A virus-borne DNA damage signaling pathway controls the lysogeny-induction switch in a group of temperate pleolipoviruses

Zhao Chen, Ying Liu, Yixuan Wang, Xincheng Du, Xiaoyuan Deng, Jialin Xiang, Yangyang Wang, Jiao Wang, Mart Krupovic, Shishen Du, et al.

3270–3287 *Nucleic Acids Research, 2023, Vol. 51, No. 7* *Published online 2 March 2023*
https://doi.org/10.1093/nar/gkad125

A virus-borne DNA damage signaling pathway controls the lysogeny-induction switch in a group of temperate pleolipoviruses

Zhao Chen[1], Ying Liu[2], Yixuan Wang[1], Xincheng Du[1], Xiaoyuan Deng[1], Jialin Xiang[1], Yangyang Wang[1], Jiao Wang[1], Mart Krupovic[2], Shishen Du[1,*] and Xiangdong Chen[1,*]

[1]State Key laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan 430072, China and [2]Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, F-75015 Paris, France

Received September 29, 2022; Revised February 07, 2023; Editorial Decision February 08, 2023; Accepted February 14, 2023

ABSTRACT

Many prokaryotic viruses are temperate and their reactivation is tightly regulated. However, except for a few bacterial model systems, the regulatory circuits underlying the exit from lysogeny are poorly understood, especially in archaea. Here, we report a three-gene module which regulates the switch between lysogeny and replicative cycle in a haloarchaeal virus SNJ2 (family *Pleolipoviridae*). The SNJ2 *orf4* encodes a winged helix-turn-helix DNA binding protein which maintains lysogeny through repressing the expression of the viral integrase gene int^{SNJ2}. To switch to the induced state, two other SNJ2-encoded proteins, Orf7 and Orf8, are required. Orf8 is a homolog of cellular AAA+ ATPase Orc1/Cdc6, which is activated upon mitomycin C-induced DNA damage, possibly through posttranslational modification. Activated Orf8 initiates the expression of Orf7 which, in turn, antagonizes the function of Orf4, leading to the transcription of int^{SNJ2}, thereby switching SNJ2 to the induced state. Comparative genomics analysis revealed that the SNJ2-like Orc1/Cdc6-centered three-gene module is common in haloarchaeal genomes, always present in the context of integrated proviruses. Collectively, our results uncover the first DNA damage signaling pathway encoded by a temperate archaeal virus and reveal an unexpected role of the widely distributed virus-encoded Orc1/Cdc6 homologs.

INTRODUCTION

Temperate viruses can undergo either the lysogenic or the lytic life cycles. During the lysogenic life cycle, the virus genome is maintained in a dormant state, known as the provirus, through either integrating into the host chromosome and replicating passively with the host or existing as an extrachromosomal episome. When entering into the replicative life cycle, the virus reproduces actively, the progeny virions are assembled and released from the host cell, in most cases, causing cell lysis. Proviral regions are frequently found in genomes of both bacteria and archaea, and play important roles in the microbial ecology and genome evolution (1,2). For instance, upon induction, temperate viruses can promote horizontal gene transfer and have significant impact on the composition and physiology of microbial communities in diverse ecological niches (3,4). Therefore, understanding the regulatory mechanisms behind the lysis-lysogeny decision of temperate viruses is of great importance.

The regulatory mechanisms of life cycle decision have been extensively explored for temperate viruses infecting bacteria (5). Many prophages (proviruses of bacteria) are activated by the host SOS response, following the DNA damage signaling pathway. The most in-depth studied is *Escherichia coli* phage lambda. The activation of lambda is regulated by many viral and host proteins. Among them, the host RecA protein which binds to single-stranded DNA (ssDNA), and the prophage master repressor CI protein responsible for the maintenance of lysogeny, are most important. Following DNA damage, RecA detects and binds to the accumulated ssDNA regions and becomes activated (RecA*). RecA* stimulates the self-cleavage of the prophage CI protein in a manner similar to the autoproteolysis of the SOS response master regulator LexA. The inactivation of CI repressor leads to the provirus excision, genome replication, virion assembly and host cell lysis (6).

Despite the fact that many temperate viruses have been detected in diverse archaeal lineages (7,8), in contrast to their bacterial counterparts, the regulatory mechanisms of their life cycle transitions remain largely unknown. Thus far, the lysis-lysogeny decision has been studied only in two

*To whom correspondence should be addressed. Tel: +86 13476029961; Fax: +86 27 68754582; Email: ssdu@whu.edu.cn
Correspondence may also be addressed to Xiangdong Chen. Tel: +86 13971423771; Fax: +86 27 68754582; Email: xdchen@whu.edu.cn

temperate archaeal viruses, the hyperthermophilic crenarchaeal virus SSV1 and the halophilic euryarchaeal virus SNJ1. In the case of SSV1, the virus-encoded transcriptional regulator F55 forms a complex with the host RadA protein, which represses the expression of the viral replicative pathway genes. Upon UV-induced host DNA damage, RadA is recruited to the single stranded DNA regions, resulting in the disassociation of the F55-RadA complex. This event causes the dissociation of F55 from its target DNA sites, leading to transcriptional derepression and active virus reproduction (9,10). In the case of halovirus SNJ1, an antitoxin MazE superfamily protein encoded by the virus functions as a master regulator repressing the viral lytic pathway genes (11). Treatment with DNA damaging agent mitomycin C (MMC) results in the reactivation of SNJ1 replication, but the mechanism has not been explored (11).

Family *Pleolipoviridae* includes non-lytic, enveloped viruses infecting halophilic archaea, and represents one of the dominant haloarchaeal viral types (12–14). There are currently three genera, namely *Alpha-*, *Beta-* and *Gammapleolipovirus* within the virus family (12,13). Viruses from distinct genera of *Pleolipoviridae* share a conserved gene cluster encoding structural proteins and a variable, largely, genus-specific region encoding various functions (15,16). Although most characterized pleolipovirus isolates do not code for the integrase and are not temperate, related proviruses are highly abundant in haloarchaeal genomes (16–20). SNJ2 (genus *Betapleolipovirus*) is a representative temperate pleolipovirus that integrates into the $tRNA^{Met}$ gene of *Natrinema* sp. J7-1 strain. Upon MMC treatment, SNJ2 is reactivated and virions are released without lysis, presumably through budding (18). Integration and excision of the SNJ2 genome is mediated by the virus-encoded integrase, Int^{SNJ2}, which represents a novel family of tyrosine recombinase (20). In this study, we investigated the regulatory mechanism of lysogeny-induction switch of SNJ2 (note that due to the non-lytic nature of pleolipoviruses, we use the term 'induction/induced' instead of 'lysis/lytic' in this study). We show that a regulatory module composed of 3 viral genes, including a gene homologous to the cellular AAA+ ATPase Orc1/Cdc6 (origin recognition complex 1/cell division control protein 6), a key component of the host replisome responsible for origin recognition and helicase loading, responds to host DNA damage and controls the switch from lysogenic to induced state of SNJ2. This three-gene regulatory module is conserved in proviruses of the genus *Betapleolipovirus*, indicating that these haloarchaeal proviruses employ a common strategy to regulate their life cycles.

MATERIALS AND METHODS

Strains, culture conditions, and transformation methods

All strains and plasmids used in this study were listed in Supplementary Tables S2 and S3, respectively. Construction of the archaeal strains, plasmids, and oligonucleotide primers used in this study were described or listed in Supplementary Information. *Natrinema* sp. J7 and other haloarchaeal strains were cultured in Halo-2 medium or 18% modified growth medium (MGM) at 45°C as previously described (21). Agar plates contained 15 g Bacto Agar (BD) per liter. Halo-2 medium contained 250 g of NaCl, 30 g of $MgCl_2{\cdot}6H_2O$, 2.5 g of lactalbumin hydrolysate (Difco Laboratories), and 2 g of Bacto yeast extract (Difco Laboratories) per liter of water. Casamino Acids medium (Hv-Ca) or MGM were prepared as described earlier (20,21). Hv-Ca contained 144 g of NaCl, 18 g of $MgCl_2{\cdot}6H_2O$, 21 g of $MgSO_4{\cdot}7H_2O$, 4.2 g of KCl, 5 g of Amicase (Sigma), 0.5 g of $CaCl_2$, and 30 ml of 1 M Tris–HCl (pH 7.5) per liter of water. 18% MGM contained 144 g of NaCl, 18 g of $MgCl_2{\cdot}6H_2O$, 21 g of $MgSO_4{\cdot}7H_2O$, 4.2 g of KCl, 5 g of peptone (Difco Laboratories), 3 g of Bacto yeast extract (Difco Laboratories), 0.5 g of $CaCl_2$, and 30 ml of 1 M Tris–HCl (pH 7.5) per liter of water. 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* were grown at 37°C in Luria–Bertani (LB) medium supplemented with ampicillin (100 μg/ml) or kanamycin (50 μg/ml) when necessary.

Transformation of haloarchaea including *Natrinema* sp. J7 and *Haloferax vocanii* H26 was performed using the modified polyethylene glycol method as described previously (22,23). DH5α, JM110 was transformed according to the $CaCl_2$ method (24).

Quantification of virus genome, attP and transcription level of genes by quantitative-PCR

The relative copy number of SNJ2 genome or *attP* was detected by qPCR method using cell culture with or without MMC treatment as the templates. For MMC treatment, strains were cultured to middle exponential phase (OD_{600} = 0.6–0.8) in Halo-2 medium and treated with a final concentration of 1 μg/ml MMC for 30 min. Cells were collected by centrifugation and washed twice in the same volume of Halo-2 medium to remove MMC. Cell pellets were resuspended in Halo-2 medium and cultured at the same condition. The protocol of MMC treatment was used throughout the study. Samples were taken at different time points post induction. To avoid the negative effect of a high salt concentration on qPCR, 10 μl of the cell suspension was added to 490 μl of distilled water, which also resulted in rapid cell lysis. The relative copy number of SNJ2 genome or *attP* was estimated by qPCR, using the iTaq Universal SYBR green Supermix (Bio-Rad) and gene-specific primers (Supplementary Table S4). The single-copy host *radA* gene was used as a reference, while *orf7* and *attP* of SNJ2 were used for viral genome. PCR was performed in a CFX96 Touch™ real-time PCR detection system (Bio-Rad, Hercules, CA, USA) with the following steps: denaturing at 95°C for 5 min (If the sample is cDNA, adjust it to 30 s), 40 cycles of 95°C 15 s, 60°C 30 s. Relative copy number of SNJ2 genome or *attP* was estimated by the comparative Ct value of *orf7*/*radA* or *attP*/*radA*, respectively.

For the analysis of transcription level of $orfX^{SNJ2}$, the cells were cultured and treated as above. RNA was extracted from samples at 0 and 9 h after induction using the Trizol reagent (Invitrogen, CA, USA). The housekeeping gene encoding RNA polymerase subunit B (*rpoB*) located on the host chromosome was used as a reference, *orf1*, *orf7*, *orf8*

of the SNJ2 genome were used as target genes. The specific primers used for qPCR were listed in Table S3. The reaction system was described previously (11), 20 μl mixtures were prepared containing 5 μl of template, 10 μl of iTaq Universal SYBR green Supermix Bio-Rad), 1 μl of primer pairs (10 μM), and 4 μl of distilled water. Amplification was performed as above. Finally, the qRT-PCR data were analyzed by the comparative Ct method with *rpoB* gene as the reference. Three independent experiments were performed, and error bars indicated the standard deviations.

Western blot analysis

For the analysis of Int^{SNJ2} expression, we first prepared an anti-Int^{SNJ2} antibodies by immunizing New Zealand white rabbits with purified Int^{SNJ2} by ABclonal Technology Co., Ltd. The J7-1-F and J7-3-F strains were cultured to middle exponential phase (OD_{600} = 0.6–0.8) in Halo-2 medium, then treated with MMC as above. Samples were collected at 0 and 9 h after MMC treatment (The OD_{600} of cultures was adjusted to ensure the same amounts of cells were taken for analysis), and crude extracts were prepared by mixing with 5× SDS-PAGE sample buffer (250 mM Tris–HCl pH 6.8; 10% SDS; 50% glycerol; 0.5% bromophenol blue) supplemented with 50 mM DTT /ml of sample buffer. After boiling at 95°C for 10 min, samples were loaded on a 12% polyacrylamide gel for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Proteins were transferred from the polyacrylamide gel onto a nitrocellulose membrane (GE Healthcare, Waukesha, WI, USA), using the Semi-Dry Electrophoretic Transfer Cell system (Bio-Rad, Hercules, CA, USA). The transferred membranes were immersed in 0.2% Ponceau S (Beijing Solarbio Science & Technology Co., Ltd) for 10 min to determine the transfer efficiency of the samples. Membranes were then washed 3 × 5 min with ddH_2O and used for immunoblotting. Briefly, the membrane was incubated in 5% skim milk blocking agent for 1h, and then incubated with primary rabbit antibodies (against Int^{SNJ2}) at 4°C, overnight. Following 3 × 10 min washing with TBST, membranes were incubated with HRP-conjugated goat anti-rabbit antibody (Beijing TransGen Biotech Co., Ltd) for 1h at RT. After washing, signal was visualized using ECL western blot substrate (Thermo Scientific, Waltham, MA, USA) and recorded by exposure to an X-ray film. Anti-Int^{SNJ2} antibodies were used for Western blot assay of Int^{SNJ2} protein.

To analyze the protein level of Orf8, the J7-1-F*Δorf8*/pFJ6-P_{phaR}-orf8-his strain was cultivated and treated as above. Samples were collected at 0, 3, 6, 9 and 12 h after MMC treatment (The OD_{600} of cultures was adjusted to ensure the same amounts of cells were taken for analysis) and crude extracts were loaded for SDS-PAGE. Procedures for Ponceau S staining and western blot were performed as above. Membrane was incubated with the primary Anti-His Mouse Monoclonal Antibody (Beijing TransGen Biotech Co., Ltd) followed by washing with TBST and finally incubation with the HRP-conjugated goat anti-mouse antibody (Beijing TransGen Biotech Co., Ltd) to detect Orf8-his.

To test whether *orf4* was cleaved or degraded, the J7-3-F/pFJ6-4-his-9-P_{Int}-amyH strain was cultivated and treated as above. Samples were collected at 0, 3, 6 and 9 h after MMC treatment (The OD_{600} of cultures was adjusted to ensure the same amounts of cells were taken for analysis) and crude extracts were loaded for SDS-PAGE. Western-blot of Orf4-his was performed as above for the detection of Orf8-his.

Amylase activity assay

All strains carrying amylase reporter plasmids were cultivated in Halo-2 medium to late exponential phase, then treated with or without MMC as above. The amylase-specific activities were measured at 16h after induction by a method described by Fuwa (25). The presence of amylase activity on plates was determined as described by Coronado et al. (26). Transformants were transferred to 5 ml of 20% HV-ca medium and grown to log phase, then 2 μl culture was spotted onto 20% Hv-Ca plates supplemented with 2% (wt/vol) soluble starch. After incubation at 45°C for about 2 days, 0.3% I_2/0.6% KI solution was added to the plates, and formation of halos around the selected transformants was examined immediately.

RNA-seq

To determine the difference in gene expression of SNJ2 with and without MMC treatment, *Natrinema* sp. J7-1 was incubated to middle exponential phase (OD_{600} = 0.6–0.8) in Halo-2 medium and treated with MMC as above. Cell samples were then collected at 0, 1, 4 and 9h after MMC treatment. RNA extraction and RNA-seq was performed in Wuhan Frasergen Bioinformatics Co. Ltd. Briefly, total RNA was extracted using Trizol reagent (Invitrogen, CA, USA) as described by the manufacturer, RNA purity and integrity was monitored by NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and a Bioanalyzer 2100 system (Agilent Technologies, CA, USA). The degradation of RNA was assessed by 1.5% agarose gel. Ribosomal RNA was removed by the Ribo-Zero Magnetic kit (Bacteria) (epicentre, MRZB12424) to enrich mRNA. Strand specific RNA-seq was performed using the NEB Next Ultra Directional RNA Library Prep Kit (NEB, E7420). After the library was constructed, Qubit 2.0 was used for preliminary quantification, and the library was diluted to 1.5 ng/μl. Then Agilent 2100 BioAnalyzer was used to detect the insert size of the library. Quantitative reverse transcription PCR (qRT-PCR) was used to accurately quantify the effective concentration of the library to ensure the quality of the library

RNA-seq libraries were sequenced using the Illumina novaseq 6000 platform. Using RSEM (27), the bowtie2 comparison results were used for statistics, and the number of Reads compared to each transcript of each sample was obtained, and the resulting data were then analyzed by Fragments Per Kilobase Per Million bases (FPKM) analysis (28) to reveal the expression levels of all genes in the *Natrinema* sp. J7-1 genome.

qPCR to test the impact of *orf4* overexpression on SNJ2 infection and replication

SNJ2 was incubated with 20 ml of early-exponential-phase ($OD_{600} \approx 0.3$) cultures of J7-3-F, J7-3-F/pFJ6-P_{phaR}-orf4 or J7-3-F/pFJ6- P_{phaR}-orf4fsm at 45°C for about 1 h. After absorption, cells were collected by centrifugation (10000 rpm, 5 min) and washed twice in the same volume of Halo-2 medium to remove the free viruses. Cell pellets were resuspended in the same volume of Halo-2 medium and cultured at 45°C for 6h. Samples were taken every half hour for the first 3h post infection. qPCR analysis was performed using the same primer pairs and procedure as described above in 'Quantification of virus genome, attP and transcription level of genes by Quantitative-PCR' The qPCR data were analyzed according to the $2^{-\Delta CT}$ method. Three independent experiments were performed, and error bars indicated the standard deviations.

Bioinformatic analysis

The *in silico*-translated protein sequences were used as queries to search for sequence homologs in the nonredundant protein database at the NCBI using BLAST (29). Searches for distant homologs and functional annotation of proteins were performed using HHpred (30) against different protein databases, including PDB, Pfam, CDD and COG. The protein secondary structure was predicted using Jpred (31), and multiple sequence alignments were performed by PROMALS3D (32). Phylogenetic analyses were carried out using PhyML 3.0 with the automatic model selection (33). To search for Orc1-encoding pleolipoproviral regions in haloarchaeal genomes, sequences of the viral hallmark spike proteins from all characterized pleolipovirus isolates (12) were used to create a HMM profile with HMMER 3 (34), and the resulting profile was used to search *Halobacteria* nonreduntant protein database downloaded (January 2022) from the NCBI. The genomic neighborhood of the identified spike protein-encoding genes was analyzed for the presence of *orc1* and other genes of the conserved gene cluster of pleolipoviruses. The exact nucleotide coordinates and attachment sites of proviruses were identified manually. For the phylogenomic analysis, all pairwise comparisons of the nucleotide sequences of the identified pleolipoproviruses were conducted using the Genome-BLAST Distance Phylogeny (GBDP) method implemented in VICTOR (35). The resulting intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR post processing for D6 formula. The structural model of SNJ2 Orf7 was obtained using AlphaFold2 (36) through ColabFold (37) and visualized using ChimeraX (38).

RESULTS

The transcription of *int*^{SNJ2} is tightly regulated and is critical for SNJ2 life cycle switch

To investigate the regulatory mechanism of SNJ2 life cycle switch, we firstly analyzed the transcription level of viral genes after MMC treatment of *Natrinema* sp. J7-1-F cells (*snj2*$^{+}$). Total RNA from cell samples at 0, 1, 4 and 9 h post-MMC treatment were extracted, and the mRNA abundance of individual genes was determined by RNA-seq. As shown in Figure 1A, nearly all SNJ2 genes were silenced before induction except for *orf10,* which was predicted to be a type IV pilin N-terminal domain-containing protein. However, 4 h post MMC induction, transcription of the *orf3-1* operon (*orf1* encodes the integrase Int^{SNJ2}), *orf7* and *orf19-25* was strongly upregulated, and continued to increase to high levels 9 h post-induction. By contrast, the transcripts of the conserved pleolipovirus gene cluster (*orf12-17*) could only be detected at 9 h post-induction. Notably, the transcription level of *orf4*, *orf8* and *orf9* was very low and did not change throughout the entire induction process (Figure 1A). To validate the RNA-seq data, the relative transcription levels of *orf1* and *orf8* were detected by qPCR at 0 and 9 h after MMC treatment. Consistent with the RNA-seq results, the transcription level of *orf1* increased significantly post induction, whereas the level of *orf8* did not change (Figure 1B). These results revealed that the transcription of SNJ2 genes is tightly regulated with the *int*^{SNJ2} operon (*orf3-1*) being expressed early and the conserved structure-related module (*orf12-17*) being a late transcript.

We previously showed that Int^{SNJ2} mediates site-specific recombination between the viral and host genomes (20). Deletion of *int*^{SNJ2} in J7-1-F strain resulted in a defective SNJ2 provirus which was unable to replicate after MMC treatment (Supplementary Figure S1), indicating that *int*^{SNJ2}-mediated excision and circularization of the SNJ2 provirus is critical for the reactivation of SNJ2 replication. Consistent with this, the *int*^{SNJ2} operon is an early transcript post MMC treatment and Western blot analysis showed that Int^{SNJ2} was not detectable before induction, but was highly expressed 9h post induction (Figure 1B and C). As a control, a protein band corresponding to Int^{SNJ2} was not detected in samples of the strain J7-3-F, from which the SNJ2 provirus has been cured. Taken together, these results suggest that the transcription of *int*^{SNJ2} is the key for SNJ2 life cycle switch, which is tightly regulated and is activated upon MMC-induced DNA damage.

SNJ2 contains a regulatory region responsible for its lysogeny-induction switch

Since Int^{SNJ2} plays a critical role in the reactivation of SNJ2, determination of the mechanism regulating its expression would likely shed light on the lysogeny-induction switch of SNJ2. Comparison of the genomes of the isolated pleolipoviruses with those of the proviruses revealed a distinct region interspersed between the conserved structural gene cluster and the integrase encoding gene in the proviral genomes (12,18). Although this region is highly divergent among pleolipoproviruses, it likely contains the regulatory genes responsible for the life cycle switch of these viruses. In the case of SNJ2, this putative regulatory region is from *orf2* to *orf11* (Figure 2A). The *int*^{SNJ2} operon (*orf3-1*) has been shown to be co-transcribed from one promoter (20). Consistently, our RNA-seq data showed similar transcriptional pattern of the three genes (Figure 1A). Thus, we speculated that this region may harbor viral factors that control *int*^{SNJ2} transcription through the promoter of the *int*^{SNJ2} operon (the intergenic region between *orf3* and *orf4*), which we denoted as the P_{Int} promoter region.

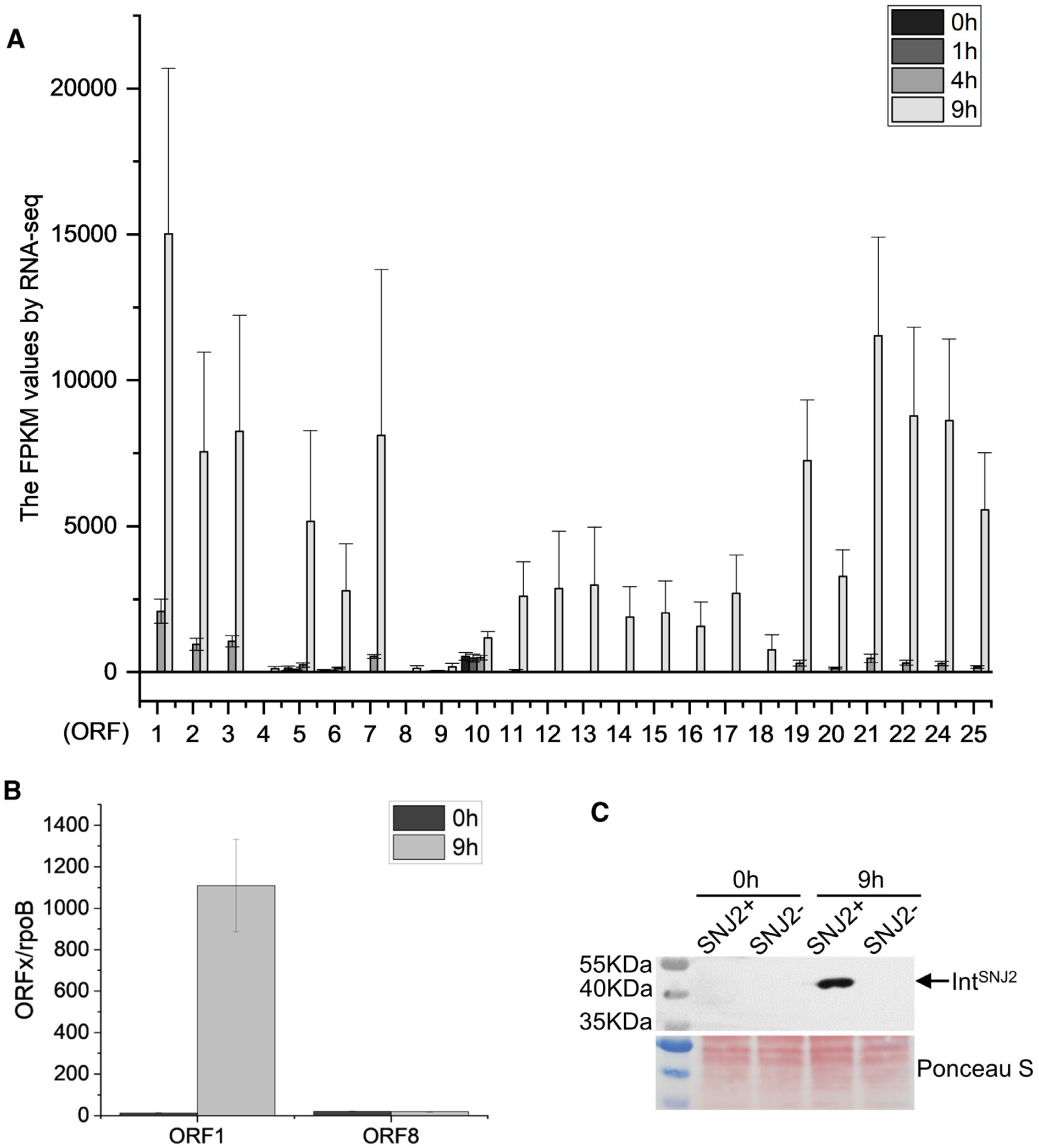

Figure 1. The transcription of int^{SNJ2} is tightly regulated. (**A**) RNA-seq data of SNJ2 *orfs* at 0, 1, 4 and 9h after MMC treatment. J7-1-F (*snj2*$^+$) strain was cultured to exponential phase (OD_{600} = 0.6–0.8) in Halo-2 medium and treated with MMC (1μg/ml). Samples were collected at 0, 1, 4 and 9 h after MMC treatment from which total RNAs were prepared and used for RNA-Seq analysis. The FPKM of each SNJ2 gene was calculated. (**B**) qPCR analysis of the relative transcriptional level of *orf1* and *orf8* in J7-1-F cells at 0h and 9h after MMC treatment. J7-1-F strain was cultured and treated as in (A). RNA was extracted from samples at 0h and 9h after suspension and used for qPCR analysis using the primer pairs *orf1*-F/ R, *orf8*-F/R and *rpoB*-F/ R. The transcription level of *orf1* and *orf8* was relative to internal reference *rpoB*. The experiment was repeated three times, and error bars indicated the standard deviations. (**C**) Western-blot assay to check the protein level of Int^{SNJ2} before and after MMC treatment. The J7-1-F and J7-3-F (*Δsnj2*) strains were cultured and treated as in (A). Samples were collected at 0h and 9h after MMC treatment and crude extracts were loaded for SDS-PAGE. Lanes 1 and 3 were samples of J7-1-F (*snj2*$^+$), whereas lanes 2 and 4 were samples of J7-3-F (*Δsnj2*), respectively. Anti-Int^{SNJ2} antibodies was used for detection of Int^{SNJ2}. Lower panel shows Ponceau S staining of the membrane as loading controls.

To test the above hypothesis, we constructed an amylase reporter system based on the pFJ6 plasmid (39), in which the α-amylase gene was placed under the P_{Int} promoter region. The reporter plasmid was transformed into the J7-1-F (*snj2*$^+$) and J7-3-F (*Δsnj2*) strains, respectively, and the amylase activity was detected 16 h post MMC treatment. As shown in Figure 2B, amylase activity was detected in J7-3-F cells carrying the reporter plasmid with or without MMC treatment, indicating that the P_{Int} promoter is constitutively active in the absence of viral proteins. By contrast, no amylase activity was detected in the J7-1-F strain without MMC treatment, but the amylase activity increased significantly upon MMC treatment (Figure 2B). These results indicate that the P_{Int} promoter is repressed by viral factors before MMC treatment, but the repression is relieved post induction. To determine if the *orf4-orf11* region contains the regulators, it was incorporated into the reporter plasmid to generate pFJ6-*orf4-11*-P_{Int}-amyH and the amylase activity was tested in J7-3-F cells. As shown in Figure 2B, amylase activity was completely inhibited in J7-3-F cells harboring the plasmid in the absence of MMC treatment but increased significantly upon MMC treatment, indicating that the *orf4-orf11* region contains both the repressor of the P_{Int} promoter and factors relieving the repression.

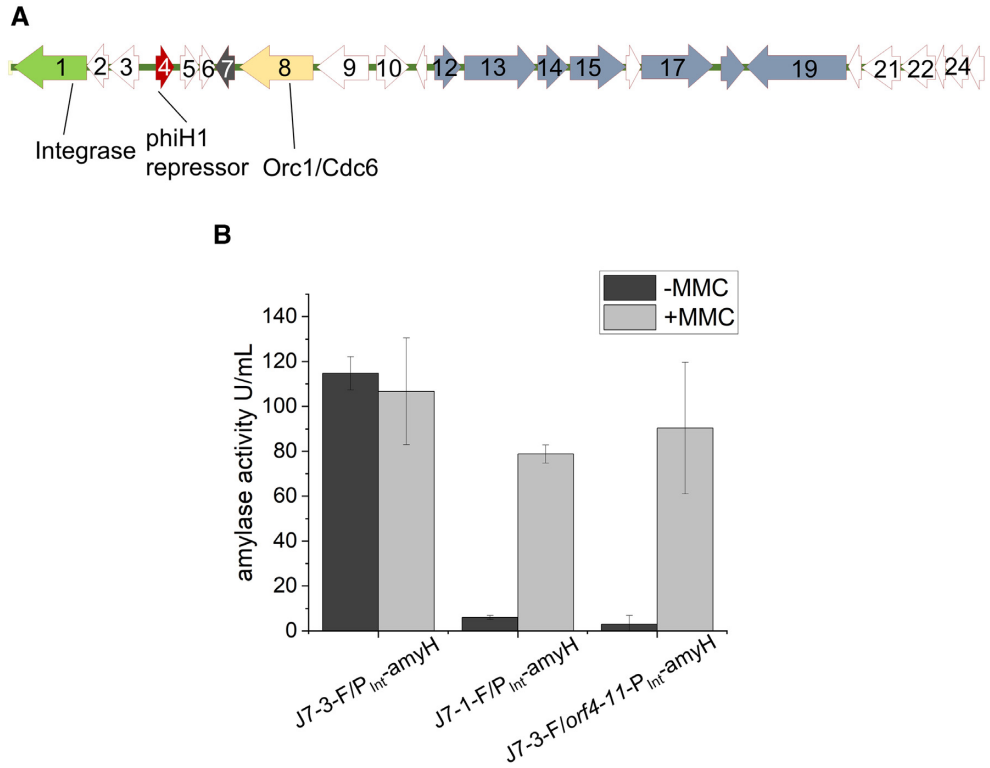

Figure 2. The region containing *orf4-11* is capable for the regulation of SNJ2 lysogeny to induction switch. (**A**) Organization of the SNJ2 provirus genome. Colors denote genes characterized in this study. *orf1* (aqua) encodes the SNJ2 integrase, *orf4* (red) encodes a φH1 repressor-like protein, *orf7* (gray) encodes a protein of unknown function, and *orf8* (golden) encodes an Orc1/Cdc6 protein. (**B**) Amylase activity assay to test the regulation of the P_{Int} promoter by the putative regulatory region of SNJ2. Reporter plasmids pFJ6-P_{Int}-amyH (without *orf4-11*) and pFJ6-*orf4-11*-P_{Int}-amyH (with *orf4-11*) were transformed into J7-1-F (*snj2*$^+$) and J7-3-F (*Δsnj2*) and amylase activity was tested by the FUWA method at 16h post MMC treatment. The experiment was repeated three times, and error bars indicated the standard deviations.

Viral factors involved in the SNJ2 life cycle switch are encoded within the *orf4-orf8* region

To identify the repressor of the P_{Int} promoter, a series of reporter plasmids containing different regions of the *orf4-11* fragment were constructed and the amylase activity was measured in strain J7-3-F (*Δsnj2*) with or without MMC treatment. As shown in Figure 3, amylase activity was nearly undetectable in cells transformed with reporter plasmids harboring *orf4*, *orf4-6*, *orf4-7* and *orf4-8* with or without MMC treatment (Figure 3), indicating that *orf4* alone is sufficient to repress the P_{Int} promoter. However, none of these constructs was sufficient for the derepression of P_{Int} upon MMC treatment. Interestingly, the construct containing *orf4-9* displayed a comparable amylase activity to that harbored the whole *orf4-11* region upon MMC treatment (Figure 3), implying that Orf9 plays an important role in the derepression of P_{Int}. Therefore, we introduced a frame-shift mutation in *orf9* to block its translation and tested if derepression of P_{Int} was abolished upon MMC treatment. Unexpectedly, amylase activity was still observed upon MMC treatment (Supplementary Figure S2A), indicating that the gene product of *orf9* is not important for the derepression of P_{Int}, but its coding sequence likely contains regulatory elements necessary for the transcription of downstream *orfs*. To test this possibility, we constructed a plasmid pFJ6-orf4-8+-P_{Int}-amyH, in which the *orf4-8* region was extended with a 297-bp-long sequence from the 3' terminus of *orf9*. As expected, amylase activity was observed in the cells carrying this construct upon MMC treatment, suggesting that a part of the promoter or transcript of *orf8* is within the *orf9* coding sequence (Figure 3).

To determine the promoter region of *orf8* (denoted as the P_8 promoter), the 114-bp-long intergenic region between *orf8* and *orf9* (P_8A) or a 411-bp-long sequence containing the intergenic region and a 297-bp-long region of *orf9* (P_8B) was placed in front of the α-amylase gene in the reporter plasmid. The resultant plasmids pFJ6-P_8A-amyH and pFJ6-P_8B-amyH were transformed into J7-3-F and amylase activity was tested. As shown in Supplementary Figure S2B, amylase activity was observed only in the J7-3-F/pFJ6-P_8B-amyH transformant, suggesting that the P_8 promoter indeed extends into the coding region of *orf9*. Taken together, these results indicate that viral factors involved in the SNJ2 life cycle transition are encoded within the *orf4-orf8* region, but the expression of *orf8* requires a part of the *orf9* coding sequence.

Prediction of the functions of SNJ2 proteins encoded by *orf4-orf8*

The putative functions of Orf4-Orf8 were predicted with a combination of BLASTP analysis and a more sensitive hidden Markov model (HMM)-based HHpred analysis (29,30). The product of *orf4* is a putative winged helix-turn-helix DNA binding protein (PDB profile 3B73_B, HHpred

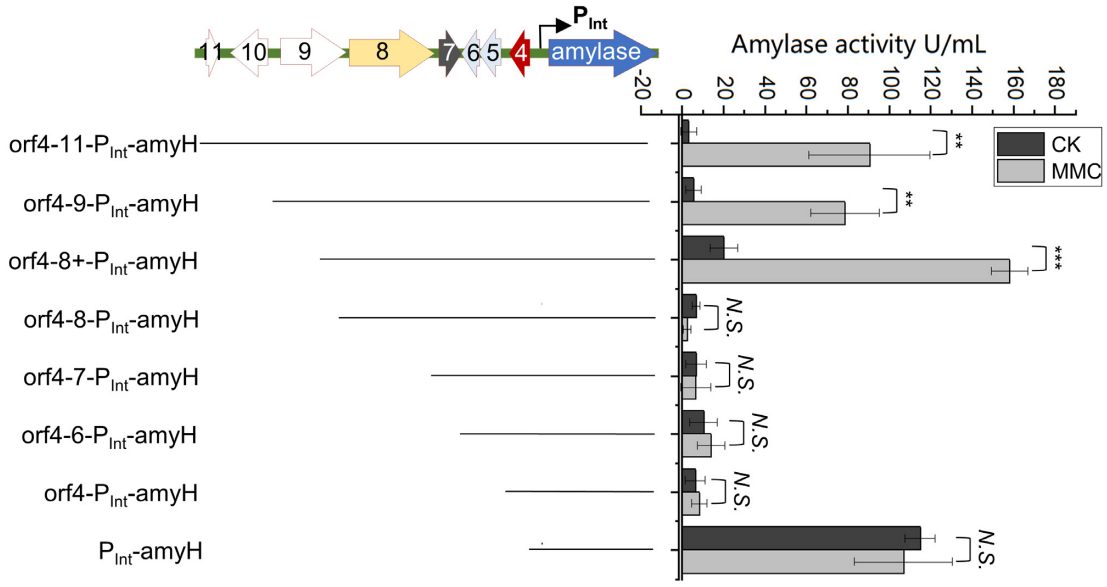

Figure 3. Determination of the viral factors required for the repression and derepression of the P_{Int} promoter. Schematic diagram of the reporter plasmids containing different part of the putative regulatory region of SNJ2 were shown on the left and the amylase activity of cells harboring these reporter plasmids with (MMC) or without (CK) treatment was shown on the right. Plasmid pFJ6-orf4-11-P_{Int}-amyH and its derivatives harboring truncated version of *orf4-11* as indicated in the diagram (*orf4-9*, *orf4-8+*, *orf4-8*, *orf4-7*, *orf4-6*, *orf4* and complete deletion) were transformed into J7-3-F (*Δsnj2*). Amylase activity was detected by the FUWA method after 16 h. The bars represent the means and standard deviations of three independent experiments. Significance testing against CK group was performed using a one-sample t test (***$P < 0.001$; **$P < 0.01$; *N.S.*, not significant ($P > 0.05$)).

probability of 98.1%), which was annotated as a φH1 repressor-like protein (18,20). Homology modelling of the structure of Orf4 showed that it likely functions as a dimer (Supplementary Figure S3) (40), similar to many other wHTH proteins (41). *orf5* encodes a putative HicB family antitoxin (PDB profile 6G1N_D, HHpred probability of 93.2%), whereas the product of *orf6* is a putative HicA mRNA interferase family toxin (PDB profile 6HPB_A, HHpred probability of 98.8%). Thus, Orf6 and Orf5 likely constitute a putative type II toxin-antitoxin (TA) module. *orf7* encodes a small protein of 93-amino-acids, rich in β-strands (containing 6 predicted β-strands and 1 α-helix), but no putative function could be confidently assigned to it. However, peptides of Orf7 had been detected in the virions by tandem mass spectrometry in previous analyses (18), indicative of a bona fide protein. *orf8* encodes an Orc1/Cdc6 family replication initiation protein (PDB profile 1FNN_B, HHpred probability of 99.9%; hereafter denoted Orc1 for simplicity). In the following sections, we investigated whether and how these gene products participate in the regulation of SNJ2 life cycle transition.

Orf4 is a transcriptional repressor of int^{SNJ2} and a palindromic sequence within the P_{Int} promoter is critical for Orf4-mediated repression of P_{Int}

As *orf4* encodes a φH1 repressor-like protein and it alone was sufficient to repress the P_{Int} promoter as shown in Figure 3, we speculated that it functions as the master repressor for the life cycle switch of SNJ2 by directly binding to the P_{Int} promoter to repress the int^{SNJ2} transcription. To test this, we attempted to purify Orf4 to determine its binding site on the P_{Int} promoter by DNase I footprinting assay *in vitro*. However, we failed to obtain functional Orf4. As an alternative approach, we used the amylase reporter system to determine its potential binding sites. We have previously shown that the P_{Int} promoter region contains two transcription start sites with two sets of putative transcription factor B recognition elements (BRE) and TATA boxes (20), as illustrated in Figure 4A. Based on these identified elements, a series of plasmids with the amylase gene under the regulation of various truncated P_{Int} promoter regions were constructed (Figure 4B). These plasmids were introduced into J7-1-F ($snj2^+$) and J7-3-F (*Δsnj2*), respectively, and the amylase activities were tested. Among the constructs, amylase activity could not be detected in those containing truncated promoters ranging from –200 bp (P_{200}) to –176 bp (P_{176}) in J7-1-F, indicating that these truncated promoters were still under the repression of Orf4 (Figure 4B). Further truncations on the promoter region, especially deletion of the region from –172 (P_{172}) to –142 bp (P_{142}), resulted in the expression of the amylase gene in J7-1-F without MMC treatment, indicative of the derepression of the promoter (Figure 4B). Further deletion of the region down to –134 bp (P_{134}) abolished the expression of amylase, presumably due to the disruption of the crucial TATA box that is essential for the recruitment of transcription factors and RNA polymerase. Consistent with this, all the truncated promoters could drive the expression of amylase in J7-3-F except for P_{134} (Figure 4B), indicating that the 5'-TTTATT-3' sequence is likely the correct TATA box of P_{Int}. Moreover, as deletion of the predicted BRE element and deletion of the -152 to -142 region right next to the TATA box did not affect the promoter activity in J7-3-F, a BRE element may be not necessary for the activity of P_{Int}. Thus, the Orf4 binding

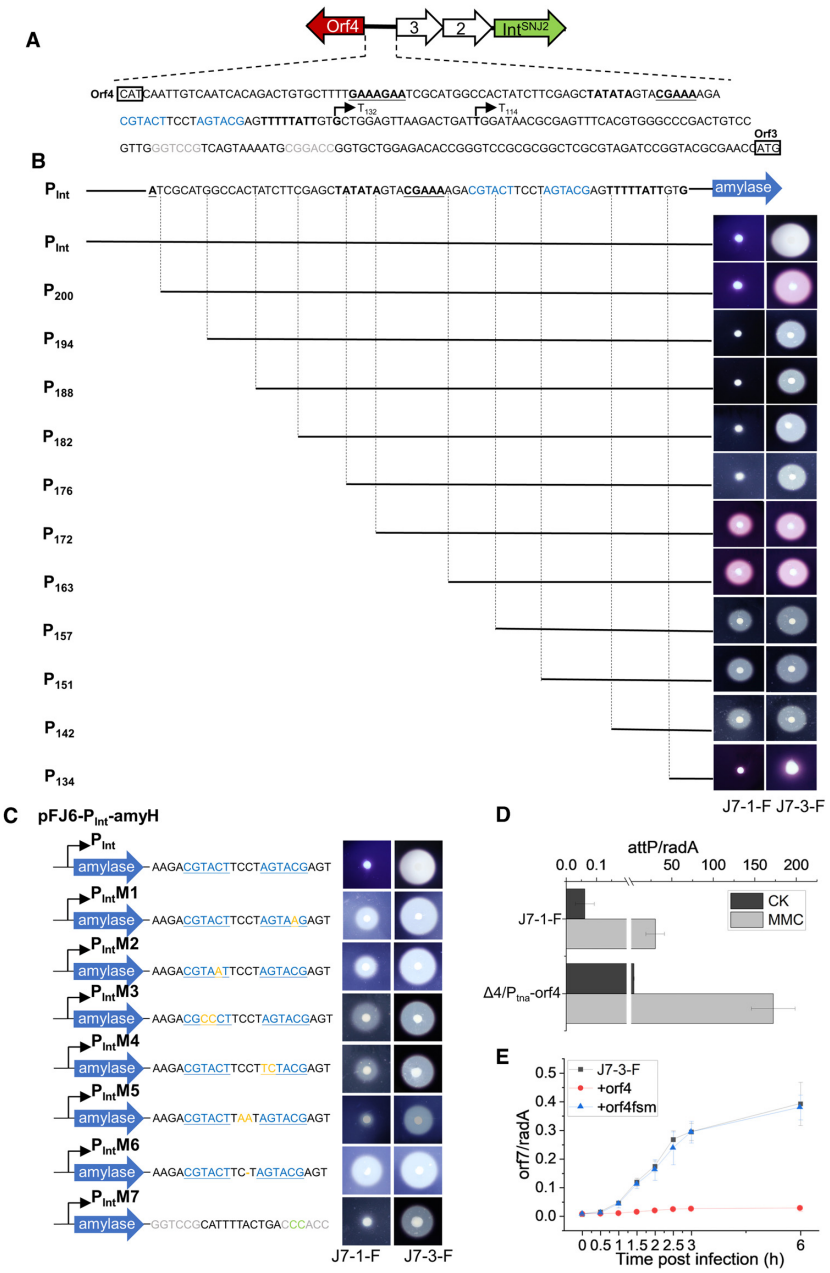

Figure 4. Orf4 is a transcriptional repressor and a palindromic sequence within the P_{Int} promoter is required for Orf4-mediated repression of P_{Int}. (**A**) Schematic diagram of the location and sequence of the P_{Int} promoter. T114 and T132 indicate the two possible transcription start sites of the *orf3-1* operon, the predicted TATA boxes are shown in bold black and the BRE motif (transcription factor B recognition element) CGAAA is underlined, the palindrome predicted by the software is shown in blue. Another palindrome within the intergenic region of *orf3-1* and *orf4* is shown in gray. The start codons of Orf3 and Orf4 are both highlighted as open boxes. (**B**) Deletion analysis of the P_{Int} promoter to identify the potential Orf4 binding sites. Reporter plasmids containing the wild-type P_{Int} (the entire intergenic region between the start codon of *orf3* and *orf4*) or 5'-truncated versions were transformed into J7-1-F (left, *snj2*$^+$) and J7-3-F (right, *Δsnj2*), respectively. Three random transformants were transferred to 5 ml of 20% MM and grown to log phase. 2μl of culture was spotted onto 20% MM plates supplemented with 2% (wt/vol) soluble starch. After 2 days, iodine solution was added to the plates and formation of haloes around the colonies was examined immediately (only one is shown here). (**C**) Mutational analysis of the predicted palindromic sequences. The palindromic sequence is colored blue and substitutions are shown in orange, whereas deletion is shown in line. P_{Int}-M1 to P_{Int}-M6 indicate different substitutions in the palindromic sequence, whereas P_{Int}-M7 represents modification to another palindromic sequence in the P_{Int} region. Amylase activity was tested as in panel (B). (**D**) qPCR analysis of SNJ2 copy number in J7-1-F and J7-1-F*Δorf4*/P_{tna}-orf4, indicated by *attP* (virus attachment site). *orf4* was deleted from the SNJ2 genome in the presence of a plasmid expressing Orf4 under the P_{tna} promoter. Three random *orf4*-knockout strains were transferred to Halo-2 medium and grown to log phase, treated with (MMC, 1μg/ml) or without (CK) MMC. Samples were taken for qPCR analysis 24 h post induction using the primer pairs attP-F/attP-R, and radA-F/radA-R. attP/radA represents the copy number of the SNJ2 virus relative to the chromosome. (**E**) Expression of Orf4 in J7-3-F (*Δsnj2*) inhibits SNJ2 replication. SNJ2 was incubated with J7-3-F, J7-3-F/P_{phaR}-orf4 or J7-3-F/ P_{phaR}-orf4f (frame shift mutation) at $OD_{600} \approx 0.3$. The cells were collected and washed in fresh Halo-2 medium to remove free viruses. The cells were then cultivated for 6 h, and samples were taken for examination every half hour for the first 3 h post infection. qPCR analyses were performed using the primer pairs Orf7-F/Orf7-R, and radA-F/radA-R. orf7/radA represents the copy number of the SNJ2 virus relative to the chromosome. The bars represent the means and standard deviations of three independent experiments.

sites are potentially located between –176 bp and –142 bp in the P_{Int} promoter, right next to the TATA box.

Intriguingly, an 18bp-long palindromic sequence 5'-CGTACTTCCTAGTACG-3' is situated adjacent to the TATA box within the –176 to –134 bp region (Figure 4A). As Orf4 is similar to the φH1 repressor which was suggested to bind the interrupted palindrome (42), we wondered if the palindromic sequence within the –176 to –134 bp region was critical for Orf4-mediated repression of P_{Int}. To test this, mutations were introduced into the palindromic sequence in the reporter plasmids. Plasmids pFJ6-P_{Int}M1-amyH and to pFJ6-P_{Int}M4-amyH contained one or two point-mutation in the left and the right arms of the palindrome, respectively. As a control, mutations were introduced into an unrelated palindromic sequence located in the 5'-UTR of the *int*^{SNJ2} operon and was shown in gray in Figure 4A (plasmid pFJ6-P_{Int}M7-amyH). These plasmids were transformed into J7-1-F (*snj2*$^+$) and J7-3-F (*Δsnj2*), respectively, and the amylase activity was determined. As shown in Figure 4C, mutations in both arms of the palindromic sequence resulted in the derepression of the amylase gene in J7-1-F cells. By contrast, amylase activity was not affected by mutations in the other palindromic sequence (Figure 4C). In J7-3-F cells (*Δsnj2*), none of the mutations affected the amylase activity from the reporter plasmids. Thus, mutations in the palindromic sequence and its adjacent area within the –176 to –134 bp region strongly reduce Orf4-mediated repression of P_{Int}, suggesting that Orf4 may bind to the palindrome.

To test if the spacer region between the palindromic sequences is important for Orf4-mediated repression of P_{Int}, the spacer sequence TCCT was replaced with TAAT (pFJ6-P_{Int}M5-amyH) or with TCT (pFJ6-P_{Int}M6-amyH). Interestingly, while replacement of TCCT with TAAT did not affect the repression of amylase expression by Orf4 in J7-1-F cells, replacement of TCCT with TCT abolished the repression (Figure 4C). This result indicates that the number of bases (i.e. the length of spacer) between the arms of the palindrome is critical for the regulation of P_{Int} by Orf4, but the composition of the spacer sequence is not. As expected, neither of the mutations affected the amylase activity when the constructs were transformed into J7-3-F (*Δsnj2*).

Notably, deletion of the nucleotides from –176 to –172 or from –172 to –163 located upstream of the palindromic sequence also resulted in the derepression of amylase gene (Figure 4B), indicating that the upstream flanking sequence of the palindrome is also important for Orf4-mediated repression of P_{Int}. It has been suggested that the φH1 repressor functions as a dimer and binds to two interrupted palindromes in the regulatory region of φH1 (42). It is possible that two Orf4 dimers bind to the palindromic sequence as well as the upstream and downstream flanking sequences within the P_{Int} promoter to repress P_{Int}. However, future studies are necessary to elucidate the binding mode of Orf4.

Despite numerous efforts, attempts to knock out *orf4* on the SNJ2 provirus in J7-1-F strain were unsuccessful. However, it could be deleted if the strain harbored a plasmid with *orf4* under the control of a tryptophan inducible promoter (P_{tna} promoter). Intriguingly, through evaluating the relative copy number of SNJ2 *attP* by qPCR, we observed more active SNJ2 replication in the resulting complemented strain than in the J7-1-F wild-type strain with and without MMC treatment (Figure 4D). These results suggest that SNJ2 life cycle is finely regulated by Orf4 and deletion of *orf4* leads to mis-regulation of the virus life cycle, which could be fatal to the cells.

To test if increasing the level of Orf4 blocks the life cycle transition of SNJ2 virus, we constructed an Orf4 overexpression plasmid, P_{phaR}-orf4(+orf4) in which *orf4* was under the regulation of the strong constitutive P_{phaR} promoter. A frame shift mutation was introduced into the plasmid P_{phaR}-Orf4fsm(+orf4fsm), which was used as a control. These two plasmids were transformed into strain J7-1-F (*snj2*$^+$) and the level of released SNJ2 viruses in cultures with or without MMC treatment was measured. We observed a slight decrease of SNJ2 copy number in the J7-1-F strain expressing Orf4 (J7-1-F/P_{phaR}-orf4) with or without MMC treatment compared to strain J7-1-F without the plasmid or with the control plasmid (J7-1-F/P_{phaR}-Orf4fsm). However, the reduction was not statistically significant (Supplementary Figure S4). Interestingly, overexpression of Orf4 in the J7-3-F strain (J7-3-F/P_{phaR}-orf4) resulted in the complete absence of released SNJ2 viruses after the strain was challenged with SNJ2 viruses (Figure 4E), whereas SNJ2 viruses were detected in culture of the J7-3-F strain or the strain harboring the mutated plasmid (J7-3-F/P_{phaR}-orf4fsm). These results indicate that Orf4 blocks the infection or replication of SNJ2 viruses.

Taken together, our results in this section confirmed that Orf4 acts as the master repressor of the life cycle of SNJ2. It may bind to the palindromic sequence **5'-CGTACTTCCTAGTACG-3'** as well as its upstream and downstream sequences in the P_{Int} promoter to suppress the *int*^{SNJ2} transcription, thereby maintaining SNJ2 in the lysogenic state.

Orf5 and Orf6 constitute a functional TA module but are not involved in the SNJ2 life cycle transition

As shown in Figure 3, derepression of the P_{Int} promoter upon MMC treatment required viral factors in the regulatory region. We first tested if Orf5 and Orf6, which constitute a putative TA module, are involved in the derepression of the P_{Int} promoter by deletion of them from the reporter plasmid. Interestingly, in the absence of *orf5* the reporter plasmid could not be transformed into J7-3-F cells (*Δsnj2*), but successful transformation could be obtained when both *orf5* and *orf6* were deleted from the reporter plasmids. Similarly, *orf5* could not be deleted in the J7-1-F strain (*snj2*$^+$), unless it was deleted simultaneously with *orf6*. Presumably, the toxin Orf6 is toxic to the host cell in the absence of the cognate antitoxin Orf5, leading to the failed transformation and failure to generate an *orf5* deletion strain. In agreement with this result, expression of *orf6* in J7-3-F (*Δsnj2*) inhibited cell growth, but simultaneous expression of *orf5* restored the growth of the strain (Supplementary Figure S5), indicating that Orf5 antagonizes the toxicity of Orf6. These results confirmed that the two genes form a functional TA module. The amylase activities were measured in J7-3-F cells containing the reporter plasmids in the absence of *orf6* and *orf5-6,* respectively. As shown in Figure 5A, cells carrying either plasmid showed high levels of amylase activity

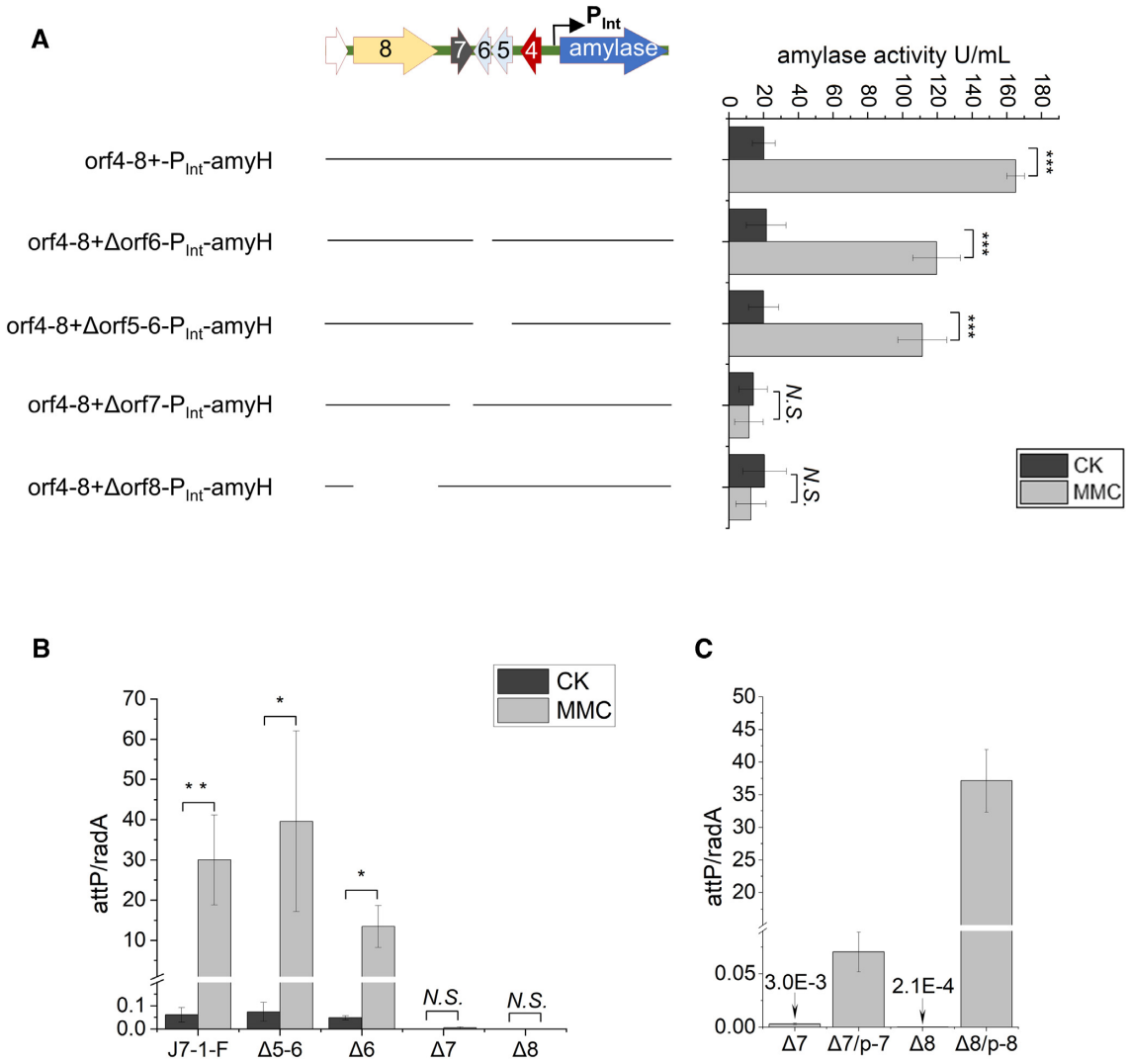

Figure 5. Orf7 and Orf8 are necessary for the derepression of the P_{Int} promoter upon MMC-induced DNA damage signals. (**A**) Amylase activity assay to determine the viral factors required for the derepression of the P_{Int} promoter. *orf6*, *orf5-6*, *orf7* and *orf8* were deleted individually in the reporter plasmid harboring orf4-8+-P_{Int}-amyH. The derived plasmids were transformed into J7-3-F. Amylase activity was detected by the FUWA method after 16 h. (**B**) qPCR analysis of SNJ2 copy number before and after MMC treatment in the indicated strains (indicated by *attP*). J7-1-F derivatives with deletion of *orf5-6*, *6*, *7* and *8* were grown in Halo-2 medium to log phase, treated with (MMC, 1μg/ml) or without (CK) induction. Samples were taken for qPCR analysis 24 h post-induction using the primer pairs attP-F/attP-R, and radA-F/radA-R. The copy number of SNJ2 virus relative to that of the chromosome was indicated by attP/radA. (**C**) qPCR analysis to test the complementation of Orf7 and Orf8. Δ7/*p*-7 and Δ8/*p*-8 represent *orf7* or *orf8* deletion strain complemented with a plasmid carrying *orf7* or *orf8* under the P_{phaR} promoter, respectively. qPCR analysis was performed as in (B) The bars represent the means and standard deviations of three independent experiments. Significance testing against CK group was performed using a one-sample *t* test (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; *N.S.*, not significant ($P > 0.05$)).

after MMC treatment, indicating that *orf5* and *orf6* are not important for the derepression of the P_{Int} promoter. Consistently, SNJ2 replication was not hindered in the *Δorf6* and *Δorf5-6* mutants upon MMC treatment, as determined by the relative copy number of the SNJ2 genome (Figure 5B). Therefore, the virus-encoded TA module does not participate in the regulation of SNJ2 lysogeny-induction switch under the tested conditions.

Orf8 is the primary viral factor activated in DNA damage response, which initiates *orf7* transcription

To determine whether Orf7 and Orf8 are required for the derepression of the P_{Int} promoter, each of them was deleted from the regulatory region in the reporter plasmid. The resultant constructs were transformed into J7-3-F cells (*Δsnj2*) and the amylase activity was tested. Excitingly, neither of the resulting constructs yielded amylase activity in J7-3-F cells post MMC treatment (Figure 5A), implying that they are critical for the derepression of the P_{Int} promoter. Consistent with this observation, deletion of either *orf7* or *orf8* in the J7-1-F strain resulted in the complete loss of SNJ2 replication upon MMC treatment, as determined by the relative copy number of the SNJ2 genome (Figure 5B). Note that deletion of either *orf7* or *orf8* also completely eliminated the spontaneous release of SNJ2, indicating that they are critical for fine-tuning the life cycle transition of SNJ2. We also checked whether

expression of these two proteins in trans would restore SNJ2 replication in the mutant strain. As shown in Figure 5C, when Orf8 was expressed under the P_{phaR} promoter from a plasmid in the J7-1-F*Δorf8* strain, replication of SNJ2 genome could be completely recovered upon MMC treatment. However, viral genome replication was only partially recovered in the J7-1-F*Δorf7* strain when it was complemented with a plasmid harboring *orf7* (Figure 5C), likely because it was not sufficiently expressed. Subsequent analysis found that transcription of *orf7* terminated downstream of its coding sequence. Thus, cloning the *orf7* coding sequence under a strong promoter was not sufficient for its expression, but including its downstream sequence (*orf7-4*) led to its successful expression and complementation (see below Figure 6F).

As shown in Figure 1A, the transcription of *orf7* was markedly upregulated in response to MMC treatment, whereas the transcription of *orf8* remained constant at a very low level. This indicates that Orf7 functions in a protein concentration-dependent manner, whereas Orf8 does not. To confirm this speculation, we overexpressed Orf8 from a plasmid under the strong P_{phaR} promoter in the J7-1-F*Δorf8* strain and checked the SNJ2 *attP* copy number. Despite the significant increase of *orf8* transcription (Figure 6A), the copy number of *attP* in the Orf8 overexpression strain was comparable to that of the wild-type strain before and after MMC treatment (Figure 6A and B).Western blot analysis of a His-tagged Orf8 in the *Δorf8* strain showed that its level did not change upon MMC treatment and it was not degraded, indicating that it is potentially activated via posttranslational modifications (PTMs).

Unlike *orf8*, the transcription of *orf7* was significantly upregulated upon MMC induction (Figure 1A). However, the transcription of *orf7* was nearly completely aborted in the J7-1-F*Δorf8* strain (Figure 6D), indicating that Orf8 is required for its transcription. In line with this, the transcription of *orf7* was recovered in the J7-1-F*Δorf8* strain when Orf8 was complemented *in trans* from a plasmid and upon MMC treatment (Figure 6D). To test if Orf8 regulates the transcription of *orf7* directly, the intergenic region between *orf7* and *orf8* (P_7 promoter) was cloned in front of the amylase gene in the reporter plasmid and amylase activity was tested in J7-3-F (*Δsnj2*), J7-1-F (*snj2*$^+$) and the J7-1-F*Δorf8* strains, respectively. As shown in Figure 6E, amylase activity increased significantly in J7-1-F cells upon MMC treatment, whereas it was not detected in neither J7-3-F nor J7-1-F*Δorf8* cells (Figure 6E), indicating that the activation of *orf7* transcription requires Orf8. To confirm Orf8 directly initiates *orf7* transcription, we incorporated *orf8* into the reporter plasmid (pFJ6-orf8-p7-amyH) and tested the amylase activity in J7-3-F cells upon MMC treatment. We observed a markedly increased amylase activity in the cells upon MMC treatment. However, if a frameshift mutation was introduced to block *orf8* translation, no amylase activity was observed (Figure 6E). Altogether, these results indicate that Orf8 is the critical viral factor that senses MMC-induced DNA damage response and initiates *orf7* transcription through the P_7 promoter.

Overexpression of Orf7 relieves Orf4-mediated repression of the P_{Int} promoter

The results above showed that expression of Orf7 is required for the derepression of the P_{Int} promoter, indicating that Orf7 antagonizes the activity of Orf4. To test if expression of Orf7 alone could counteract the function of Orf4 and lead to the derepression of the P_{Int} promoter, we constructed an amylase reporter plasmid, pFJ6-P_{tna}-*orf7-4*-P_{Int}-amyH, in which the P_7 promoter of Orf7 was replaced by the tryptophan-inducible promoter P_{tna}. The plasmid was transformed into J7-3-F cells and amylase activity in the presence or absence of 2mM Trp was determined. As shown in Figure 6F, in the absence of Trp, no amylase activity could be detected around the colonies containing the reporter plasmid. However, in the presence of Trp, strong amylase activity was observed, indicating that induction of Orf7 abolished the repression of the P_{Int} promoter by Orf4. As expected, when translation of *orf7* was blocked by a frameshift mutation in plasmid pFJ6-P_{tna}-orf7fsm-4-P_{Int}-amyH, no amylase activity was observed (Figure 6F). As a control, a reporter plasmid carrying Orf7 under its own promoter (requires Orf8 for transcriptional activation) could not derepress the P_{Int} promoter. Taken together, these results suggested that overexpression of Orf7 can bypass the requirement of Orf8 and counteract the function of Orf4 to derepress the P_{Int} promoter.

The three-gene regulatory module exemplified by SNJ2 *orf4-orf7-orf8* is widespread among proviruses of the genus *betapleolipovirus*

The main chromosome of *Natrinema* sp. J7 encodes seven other Orc1 homologs (Supplementary Figure S6) but they apparently could not compensate for the loss of the virus-derived Orc1 (Orf8 of SNJ2). As shown in Figure 6D, deletion of *orf8* caused the provirus to lose the ability to respond to the DNA damage signals, even if other *orc*1 genes are still present on the chromosome. Correspondingly, expression of P_7-amyH was successful only in J7-1-F cells, but not in J7-3-F or J7-1-F*Δorf8* cells after MMC treatment (Figure 6E). Thus, it appeared that viral Orf7 and Orc1 are functionally coupled. Unlike the commonly conserved repressor gene *orf4*, only several proviruses reported previously contain the *orc1* gene (18,20,43). To study the conservation and co-occurrence of the three genes, namely, genes encoding the repressor Orf4, Orf7 and Orc1(Orf8) that are critical for the lysogeny-induction switch of SNJ2, we searched for their homologs in the population of pleolipoproviruses. To this end, we first used the HMM-based searches against haloarchaeal protein sequences (class Halobacteria) at the non-redundant protein database at the National Center for Biotechnology Information (NCBI), using the HMM profile built from the pleolipoviral hallmark spike protein as a query (12,44). The genomic neighborhoods of the identified spike protein-encoding genes were analyzed for the presence of *orc1* and other genes of the conserved gene cluster of pleolipoviruses (see Materials and Methods). As a result, 31 Orc1-encoding proviruses were identified in haloarchaeal strains belonging to orders Natrialbales ($n = 15$), Halobacteriales ($n = 15$) and Haloferacales ($n = 1$), respec-

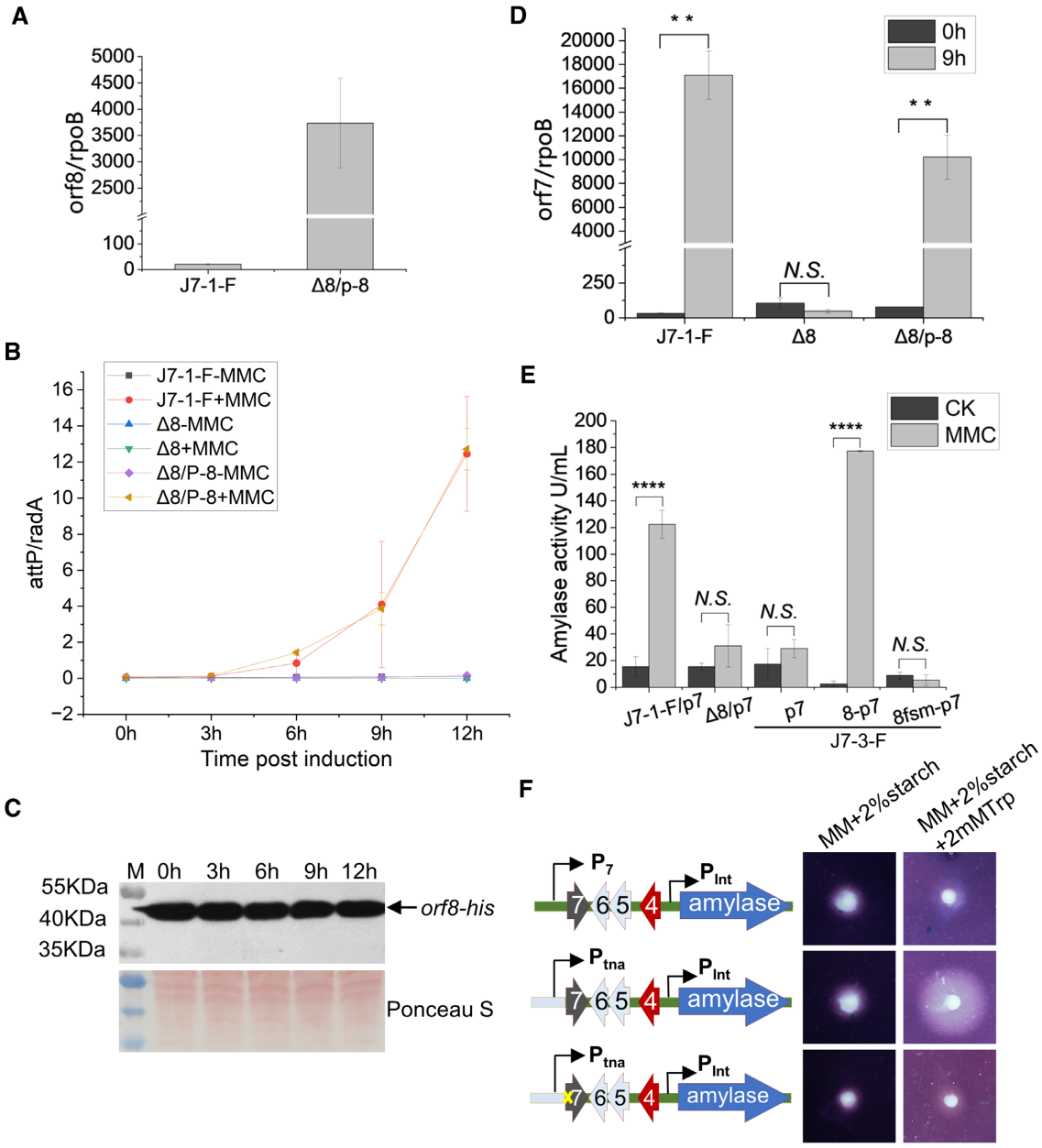

Figure 6. Orf8 responds to DNA damage signals and activates transcription of Orf7. (**A**) The transcription level of *orf8* was significantly increased when Orf8 was overexpressed from a plasmid. The expression level of *orf8* in J7-1-F and Δ8/*p*-8 at 9h after MMC treatment was estimated by qPCR normalized to the level of *rpoB* RNA. (**B**) MMC treatment was required for Orf8 to reactivate the replication of SNJ2. Strain J7-1-F(triangle) and Δ8/*p*-8 (circle) were cultured to middle exponential phase (OD_{600} = 0.6–0.8) in Halo-2 medium and treated with (solid) or without (hollow) MMC (1μg/ml) for 30 min. Samples were taken every three hours post induction and the relative copy number of SNJ2 was measured by qPCR using the primer pairs attP-F/attP-R and radA-F/radA-R. (**C**) Western blot analysis of Orf8-his protein in strain Δ8/*p*-8-his post MMC treatment. Samples were collected at 0, 3, 6, 9 and 12h after MMC treatment (OD_{600} was adjusted to take the same amounts of samples) and crude extracts were loaded for SDS-PAGE. Anti-his antibodies were used for detection of Orf8-his. Lower panel shows Ponceau S staining of the membrane as loading controls. (**D**) Relative expression levels of *orf7* in J7-1-F, J7-1-FΔ8 and Δ8/*p*-8 at 0 and 9 h after MMC treatment. The expression level of *orf7* in the samples was estimated by qPCR with the obtained data normalized to the level of *rpoB* RNA. (**E**) Amylase activity test to check the regulation of the P_7 promoter by Orf8. Plasmid pFJ6-P_7-amyH, pFJ6-orf8-P_7-amyH and pFJ6-orf8f-P_7-amyH with amylase gene under the control of the P_7 promoter were transformed into the indicated strains: J7-1-F (*snj2*$^+$), J7-3-F (*Δsnj2*) and J7-1-FΔ8 (*Δorf8*). Strains were cultured and then treated with (MMC) or without (CK) induction and amylase activity were detected by FUWA method after 16 h. (**F**) Overexpression of Orf7 antagonizes Orf4-mediated repression of the P_{Int} promoter. Reporter plasmids were transformed into J7-3-F and amylase activity was tested as in Figure 4B (Only one representative is shown here). The yellow cross represents the introduction of a frameshift mutation in *orf7*.Error bars: standard derivations of three independent experiments. Significance testing against 0h group was performed using a one-sample *t* test (****$P < 0.0001$; **$P < 0.01$; N.S., not significant ($P > 0.05$)).

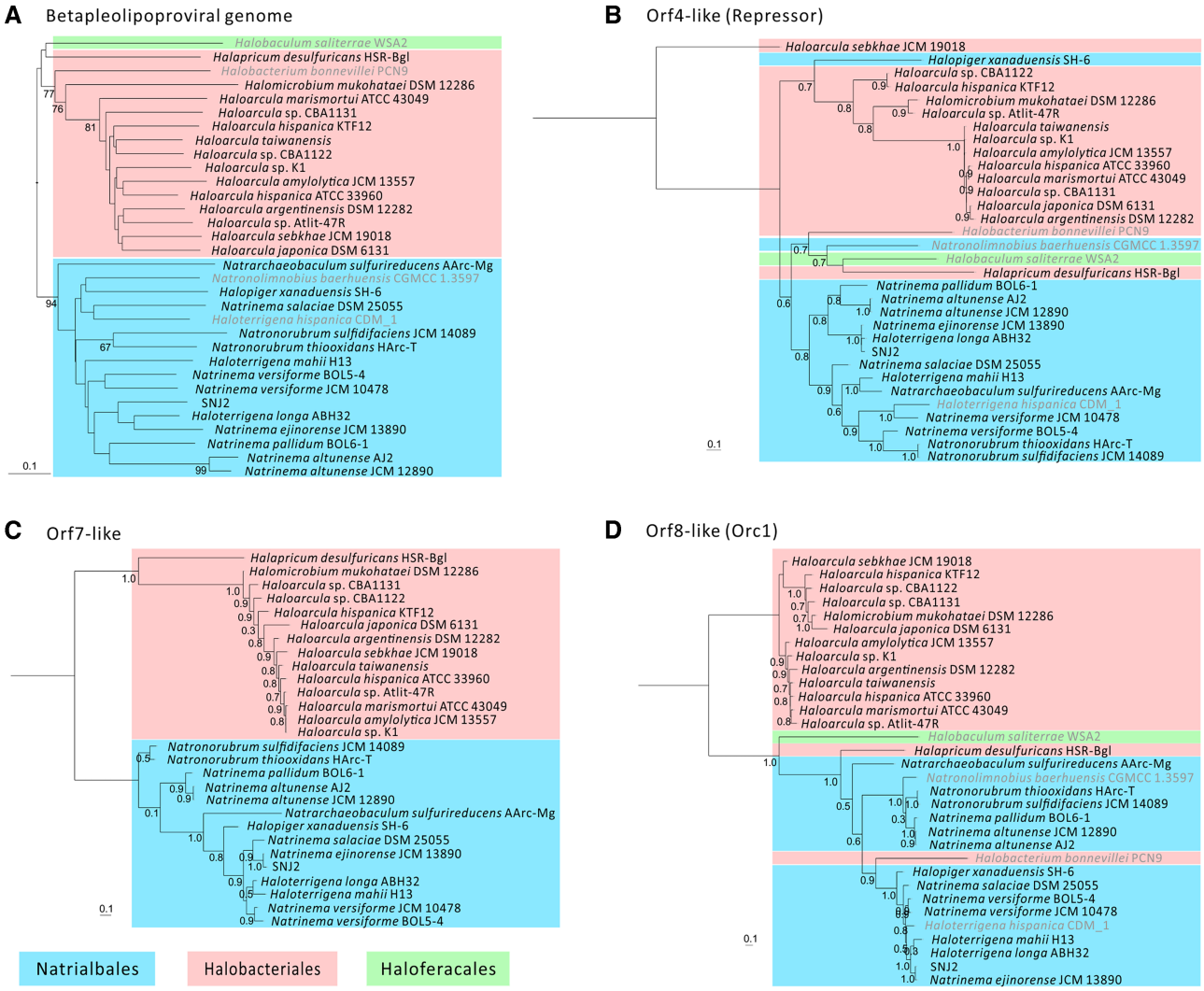

Figure 7. Phylogenetic analyses of pleolipoproviruses containing the SNJ2-like three-gene regulatory module. (**A**) Phylogenomic tree of pleolipoproviruses based on whole genome VICTOR(35) analysis at the amino acid level. The bootstrap support values from 100 replications are indicated in the nodes, and the branch length is scaled in terms of the Genome BLAST Distance Phylogeny (GBDP) distance formula D6. The maximum likelihood phylogenetic trees of (**B**). Orf4-like repressors, (**C**) Orf7-like proteins and (**D**) Orf8-like Orc1 proteins encoded by the pleolipoproviruses. The trees are mid-rooted, and numbers at the nodes represent support values. The scale bars represent the number of substitutions per site. Proviruses are indicated by names of their host strains, and displayed with colored blocks according to the orders of their host strains, namely *Natrialbales*, *Halobacteriales* and *Haloferacales*, respectively. Proviruses that lack the Orf7-like protein encoding gene are indicated by grey in the trees.

tively. All these proviruses are related to members of the genus *Betapleolipovirus* (Table S1). To assess the relationship between these viruses and SNJ2, phylogenomic analysis was performed using Genome BLAST Distance Phylogeny method implemented in VICTOR (35). The resultant tree revealed that proviruses infecting those from Natrialbales and Halobacteriales fall into two well-supported clades, indicating that the genetic divergence between the proviruses correlates with their respective host ranges (Figure 7 and Supplementary Figure S7). Interestingly, *orf7*-like genes could be identified in 27 of the Orc1 encoding-proviruses, consistent with the functional coupling between Orf7-like and Orc1 proteins in these proviruses. The 4 proviruses that lack the *orf7*-like gene may have replaced *orf7* with analogous but nonorthologous gene(s) or the pathway (and proviruses) may have been inactivated. The genetic arrangement of *orf4*-, *orf7*- and *orc1*-like genes is conserved in all 27 proviruses, resembling that in SNJ2 (Supplementary Figure S7). Phylogenetic analyses of the three gene products revealed a largely congruent pattern that was also consistent with the phylogenomic tree (Figure 7). Despite a few cases in the Repressor and Orc1 trees, gene exchange appears to be largely restricted among proviruses belonging to distinct clades, likely due to the typically narrow host range of pleolipoviruses (12).

Comparative genomic analysis of the pleolipoproviruses revealed that the integrase and the putative repressor are encoded by nearly all identified pleolipoproviruses. Moreover, the genetic arrangement of the integrase and repressor genes in these proviruses resembles that in SNJ2 (Figure 7) (18,20), implying a similar regulatory mechanism for the maintenance of the lysogenic state. To test this pos-

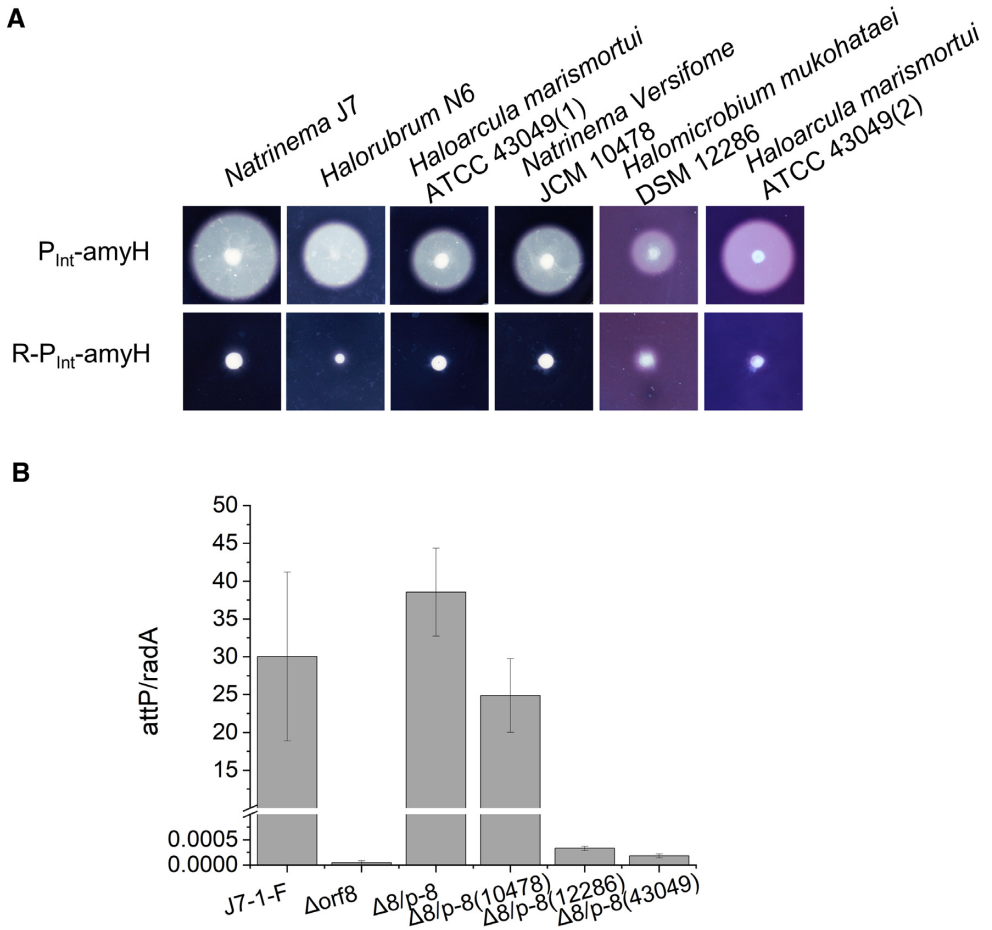

Figure 8. Functional characterization of Orf4 and Orf8 homologues from SNJ2-like proviruses. (**A**) Orf4-like proteins (R: repressor) from SNJ2-like proviruses repress corresponding P_{Int} promoter activity. Reporter plasmids harboring the P_{Int} promoters or P_{Int} promoters with Orf4-like repressors from diverse haloarchaeal species were transformed into *Haloferax volcanii* H26. Amylase activity of three independent transformants for each construct was tested as in Figure 4B (Only one is shown here). (**B**) Functional characterization of Orf8 homologues from different SNJ2-like proviruses. *orf8* orthologs from SNJ2-like proviruses from *Natrinema versifome* JCM10478, *Halomicrobium mukohataei* DSM 12286 and *Haloarcula marismortui* ATCC43049 were cloned into pFJ6 plasmid and expressed under the P_{phaR} promoter. These plasmids were transformed into J7-1-FΔ*orf8*. Transformants were cultured in Halo-2 and then treated with MMC. Complementation of *orf8* deletion by its homologs was checked by the relative level of *attP* of SNJ2 after MMC treatment. J7-1-F was included as a positive control. The bars represent the means and standard deviations of three independent experiments.

sibility, we checked the repression of the P_{Int} promoters of 5 SNJ2-like proviruses by their cognate putative repressors using the amylase reporter system described above. The five selected proviruses infect haloarchaeal species belonging to four genera, including *Haloarcula marismortui* ATCC 43049, *Natrinema versifome* JCM 10478, *Halomicrobium mukohataei* DSM 12286 and *Halorubrum* sp. N6 (unpublished lab strain), respectively. As shown in Figure 8A, all five putative repressors were able to repress the promoter of their cognate integrases, confirming that temperate pleolipoproviruses use a common mechanism to maintain lysogeny across different genera of haloarchaea.

To test if other proviral Orc1 homologs also respond to MMC-induced DNA damage, we checked if they could complement the function of Orf8 of SNJ2 in the J7-1-F*Δorf8* strain. Among the three tested Orc1 homologs, only the one encoded by the provirus from *Natrinema versifome* JCM10478 could successfully restore the replication of SNJ2*Δorf8* after MMC treatment, whereas the other two from *Halomicrobium mukohataei* DSM 12286 and *Haloarcula marismortui* ATCC43049 failed (Figure 8B). This result suggests that the viral Orc1s and Orf7-like proteins coevolve and thus divergent orthologs can no longer substitute the SNJ2 Orc1 (they belong to distinct clades as shown in Figure 7), or because of improper protein folding in a heterologous cell. Altogether, we showed that the three-gene regulatory module is encoded by a group of betapleolipoproviruses, suggesting that the SNJ2-like lysogeny-induction switch mechanism is also employed by these proviruses.

DISCUSSION

Archaeal viruses are known to play important ecological roles in various ecosystems (45,46), and many archaeal viruses are temperate (18,47–53). However, the regulatory mechanisms that govern their life cycles remain poorly understood, hindering our appreciation of their ecological impacts. Here, we uncovered the molecular mechanism behind the lysogeny-induction switch of the temperate haloarchaeal virus SNJ2. The identified three-gene regulatory module is conserved in a group of proviruses belonging to

the genus *Betapleolipovirus*, suggestive of a common regulatory strategy employed by these proviruses. Thus, SNJ2 may serve as a prototype for the study of lysogeny-induction switch of temperate betapleolipoviruses.

SNJ2 could maintain in host cells as a carrier state for several passages, but only if the int^{SNJ2} gene is intact (14,19,20). As shown in this study, deletion of int^{SNJ2} from the provirus completely abolishes virus replication. Presumably, Int^{SNJ2}-mediated provirus excision from host chromosome and subsequent circularization are prerequisite for virus replication. To maintain a lysogenic state, proviruses often suppress the promoter(s) of the lytic pathway genes by the virus encoded repressor(s) (5). In this study, we showed that the transcription of int^{SNJ2} is tightly regulated by the virus-encoded repressor Orf4. A palindromic sequence 5'-CGTACTTCCTAGTACG-3' as well as its upstream flanking sequences adjacent to the TATA box of the P_{Int} promoter are critical for Orf4-mediated repression. In the temperate haloarchaeal virus φH1, the repressor protein was suggested to bind to two short interrupted palindromes upstream of its start codon, and it was suggested to be a part of an auto-regulation system (42). Orf4 potentially functions in a similar manner by binding to the palindromic sequence and adjacent area. As the palindrome is adjacent to the TATA box, the binding of Orf4 may interfere with the RNA polymerase or transcription factor B (TFB) association, thereby repressing int^{SNJ2} transcription. However, the binding mode of Orf4 to the P_{Int} promoter requires further studies. The integrase and a putative repressor are encoded in nearly all pleolipoproviruses, and in a few cases, two consecutive repressor genes are encoded (54). The repressor gene(s) is located in the upstream and in opposite transcriptional direction of the corresponding integrase gene in these proviruses, resembling the gene arrangement in SNJ2 (18,20). Our assays with the putative repressors encoded by 5 pleolipoproviruses infecting host species belonging to 4 haloarchaeal genera, *Haloarcula*, *Natrinema*, *Halomicrobium* and *Halorubrum*, confirmed their repression of the corresponding promoter regions of viral integrase genes (Figure 7A). These results indicate that the mechanism for the maintenance of lysogeny is conserved in pleolipoproviruses.

To inactivate the repressor and switch from lysogenic to induced virus life cycle, SNJ2 employs a circuit not previously seen in other reported temperate viruses. Unlike the autoproteolysis of the CI repressor in lambda phage or the inactivation of the regulator through dissociation of a protein complex in archaeal SSV1 virus upon induction (9,10,55,56), the SNJ2 repressor could not directly respond to the host DNA damage. We found that two other viral factors: the Orc1-like protein Orf8, and Orf7, are required for the SNJ2 lysogeny-induction switch. Orf8 responds to host DNA damage and upregulates the transcription of *orf7*, whose product then acts as the inactivator of the repressor Orf4, thereby derepressing int^{SNJ2} transcription (Figure 9). The level of Orf8 did not change after MMC treatment and overexpression of Orf8 could not activate *orf7* transcription, indicating that it may be activated by some kind of posttranslational modifications. Modification site prediction indicated that Lys-53, a residue in the conserved walker A motif of the SNJ2 Orc1 AAA+ domain, could be acetylated. We speculated that acetylation/deacetylation of Lys-53 may be responsible for the activation of Orf8. However, direct evidence for posttranslational modifications of Orf8 is missing so far. Future studies to check the posttranslational modification of SNJ2 Orc1 before and post MMC treatment would be key to uncover its working mechanism.

Orc1/Cdc6 superfamily proteins are widely present in archaea and normally function as a key component of the host replisome responsible for origin recognition and helicase loading. Although most archaea contain more than one Orc1 homologs and replication origins, haloarchaea contain a particularly high number (usually more than 10) of Orc1 encoding genes that outnumbers the replication origins (57). The exact functions of most Orc1 homologs remain unclear in haloarchaea, although some were proven to function as *bona fide* replication initiation proteins, such as the Orc1 homologs in *Halobacterium salinarum* NRC-1 and *Haloarcula hispanica* (57–59). A recent study revealed that one of the Orc1 homologs in a hyperthermophilic crenarchaeon *Saccharolobus islandicus*, Orc1-2, acts as a global regulator of DNA damage response (60), whereas the other two Orc1 homologs function as replication initiation proteins (61). Here we showed that the Orc1 protein encoded by SNJ2 is involved in DNA damage response, the first example in the phylum *Euryarchaeota*. Deletion of *orf8* from the provirus did not result in obvious growth defect of the host cell under normal growth condition. However, the mutant strain appeared more sensitive than the wild-type strain to MMC treatment (Supplementary Figure S8), implying the participation of SNJ2 Orc1 in regulating the host DNA damage response (DDR). The phenomenon is similar to the hypersensitivity of *S. islandicus Δorc1-2* mutant to the NQO (4-nitroquinoline 1-oxide)-induced DNA damage (60). Intriguingly, the activation of *S. islandicus* Orc1-2 is also suggested to be dependent on PTMs. In *S. islandicus*, activated Orc1-2 upregulates the transcription of DDR genes through binding to the promoters of DDR genes (60). We showed that the activated SNJ2 Orc1 upregulates *orf7* transcription through the promoter region of *orf7* (Figure 6E). Notably, in the case of SNJ2 Orc1 its expression level was not affected by DNA damage (Figure 1A), however, the expression of *S. islandicus* Orc1-2 was strongly upregulated upon UV irradiation (62,63), suggesting that they might function differently. Further tests to determine the SNJ2 Orc1 binding site on the P_7 promoter, as well as to those of potentially associated host DDR genes will allow functional comparisons between SNJ2 Orc1 and *S. islandicus* Orc1-2.

It is noteworthy that the main chromosome of *Natrinema* sp. J7 encodes seven Orc1 homologs in addition to SNJ2 Orc1, but none of them could compensate for the loss of the viral Orc1 to regulate viral *orf7* transcription (Figure 6D and E). This observation suggests that SNJ2 Orc1 may be the only Orc1 that is involved in DNA damage response, or the host Orc1 proteins(s) is involved but merely not regulating the provirus. Regardless, as shown in Supplementary Figure S8, the integrated SNJ2 provirus confers an increased fitness to the host cell upon DNA damage. Interestingly, several haloarchaeal head-tailed (pro)viruses were reported to code for Orc1-like proteins (43,49,64,65), and these viruses all encode integrase gene therefore having a

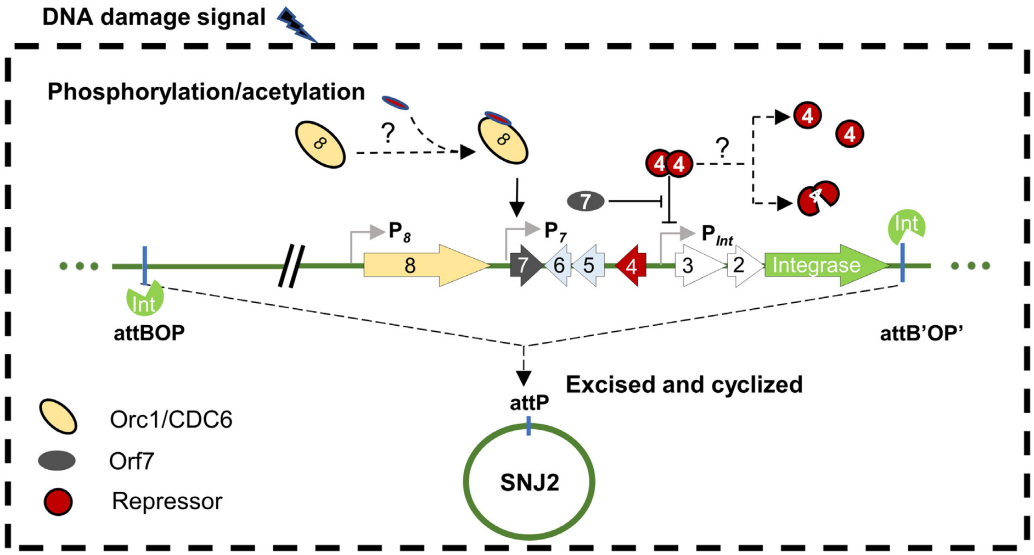

Figure 9. Schematic diagram of the regulation of SNJ2 life cycle transition. In the lysogenic state, SNJ2 proviruses integrate into the host genome as a linear form by integrase (aqua). Reactivation of viral replication requires excision of SNJ2 proviral genome from the host chromosome and subsequent circularization. However, the expression of integrase is repressed by the repressor Orf4 (red), thereby maintaining SNJ2 in the lysogenic state. Upon host DNA damage (black lightning shape), the Orc1/Cdc6 protein encoded by *orf8* (golden) is activated, possibly by posttranslational modification (crimson). Activated Orf8 turns on the expression of *orf7* (gray), which antagonizes the function of Orf4, leading to the derepression of integrase expression. Accumulation of Int^{SNJ2} mediates the excision and circularization of SNJ2 provirus, ultimately leading to the reactivation of viral replication.

potential temperate life cycle. Perhaps these Orc1-like proteins are also involved in DNA damage response and play a critical role in the life cycle switch.

Overexpression of Orf7 led to the derepression of the P_{Int} promoter in strains expressing the repressor Orf4, suggesting that accumulation of Orf7 overwhelms the repression of Orf4. Thus, a balance between Orf7 and Orf4 fine tunes the life cycle of SNJ2. An attractive and simple mechanism is that expression of Orf7 leads to the cleavage or degradation of Orf4. However, we were unable to detect a His-tagged Orf4 protein or its degradation product by western blot even though the protein fusion was functional (Supplementary Figure S9), presumably because the protein level of His-tagged Orf4 was below the detection limit. An alternative possibility is that Orf7 interacts directly with Orf4, interfering with its dimerization or binding to the P_{Int} promoter, thus relieving the repression. However, we did not observe an interaction between the two proteins so far. Thus, how Orf7 works is not clear yet. Homology searches using BLASTP or HHpred failed to predict the function of Orf7 due to the lack of sequence similarity to other characterized proteins. However, among pleolipoviruses, Orf7 homologs display considerable sequence conservation with distinct conservation patterns discernible for Natrialbales and Halobacteriales proviruses (Supplementary Figure S10A), further showing that Orf7-encoding proviruses have coevolved with their hosts with minimal horizontal transfer of the *orf7* gene. Structural modeling of Orf7 using AlphaFold2 (36,37) produced a high-quality model showing a fold consisting of a slightly twisted 6-stranded β-sheet with an α-helix packed against it (Supplementary Figure S10B). Unfortunately, structure-based searches against the PDB database using DALI did not reveal significant matches.

The survey of the pleolipoproviruses in haloarchaeal genomes revealed that only ∼20% of proviruses belonging to the genus *Betapleolipovirus* contain the coupled *orf7*-like and *orc1* genes. Interestingly, these two genes were not identified in proviral members of the genus *Alphapleolipovirus* (no provirus from the genus *Gammapleolipovirus* was identified during our searches). It is clear that proviruses from the family *Pleolipoviridae* employ distinct mechanisms to regulate their life cycle switch. Moreover, different mechanisms could be employed by closely related betapleolipoproviruses, suggestive of a possible horizontal acquisition of the *orc1-orf7-like* gene module of the virus. The ultimate provenance of the *orc1-orf7-like* gene module (e.g. of MGE or cellular origin) remains unclear.

In addition to the Orf4–Orf7–Orf8 regulatory module, we identified a type II TA system in the SNJ2 genome. Although this TA system is functional (Supplementary Figure S5), it is not involved in virus life cycle switch (Figure 5A and B). BLASTP searches revealed that homologs of the SNJ2 TA module are widely encoded by haloarchaea. SNJ2 likely acquired the system from the host and may use it for stable provirus maintenance.

Collectively, our study reveals the molecular mechanism underlying the lysogeny-induction switch of the temperate haloarchaeal virus SNJ2, which is regulated by a virus encoded DNA damage signaling pathway (Figure 9). The SNJ2 Orf4 is a repressor of the int^{SNJ2} transcription, maintaining the virus in the lysogenic state. Following host DNA damage, SNJ2-encoded Orc1 is activated, potentially via posttranslational modifications, to initiate the expression of Orf7. Accumulation of Orf7 antagonizes the function of Orf4, leading to the derepression of the int^{SNJ2} transcription, which ultimately reactivates the replication cycle of SNJ2. Such a regulatory strategy for the lysogeny/induction

switch appears to be common among a group of betapleolipoproviruses. Through study of the life cycle transition of SNJ2, we discovered for the first time that an Orc1 superfamily protein is involved in DNA damage response in a euryarchaeon, and employed by a temperate virus for life cycle regulation. SNJ2 Orf7 likely represents a novel protein family, and how the protein mediates the derepression of int^{SNJ2} remains to be investigated.

DATA AVAILABILITY

The accession number for the RNA-seq data of Natrinema sp. J7-1 in this paper is GEO: GSE224586.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank members of the Chen and Du laboratories for comments and advice in preparing the manuscript. We thank Prof. Yuping Huang (College of Life Sciences, Wuhan University) for kindly providing the pNBK-F plasmid. We thank Dr Hua Xiang (Institute of Microbiology, Chinese Academy of Sciences) for kindly providing the pRF plasmid.
Author contributions: Z.C., X.C. and S.D. designed the research. Z.C., Y.W., X.D., X.D., J.X., Y.W. and J.W., performed the research. Z.C., Y.L., M.K., X.C. and S.D. analyzed data and wrote the manuscript.

FUNDING

National Natural Science Foundation of China [32270167]; National Foundation for Fostering Talents of Basic Sciences [J1103513]; Research (Innovative) Fund of Laboratory Wuhan University (to X.C.).
Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Touchon,M., Bernheim,A. and Rocha,E.P. (2016) Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.*, **10**, 2744–2754.
3. Howard-Varona,C., Hargreaves,K.R., Abedon,S.T. and Sullivan,M.B. (2017) Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.*, **11**, 1511–1520.
4. 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.
5. Brady,A., Felipe-Ruiz,A., del Sol,F.G., Marina,A., Quiles-Puchalt,N. and Penades,J.R. (2021) Molecular basis of lysis-lysogeny decisions in gram-positive phages. *Annu. Rev. Microbiol.*, **75**, 563–581.
6. Oppenheim,A.B., Kobiler,O., Stavans,J., Court,D.L. and Adhya,S. (2005) Switches in bacteriophage lambda development. *Annu. Rev. Genet.*, **39**, 409–429.
7. Prangishvili,D. (2013) The wonderful world of archaeal viruses. *Annu. Rev. Microbiol.*, **67**, 565–585.
8. Prangishvili,D., Bamford,D.H., Forterre,P., Iranzo,J., Koonin,E.V. and Krupovic,M. (2017) The enigmatic archaeal virosphere. *Nat. Rev. Microbiol.*, **15**, 724–739.
9. Fusco,S., Aulitto,M., Iacobucci,I., Crocamo,G., Pucci,P., Bartolucci,S., Monti,M. and Contursi,P. (2020) The interaction between the F55 virus-encoded transcription regulator and the RadA host recombinase reveals a common strategy in Archaea and Bacteria to sense the UV-induced damage to the host DNA. *Bba-Gene Regul. Mech.*, **1863**, 194493.
10. Fusco,S., She,Q.X., Fiorentino,G., Bartolucci,S. and Contursi,P. (2015) Unravelling the Role of the F55 Regulator in the Transition from Lysogeny to UV Induction of Sulfolobus Spindle-Shaped Virus 1. *J. Virol.*, **89**, 6453–6461.
11. Chen,B.B., Chen,Z., Wang,Y.C., Gong,H., Sima,L.S., Wang,J., Ouyang,S.S., Gan,W.Q., Krupovic,M., Chen,X.D. *et al.* (2020) ORF4 of the Temperate Archaeal Virus SNJ1 Governs the Lysis-Lysogeny Switch and Superinfection Immunity. *J. Virol.*, **94**, e00841-20.
12. Demina,T.A. and Oksanen,H.M. (2020) Pleomorphic archaeal viruses: the family Pleolipoviridaeis expanding by seven new species. *Arch. Virol.*, **165**, 2723–2731.
13. Dyall-Smith,M., Pfeiffer,F., Chiang,P.W. and Tang,S.L. (2021) Genome sequence of Hardyhisp2, a Gammapleolipovirus infecting haloarcula hispanica. *Microbiol. Resour. Ann.*, **10**, 00226–00221.
14. 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**, 6384–6384.
15. Krupovic,M., Cvirkaite-Krupovic,V., Iranzo,J., Prangishvili,D. and Koonin,E.V. (2018) Viruses of archaea: structural, functional, environmental and evolutionary genomics. *Virus Res.*, **244**, 181–193.
16. 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.
17. Bath,C., Cukalac,T., Porter,K. and Dyall-Smith,M.L. (2006) His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology*, **350**, 228–239.
18. Liu,Y., Wang,J., Liu,Y., Wang,Y.C., Zhang,Z.Q., Oksanen,H.M., Bamford,D.H. and Chen,X.D. (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.
19. 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.
20. Wang,J., Liu,Y.C., Liu,Y., Du,K.X., Xu,S.Q., Wang,Y.C., Krupovic,M. and Chen,X.D. (2018) A novel family of tyrosine integrases encoded by the temperate pleolipovirus SNJ2. *Nucleic Acids Res.*, **46**, 2521–2536.
21. 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.
22. Charlebois,R.L., Lam,W.L., Cline,S.W. and Doolittle,W.F. (1987) Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaebacterium. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 8530–8534.
23. Cline,S.W., Lam,W.L., Charlebois,R.L., Schalkwyk,L.C. and Doolittle,W.F. (1989) Transformation methods for halophilic archaebacteria. *Can. J. Microbiol.*, **35**, 148–152.
24. 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**, 570–572.
25. Fuwa,H. (1954) A new method for microdetermination of amylase activity by the use of amylose as the substrate. *J. Biochem.*, **41**, 583–603.
26. Coronado,M.J., Vargas,C., Mellado,E., Tegos,G., Drainas,C., Nieto,J.J. and Ventosa,A. (2000) The alpha-amylase gene amyH of the moderate halophile *Halomonas meridiana*: cloning and molecular characterization. *Microbiol-Sgm*, **146**, 861–868.
27. Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.*, **12**, 323.
28. Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., van Baren,M.J., Salzberg,S.L., Wold,B.J. and Pachter,L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.

29. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J.H., 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.
30. 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.
31. Drozdetskiy,A., Cole,C., Procter,J. and Barton,G.J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.*, **43**, W389–W394.
32. Pei,J.M., Kim,B.H. and Grishin,N.V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.*, **36**, 2295–2300.
33. 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.
34. Eddy,S.R. (2011) Accelerated profile HMM Searches. *PLoS Comput. Biol.*, **7**, e1002195.
35. Meier-Kolthoff,J.P. and Goker,M. (2017) VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics*, **33**, 3396–3404.
36. Mirdita,M., Schutze,K., Moriwaki,Y., Heo,L., Ovchinnikov,S. and Steinegger,M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*, **19**, 679–682.
37. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
38. Pettersen,E.F., Goddard,T.D., Huang,C.R.C., Meng,E.E.C., Couch,G.S., Croll,T.I., Morris,J.H. and Ferrin,T.E. (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.*, **30**, 70–82.
39. Wang,Y.C., Chen,B.B., Sima,L.S., Cao,M.Z. and Chen,X.D. (2017) Construction of expression shuttle vectors for the haloarchaeon *Natrinema* sp J7 based on its chromosomal origins of replication. *Archaea*, **2017**, 4237079.
40. Waterhouse,A., Bertoni,M., Bienert,S., Studer,G., Tauriello,G., Gumienny,R., Heer,F.T., de Beer,T.A.P., Rempfer,C., Bordoli,L. *et al.* (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.*, **46**, W296–W303.
41. Aravind,L., Anantharaman,V., Balaji,S., Babu,M.M. and Iyer,L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231–262.
42. Ken,R. and Hackett,N.R. (1991) Halobacterium halobium strains lysogenic for phage phi H contain a protein resembling coliphage repressors. *J. Bacteriol.*, **173**, 955–960.
43. Dyall-Smith,M., Pfeiffer,F., Chiang,P.W. and Tang,S.L. (2021) The novel halovirus Hardycor1, and the presence of active (induced) proviruses in four haloarchaea. *Genes (Basel)*, **12**, 149.
44. El Omari,K., Li,S., Kotecha,A., Walter,T.S., Bignon,E.A., Harlos,K., Somerharju,P., De Haas,F., Clare,D.K., Molin,M. *et al.* (2019) The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. *Nat. Commun.*, **10**, 846.
45. Danovaro,R., Dell'Anno,A., Corinaldesi,C., Rastelli,E., Cavicchioli,R., Krupovic,M., Noble,R.T., Nunoura,T. and Prangishvili,D. (2016) Virus-mediated archaeal hecatomb in the deep seafloor. *Sci. Adv.*, **2**, e1600492.
46. Roux,S., Enault,F., Ravet,V., Colombet,J., Bettarel,Y., Auguet,J.C., Bouvier,T., Lucas-Staat,S., Vellet,A., Prangishvili,D. *et al.* (2016) Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environ. Microbiol.*, **18**, 889–903.
47. Witte,A., Baranyi,U., Klein,R., Sulzner,M., Luo,C., Wanner,G., Kruger,D.H. and Lubitz,W. (1997) Characterization of *Natronobacterium magadii* phage Phi Ch1, a unique archaeal phage containing DNA and RNA. *Mol. Microbiol.*, **23**, 603–616.
48. Prangishvili,D., Vestergaard,G., Haring,M., Aramayo,R., Basta,T., Rachel,R. and Garrett,R.A. (2006) Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *J. Mol. Biol.*, **359**, 1203–1216.
49. 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.
50. Mochizuki,T., Sako,Y. and Prangishvili,D. (2011) Provirus induction in hyperthermophilic archaea: characterization of *Aeropyrum pernix* spindle-shaped virus 1 and *Aeropyrum pernix* ovoid virus 1. *J. Bacteriol.*, **193**, 5412–5419.
51. Zhang,Z.Q., Liu,Y., Wang,S., Yang,D., Cheng,Y.C., Hu,J.N., Chen,J., Mei,Y.J., 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.
52. Contursi,P., Fusco,S., Cannio,R. and She,Q.X. (2014) Molecular biology of fuselloviruses and their satellites. *Extremophiles*, **18**, 473–489.
53. SCHNABEL,H. (1984) An immune strain of Halobacterium halobium carries the invertible L segment of phage 1H as a plasmid. *Proc. Nati. Acad. Sci. U.S.A.*, **81**, 1017–1020.
54. Atanasova,N.S., Heinio,C.H., Demina,T.A., Bamford,D.H. and Oksanen,H.M. (2018) The unexplored diversity of Pleolipoviruses: the surprising case of two viruses with identical major structural modules. *Genes-Basel*, **9**, 131.
55. Sauer,R.T., Ross,M.J. and Ptashne,M. (1982) Cleavage of the lambda and P22 repressors by Reca protein. *J. Biol. Chem.*, **257**, 4458–4462.
56. Galkin,V.E., Yu,X., Bielnicki,J., Ndjonka,D., Bell,C.E. and Egelman,E.H. (2009) Cleavage of bacteriophage lambda cI repressor involves the RecA C-terminal domain. *J. Mol. Biol.*, **385**, 779–787.
57. Wu,Z.F., Liu,H.L., Liu,J.F., Liu,X.Q. and Xiang,H. (2012) Diversity and evolution of multiple orc/cdc6-adjacent replication origins in haloarchaea. *BMC Genomics*, **13**, 478.
58. Berquist,B.R. and DasSarma,S. (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, Halobacterium sp. strain NRC-1. *J. Bacteriol.*, **185**, 5959–5966.
59. Wu,Z., Liu,J., Yang,H., Liu,H. and Xiang,H. (2014) Multiple replication origins with diverse control mechanisms in *Haloarcula hispanica*. *Nucleic Acids Res.*, **42**, 2282–2294.
60. Sun,M., Feng,X., Liu,Z., Han,W., Liang,Y.X. and She,Q. (2018) An Orc1/Cdc6 ortholog functions as a key regulator in the DNA damage response in Archaea. *Nucleic Acids Res.*, **46**, 6697–6711.
61. Samson,R.Y., Xu,Y.Q., Gadelha,C., Stone,T.A., Faqiri,J.N., Li,D.F., Qin,N., Pu,F., Liang,Y.X., She,Q.X. *et al.* (2013) Specificity and function of archaeal DNA replication initiator proteins. *Cell Rep.*, **3**, 485–496.
62. Frols,S., Gordon,P.M., Panlilio,M.A., Duggin,I.G., Bell,S.D., Sensen,C.W. and Schleper,C. (2007) Response of the hyperthermophilic archaeon Sulfolobus solfataricus to UV damage. *J. Bacteriol.*, **189**, 8708–8718.
63. Gotz,D., Paytubi,S., Munro,S., Lundgren,M., Bernander,R. and White,M.F. (2007) Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome Biol.*, **8**, R220.
64. Dyall-Smith,M., Palm,P., Wanner,G., Witte,A., Oesterhelt,D. and Pfeiffer,F. (2019) *Halobacterium salinarum* virus ChaoS9, a novel halovirus related to PhiH1 and PhiCh1. *Genes-Basel*, **10**, 194.
65. Pagaling,E., Haigh,R.D., Grant,W.D., Cowan,D.A., Jones,B.E., Ma,Y., Ventosa,A. and Heaphy,S. (2007) Sequence analysis of an Archaeal virus isolated from a hypersaline lake in Inner Mongolia, China. *BMC Genomics*, **8**, 410.