017) Mechanisms of gene flow in archaea. *Nat. Rev. Microbiol.*, **15**, 492–501.
82. Iranzo,J., Koonin,E.V., Prangishvili,D. and Krupovic,M. (2016) Bipartite network analysis of the archaeal virosphere: evolutionary connections between viruses and capsidless mobile elements. *J. Virol.*, **90**, 11043–11055.
83. DeMaere,M.Z., Williams,T.J., Allen,M.A., Brown,M.V., Gibson,J.A., Rich,J., Lauro,F.M., Dyall-Smith,M., Davenport,K.W., Woyke,T. *et al.* (2013) High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 16939–16944.
84. Krupovic,M., Gonnet,M., Hania,W.B., Forterre,P. and Erauso,G. (2013) Insights into dynamics of mobile genetic elements in hyperthermophilic environments from five new Thermococcus plasmids. *PLoS One*, **8**, e49044.
85. Krupovic,M. (2012) Recombination between RNA viruses and plasmids might have played a central role in the origin and evolution of small DNA viruses. *Bioessays*, **34**, 867–870.
86. Erdmann,S., Tschitschko,B., Zhong,L., Raftery,M.J. and Cavicchioli,R. (2017) A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. *Nat. Microbiol.*, **2**, 1446–1455.