

# SCIENTIFIC REPORTS



OPEN

## Genes essential for the morphogenesis of the Shiga toxin 2-transducing phage from *Escherichia coli* O157:H7

Received: 16 September 2016

Accepted: 16 November 2016

Published: 14 December 2016

Shakhinur Islam Mondal<sup>1,2,\*</sup>, Md Rakibul Islam<sup>3,\*</sup>, Akira Sawaguchi<sup>4</sup>, Md Asadulghani<sup>5</sup>, Tadasuke Ooka<sup>6</sup>, Yasuhiro Gotoh<sup>7</sup>, Yasuhiro Kasahara<sup>8</sup>, Yoshitoshi Ogura<sup>7</sup> & Tetsuya Hayashi<sup>7</sup>

Shiga toxin 2 (Stx2), one of the most important virulence factors of enterohaemorrhagic *Escherichia coli* (EHEC), is encoded by phages. These phages (Stx2 phages) are often called lambda-like. However, most Stx2 phages are short-tailed, thus belonging to the family *Podoviridae*, and the functions of many genes, especially those in the late region, are unknown. In this study, we performed a systematic genetic and morphological analysis of genes with unknown functions in Sp5, the Stx2 phage from EHEC O157:H7 strain Sakai. We identified nine essential genes, which, together with the terminase genes, determine Sp5 morphogenesis. Four of these genes most likely encoded portal, major capsid, scaffolding and tail fiber proteins. Although exact roles/functions of the other five genes are unknown, one was involved in head formation and four were required for tail formation. One of the four tail genes encoded an unusually large protein of 2,793 amino-acid residues. Two genes that are likely required to maintain the lysogenic state were also identified. Because the late regions of Stx2 phages from various origins are highly conserved, the present study provides an important basis for better understanding the biology of this unique and medically important group of bacteriophages.

Bacteriophages (phages) are the most abundant living entities on earth. Some bacteriophages, such as lambda, P2, and T-numbered phages, have contributed to advances in genetics and molecular biology as model organisms. In principle, phages have been classified based on their morphology and are sometimes further subgrouped into families based on sequence homologies to prototype phages, such as lambda-like and P2-like phage families. Based on the recent advances in sequencing technologies, the number of sequenced phage genomes has been exponentially increasing, and a huge number of prophages integrated in bacterial genomes have also been identified. The data from these sequencing efforts have unveiled a surprisingly high level of genetic diversity among bacteriophages. Studies have also shown that bacteriophages, particularly those that can be integrated into the host chromosomes (called temperate phages), play more critical roles than was previously recognized in the evolution of a wide range of bacteria by carrying various genes in host cells<sup>1-3</sup>. In addition, this expanding sequence information about phages/prophages has identified a huge number of phage-encoded genes with unknown functions<sup>4,5</sup>. However, limited efforts have been devoted to functional analyses of such newly identified genes, leaving their functions uncharacterized even in phages of medical or industrial importance. Among such cases are Shiga toxin 2 (Stx2)-transducing phages (Stx2 phage) of enterohemorrhagic *Escherichia coli* (EHEC).

<sup>1</sup>Division of Microbiology, Department of Infectious Diseases, Faculty of Medicine, University of Miyazaki, 5200 Kihara, Kiyotake, Miyazaki 889-1692, Japan. <sup>2</sup>Genetic Engineering and Biotechnology Department, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh. <sup>3</sup>Biochemistry and Molecular Biology Department, University of Dhaka, Dhaka 1000, Bangladesh. <sup>4</sup>Department of Anatomy, Ultrastructural Cell Biology, Faculty of Medicine, University of Miyazaki, 5200 Kihara, Kiyotake, Miyazaki 889-1692, Japan. <sup>5</sup>Biosafety & BSL3 Laboratory, ICDDR,B, Mahakhali, Dhaka 1212, Bangladesh. <sup>6</sup>Department of Microbiology, Graduate School of Medical and Dental Sciences, Kagoshima University, 8-35-1 Sakuragaoka, Kagoshima 890-8544, Japan. <sup>7</sup>Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan. <sup>8</sup>Institute of Low Temperature Science, Hokkaido University, Kita 19, Nishi 8, Kita-ku, Sapporo 060-0819, Japan. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.H. (email: thayash@bact.med.kyushu-u.ac.jp)

EHEC strains cause a range of diseases, from mild diarrhea to severe, life-threatening infections, such as hemorrhagic colitis and hemolytic uremic syndrome<sup>6,7</sup>. While multiple virulence factors are required for EHEC pathogenesis, the key virulence factors involved in the onset of severe diseases are Stxs<sup>6,8</sup>, highly potent cytotoxins that inhibit protein synthesis in mammalian host cells by depurinating a specific adenine residue of the 28S rRNA<sup>9</sup>. Stxs are classified into two subtypes: Stx1 and Stx2. There are several variants in each subtype, all of which are encoded by bacteriophages<sup>10–12</sup>. The Stx1 gene in *Shigella dysenteriae* is also thought to be encoded by a prophage remnant<sup>13</sup>. Although EHEC strains produce one or more such Stxs, it is thought that Stx2-producing strains cause more severe infections than Stx1-producing strains<sup>14,15</sup>. The EHEC strains of the O157 serotype, the most prevalent and widely studied EHEC serogroup, produce Stx1, Stx2a, and Stx2c, alone or in combination<sup>16,17</sup>. While a high level of Stx2a is produced upon phage induction, the level of Stx2c produced by EHEC O157 is generally very low, even after phage induction, and the expression of Stx1 production is iron-regulated and not linked to phage induction<sup>18–22</sup>.

Stx phages show remarkable variations in genome size, genome composition, morphology, and host specificity<sup>23–27</sup> but are generally regarded as lambda-like phages because they contain many genes homologous to those of the archetypal bacteriophage lambda<sup>11,28–31</sup>. Among the sequenced Stx phages of O157 strains<sup>11,18,26,28–30</sup>, the Stx1 and Stx2c phages are highly similar to lambda phage and share many early and late genes with lambda phage. Consistent with this conservation, they have morphologies similar to that of lambda phage, which is constituted by a hexagonal head and a long flexible and non-contractile tail, a feature of the *Siphoviridae* family. However, Stx2a phages of O157 strains, such as Sp5 of strain Sakai<sup>32</sup>, 933W of strain EDL933<sup>28</sup>, E86654-Stx2 of E86654<sup>33</sup>, have short tails and are thus morphologically classified into the *Podoviridae* family. In fact, the Stx2a phages encode many genes that are absent in phage lambda. In particular, most genes in the late region, which are required for phage morphogenesis, are distinct not only from those in phage lambda and from those in other well-studied phages belonging to the *Podoviridae* family. Thus, the functions of these genes are largely unknown, except for a gene located just downstream of the *stx2* gene (z1466 in phage 933W, corresponding to ECs1207 of O157 strain Sakai), which was recently found to encode an esterase<sup>34</sup>. However, due to the difficulties in accurately enumerating plaque numbers and in obtaining high-titer phage lysates for Stx2a phages, conventional biochemical and genetic approaches to analyze the functions of each gene are not easily applicable. Although the former problem was partially solved by recent development of a sensitive and simple plaque formation method<sup>35</sup>, it is still difficult to perform biochemical analyses using purified phage particles, such as protein composition determination, because Stx2a phages yield only a low titer of phage lysate (up to 10<sup>6</sup> PFU/ml). Therefore, a different approach needs to be employed to analyze the functions of Stx2a phage genes.

Here, we describe the results of a systematic genetic and morphological analysis of the genes of unknown function encoded by a representative Stx2a phage, Sp5 of O157:H7 strain Sakai<sup>32,36</sup>. We constructed a nearly complete set of in-frame deletion mutants of all Sp5 genes with unknown functions, systematically analyzed the excision, circularization, plaque formation, and lysogenization capacities of each mutant, and newly identified nine genes essential for infectious Sp5 phage particle formation. The results of electron microscopic examinations, including observation using an immunogold labeling technique, are also presented to assign the proteins' functions in Sp5 morphogenesis.

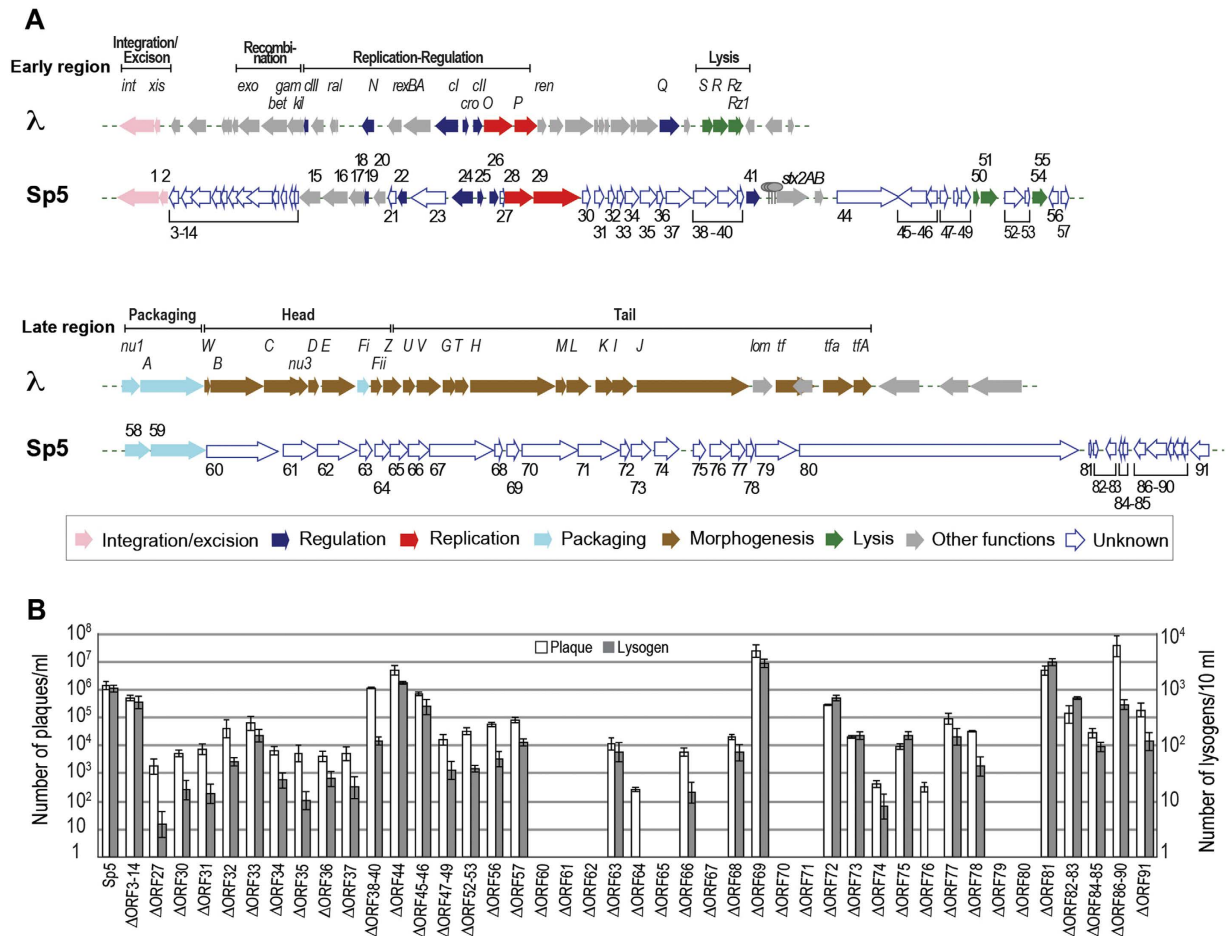
## Results and Discussion

**Construction of a set of Sp5 mutants to determine the functions of unknown genes.** The Sp5 genome is 62,708 bp long<sup>36</sup>. Among 91 protein-coding genes (open reading frames – ORFs) identified on the Sp5 genome (ECs1160 to ECs1251 in the O157 Sakai genome annotation; for convenience, these genes are referred to as ORF1 to ORF91 in this article; note that five small ORFs annotated in the original report of the genome sequence for O157 Sakai were excluded from the list of ORFs and four previously not-annotated ORFs were included; see Supplementary Table S1 for more details), the functions of 24 genes were reliably predicted based on high sequence homologies to functionally characterized genes from other phages, especially those from lambda phage<sup>37</sup>. These ORFs included genes for integration, replication, and gene regulation, and those encoding two terminase subunits (Fig. 1A and Supplementary Table S1). However, the remaining 67 ORFs lacked reliable database matches, so their functions were unpredictable. In particular, most genes in the late region were unique, and we were unable to confidently predict their functions<sup>32</sup>.

To characterize these genes with unknown functions, we constructed marker-free deletion mutants for these genes on the K-12 MG1655 genetic background. The parent phage used for mutant construction was an Sp5 derivative that was transferred to K-12 MG1655 from O157 Sakai, the *stx2* gene of which was replaced with the chloramphenicol resistance (*Cp<sup>r</sup>*) gene<sup>32</sup>. We used this *Cp<sup>r</sup>*-marked phage as the wild-type (WT) Sp5 throughout this study. As all deletions introduced were in-frame and marker-free deletions, polar effects were mostly avoidable. Several genes were deleted *en bloc*. Although we could not obtain deletion mutants of two ORFs in the early region, ORF21 (ECs1180) and ORF23 (ECs1184), we finally constructed 43 mutants and subjected them to the subsequent functional analyses (Fig. 1A).

The exact reason(s) for our failure to delete ORF21 and ORF23 are currently unknown, but the deletion of these genes appears to be lethal to the host organism. They are most likely required for the stable maintenance of the lysogenic state of Sp5. In the database search, both ORFs showed high homologies to many hypothetical proteins of prophages from EHEC and other bacteria. No conserved domains were found in these two ORFs.

**Excision and circularization of mutant phage genomes.** We first determined whether each mutant could be excised from the host chromosome and could form an extrachromosomal circular intermediate. For this purpose, K-12 MG1655 lysogens for each mutant were treated with mitomycin C (MMC), and their total cellular DNA was isolated. Then, circularized phage genomes were detected by PCR using a pair of primers that specifically amplified the attachment site (*attP*)-containing region of Sp5, which is created upon prophage



**Figure 1. Genomic Structures of Sp5 and the plaque formation and lysogenization capacities of mutant Sp5 phages.**

(A) The genomic organizations of lambda phage and Sp5, the Stx2 phage of EHEC O157:H7 strain Sakai. Early and late regions are shown separately. Genes of known or reliably predicted functions<sup>32,56</sup> are colored according to the functional category shown in the inset. The open arrows in Sp5 are the genes of unknown function that were analyzed in this study. ORFs deleted *en bloc* are indicated. (B) The plaque formation and lysogenization capacities of the WT and mutant Sp5 phages are shown. Error bars represent the standard deviations of three independent experiments.

excision (Supplementary Table S2). In this analysis, we detected a single PCR product of an expected size for all K-12 lysogens, indicating that all mutant phages could be excised from the host chromosome to form a circular intermediate (Supplementary Fig. S1).

**Plaque formation and lysogenization of mutant phages.** To identify the essential genes of Sp5, plaque formation and lysogenization assays were performed for each of the mutant phages. For the plaque formation assay, phage lysates were prepared by MMC treatment of each lysogen, and the plaque-forming units in each lysate were determined using a modified double agar overlay method that we previously developed to easily and accurately enumerate plaques of Stx2a phages<sup>35</sup>. This analysis revealed that nine out of the 43 mutants examined formed no plaques (Fig. 1B).

For the lysogenization assay, phage lysates from each mutant Sp5 phage lysogen were incubated with K-12 MG1655 cells, and the efficiency of lysogenization was determined by counting  $Cp^f$  colonies. This assay revealed that all nine mutants that were deficient in plaque formation were also unable to form lysogens (Fig. 1B), confirming their inability to infect K-12 cells. We could not detect any lysogens from two additional mutants,  $\Delta$ ORF64 and  $\Delta$ ORF76, under standard conditions. However, several  $Cp^f$  colonies were obtained when concentrated lysates were used. Therefore, these two mutants retained the capacity to infect and propagate in K12, but their capacity to lysogenize was significantly reduced. The biological significance of this somewhat unexpected behavior of  $\Delta$ ORF64 and  $\Delta$ ORF76 is currently unknown, but these two genes may be required for the efficient infection and/or lysogenization of Sp5. In contrast, three mutants,  $\Delta$ ORF69,  $\Delta$ ORF81, and  $\Delta$ ORF86–90 produced several times more plaques and lysogens than the WT (only a higher plaque formation ability was detected for  $\Delta$ ORF86–90), suggesting that ORF69, ORF81, and some ORFs in the ORF86–90 gene cluster may have some regulatory functions.

ORFs (ECs)	Excision	Lysogenization	Plaque formation	Complementation <sup>a</sup>	Packaging	Superfamily/Conserved domains
ORF21 (ECs1180)	NA	NA	NA	NA	NA	No
ORF23 (ECs1184)	NA	NA	NA	NA	NA	No
ORF60 (ECs1221)	+	–	–	+	–	No
ORF61 (ECs1222)	+	–	–	+	–	No
ORF62 (ECs1223)	+	–	–	+	–	P22 coat protein superfamily
ORF65 (ECs1226)	+	–	–	+	–	No
ORF67 (ECs1228)	+	–	–	+	+	phage tail fiber C-terminus domain, phage tail fiber repeat 2, multiple collagen like domains
ORF70 (ECs1232)	+	–	–	+	+	No
ORF71 (ECs1233)	+	–	–	+	+	No
ORF79 (ECs1241)	+	–	–	+	+	No
ORF80 (ECs1242)	+	–	–	+	+	predicted ATPase, AAA <sup>+</sup> superfamily

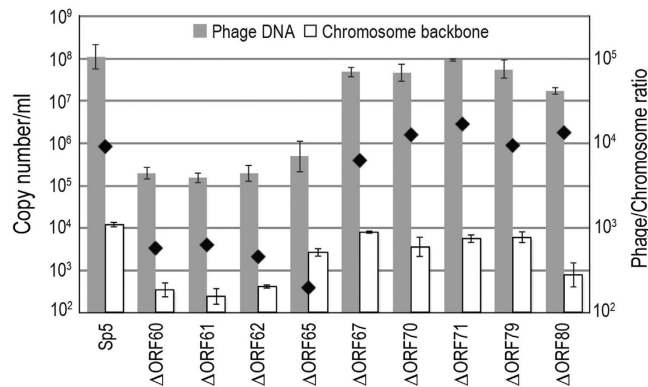
**Table 1. The essential genes identified in this study, the phenotypes of their deletion mutants, and the results of bioinformatics analysis of each gene.** <sup>a</sup>Complementation was examined in lysogenization assays. NA, not applicable; these two ORFs are not deletable.

These results indicate that nine ORFs (ORFs60, 61, 62, 65, 67, 70, 71, 79, and 80; all are located in the late region) are essential for Sp5 infectivity (Fig. 1 and Table 1). Therefore, our following analysis focused on these nine genes.

**Complementation analysis of the nine newly identified essential genes.** To verify that the products of the nine genes are required for Sp5 infection and propagation, we examined whether the deficiencies in lysogenization of the nine mutants were each able to be complemented *in trans* with plasmid-expressed gene products. In this analysis, we employed lysogenization assays because, if relevant gene products are provided during the phage induction process in K-12 Sp5 lysogens, phage particles which package the Sp5 genome, inject them into recipient cells, and generate Sp5 lysogens, could be produced. In this complementation analysis, lysogenization assays should be more sensitive than plaque formation assays where each gene product needs to be efficiently supplied not only in the phage induction process but also in recipient cells (during the plaque formation process). In fact, when each mutant phage was induced in K-12 cells that contained a plasmid encoding a relevant cloned gene, all phage lysates yielded Sp5 lysogens in K-12 (Table 1). These results indicate that the nine genes encode proteins essential for producing infectious Sp5 phage particles and also that the constructed deletion mutants do not carry any other fortuitous mutations.

**Identification of genes required for genome packaging.** The process of phage particle formation for double-stranded (ds) DNA phages begins with head formation and packaging genomic DNA therein. To determine whether the nine mutant phages can package their genomes, we quantified their packaged (and thus resistant to DNase treatment) genomic DNA using quantitative PCR (qPCR), as described in the Methods. After MMC treatment, WT Sp5 yielded approximately 10<sup>8</sup> copies of DNase-resistant genomes per one-milliliter culture (Fig. 2). Compared to the WT phages, remarkably lower copy numbers of genomic DNA were detected in the phage lysates from four mutants ( $\Delta$ ORF60,  $\Delta$ ORF61,  $\Delta$ ORF62, and  $\Delta$ ORF65), whereas those from five mutants ( $\Delta$ ORF67,  $\Delta$ ORF70,  $\Delta$ ORF71,  $\Delta$ ORF79, and  $\Delta$ ORF80) were compatible with those from the WT phages. In this assay, a low amount of chromosomal DNA was detected in all phage lysates examined (open bars in Fig. 2), likely due to the incomplete digestion of unpackaged DNA, the non-specific packaging/incorporation of chromosomal DNA into phage particles or membrane vesicles. Therefore, the phage DNA detected in the DNase-treated lysates from the former four mutants are also likely generated by similar mechanisms. In fact, the ratios of phage DNA to chromosomal DNA in the lysates from the four mutants were 100–1,000 times lower than that from the WT phage, while those from the latter five mutants were in the same range as the WT (indicated by diamonds in Fig. 2).

These results indicate that the mutants of ORF60, ORF61, ORF62, and ORF65 are deficient in genomic DNA packaging; therefore, these genes are required for the head formation and/or genome packaging processes. The results also indicate that head formation and genome packaging proceed properly in the mutants of ORF67, ORF70, ORF71, ORF79, and ORF80; thus, these five genes are required for tail formation (Table 1).



**Figure 2. Quantification of packaged phage DNA in WT and mutant Sp5 phages.** Packaged (and thus DNase-resistant) phage DNA was quantified via qPCR. For normalization, DNA preparations corresponding to those obtained from 1 ml of phage lysate were used as the template in all cases. The chromosomal DNA from a chromosome backbone region (CB) was quantified as a control. The average copy numbers of each genome are shown. Error bars represent the standard deviations of three independent experiments. The ratios of phage DNA and chromosomal DNA are indicated with black diamonds.

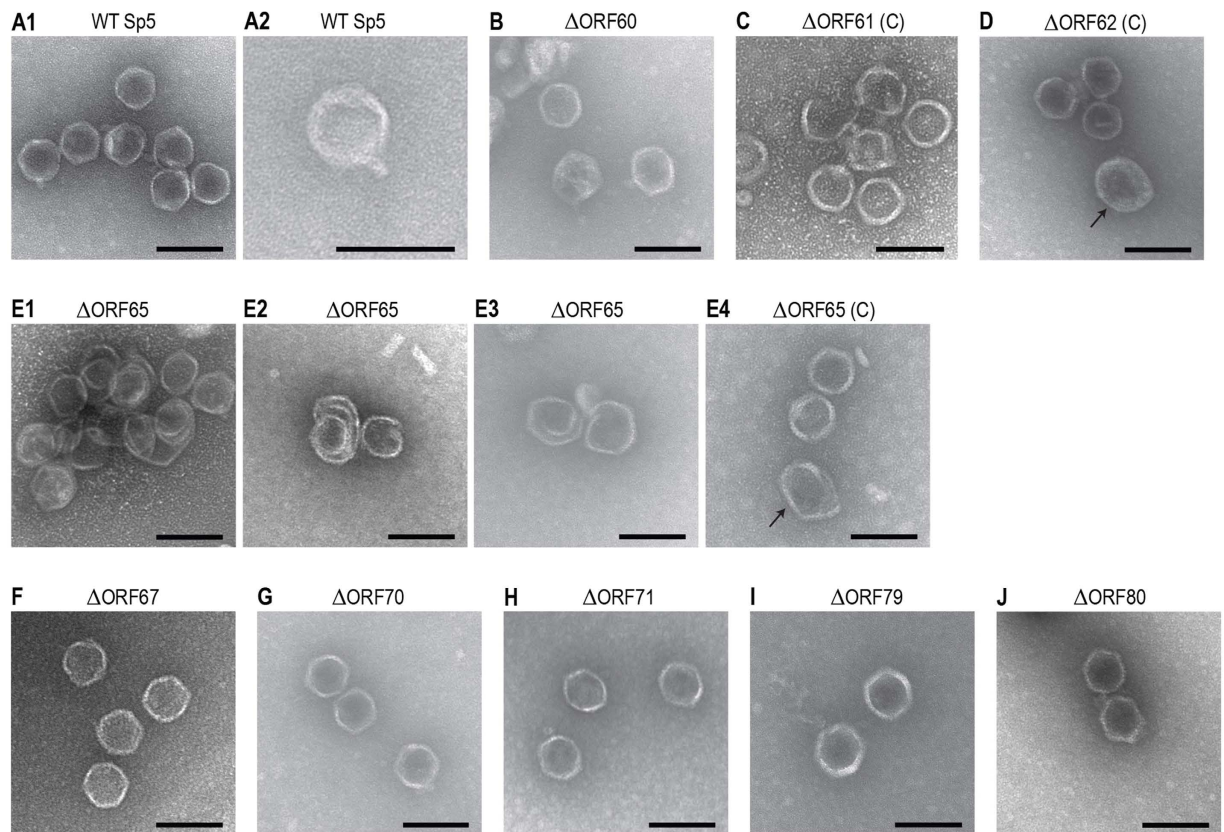
**Bioinformatics and morphological analyses of the nine essential genes.** The first step in the head formation of dsDNA phages is the assembly of an intermediate called the prohead or procapsid. The prohead is a DNA-free head structure into which phage DNA is subsequently packaged. The prohead mainly consists of the portal protein, the major head protein, and the scaffolding protein, and its assembly process is well conserved among dsDNA phages and starts with the copolymerization of scaffolding proteins and major capsid proteins<sup>38</sup>, while, in HK97 and T5, the N-terminal domain of their major head proteins may perform a scaffolding function<sup>39,40</sup>. Coupled with genomic DNA packaging through the portal protein by the action of the terminase complex, scaffolding proteins are degraded by prohead proteases (e.g.,  $\lambda$  and T4) or leave the procapsid without being degraded (e.g., P22 and  $\varphi$ 29), providing the space for genomic DNA<sup>38,41</sup>. After DNA packaging, the terminase complex leaves the portal vertex, and head completion protein(s) attach there instead. Although the tail and head are formed separately and are finally joined together in phages of the *Siphoviridae* and *Myoviridae* families<sup>38,41</sup>, the tails of the *Podoviridae* family phages are formed on the portal vertex<sup>42</sup>.

To explore the functions/roles of the nine newly identified essential genes of Sp5 in its morphogenesis, we re-examined the conserved domains in each gene product and their sequence homologies to known proteins. Then, we performed electron microscopic (EM) analyses of the mutants of each gene and those complemented with plasmid-cloned genes (Fig. 3). In addition, we conducted immune-EM examinations via the immuno-gold labeling technique using antibodies specific to three gene products (ORF61, ORF62, and ORF67) (Fig. 4). We generated recombinant genes fused to the glutathione S-transferase (GST) gene for each of the nine genes to obtain purified recombinant proteins that could be used for specific antibody production. However, we were only able to obtain specific antibodies against three ORFs due to the high insolubility of other recombinant proteins. In this section, we first describe the characterization of ORF62 and ORF67 and then that of the other ORFs because ORF62 and ORF67 were confidently identified as the major capsid protein and the tail fiber protein, respectively, and their antibodies can be useful probes to detect head (or head-like) and tail structures.

**ORF62 (ECs1223):** A conserved domain search indicated that ORF62 belongs to the P22 coat protein superfamily and contains the DUF4043 and capsid\_maj\_N4 domains. However, ORF62 exhibited very low or no significant homology to the P22 coat (Gp5) and N4 (Gp56) proteins, while it showed high sequence similarities (up to 85% sequence identity) to hypothetical or “putative virion structural” proteins of many uncharacterized phages. In the EM analysis of  $\Delta$ ORF62, although a large number of outer membrane vesicle (OMV)-like particles were observed, no phage-head like particles were detected (data not shown). By complementation, heads with regular hexagonal outlines and of approximately 64 nm in diameter, which were similar to the WT (Fig. 3A), were observed (Fig. 3D), indicating that normal head formation was restored. The anti-ORF62 antibody was bound to the whole surface of the WT Sp5 head (Fig. 4A), but this binding was abolished in the  $\Delta$ ORF62 mutant (data not shown). These findings indicate that ORF62 is most likely the major capsid protein of Sp5.

**ORF67 (ECs1228):** Although a close homologue of ORF67 in phage 933W (L0121) was annotated as “tail-fiber-like”<sup>28</sup>, ORF67 exhibited no significant homologies to characterized phage proteins. However, it contains the phage tail fiber C-terminus domain and the phage tail fiber repeat 2 (Table 1). Multiple collagen-like domains with the repeating collagen signature sequence (Gly-X-Y)<sub>n</sub> are also present in ORF67, as in the tail fiber proteins from several phages<sup>43</sup>. In the EM examination of  $\Delta$ ORF67, apparently intact phage head structures were observed (Figs 3F and 4D). In the immuno-EM analysis of the WT phage using the anti-ORF67 antibody, multiple gold particles were bound to the tail ends of the phage particles (Fig. 4E). The binding of the anti-ORF67 antibody was not observed for the  $\Delta$ ORF67 mutant (data not shown). These observations indicate that ORF67 most likely encodes the tail fiber protein of Sp5, which probably recognizes the YaeT (BamA) protein, a proposed *E. coli* surface receptor for Sp5<sup>35</sup> and other short-tailed Stx2 phages<sup>44</sup>.

**ORF60 (ECs1221):** No conserved domains were found in ORF60. Although it showed remarkable sequence similarities (up to 79% identity) to proteins of many uncharacterized bacteriophages from various Gram-negative

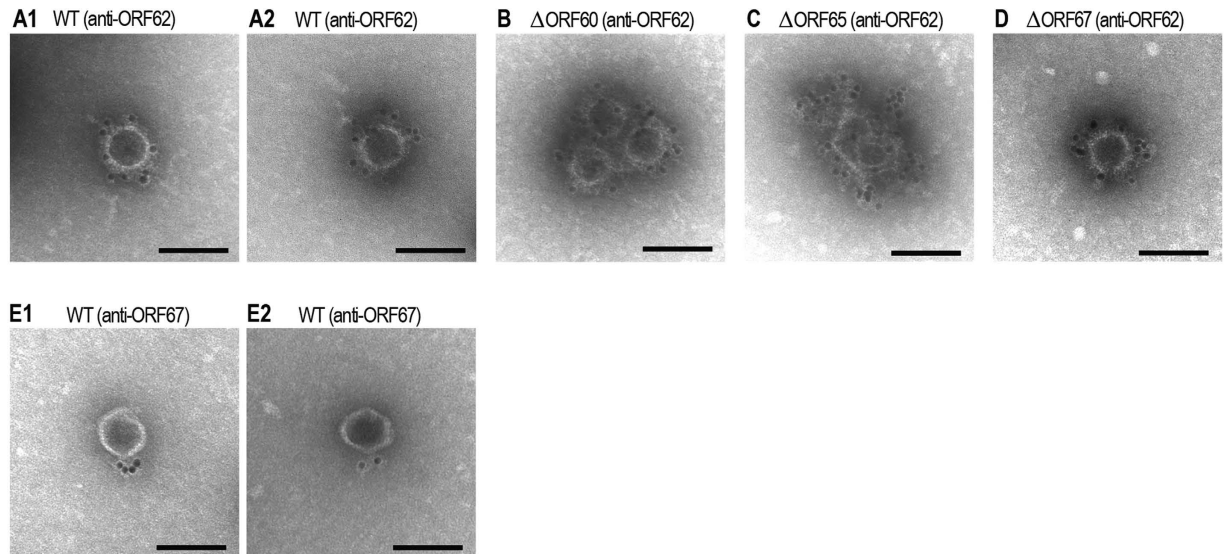


**Figure 3. Electron microscopic analysis of the negative stained WT and mutant Sp5 phages.** (A) WT Sp5. Short tail-like structures were occasionally observed, as shown in panel A2. (B–J) Sp5 mutants. Mutants complemented with the plasmid-encoded cloned genes (indicated by (C) added after mutant names) are also shown when clear morphological differences were observed compared to the non-complemented mutants. Refer to the text for detailed descriptions of the morphologies of each mutant. An arrow in panel D indicates an outer membrane vesicle-like particle. Some amounts of aberrant phage heads were observed in every sample from complemented head mutants as indicated by an arrow in panel E4. Electron micrographs were taken at magnifications of 60 K using a transmission electron microscope and were negatively contrasted with 2% uranyl acetate dihydrate. Bar, 100 nm.

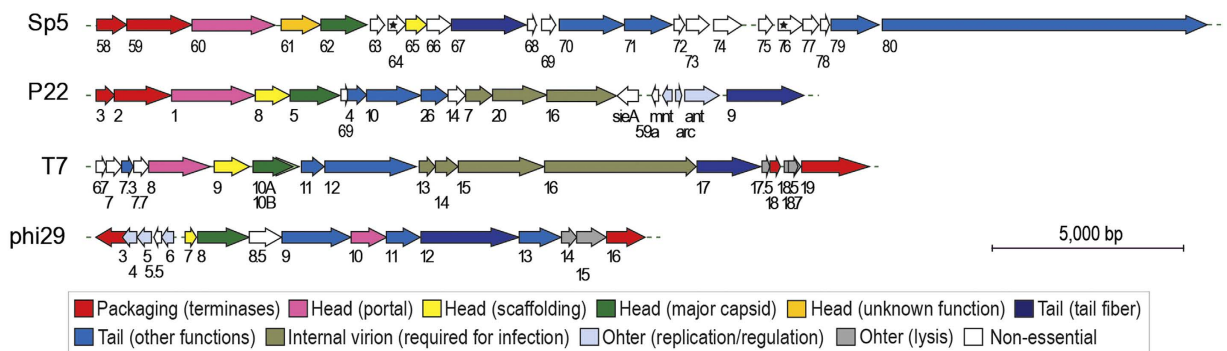
bacteria, all were annotated as hypothetical or “putative portal” proteins. In the EM analysis of  $\Delta$ ORF60, in addition to many OMV-like particles, a significant number of apparently normal head-like structures were observed (Fig. 3B). Although clear increase in the proportion of normal head-like structures was not observed by complementation (data not shown), the anti-ORF62 antibody was bound to the surfaces of these head-like structures (Fig. 4B). This finding suggests that the  $\Delta$ ORF60 mutant can produce a head-like structure that is difficult to distinguish from the normal head of the WT. As no binding of the anti-ORF67 antibody was detected for the  $\Delta$ ORF60 mutant (data not shown), tails were likely not formed on these head-like structures. As mentioned above, this mutant is defective in genomic DNA packaging (Fig. 2). Several phages, such as the lambda, P22, and phi29 phages, still produce apparently normal head-like structures (prohead or procapsid) in the absence of portal protein<sup>41,45–47</sup>, so the biological and morphological features of  $\Delta$ ORF60 indicate that ORF60 most likely encodes a portal protein of Sp5. This conclusion is further supported by the genomic location of ORF60; it is located just downstream of the terminase genes, like in many phages (Fig. 5).

ORF61 (ECs1222): No significant homology to known characterized proteins and no conserved domains were found for ORF61. In the EM analysis of  $\Delta$ ORF61, no phage-head-like particles were detected (data not shown), and normal head structures were formed by complementation (Fig. 3C). These results indicate that ORF61 is required for phage head formation. However, the anti-ORF61 antibody did not bind to either WT phage particles or any of the mutants constructed in this study (data not shown). Thus, we cannot assign a precise role to ORF61 in head formation.

ORF65 (ECs1226): Bioinformatics analysis provided no information to aid in the functional assignment of ORF65. The EM analysis showed that the  $\Delta$ ORF65 mutant produced a large number of highly deformed particles, which often showed spiral structures (Figs. 3E1–3E3). A significant number of normal head-like structures were produced by complementation (Fig. 3E4). The anti-ORF62 antibody was bound almost evenly to the spiral structures (Fig. 4C), indicating that these structures are made of ORF62, the major capsid protein. Such open spiral structures were observed for the scaffolding protein-less mutants of phages lambda, P22, and phi29<sup>48–51</sup>. Binding



**Figure 4. Immunoelectron microscopic analysis of the WT and mutant Sp5 phages.** Immunoelectron microscopic analyses were performed using the anti-ORF62 (A–D) or anti-ORF67 antibodies (E). WT and mutant Sp5 phages were initially treated with primary antibodies and then treated with secondary antibody-conjugated gold particles. Bar, 100 nm.



**Figure 5. Comparison of the Sp5 late region with three well-characterized Podoviridae family phages.** The genes (or ORFs) of each phage are numbered according to Asadulghani *et al.*<sup>32</sup> for Sp5, Pedulla *et al.*<sup>4</sup> for P22, Dunn & Studier<sup>68</sup> for T7, and accession number NC\_008720 for phi29. Essential genes are colored, and open arrows indicate non-essential genes. Genes with the same or similar functions are depicted in the same color, as indicated in the inset. Head completion/neck proteins are included in the “tail formation” category. Note that the ORF64 and ORF76 mutants (marked by asterisks) showed markedly reduced infectivity but were still infectious (see Fig. 1 and the main text). In T7, gp10B (light green in the figure) encodes a minor capsid protein.

of the anti-ORF67 antibody to the  $\Delta$ ORF65 mutant was not detected (data not shown). These findings indicate that ORF65 is required for head formation and may play a scaffolding function.

ORF70 (ECs1232), ORF71 (ECs1233), and ORF79 (ECs1241): No information to aid in the functional assignments of these three ORFs was obtained from the bioinformatics analyses. ORF71 contains DUF1983 and COG4733 domains. Although both domains have been found in various bacteriophage tail component proteins, including the host specificity protein (gpJ) of lambda phage<sup>52</sup>, their functions are unknown. In the EM examination of their mutants, head structures indistinguishable from that of the WT were observed (Fig. 3G–I). No binding of the anti-ORF67 antibody was detected for any of the three mutants. These results and their intact capacities to package phage DNA (Fig. 2) indicate that these ORFs participate in the Sp5 tail assembly process although their exact roles/functions are unknown.

ORF80 (ECs1242): This ORF encodes a very large protein consisting of 2,793 amino acid residues. ORF80 also showed no significant homologies to known characterized proteins, but contains the COG1483 (predicted ATPase, AAA+ superfamily) domain. The  $\Delta$ ORF80 mutant produced head structures indistinguishable from that of the WT phage (Fig. 3J), and binding of the anti-ORF67 antibody was not detected (data not shown). Thus, ORF80 participates in the Sp5 tail assembly process. Like ORF70, ORF71, and ORF79, the exact role/function of this extremely large ORF is unknown. However, the presence of the COG1483 domain suggests that ORF80 may be involved in some energy-requiring step in tail formation.

Strains, phage, and plasmids	Features	References or sources
Bacterial strains:		
<i>E. coli</i> MG1655	F <sup>-</sup> , lambda <sup>-</sup> , rph-1	69
<i>E. coli</i> BL21 (DE3)	F <sup>-</sup> , ompT, hsdS (r <sup>B-</sup> , m <sup>B-</sup> ), gal	64,65
<i>E. coli</i> Sm10λpir	Permissive strain for replication of pABB-CSR2	61
Phage		
Sp5	Δstx2::cat-cassette	32
Plasmids:		
pKD4	FRT-flanked Tn5 neomycin phospho-transferase, Ap <sup>r</sup> , Km <sup>r</sup>	70
pKD46	Lambda red helper plasmid, Ap <sup>r</sup>	70
pDONR201	Gateway cloning vector	Invitrogen
pABB-CRS2	Suicide vector	60
pDONR-ORF#	pDONR201-derivatives containing a deletion target ORF (# indicates the ORF number)	This study

**Table 2. Bacterial strains, phage, plasmids used in this study.**

To briefly summarize this series of analyses, we identified the genes which most likely encode the portal (ORF60), major capsid (ORF62), scaffolding (ORF65), and tail fiber (ORF67) proteins of Sp5. Although the exact roles/functions of the other five essential genes are unknown, ORF61 is involved in head formation, and ORF70, ORF71, ORF79, and ORF80 are required for tail formation. The apparent absence of a protease indicates that, during the head formation process, the scaffolding protein of Sp5 leave the proheads without degradation, as has been shown for phages P22 and φ29<sup>53,54</sup>. The Sp5 gene product that corresponds to head completion proteins was not identified in this study.

**Organization of the Sp5 genes essential for its morphogenesis and comparison with other well-characterized *Podoviridae* family phages.** Through a systematic analysis of Sp5 genes of unknown function, we identified nine essential genes required for its morphogenesis. Together with ORF58 and ORF59, which, respectively encode the small and large terminase subunits, these genes determine the essential processes of Sp5 morphogenesis. The number of essential late genes (11 ORFs) in Sp5 are roughly comparable to those in other well-characterized phages of the *Podoviridae* family: P22, T7, and phi29 (Fig. 5), but much smaller than the members in the *Siphoviridae* and *Myoviridae* families (21 and 18 genes in lambda and T5 belonging to *Siphoviridae*, respectively; 24 and 49 genes in P2 and T4 belonging to *Myoviridae*, respectively). However, the Sp5 late region is markedly longer than those of other phages. This difference is mainly due to the presence of an extremely large ORF (ORF80) but is also partly due to a significantly larger number of non-essential genes that are dispersed in the late region. Among these genes, ORF68 encodes a small polypeptide of 89 amino acids that comprise an apparent phage fiber C-terminus domain<sup>55,56</sup>. Although we found no sequences that can potentially induce a programmed frame shift to produce a tail fiber with an altered C-terminal sequence by fusing ORF67 and ORF68, ORF68 may represent an intragenomic reservoir sequence to create a variation in the tail fiber protein. The functions of ORF64 and ORF76 are also interesting. Both genes may be required for the efficient infection and/or lysogenization of Sp5 because their deletion mutants showed significantly reduced infection and lysogenization capacities (Fig. 1).

## Conclusion

Through systematic genetic and morphological analyses of the Sp5 genes of unknown function, we identified nine genes essential for its morphogenesis and two early genes that are likely required for the stable maintenance of the lysogenic state of Sp5. Among the nine essential late genes, we were able to assign the most likely functions to four genes, which encoded the portal, major capsid, scaffolding, and tail fiber proteins. The exact roles/functions of the other five genes are unknown, but one is required for head formation and the other four genes are required for tail formation. One of the four genes encodes an unusually large protein of 2,793 amino acid residues. These nine genes, together with two genes encoding the large and small subunits of terminase, determine the essential processes for Sp5 morphogenesis. As many Stx2 phages from various EHEC strains (O157 and non-O157 EHECs) contain the late regions same or very similar to that of Sp5 and share the 11 essential genes and most non-essential late genes, the present study is an important first step in understanding the biology of this unique and medically important, but not yet well-characterized, group of bacteriophages.

## Methods

**Bacterial strains, culture media and bacteriophages.** The bacterial strains, bacteriophages, and plasmids used are listed in Table 2. Bacteria were grown under aerobic conditions at 37 °C in Lysogeny broth (LB) or on LB agar. When necessary, culture media or plates were supplemented with 40 μg/ml chloramphenicol (Cp),



100 µg/ml ampicillin (Ap), 25 µg/ml kanamycin (Km) or 100 µg/ml spectinomycin (Spec). *E. coli* K-12 MG1655 was used as the host for the Sp5 phage. A derivative of Sp5, in which *stx2* was replaced by the *cat* gene<sup>32</sup>, was used as the wild-type phage, and this derivative (Sp5*stx2::cat*) is referred to as the WT Sp5 in this article.

**Bioinformatics analysis of ORFs.** Homology searches were performed using BLASTP through the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/>). The conserved domains and motifs were searched using Conserved Domain Database (CDD)<sup>57</sup>, InterProScan<sup>58</sup>, and SMART<sup>59</sup>.

**Mutant construction.** To avoid the polar effects that may be caused by deletion and/or marker insertion, marker-free in-frame deletion mutants of Sp5 were constructed under the K-12 MG1655 background as described previously<sup>60</sup>, with some modifications. Flanking regions with nine nucleotides from each end of the target ORFs were amplified using the primers listed in Supplementary Table S2 and the genomic DNA of K-12 Sp5 lysogen as a template to obtain two amplicons. These two amplicons were combined together and used as the template to generate an in-frame deleted PCR product of the desired ORF by overlap extension PCR (OE-PCR) using adapter-PCR primers (Supplementary Table S2). The resulting PCR product was purified by gel extraction and cloned into pDONR201 (Gateway cloning system, Invitrogen) by a BP-reaction to obtain pDONR201-ΔORF#. The recombinant product was introduced into DH5α by transformation. Clones were selected on Km plates along with negative selection for the *ccdB* death cassette, which remains in pDONR201 if the expected recombination does not occur. In-frame deletion in each clone was confirmed by PCR using M13 primers and capillary sequencing. The ORF containing an in-frame deletion was cloned into the *sacB*-containing positive suicide vector pABB-CRS2<sup>61</sup> by the LR recombination reaction (Gateway cloning system, Invitrogen) to obtain pABB-CSR2-ΔORF#. This construct was transformed into *E. coli* SM10 λpir and selected on Ap plates. The plasmid was transconjugated into an Sp5 lysogen of K-12 MG1655, and its resistance to Ap and Cp was confirmed on Ap and Cp plates. Integration of pABB-CSR2-ΔORF# by a single crossover was confirmed by PCR using primers listed in Supplementary Table S2. Confirmed transconjugants were grown without selection until the late logarithmic phase in LB at 37 °C to allow the excision of the *sacB*-containing vector and were then spread on LB agar plates containing 5% sucrose. Finally, the in-frame deletion in each sucrose-resistant and Ap-sensitive colony was confirmed by PCR using respective ORF primers (Supplementary Table S2) and capillary sequencing.

**Preparation of phage lysates.** Phage lysates were prepared as described previously<sup>32</sup>, with a few modifications. Briefly, K-12 Sp5 lysogens were grown overnight at 37 °C in Cp-containing LB. The cell concentration was adjusted to an optical density at 600 nm (OD<sub>600</sub>) of 0.3–0.4 with LB. Diluted cultures were grown for 50–55 min at 37 °C, and MMC was added to a final concentration of 1.0 µg/ml. After 6–7 h of incubation, the culture was treated with a few drops of chloroform and centrifuged at 14,000xg for 50 min at 4 °C. The supernatant was filter-sterilized with Minisart 0.22-µm pore size filters (Sartorius) and incubated with 20% polyethylene glycol 8000/10% NaCl for phage precipitation. Precipitated phage particles were pelleted by centrifugation at 24,000xg for 1 h at 4 °C and were resuspended in SM buffer (0.58% NaCl, 0.2% MgSO<sub>4</sub>·7 H<sub>2</sub>O, 50 mM Tris-HCl (pH 7.5), and 0.01% gelatin) in a 1/100 volume of the original supernatant. The resuspended phage lysate was mixed with a 0.1 volume of chloroform and briefly centrifuged. The supernatant was collected and incubated at 37 °C for 20 min to remove the residual chloroform and was used for analyses as “phage lysate”.

**Plaque formation and lysogenization assays.** Plaque formation and lysogenization assays were performed as described previously<sup>35</sup>. In the lysogenization assay, several colonies were randomly selected, and the lysogenization of Sp5 in these colonies was confirmed by colony PCR using primers Sp5\_LysF and Sp5\_LysR (Supplementary Table S2).

**Complementation assay.** Complementation experiments were performed using pTB101, an expression vector containing the *tac* promoter and *lacI*<sup>62</sup>. The entire coding region with the presumed Shine-Dalgarno (SD) sequence of the target ORF was PCR-amplified with a pair of primers, each containing EcoRI and SalI restriction sites (Supplementary Table S2). The amplicons were gel-purified, digested with the two enzymes and ligated into the EcoRI/SalI-double digested pTB101. After confirming the sequence of the cloned ORF by PCR and sequencing, the construct was transformed into K-12 containing the respective Sp5 mutants. The transformed cells were grown at 37 °C in LB containing Cp and Spec. The cloned ORF was induced with isopropyl-beta-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.025 mM 30 min after the MMC induction. After 6–7 h of incubation at 37 °C, the phage lysate was prepared and used for lysogenization assay and for EM analysis.

**Detection of excised phage genomes and quantification of DNase-resistant phage DNA in phage lysates.** To detect the excised and circularized Sp5 genomes, qualitative PCR was performed using primers specific to each of the attachment site (*attP*)-flanking regions (Supplementary Table S2). The template DNA was prepared using a simplified alkaline DNA preparation method<sup>63</sup>. Briefly, after 1-h treatment with MMC, bacterial cells were collected by centrifugation at 8000xg for 5 min at 4 °C. The cell pellet was lysed with 25 mM NaOH for 10 min at 95 °C and then neutralized with 80 mM Tris-HCl. After a brief centrifugation, one microliter of the supernatant was used as a PCR template.

To quantify the DNase-resistant Sp5 DNA in phage lysates, each lysate prepared as described above was treated with DNase I (50 µg/ml) (Sigma-Aldrich) and RNase A (50 µg/ml) (Sigma-Aldrich) for 1 h at 37 °C. The sample was then treated with proteinase K (100 µg/ml; Wako Pure Chemical Industries, Ltd.) for 1 h at 50 °C, and the DNA was isolated via phenol:chloroform extraction and ethanol precipitation and quantified by qPCR using the StepOnePlus™ Real-Time PCR system and TaqMan Fast Advanced Master Mix (Applied Biosystems). The concentrations of the TaqMan probes and amplification primers used for qPCR were 250 nM and 400 nM, respectively. For TaqMan probes, FAM (5′) and TAMRA (3′) were used as the reporter and quencher dyes, respectively.

Primer Express™ (PE Biosystems) was used to design TaqMan probes and primers (Supplementary Table S2). The qPCR was performed for 45 cycles at 95 °C for 15 sec and 60 °C for 1 min. Phage DNA isolated from 1 ml of culture was used as a template. Standard curves with a dynamic range from 10<sup>2</sup> to 10<sup>8</sup> copies/μl were generated for each amplicon. Residual chromosomal backbone DNA in DNase-treated phage lysates was also quantified as a control.

**Protein expression, purification and antibody preparation.** Full-length target ORFs were PCR-amplified and cloned into the pGEX-6p-1 expression vector (GE Healthcare) to generate genes encoding GST-fusion proteins. The sequences of cloned ORFs were confirmed by DNA sequencing. The recombinant plasmids were introduced into *E. coli* BL21 (DE3) cells<sup>64,65</sup>, which were grown at 37 °C in Ap-containing LB. To induce protein expression, IPTG was added at a concentration of 0.1 mM when the OD<sub>600</sub> reached 0.6, and the cells were then cultured for 3 h at 25 °C. The cells were harvested by centrifugation at 5000xg at 4 °C, resuspended in phosphate-buffered saline (PBS) (pH 7.3), lysed by sonication and then separated into soluble and insoluble fractions by centrifugation at 14,000xg at 4 °C. Proteins in each fraction were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). GST-fusion proteins in soluble fractions were affinity-purified using Glutathione Sepharose 4B (GE Healthcare) according to the manufacturer's instructions, and the purities of those proteins were confirmed by SDS-PAGE. Rabbit antibodies against purified proteins were prepared by Sigma-Aldrich Japan. The titers and specificities of each antiserum were determined via enzyme-linked immunosorbent assays or Western blotting using horseradish peroxidase-conjugated goat anti-rabbit IgG antibodies (Bio-Rad).

**Electron microscopy and immunoelectron microscopy.** A drop of phage suspension was loaded onto a formvar-carbon-coated copper grid and allowing it to sediment for 1 min at room temperature (RT). Excess liquid was removed with Whatman blotting paper, and the copper grids were then negatively contrasted with 2% uranyl acetate. Phage morphologies were analyzed using the HT7700 transmission EM (Hitachi) operating at 80 kV.

Monodisperse colloidal gold (CG) 8 nm in diameter was prepared as described previously<sup>66</sup>. The conjugation of CG particles to goat anti-rabbit IgG (Vector Laboratories) was performed as previously described<sup>67</sup>. For the immunoelectron microscopy analysis, a drop of phage suspension was placed onto a 150-mesh gold grid coated with Formvar film and incubated for 10 min at RT, and excess liquid was removed using Whatman blotting paper. The grids were incubated with 1% bovine serum albumin (BSA) and were incubated with primary antibodies (1:500) for 1 h at RT. After being rinsed with distilled water (DW), the grids were incubated with goat anti-rabbit IgG-CG conjugate (1/50 diluted with 1% BSA) for 40 min at RT. Each grid was then washed with DW and fixed with 1% glutaraldehyde-cacodylate-buffer (pH 7.4). The samples were negatively stained with 2% uranyl acetate and examined with Hitachi HT7700.

## References

- Hatfull, G. F. Bacteriophage genomics. *Curr Opin Microbiol* **11**, 447–453 (2008).
- Rodriguez-Valera, F. *et al.* Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**, 828–836 (2009).
- Chen, J. & Novick, R. P. Phage-mediated intergeneric transfer of toxin genes. *Science* **323**, 139–141 (2009).
- Pedulla, M. L. *et al.* Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**, 171–182 (2003).
- Pope, W. H. *et al.* Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *eLife* **4**, 06416, 10.7554/eLife06416 (2015).
- Hunt, J. M. Shiga toxin-producing *Escherichia coli* (STEC). *Clin Lab Med* **30**, 21–45 (2010).
- Davis, T. K., Van De Kar, N. C. & Tarr, P. I. Shiga Toxin/Verocytotoxin-Producing *Escherichia coli* Infections: Practical Clinical Perspectives. *Microbiol Spectr* **2**, EHEC-0025–2014 (2014).
- Mauro, S. A. & Koudelka, G. B. Shiga toxin: expression, distribution, and its role in the environment. *Toxins (Basel)* **3**, 608–625 (2011).
- Croxen, M. A. *et al.* Recent Advances in Understanding Enteric Pathogenic *Escherichia coli*. *Clin Microbiol Rev* **26**, 822–880 (2013).
- O'Brien, A. D. *et al.* Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* **226**, 694–696 (1984).
- Allison, H. E. Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. *Future Microbiol* **2**, 165–174 (2007).
- Los, J. M., Los, M. & Wegrzyn, G. Bacteriophages carrying Shiga toxin genes: genomic variations, detection and potential treatment of pathogenic bacteria. *Future Microbiol* **6**, 909–924 (2011).
- McDonough, M. A. & Buttermont, J. R. Spontaneous tandem amplification and deletion of the shiga toxin operon in *Shigella dysenteriae* 1. *Mol Microbiol* **34**, 1058–1069 (1999).
- Boerlin, P. *et al.* Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. *J Clin Microbiol* **37**, 497–503 (1999).
- Donohue-Rolfe, A., Kondova, I., Oswald, S., Hutto, D. & Tzipori, S. *Escherichia coli* O157:H7 strains that express Shiga toxin (Stx) 2 alone are more neurotropic for gnotobiotic piglets than are isotypes producing only Stx1 or both Stx1 and Stx2. *J Infect Dis* **181**, 1825–1829 (2000).
- Ferens, W. A. & Hovde, C. J. *Escherichia coli* O157:H7: Animal Reservoir and Sources of Human Infection. *Foodborne Pathog Dis* **8**, 465–487 (2011).
- Nguyen, Y. & Sperandio, V. Enterohemorrhagic *E. coli* (EHEC) pathogenesis. *Front Cell Infect Microbiol* **2**, 90, 10.3389/fcimb00090 (2012).
- Ogura, Y. *et al.* The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. *Sci Rep* **5**, 16663, 10.1038/srep16663 (2015).
- de Sablet, T. *et al.* Differential expression of *stx2* variants in Shiga toxin-producing *Escherichia coli* belonging to seropathotypes A and C. *Microbiology* **154**, 176–186 (2008).
- Kawano, K., Okada, M., Haga, T., Maeda, K. & Goto, Y. Relationship between pathogenicity for humans and *stx* genotype in Shiga toxin-producing *Escherichia coli* serotype O157. *Eur J Clin Microbiol Infect Dis* **27**, 227–232 (2008).
- Kulasekara, B. R. *et al.* Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. *Infect Immun* **77**, 3713–3721 (2009).

22. Calderwood, S. B. & Mekalanos, J. J. Iron regulation of Shiga-like toxin expression in *Escherichia coli* is mediated by the fur locus. *J Bacteriol* **169**, 4759–4764 (1987).
23. Ogura, Y. *et al.* Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc Natl Acad Sci USA* **106**, 17939–17944 (2009).
24. Garcia-Aljaro, C., Muniesa, M., Jofre, J. & Blanch, A. R. Genotypic and phenotypic diversity among induced, *stx2*-carrying bacteriophages from environmental *Escherichia coli* strains. *Appl Environ Microbiol* **75**, 329–336 (2009).
25. Ohnishi, M. *et al.* Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. *Proc Natl Acad Sci USA* **99**, 17043–17048 (2002).
26. Schmidt, H. Shiga-toxin-converting bacteriophages. *Res Microbiol* **152**, 687–695 (2001).
27. Smith, D. L. *et al.* Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics* **13**, 311, 10.1186/13–311 (2012).
28. Plunkett G., I., Rose D. J., Durfee & T. J., Blattner FR. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J. Bacteriol*, 1767–1678 (1999).
29. Makino, K. *et al.* Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. *Genes Genet Syst* **74**, 227–239 (1999).
30. Yokoyama, K. *et al.* Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak. *Gene* **258**, 127–139 (2000).
31. Cooper, K. K. *et al.* Comparative genomics of enterohemorrhagic *Escherichia coli* O145:H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157:H7. *BMC Genomics* **15**, 17, 10.1186/15–17 (2014).
32. Asadulghani, M. *et al.* The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog* **5**, e1000408, 10.1371/pat.1000408 (2009).
33. Allison, H. E. *et al.* Immunity profiles of wild-type and recombinant shiga-like toxin-encoding bacteriophages and characterization of novel double lysogens. *Infect Immun* **71**, 3409–3418 (2003).
34. Nubling, S. *et al.* Bacteriophage 933W encodes a functional esterase downstream of the Shiga toxin 2a operon. *Int J Med Microbiol* **304**, 269–274 (2014).
35. Islam, M. R. *et al.* A sensitive and simple plaque formation method for the Stx2 phage of *Escherichia coli* O157:H7, which does not form plaques in the standard plating procedure. *Plasmid* **67**, 227–235 (2012).
36. Hayashi, T. *et al.* Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**, 11–22 (2001).
37. Lucchini, S., Desiere, F. & Brussow, H. The structural gene module in *Streptococcus thermophilus* bacteriophage phi Sfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. *Virology* **246**, 63–73 (1998).
38. Fokine, A. & Rossmann, M. G. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **4**, e28281, 10.4161/bact.28281 (2014).
39. Duda, R. L., Martincic, K. & Hendrix, R. W. Genetic basis of bacteriophage HK97 prohead assembly. *J Mol Biol* **247**, 636–647 (1995).
40. Huet, A., Conway, J. F., Letellier, L. & Boulanger, P. *In vitro* assembly of the T=13 procapsid of bacteriophage T5 with its scaffolding domain. *J Virol* **84**, 9350–9358 (2010).
41. Aksyuk, A. A. & Rossmann, M. G. Bacteriophage Assembly. *Viruses* **3**, 172–203 (2011).
42. Casjens, S. & Hendrix, R. The Viruses in *The Bacteriophages* (ed. Calendar, R.) 15–91 (Springer US, 1988).
43. Smith, M. C. *et al.* Bacteriophage collagen. *Science* **279**, 1834 (1998).
44. Smith, D. L. *et al.* Short-tailed *stx* phages exploit the conserved YaeT protein to disseminate Shiga toxin genes among enterobacteria. *J Bacteriol* **189**, 7223–7233 (2007).
45. Becker, A., Murialdo, H. & Gold, M. Studies on an *in vitro* system for the packaging and maturation of phage lambda DNA. *Virology* **78**, 277–290 (1977).
46. Moore, S. D. & Prevelige, P. E. Jr. Bacteriophage P22 portal vertex formation *in vivo*. *J Mol Biol* **315**, 975–994 (2002).
47. Fu, C. Y. & Prevelige, P. E. Jr. *In vitro* incorporation of the phage Phi29 connector complex. *Virology* **394**, 149–153 (2009).
48. Kemp, C. L., Howatson, A. F. & Siminovitch, L. Electron microscopy studies of mutants of lambda bacteriophage. I. General description and quantitation of viral products. *Virology* **36**, 490–502 (1968).
49. Thuman-Commike, P. A., Greene, B., Malinski, J. A., King, J. & Chiu, W. Role of the scaffolding protein in P22 procapsid size determination suggested by T = 4 and T = 7 procapsid structures. *Biophys J* **74**, 559–568 (1998).
50. Padilla-Meier, G. P. *et al.* Unraveling the role of the C-terminal helix turn helix of the coat-binding domain of bacteriophage P22 scaffolding protein. *J Biol Chem* **287**, 33766–33780 (2012).
51. Fu, C. Y., Morais, M. C., Battisti, A. J., Rossmann, M. G. & Prevelige, P. E. Jr. Molecular dissection of  $\phi$ 29 scaffolding protein function in an *in vitro* assembly system. *J Mol Biol* **366**, 1161–1173 (2007).
52. Roessner, C. A. & Ihler, G. M. Proteinase sensitivity of bacteriophage lambda tail proteins gpJ and pH in complexes with the lambda receptor. *J Bacteriol* **157**, 165–170 (1984).
53. King, J. & Casjens, S. Catalytic head assembling protein in virus morphogenesis. *Nature* **251**, 112–119 (1974).
54. Nelson, R. A., Reilly, B. E. & Anderson, D. L. Morphogenesis of bacteriophage phi 29 of *Bacillus subtilis*: preliminary isolation and characterization of intermediate particles of the assembly pathway. *J Virol* **19**, 518–532 (1976).
55. Mitraiki, A., Miller, S. & van Raaij, M. J. Review: conformation and folding of novel beta-structural elements in viral fiber proteins: the triple beta-spiral and triple beta-helix. *J Struct Biol* **137**, 236–247, (2002).
56. Hendrix, R. & Casjens, S. Bacteriophage lambda and its genetic neighborhood In *The Bacteriophages* (ed. Calendar, R.). 409–445 (Oxford University Press, 2006).
57. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**, D222–226 (2015).
58. Hunter, S. *et al.* InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**, D211–215, (2009).
59. Letunic, I., Doerks, T. & Bork, P. SMART 6: recent updates and new developments. *Nucleic Acids Res* **37**, D229–232 (2009).
60. Sekiya, K. *et al.* Supermolecular structure of the enteropathogenic *Escherichia coli* type III secretion system and its direct interaction with the EspA-sheath-like structure. *Proc Natl Acad Sci USA* **98**, 11638–11643 (2001).
61. Donnenberg, M. S. & Kaper, J. B. Construction of an eae deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. *Infect Immun* **59**, 4310–4317 (1991).
62. Tobe, T., Sasakawa, C., Okada, N., Honma, Y. & Yoshikawa, M. *vacB*, a novel chromosomal gene required for expression of virulence genes on the large plasmid of *Shigella flexneri*. *J Bacteriol* **174**, 6359–6367 (1992).
63. Beige, J. *et al.* Clinical evaluation of a *Mycobacterium tuberculosis* PCR assay. *J Clin Microbiol* **33**, 90–95 (1995).
64. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113–130 (1986).
65. Grodberg, J. & Dunn, J. J. *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol* **170**, 1245–1253 (1988).
66. Slot, J. W. & Geuze, H. J. A new method of preparing gold probes for multiple-labeling cytochemistry. *Eur J Cell Biol* **38**, 87–93 (1985).
67. De Mey, J., Moeremans, M., Geuens, G., Nuydens, R. & De Brabander, M. High resolution light and electron microscopic localization of tubulin with the IGS (immuno gold staining) method. *Cell Biol Int Rep* **5**, 889–899 (1981).
68. Dunn, J. J. & Studier, F. W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol* **166**, 477–535 (1983).

69. Jensen, K. F. The *Escherichia coli* K-12 “wild types” W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. *J Bacteriol* **175**, 3401–3407 (1993).
70. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97**, 6640–6645 (2000).

### Acknowledgements

We thank Akemi Yoshida, Miki Shimbara, Yumiko Takeshita, and Nobuko Fuji for their technical assistance. Shakhinur Islam Mondal and Md Rakibul Islam were supported by scholarships from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

### Author Contributions

S.I.M., M.R.I. and T.H. conceived of and designed the experiments. S.I.M. and M.R.I. conducted all experiments with supports of A.S. (electron microscopic examinations), T.O., Y.G., Y.K. and Y.O. (other experiments). M.A. provided tools and bacterial strains. S.I.M. and T.H. wrote the manuscript. All authors reviewed the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Mondal, S. I. *et al.* Genes essential for the morphogenesis of the Shiga toxin 2-transducing phage from *Escherichia coli* O157:H7. *Sci. Rep.* **6**, 39036; doi: 10.1038/srep39036 (2016).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016