


# Genome Plasticity by Insertion Sequences Learned From a Case of Radiation-Resistant Bacterium *Deinococcus geothermalis*

Bioinformatics and Biology Insights  
Volume 15: 1–10  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11779322211037437



Chanjae Lee<sup>1\*</sup>, Min K Bae<sup>1,2\*</sup>, Nakjun Choi<sup>1,3</sup>, Su Jeong Lee<sup>1</sup>  
and Sung-Jae Lee<sup>1</sup> 

<sup>1</sup>Department of Biology, Kyung Hee University, Seoul, Korea. <sup>2</sup>Department of Geography, Kyung Hee University, Seoul, Korea. <sup>3</sup>Department of Life Sciences, Korea University, Seoul, Korea.

**ABSTRACT:** The genome of the radiation-resistant bacterium *Deinococcus geothermalis* contains 19 types of insertion sequences (ISs), including 93 total transposases (Tpases) in 73 full-length ISs from the main chromosome and 2 mega plasmids. In this study, 68 ISs from the *D. geothermalis* genome were extracted to implicate the earlier genome before its mutation by transposition of ISs. The total size of eliminated ISs from genome was 78.85 kb. From these in silico corrections of mutation by the ISs, we have become aware of some bioinformatics factualness as follows: (1) can reassemble the disrupted genes if the exact IS region was eliminated, (2) can configure the schematic clustering of major DDE type Tpases, (3) can determine IS integration order across multiple hot spots, and (4) can compare genetic relatedness by the partial synteny analysis between *D. geothermalis* and *Deinococcus* strain S9. Recently, we found that several IS elements actively transferred to other genomic sites under hydrogen peroxide-induced oxidative stress conditions, resulting in the inactivation of functional genes. Therefore, the single species genome's mobilome study provides significant support to define bacterial genome plasticity and molecular evolution from past and present progressive transposition events.

**KEYWORDS:** *Deinococcus geothermalis*, insertion sequence (IS), in silico correction, genomic plasticity, transposase (Tpase), transposition

**RECEIVED:** May 17, 2021. **ACCEPTED:** July 12, 2021.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Grant of the National Research Foundation of Korea funded by the Korean government (MSIT; 2020R1F1A1A070779).

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Sung-Jae Lee, Department of Biology, Kyung Hee University, 26 Kyungheedaero, Dongdaemun-gu, Seoul 02447, Korea.  
Email: sungjaelee@khu.ac.kr

## Introduction

Bacterial genomes contain a number of transposable elements (TEs), including transposons (Tn) and insertion sequences (ISs), which are separated into active and inactive forms depending on the occurrence of transposition.<sup>1</sup> There are several potential IS transposition mechanisms that depend on 4 functional transposases (Tpases): (1) RNase H-like Tpase, characterized by their DDE and DEDD motifs, (2) HuH single-stranded DNA Tpase, (3) serine Tpase, and (4) tyrosine Tpase.<sup>2-4</sup> The majority of Tpase in bacterial ISs contain the DDE motif. In general, the genome sequence of a TE contains 2 iterative sequences generated by the transposition event: the direct repeat (DR) sequence at both ends of the IS element and the terminal inverted repeat (TIR) sequence located directly within 2 DR sequences.<sup>5,6</sup> TIR sequences generate a stem structure that induces site recognition, and Tpase and DR sequences are determined by a cleavage site.<sup>4</sup> Thus, unique repeat sequences along with the type of Tpase are key criteria for the classification of the IS family. At present, there are more than 4500 ISs and they are categorized into 29 IS families.<sup>3</sup> Many types of ISs are replicated multiple times in genomic DNA (chromosome and plasmids) due to transposition. Particular IS elements are duplicated at the original site and

then moved to a new location in the genome in a manner known as the copy-and-paste mode of transposition. Certain other IS elements move to a new site in the genome after being excised at the original site in a manner known as the cut-and-paste mode of transposition.<sup>1,7</sup> These transposition modes depend on the Tpase mechanism. Many IS family members are distributed throughout bacterial genera and exist as a single or multiple copies, resulting in genomic plasticity driven by gene disruption, promoter malfunction, and gene activation.<sup>1,3</sup> An understanding of these modes of action has facilitated the development of computational identification pipeline platforms for IS determination with distinct strategies; for example, conserved sequences of Tpase or short repeat sequences (eg, ISfinder, ISQuest, and Oasis).<sup>8-12</sup> However, there are many unknown transposition machineries with the unique structural scheme of ISs.<sup>13</sup>

According to some reports, wild-type *Deinococcus* sp. have multiple ISs transpositions and accumulate Tpase gene duplications through the movement of TEs under various stress conditions (eg, UV or gamma-ray irradiation, chemicals, reactive oxygen species [ROS] through oxidative reactions).<sup>14-16</sup> Based on the ISfinder platform, the genome of the radiation-resistant bacterium *Deinococcus geothermalis* contains 19 types of ISs and 73 ISs.<sup>8</sup> For example, the ISDge2 type member of the IS1 family includes 15 copies within the genome.

\*Chanjae Lee and Min K Bae contributed equally.



**Table 1.** Genomic DNA features after elimination of ISs and recovered ORFs.

GENOMES	SIZE (MB)/PROTEINS	NO. OF TPASES/IS	ELIMINATED IS SIZE	RECOVERED ORF NUMBER (CODE OF GENES)
Chromosome	2.47/2310	33/30	34 627 bp (1.38%)	8 (Dgeo_0307, 0806, 0926, 1674, 1806, 2109, 2196, 2207)
Plasmid pDGE001	0.57/490	38/28	30 186 bp (5.26%)	9 (Dgeo_2373, 2380, 2384, 2431, 2432, 2439, 2658, 2684, 2699)
Plasmid pDGE002	0.21/203	22/15	14 037 bp (6.67%)	2 (Dgeo_2937, 3099)
Total	3.25/3003	93/73	78 850 bp (2.43%)	19

Abbreviations: IS, insertion sequences; ORF, open reading frame.

Despite wild-type *D. geothermalis* genome already containing 10 and 4 copies of *ISDge5* and *ISDge7* elements, respectively, the active transposition of *ISDge5* and *ISDge7* IS elements occurred under oxidative stress of hydrogen peroxide treatment in a *Dps* gene (DNA protection proteins from starved cell) disrupted mutant strain.<sup>17</sup> In addition, active transposition of *ISDge6* from the LysR family regulator disrupted mutant ( $\Delta$ *dge\_2840*) and *ISDge11* from the wild type by oxidative stress was first reported in our previous work.<sup>18</sup> From these observations, we have motivated the question: Can we look at the earlier genome contents of the bacterium before transposition of ISs in terms of genomic plasticity? Thus, we performed an in silico correction to imagine past events by the recovery of mutation of genomic contents by ISs through the elimination of full-length ISs from the genome sequence of wild-type *D. geothermalis* DSM11300<sup>T</sup>. Significant information can be obtained with regard to the types of IS elements and the conserved genetic sequences involved in the translocation and integration of ISs into the genome (eg, DR and TIR sequences).

The use of this in silico correction of genomic mutation by transposition of ISs has resulted in a number of interesting findings: 21 disordered genes by the transposition of IS elements were recovered through the elimination of full-length IS elements and the genomic plasticity of the *Deinococcus* species in comparison with the most closely related *Deinococcus* genomes was explained.

## Materials and Methods

### Data summary

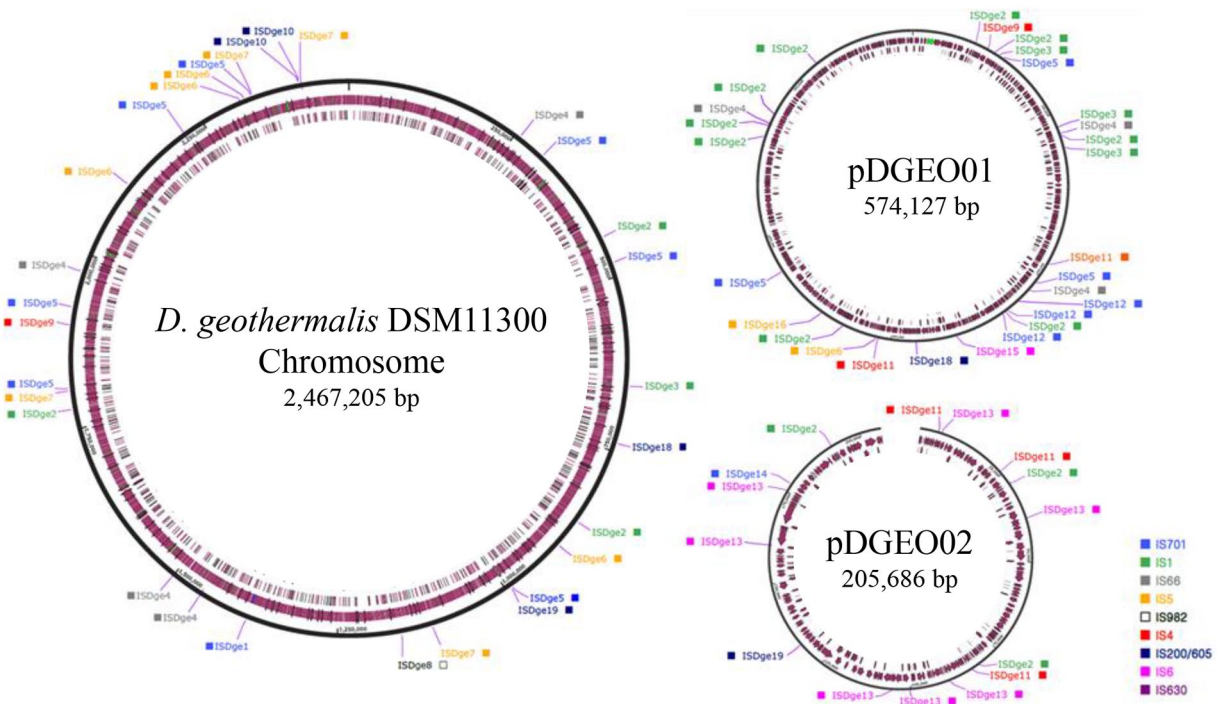
The genome database of the National Center for Biotechnology Information (NCBI) was the source for the genome data. Genome sequence information for *D. geothermalis* was obtained for DSM11300<sup>T</sup>, NC\_008025.1 for chromosome, NC\_008010.2 for plasmid 1, NC\_009939 for plasmid 2, from NCBI GenBank databases and the classification of ISs followed the ISfinder platform (<https://isfinder.biotoul.fr>). All data are available with the online version of this article.

### Distribution of ISs in the *D. geothermalis* genome

The genome sequence information for *D. geothermalis* was obtained from DSM11300<sup>T</sup>. The genomic size, numbers of located Tase and IS elements, IS extracted size, and recovered open reading frame (ORF) number, and genes are summarized in Table 1. Insertion sequence classification and basic information about TIR sequences and DR sequences were gathered using the ISfinder platform.<sup>8</sup> The *D. geothermalis* genome contains 73 full-length ISs clustered into 19 types and three *ISDge* types, *ISDge2*, *ISDge3*, and *ISDge10*, contain second Tase-related genes in complete IS elements. A free software version of the “SnapGene Viewer” program (<https://www.snapgene.com/snapgene-viewer/>) was used for marking the loci of ISs (Figure 1).

### In silico correction of genomic mutation by ISs

First, we needed to identify the full-length region for typical IS elements through TIR and DR sequences. The entire region of the IS element was then eliminated and kept as a single copy of the DR sequence (Figure S1). The DNA sequence-converting work was performed by the nucleic acid sequence massager of Attotron Biotechnologies Corporation (<http://www.mathadict.net/dnatranslate3.htm>). Next, we analyzed the connected DNA sequences between upstream and downstream regions of the IS element in the combined ORF production after extraction of the full IS region. For this, we used an available translation tool: Translation of ExPASy from Swiss Institute of Bioinformatics (SIB; <https://web.expasy.org/translate/>). Nucleotide sequences and/or amino acid sequences were aligned and compared among genus *Deinococcus* species to determine the identity and sequence-pairing using BLAST from NCBI and ClustalW (for multiple sequence alignments; <https://www.genome.jp/tools-bin/clustalw>). A 100% match of combined ORF sequences with other genes and proteins from *Deinococcus* species indicated that the genes were restored in an in silico manner.



**Figure 1.** Genomic distribution of 73 IS elements in *D. geothermalis* DSM11300<sup>T</sup> (chromosome and 2 mega plasmids). The genome includes 9 IS family members. A list of the IS family members among 4 *Deinococcus* species is compared in supplementary data Table S1, and all 73 ISs are shown in supplementary data Table S2. The information was based and modified from data supplied by ISfinder. IS indicates insertion sequences.

### Synteny analysis among *Deinococcus* species

We used the SyntTax program (<https://archaea.i2bc.paris-saclay.fr/SyntTax>),<sup>19</sup> SynMap2 (<https://genomeevolution.org/CoGe/SynMap.pl>) and <https://genomeevolution.org/CoGe/GEvo.pl>),<sup>20</sup> and NCBI BLAST to define the similarity of genomic contents and target-gene identity and loci among *Deinococcus* species. The genome sequence of *D. geothermalis* DSM11300<sup>T</sup> was compared to that of *Deinococcus* sp. S9 strain, which has 101 contigs in the NCBI genome Genbank (SKCF00000000.1). YASS (<https://bioinfo.lifl.fr/yass/index.php>) was also used to compare contigs of strain S9 and the *D. geothermalis* genome.<sup>21</sup> *Deinococcus* sp. S9 strain showed 99.06% identity of the 16S rRNA sequence to *D. geothermalis*.<sup>22</sup>

## Results

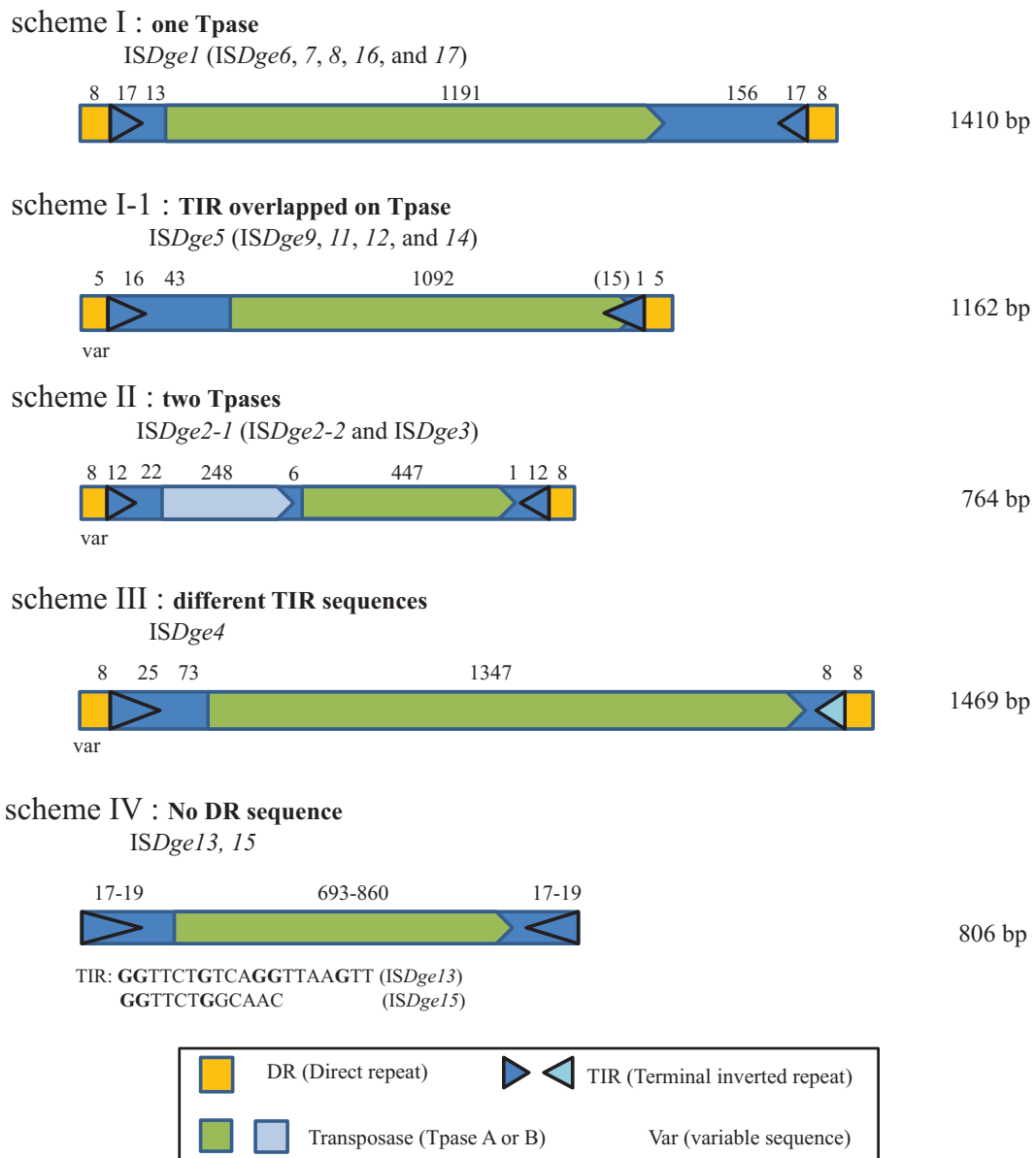
### Distribution and classification of insertion sequences within the genome

ISs of the radiation-resistant bacterium *D. geothermalis* are distributed on a single chromosome and 2 mega plasmids. Table 1 shows a summary of the genome size, IS number, eliminated IS size, and the number of recovered ORFs in *D. geothermalis* DSM11300<sup>T</sup>. The *D. geothermalis* genome includes 9 ISs family members, while other *Deinococcus* species, *Deinococcus radiodurans*, *Deinococcus deserti*, and *Deinococcus* strain S9, have 7, 5, and 13 IS families, respectively (Table S1). According to the ISfinder proposed classification, 93 Tpsases belonged to full-length 73 ISs according to the names of 19 *Dgeo* IS types

(Table S2). The highest gene copy number of IS type is *ISDge2* (15 copies), followed by *ISDge5* (10 copies). Our previous studies have found that multi-copy IS elements, *ISDge5*, *ISDge6*, *ISDge7*, and *ISDge11* members, used a replicative transposition action mode and their movement was actively triggered by oxidative stress resulting in enhancing the copy number through active transposition.<sup>17,18</sup> There are 6 single IS members (*ISDge1*, *ISDge8*, *ISDge10*, *ISDge14*, *ISDge15*, and *ISDge17*). Interestingly, all 7 IS copies of *ISDge13* type are found in plasmid pDGEO02. The loci of 73 ISs within the genomic DNA are indicated in a genome-wide schematic (Figure 1).

### Structural features of 19 *ISDge* elements

Figure 2 and Table S2 illustrate the total ISs profile, including (1) family classification, (2) Tpsase amino acid length, its catalytic motif and the number of copies of Tpsase genes, (3) TIR, and together with (4) DR sequences as a crucial result of the transposition in the genome of *D. geothermalis*. We propose an upper level of IS clustering and classify 4 structural schemes of major DDE motif IS elements following ORF (mainly Tpsase) number, the identity of TIR sequence, and DR sequence presence or absence. There are 2 typical ISs structures (scheme I–II) and 2 distinct structures (scheme III–IV) (Table 2). Scheme I has a single Tpsase gene and constant DR sequences in 6 IS types (*ISDge1*, 6, 7, 8, 16, and 17). Scheme I-1 has 5 IS types (*ISDge5*, 9, 11, 12, and 14), all of which have a single Tpsase gene and an overlapping right TIR sequence in the



**Figure 2.** Four structural schemes of ISs containing DDE motif Tphase, consisting of two classic IS types (scheme I-II) and two unique IS types (scheme III-IV). ISs are typically composed of transposase (Tphase) and terminal inverted repeat sequence (TIR), and additional direct repeat sequence (DR) as a result of transposition event. Interesting and unique ISs structures have included varying TIR sequences and no DR sequences. DR indicates direct repeat; IS, insertion sequences; TIR, terminal inverted repeat; DDE, Asp-Asp-Glu.

ORF region of the Tphase. Scheme II ISs (*ISDge2* and *ISDge3*) have 2 Tphases-related genes which include IS1 family members composed of 148 aa-long Tphase A and 82 aa-long Tphase B with same transcriptional direction. *ISDge2* is divided into 2 subtypes according to the presence or absence of DR sequences (Figure 3). Scheme III contains only *ISDge4* (an IS66 family member) with 8 copies in the genome; although this scheme has a typical structural composition, interestingly, its TIR sequence is not identical. Scheme IV, composed of *ISDge13* and *ISDge15* (an IS6 family), has well conserved TIR sequence contained first 2 “GG” and seventh “G” and does not have both DR sequences in the IS border region.<sup>23</sup> Without DR sequences, the mechanism of IS element integration into DNA is unclear; one predicted hypothesis is that this transposition

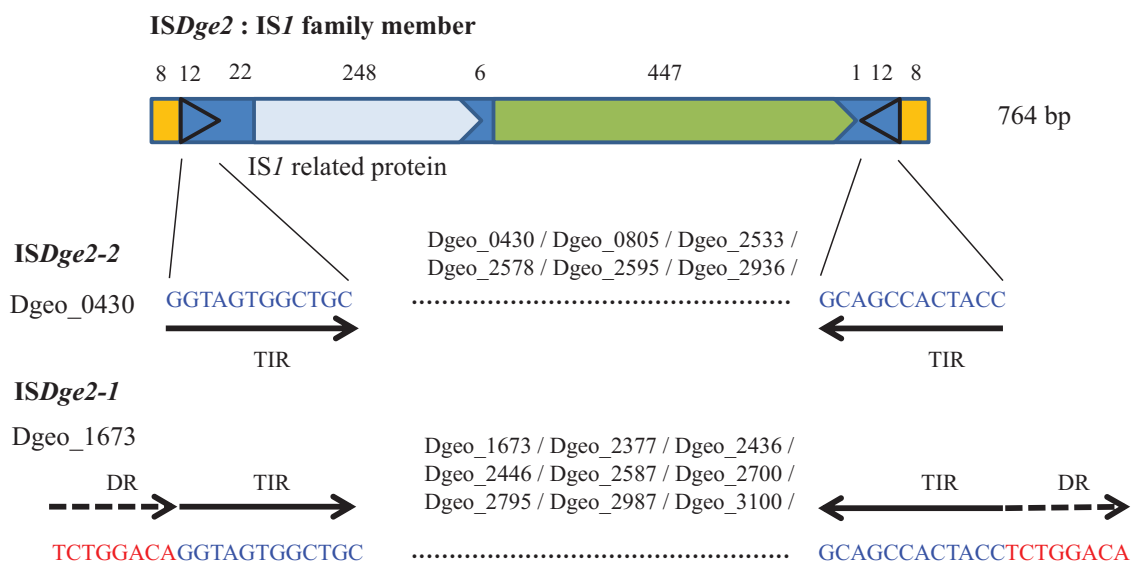
should be accompanied by the cointegration mechanism of IS6/IS26 family members.<sup>24</sup>

Based on these 4 clusterings of IS structural schemes containing DDE motif, IS701 family members, *ISDge1*, *ISDge5*, *ISDge12*, and *ISDge14* were separated into scheme I and scheme I-1. *ISDge1* has 13.7% to 14.3% amino acid sequence identity of Tphase to other IS701 family members and contained a non-overlapping TIR sequence from the C-terminal region of Tphase. Tphase of *ISDge5* showed 67.2% to 68.6% amino acid sequence identity to *ISDge12* and *ISDge14* Tphase. The amino acid sequence identity of Tphase was 94.49% between *ISDge12* and *ISDge14*. The right border TIR sequences of *ISDge5*, *ISDge12*, and *ISDge14* are distinctly overlapped into the C-terminal region of Tphase. Thus, *ISDge1* is distinctly separated from

**Table 2.** Clustering of 4 structural schemes of ISs containing DDE motif Tpnase in *D. geothermalis*.

SCHEME	IS TYPE	SIZE (NT)				COPY NUMBER
		DR	TIR	TPASE	FULL LENGTH	
Scheme I	ISDge1	8	17	1191	1410	1
	ISDge6	3	22/23	1011	1123	5
	ISDge7	2	8	798	880	4
	ISDge8	5	19	816	917	1
	ISDge16	5	19	858	920	1
	ISDge17	2	6	1026	1220	1
Scheme I-1	ISDge5	5	16	1092	1162	10
	ISDge9	9	17	1059	1145	2
	ISDge11	8/9	22	996	1192/1194	5
	ISDge12	2/3	3	1092	1149/1151	2
	ISDge14	4	23	1092	1160	1
Scheme II	ISDge2-1	8	12	248/447	764	9
	ISDge2-2	0	12	248/447	756	6
	ISDge3	2/8	13	248/447	754/766	4
Scheme III	ISDge4	8	17/8	1347	1469	8
Scheme IV	ISDge13	0	19	757	806	7
	ISDge15	0	17	693	747	1

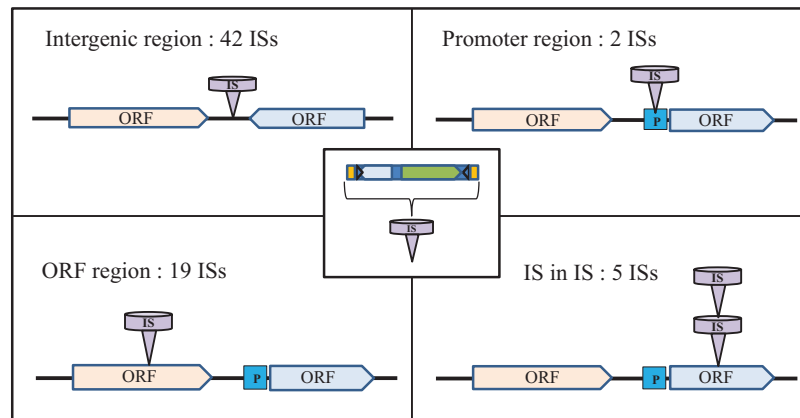
Abbreviations: DDE, Asp-Asp-Glu; DR, direct repeat; IS, insertion sequences; TIR, terminal inverted repeat.

**Figure 3.** Structural features of two ISDge2 subtypes. Both subclass members have an identical TIR sequence; however, members of the ISDge2-2 subtype have no DR sequences.

DR indicates direct repeat; IS, insertion sequences; TIR, terminal inverted repeat.

others. Therefore, this clustering of many DDE motif IS family members based on the structural scheme should help sort for more systematic management after a detailed schematic analysis of all DDE motif IS-related family members originated from prokaryotes for the last 4 decades.

*D. geothermalis* genome contains an additional IS family member of IS200/IS605. ISDge10 is an IS605 group member with 2 Tpnase-related ORFs. ISDge18 and ISDge19 as both IS1341 group members contain a single Tpnase B. In general, IS200/IS605 family member ISs lack conserved TIR sequences



**Figure 4.** Four patterns of IS integrational regions (ie, intergenic, promoter, ORF, and IS regions). The list of transposase gene regions for IS integration is specified in supplementary data Table S3. IS indicates insertion sequences; ORF, open reading frame.

and not shown DR sequence at the out of IS border region. They occasionally used a single-strand circular DNA intermediate. This architectural feature, distinct from classical ISs, has occurred in the transposition mode described as a “peel and paste.”<sup>25</sup>

#### Correction of genes interrupted by ISs

Based on total IS elements in *D. geothermalis* genome of Table 2 and Table S2, DNA sequence elimination of 73 IS elements was performed from the genomic loci. After eliminating the ISs, 78.85 kb of the *D. geothermalis* genome had been extracted from the main chromosome and 2 mega plasmids; this correlated to 2.43% of the total genome size which involves 19 ORFs (Table 1). An annotated T<sub>p</sub>ase gene encoding region for a specific IS element was used in this work, which was helpful in identifying the primary T<sub>p</sub>ase genes (ie, those with the most copies of ISD<sub>g</sub>e2 as 15 copies). Therefore, the “Dgeo\_0430 region” indicates the full IS element length comprising T<sub>p</sub>ase Dgeo\_0430.

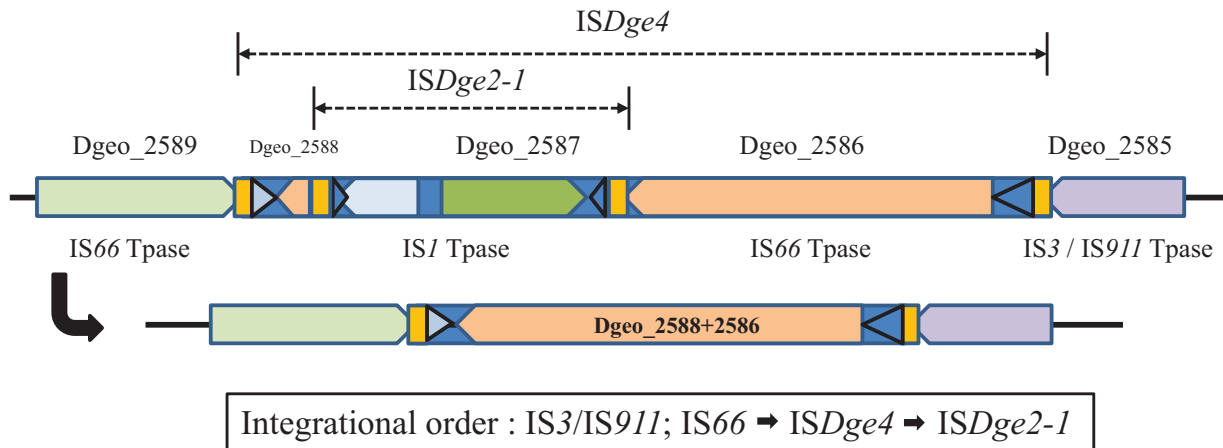
Four restoration patterns from disorders by the IS integration were observed in this analysis (Figure 4). First, an extract that does not affect the expression of surrounding genes and is harmless. In total, 42 IS elements are integrated between the intergenic regions where the gene’s 3’ terminal meets, indicating this non-effective phenomenon (Table S3). Second, 2 genes were silenced due to IS integration within a promoter region (ie, Dgeo\_0464 region integrated into the promoter region of the SAM-dependent methyltransferase Dgeo\_0463 and Dgeo\_3074 region ISD<sub>g</sub>e13 type integrated into the promoter region of a hypothetical protein Dgeo\_3073). Third, there were 19 cases of gene fusion or extension of ORF amino acid length by IS elimination. When a full-length IS region including T<sub>p</sub>ase was eliminated, a portion of the gene fragment became connected into the adjacent ORF, which occurred in 19 ORFs (8 chromosome genes, 9 pDGEO01 genes, 2 pDGEO02 genes) (Figure S1). These corrections of the interrupted genes by elimination of ISs have included 3 transporters, such as 2 ABC transporters

included for D-xylose and one MFS transporter for a metabolite:H<sup>+</sup> symporter (MHS) family; 8 enzymes, such as a diguanylate cyclase, a 4-hydroxybenzoate-3-monooxygenase, a sugar phosphate isomerase/epimerase, a type III restriction endonuclease, an ATPase, a glycosyl hydrolase, a monooxygenase, and an endoxylanase; 3 functional proteins, such as a *sufE* family, a DUF11 domain containing protein, and a secretion system protein; and 5 hypothetical proteins. The fourth is “IS in IS” pattern, detected in 5 cases (Table S3). This is another IS element embedded in the IS inserted in the ORF (see the section of “Prediction of IS integration order”).

#### Prediction of IS integration order

An in silico correction of eliminated IS elements resulted in the discovery of several IS integration hotspots in the *D. geothermalis* genome. The positions of these ISs are sometimes proximal to one another and even overlap. Using IS structural analysis, we performed an analysis of the order of genomic integration among the different IS elements. As an interesting example, the Dgeo\_2589-2585 region within plasmid pDGEO01 includes a variety of IS element types (eg, IS1, IS66, IS3/IS911 families) (Figure 5). The Dgeo\_2586 region may have been interrupted by the integration of the Dgeo\_2587 region as a full-length IS element. Elimination of this ISD<sub>g</sub>e2-1 region resulted in a combination of the disrupted proteins Dgeo\_2586 and Dgeo\_2588 into the full 448 aa-long IS66 family T<sub>p</sub>ase. Thus, the Dgeo\_2586 + 2588 region as an ISD<sub>g</sub>e4 type IS element was integrated into the C-terminal region of IS3/IS911 T<sub>p</sub>ase and IS66 T<sub>p</sub>ase of a genomic DNA locus and followed further integration by the ISD<sub>g</sub>e2-1 type of Dgeo\_2587 region. Therefore, the integration order is as follows: IS3/IS911 (Dgeo\_2585)-IS66 (Dgeo\_2589) > IS66 (ISD<sub>g</sub>e4 of Dgeo\_2586 + 2588 region) > IS1 (ISD<sub>g</sub>e2-1 of Dgeo\_2587 region).

The Dgeo\_2594-2596 region in plasmid pDGEO01 is somewhat different. The Dgeo\_2594 region is an IS4 family member; its N-terminal region has been truncated by the IS66



**Figure 5.** An example of IS integrational order within a genome hot spot. The region Dgeo\_2589-2585 of plasmid pDGEO01 contained at least 5 Tpase-related genes. The complete structure of *ISDge4* was first integrated into this region and then *ISDge2-1* type element was finally transposed. When *ISDge2-1* element was eliminated, Dgeo\_2588 and 2586 ORFs were combined, resulting in a perfect *IS66* Tpase. IS indicates insertion sequences; ORF, open reading frame.

family IS integration. The C-terminal region of Dgeo\_2596 as a member of the *IS66* family is disrupted by a full-length *ISDge2-2* element of the Dgeo\_2595 region as a member of the *IS1* family. Thus, we can predict the order of integration among these IS elements. Specifically, *IS4* family IS was first integrated into the genomic DNA locus, followed by *IS66* family IS into the N-terminal region of *IS4*, and finally, *IS1* family IS integrated into the adjacent *IS66* C-terminal region: *IS4* (Dgeo\_2594) > *IS66* (*ISDge4*) > *IS1* (*ISDge2-2*). We conclude that *ISDge2* type *IS1* family members would be the recent elements that were transposed in the plasmid pDGEO01 of *D. geothermalis*.

Elimination of the Dgeo\_2987 region was resulted in the extension of the C-terminal region of *IS3* family Tpase Dgeo\_RS15785 to 31 aa-long, making a complete ORF together with Dgeo\_RS15790 of a 155 aa-length region. Thus, *IS3* family member was also newly found in the *D. geothermalis* genome after eliminating the integrated IS member.

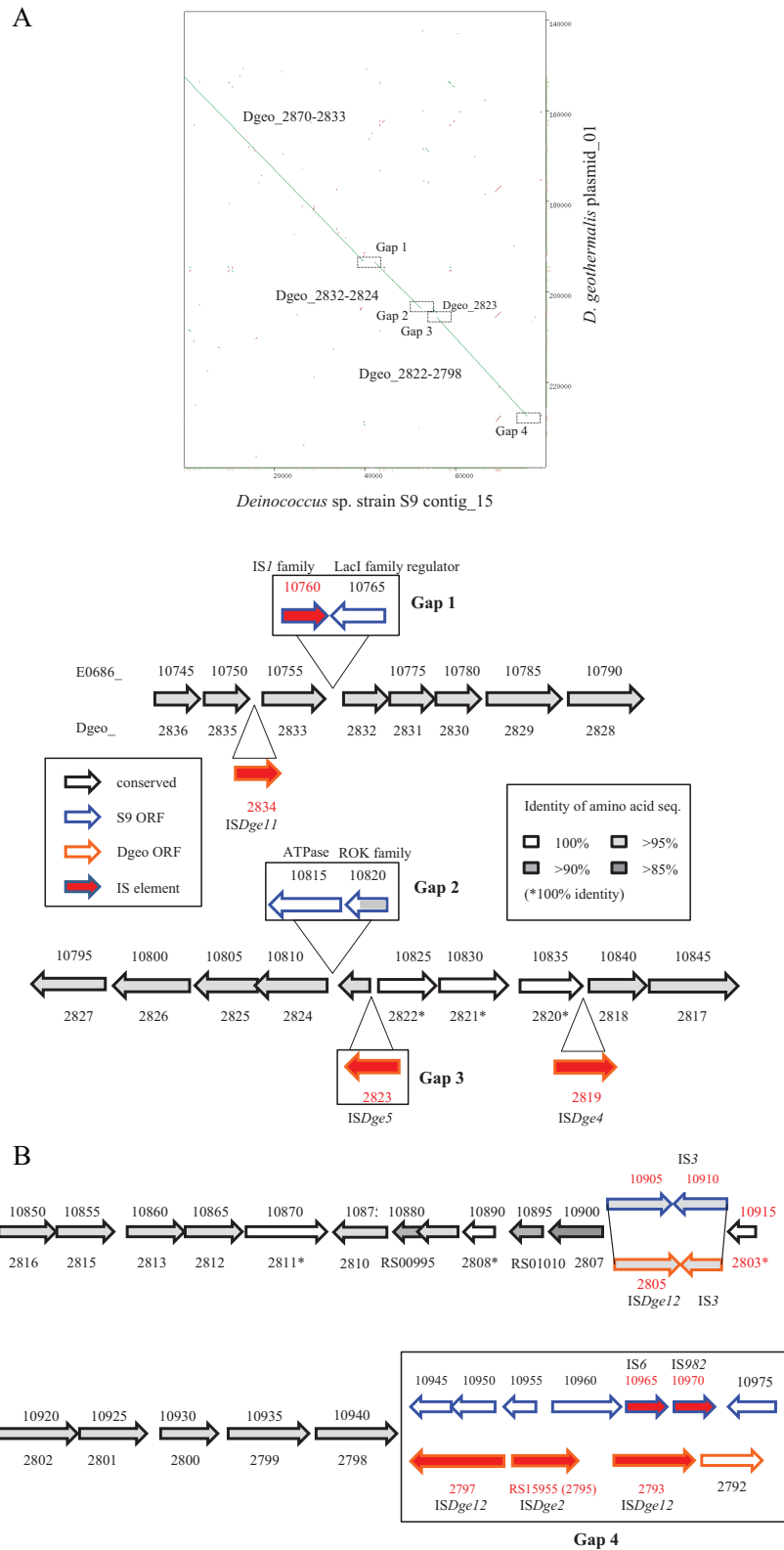
#### Lesson from comparative genomics for genome plasticity by ISs

DNA sequence similarity between genomic DNAs of 2 bacterial species was analyzed using SynMap.<sup>26</sup> The SynMap data analysis of fully assembled genomes showed the relationship of gene arrangement similarity between *D. geothermalis* and *D. deserti* (Figure S2A). The “X”-like pattern represents genome inversion, including insertion, deletion, and recombination.<sup>27</sup> Seven spots were selected, and GEvo analysis was performed to define gene arrangements match and inversion breakpoints. Spots 2, 4, and 6 of the superior axis were well-matched gene arrangements. However, Spot 7 was relatively less matched and Spots 1, 3, and 5 of the inferior axis had many genetic inversions (Figure S2B). Talwar et al<sup>22</sup> recently reported genomic sequencing data for the *Deinococcus* sp. strain S9, isolated from

microbial mat deposits of hot springs located atop the Himalayas ranges at Manikaran. While *D. geothermalis* and S9 strain are quite consistent within the 16S rRNA sequence of 99.6% identity, the similarity of their genomic DNA and gene arrangement between 2 organisms remained unknown. Although 101 contigs of the strain S9 genome cover only two-thirds of the *D. geothermalis* genome, the S9 strain genome already contains 13 IS family members and 53 ISs (Table S1). When the full-length IS regions were eliminated from the *D. geothermalis* DSM11300<sup>T</sup> genomic sequence, the reconstructed functional ORFs (eg, Dgeo\_2196 and Dgeo\_2558-2660 fused protein) were determined to be 100% identical to strain S9 proteins. Thus, we detected the gene arrangement of the longest Contig 15, which was marked a yellow box in Spot 7 from GEvo analysis, of strain S9 in plasmid\_01 between *D. geothermalis* and *Deinococcus* strain S9 (Figure 6A and B). Most genes were conserved with over 95% amino acid sequence identity; however, unique IS elements and additional genes were integrated into the genome by following evolutionary tracks. In addition, of particular interest, the *IS701* family *ISDge12* of Dgeo\_2805 region and a partial *IS3* family were conserved in both genomes. Surprisingly, the genome of strain S9 had the most IS family members in all of the *Deinococcal* ISs that originated in habitats quite different from polar regions, deserts, soil, river, and mountain (Table S1). Defining transposition events of IS elements between 2 different habitat organisms, involving different evolutionary histories, is of particular interest. Therefore, we hope that the *Deinococcus* mobilome study findings will provide us with a better understanding of strain adaptation causing genome plasticity in a variety of environments.

#### Discussion

The current research environment provides free access to GenBank for bioinformation, such as DNA sequences and genetic contents, in the “Post era of Genomics.” Despite the





growing understanding of the big data in genome sequences and new technologies in genetic engineering, including gene editing, mobile genetic element (MGE) transposition has resulted in many miss-annotated genetic predictions and gene fragmentation. Bacterial genomes contain many TEs, including Tn and IS, therefore, using various genome-survey pipelines during MGE integration, the available genomic information, especially from bacteria, can detect many scars caused by gene fragmentation and mutations.<sup>8-10,28</sup> These TE contribute to gene activation or inactivation through structural gene disruption and promoter disorder, resulting in genetic variation through plasticity and evolutionary aspects of bacterial genomes.<sup>1,3,6</sup>

Case studies on the effects of TE on bacterial genomes have been published for a variety of bacteria, including *Escherichia coli*,<sup>29,30</sup> *Acinetobacter baumannii*,<sup>31,32</sup> *Bacillus*,<sup>33,34</sup> *Klebsiella pneumoniae*,<sup>32</sup> *Mycobacterium*,<sup>12,35,36</sup> *Lactococcus garvieae*,<sup>37</sup> *Burkholderia multivorans*,<sup>38</sup> *Pseudomonas aeruginosa*,<sup>11</sup> *Pseudomonas stutzeri*,<sup>39</sup> *Thermus* spp.,<sup>40</sup> *D. radiodurans*,<sup>14-16,41</sup> and *D. geothermalis*.<sup>17</sup>

Taking *L. garvieae* as an example, 77 published genomes from 12 strains reveal 15 types of IS elements and 4 families of ISs (IS3, IS982, IS6, and IS21). Interestingly, 2 strains (8831 and TRF1) contain no ISs within their genomic DNA.<sup>36</sup> For *P. stutzeri*, 12 types of ISs and their presence vary among available sequenced strains. For example, the NCTC10475 strain has no ISs. However, strain A1501 has 56 ISs.<sup>39</sup> A mobilome study of *Thermus* reported a significantly different number of IS copies per mega base pair genome size of 28 *Thermus* spp. In the 3 *T. thermophilus* strains HB27, HB8, and NAR1, many IS type elements distributed to rearrangement on genomes, including chromosomes and plasmids.<sup>40</sup> In particular, *ISTh7* type IS element was full length and present in its active form in all 3 *T. thermophilus* genomes.

Using the IS element finding pipelines, the total number of ISs can be calculated and the IS family members can be easily configured. Nevertheless, the composition and structure of IS elements must first be defined to characterize genome-plasticity studies properly. Researchers recently focused on bioinformatic analyses of TE integration among pathogenic bacteria better to understand their role in bacterial adaptation and evolution.<sup>12</sup> For this reason, we can realize genome variation and environmental adaptation repertoires much better (eg, antibiotic resistance, toxicity, pathogenicity through transpositional genomic plasticity).

The *D. geothermalis* genome contains a relatively higher number of IS copies compared with other species of genus *Deinococcus* (Table 2). Our recently reported data showed that removing the putative DNA binding protein, Dgeo\_0257, resulted in active transposition of the IS of *ISDge7* to another site by oxidative stress.<sup>17</sup> These transposition events were identified following the isolation of colorless colonies, where phytoene desaturase, a key enzyme for carotenoid biosynthesis, was disrupted by IS transposition. In addition, an *ISDge5* member

was also integrated into 2 genes (a Tase and a hypothetic protein). Deletion of the LysR family regulator gene Dgeo\_2840 resulted in the transposition of *ISDge6* type IS into carotenoid biosynthesis enzymes through replicative transposition in an oxidative-stress induced manner.<sup>18</sup> Although the genome of a strain of wild-type *D. geothermalis* already contains 10 copies of *ISDge5*, 5 copies of *ISDge6*, and 4 copies of *ISDge7*, active transposition of their ISs was caused by oxidative stress. In other words, hydrogen peroxide treatment under conditions without specific DNA-binding proteins, Dgeo\_0257 and Dgeo\_2840, resulted in further gene disruption caused by active transposition of distinct IS element (Table 2).<sup>17,18</sup> These active transposition events within *Deinococcus* genome were presumably caused by gene disruption and rearrangement, resulting in unexpected bacterial cell damage. These events can affect bio-engineered bacterial strains; for example, the systematic application of toxic materials bioremediation, including heavy metals and radionuclide waste of low-energy level and the production of certain substances, including amino acids, antibacterial substances, and metabolic intermediates. The *D. geothermalis* strain was chosen as a suitable bacterium for utilization due to its strong accumulating capacity of metal ions.<sup>41,42</sup> Therefore, it is necessary to determine whether a specific DNA-binding protein-dependent manner of IS type selection for transposition will occur. Based on this, if you can control IS transposition in usable industrial bacteria, you can hypothesize that the genetically engineered bacterial strains could be well maintained for application purposes. Many research groups have recently studied gene-editing technologies using the CRISPR-Cas systems; this application can also be applied to control the transposition of IS elements. For example, Nyerges et al<sup>43</sup> used the CRISPR-Cas control system in *E. coli* and achieved great success in the downregulation of transposition events. Our team will also challenge the protection effects of genome plasticity by the ISs on *D. geothermalis* as a genetically modified bacterial strain that may be used for industrial applications. Therefore, this type of approach is beneficial for maintaining engineered bacterial strains and understanding basic transposition studies, including investigating IS distribution and the active mechanism of IS transposition.

## Conclusion

Because the genomic content of *D. geothermalis* harbors genes encoding to reduce various metals and their toxicity, *D. geothermalis* strains have been applied to the bioremediation of toxic metals and radioactive wastes over the past 2 decades. However, their performance has not expanded significantly. We recently reported if transposable genetic elements, such as an IS, move to other sites, it causes gene inactivation when oxidative stress occurs in *D. geothermalis*.<sup>17,18</sup> Thus, this active transposition is a significant impediment to genetic engineering and the maintenance of its efficacy. This study highlights a simplified bioinformatic approach to eliminate ISs and demonstrates how

transposition affected genomic plasticity by recovering structural genes disrupted by ISs. In addition, this work helps us to better elucidate the behavioral modes of specific Tpsases in radiation-resistant bacteria, especially.

## Acknowledgements

S.-J.L. thanks Kyung Hee University's support.

## Author Contributions

CL, MKB, NC, SJL, and S-JL contributed to study design and data analysis; MKB and S-JL helped in writing the article. All authors reviewed and approved the final article.

## ORCID iD

Sung-Jae Lee  <https://orcid.org/0000-0002-3997-9020>

## Supplemental Material

Supplemental material for this article is available online.

## REFERENCES

- Vandecraen J, Chandler M, Aertsen A, Van Houdt R. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit Rev Microbiol.* 2017;43:709-730.
- Mahillon J, Chandler M. Insertion sequences. *Microbiol Mol Biol Rev.* 1998;62:725-774.
- Siguier P, Goubeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev.* 2014;38:865-891.
- Hickman AB, Dyda F. Mechanisms of DNA transposition. *Microbiol Spectr.* 2015;3:MDNA3-0034-2014.
- Guerillot R, Siguier P, Goubeyre E, Chandler M, Glaser P. The diversity of prokaryotic DDE transposases of the mutator superfamily, insertion specificity, and association with conjugation machineries. *Genome Biol Evol.* 2014;6:260-272.
- Siguier P, Goubeyre E, Varani A, Ton-Hoang B, Chandler M. Everyman's guide to bacterial insertion sequences. *Microbiol Spectr.* 2015;3:MDNA3-0030-2014.
- Chandler M, Fayet O, Rousseau P, Ton Hoang B, Duval-Valentin G. Copy-out-paste-in transposition of IS911: a major transposition pathway. *Microbiol Spectr.* 2015;3:MDNA3-0031-2014.
- Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 2006;34:D32-D36.
- Robinson DG, Lee M-C, Marx CJ. Oasis: an automatic program for global investigation of bacterial and archaeal insertion sequences. *Nucleic Acid Res.* 2012;40:e174.
- Biswas A, Gauthier DT, Ranjan D, Zubair M. ISQuest: finding insertion sequences in prokaryotic sequence fragment data. *Bioinformatics.* 2015;31:3406-3412.
- Al-Nayyef H, Guyeux C, Bahi JM. A pipeline for insertion sequence detection and study for bacterial genome. arXiv:1706.08267v1, 2017, <https://arxiv.org/pdf/1706.08267.pdf>.
- Durrant MG, Li MM, Siranosian BA, Montgomery SB, Bhatt AS. A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe.* 2020;27:140-153.
- Siguier P, Goubeyre E, Chandler M. Known knowns, known unknowns and unknown unknowns in prokaryotic transposition. *Curr Opin Microbiol.* 2017;38:171-180.
- Narumi I, Cherdchu K, Kitayama S, Watanabe H. The *Deinococcus radiodurans* *worA* gene: identification of mutation sites in two mitomycin-sensitive strains and the first discovery of insertion sequence element from deinobacteria. *Gene.* 1997;198:115-126.
- Mennecier S, Servant P, Coste G, Bailone A, Sommer S. Mutagenesis via IS transposition in *Deinococcus radiodurans*. *Mol Microbiol.* 2006;59:317-325.
- Pasternak C, Dulerio R, Ton-Hoang B, et al. ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB. *Mol Microbiol.* 2013;88:443-455.
- Lee C, Choi N, Bae MK, Choo K, Lee S-J. Transposition of insertion sequences was triggered by oxidative stress in radiation-resistant bacterium *Deinococcus geothermophilus*. *Microorganisms.* 2019;7:E446.
- Lee C, Choo K, Lee S-J. Active transposition of insertion sequences by oxidative stress in *Deinococcus geothermophilus*. *Front Microbiol.* 2020;11:558747.
- Oberto J. SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics.* 2013;14:4.
- Haug-Baltzell A, Stephenes SA, Davey S, Scheidegger CE, Lyons E. SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics.* 2017;33:2197-2198.
- Noe L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res.* 2005;33:w540-w543.
- Talwar C, Singh AK, Singh DN, et al. Draft genome sequence of *Deinococcus* sp. S9, isolated from microbial mat deposits of hot springs located atop the Himalaya ranges at Manikaran, India. *Microbiol Res Announc.* 2019;8:e00316-19.
- Harmer CJ, Hall RM. An analysis of the IS6/IS26 family of insertion sequences: is it a single family? *Microb Genom.* 2019;5:e000291.
- Varani A, He S, Siguier P, Ross K, Chandler M. The IS6 family, a clinically important group of insertion sequences including IS26. *Mob DNA.* 2021;12:11.
- He S, Corneloup A, Guynet C, et al. The IS200/IS605 family and "peel and paste" single-strand transposition mechanism. *Microbiol Spectr.* 2015;3:MDNA3-0039-2014.
- Blesa A, Sanchez M, Sacristan-Horrajada E, Fuente SG, Peiro R, Berenguer J. Into the *Thermus* mobilome: presence, diversity and recent activities of insertion sequences across *Thermus* spp. *Microorganisms.* 2019;7:25.
- Eisen JA, Heidelberg JF, White O, Salzberg SL. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* 2000;1:RESEARCH0011.
- Touchon M, Rocha EPC. Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol.* 2007;24:969-981.
- Sawyer SA, Dykhuizen DE, DuBose RF, et al. Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. *Genetics.* 1987;115:51-63.
- Sousa A, Bourgard C, Wahl LM, Gordo I. Rates of transposition in *Escherichia coli*. *Biol Lett.* 2018;9:20130838.
- Wright MS, Mountain S, Beerli K, Adams MD. Assessment of insertion sequence mobilization as an adaptive response to oxidative stress in *Acinetobacter baumannii* using IS-seq. *J Bacteriol.* 2017;199:e00833-16.
- Adams MD, Bishop B, Wright MS. Quantitative assessment of insertion sequence impact on bacterial genome architecture. *Microb Genom.* 2016;2:e000062. doi:10.1099/mgen.0.000062.
- Qiu N, He J, Wang Y, Cheng G, Li M, Yu Z. Prevalence and diversity of insertion sequences in the genome of *Bacillus thuringiensis* YBT-1520 and comparison with other *Bacillus cereus* group members. *FEMS Microbiol Lett.* 2010;310:9-16.
- Fayad N, Awad MK, Mahillon J. Diversity of *Bacillus cereus* sensu lato mobilome. *BMC Genomics.* 2019;20:436.
- Thabet S, Souissi N. Transposition mechanism, molecular characterization and evolution of IS6110, the specific evolutionary marker of *Mycobacterium tuberculosis* complex. *Mol Biol Rep.* 2017;44:25-34.
- Gonzalo-Asensio J, Perez I, Aguilo N, et al. New insights into the transposition mechanisms of IS6110 and its dynamic distribution between *Mycobacterium tuberculosis* complex lineages. *PLoS Genet.* 2018;14:e1007282.
- Eraclio G, Ricci G, Fortina MG. Insertion sequence elements in *Lactococcus garvieae*. *Gene.* 2015;555:291-296.
- Ohtsubo Y, Genka H, Komatsu H, Nagata Y, Tsuda M. High-temperature-induced transposition of insertion elements in *Burkholderia multivorans* ATCC17616. *Appl Environ Microbiol.* 2005;71:1822-1828.
- de Sousa LP. Mobile genetic elements in *Pseudomonas stutzeri*. *Curr Microbiol.* 2020;77:179-184.
- Pasternak C, Ton-Hoang B, Coste G, Bailone A, Chandler M, Sommer S. Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence. *PLoS Genet.* 2010;6:e1000799.
- Brim H, Venkateswaran A, Kostandarites HM, Fredrickson JK, Daly MJ. Engineering *Deinococcus geothermophilus* for bioremediation of high-temperature radioactive waste environments. *Appl Environ Microbiol.* 2003;69:4575-4582.
- Makarova KS, Omelchenko MV, Gaidamakova EK, et al. *Deinococcus geothermophilus*: the pool of extreme radiation resistance genes shrinks. *PLoS ONE.* 2007;2:e955.
- Nyerges Á, Bálint B, Cseklye J, Nagy I, Pál C, Fehér T. CRISPR-interference-based modulation of mobile genetic elements in bacteria. *Synth Biol (Oxford).* 2019;4:ysz008.