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. 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. The majority of Tpase in bacterial ISs contain the DDE motif. In general, the genome sequence of a TE contains more than 4500 ISs and they are categorized into 29 IS families. At present, there are more multiple ISs transpositions and accumulate Tpase gene duplications through the movement of TEs under various stress conditions. Several 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. An understanding of these modes of action has facilitated the development of computational identification pipelines for IS determination with distinct strategies; for example, conserved sequences of Tpase or short repeat sequences are determined by a cleavage site. 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. 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. 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. 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). However, there are many unknown transposition machineries with the unique structural scheme of ISs.

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). Based on the ISfinder platform, the genome of the radiation-resistant bacterium Deinococcus geothermalis contains 19 types of ISs and 73 ISs. For example, the ISDge2 type member of the IS1 family includes 15 copies within the genome.
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. In addition, active transposition of ISDge6 from the LysR family regulator disrupted mutant (Δdgeo_2840) and ISDge11 from the wild type by oxidative stress was first reported in our previous work. 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* DSM11300T. 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).

### 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 from DSM11300T, 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 DSM11300T. The genomic size, numbers of located Tpase 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. The *D. geothermalis* genome contains 73 full-length ISs clustered into 19 types and three ISDge types, ISDge2, ISDge3, and ISDge10, contain second Tpase-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 ORF sequence. The DNA sequence-converting work was performed by the nucleic acid sequence massager of Attotron Biotechnologies Corporation (http://www.mathdict.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.

| GENOMES | SIZE (MB)/PROTEINS | NO. OF TPASES/IS | ELIMINATED IS SIZE | RECOVERED ORF NUMBER |
|---------|--------------------|-----------------|-------------------|---------------------|
| Chromosome | 2.47/2310         | 33/30           | 34 627 bp (1.38%) | 8 (Dgeo_0307, 0806, 0926, 1674, 1806, 2109, 2196, 2207) |
| Plasmid pDGEO01 | 0.57/490         | 38/28           | 30 186 bp (5.26%) | 9 (Dgeo_2373, 2380, 2384, 2431, 2432, 2439, 2658, 2684, 2699) |
| Plasmid pDGEO02 | 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.

### Table 1. Genomic DNA features after elimination of ISs and recovered ORFs.
Synteny analysis among Deinococcus species

We used the SyntTax program (https://archaea.i2bc.paris-saclay.fr/SyntTax), SynMap2 (https://genomevolution.org/CoGe/SynMap.pl and https://genomevolution.org/CoGe/GEvo.pl), 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 DSM11300T 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. Deinococcus sp. S9 strain showed 99.06% identity of the 16S rRNA sequence to D. geothermalis.22

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 DSM11300T. The D. geothermalis genome includes 9 IS 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 Tpases 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.17,18 There are 6 single IS members (ISDge1, ISDge6, ISDge10, ISDge14, ISDge15, and ISDge17). Interestingly, all 7 IS copies of ISDge13 type are found in plasmid pDGE002. 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) Tpase amino acid length, its catalytic motif and the number of copies of Tpase 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 Tpase) 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 Tpase gene and constant DR sequences in 6 IS types (ISDge1, 6, 7, 8, 16, and 17). Scheme I has a single Tpase 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 Tpase gene and an overlapping right TIR sequence in the
Bioinformatics and Biology Insights

ORF region of the Tpase. Scheme II ISs (ISDge2 and ISDge3) have 2 Tpases-related genes which include IS1 family members composed of 148 aa-long Tpase A and 82 aa-long Tpase 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. 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.24

Based on these 4 clusterings of IS structural schemes containing DDE motif Tpase, consisting of two classic IS types (scheme I-II) and two unique IS types (scheme III-IV). ISs are typically composed of transposase (Tpase) 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.

Figure 2. Four structural schemes of ISs containing DDE motif Tpase, consisting of two classic IS types (scheme I-II) and two unique IS types (scheme III-IV). ISs are typically composed of transposase (Tpase) 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.
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 Tpase-related ORFs. ISDge18 and ISDge19 as both IS1341 group members contain a single Tpase B. In general, IS200/IS605 family member ISs lack conserved TIR sequences.

Table 2. Clustering of 4 structural schemes of ISs containing DDE motif Tpase 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.

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

**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 Tpase gene encoding region for a specific IS element was used in this work, which was helpful in identifying the primary Tpase genes (ie, those with the most copies of ISDge2 as 15 copies). Therefore, the “Dgeo_0430 region” indicates the full IS element length comprising Tpase 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 ISDge13 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 Tpase was eliminated, a portion of the gene fragment became connected into the adjacent ORF, which occurred in 19 ORFs (8 chromosome genes, 9 pDGE001 genes, 2 pDGE002 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\(^+\) symporter (MHS) family; 8 enzymes, such as a diguanylate cyclase, a 4-hydroxybenzoate-3-monoxygenase, 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 hypothetic 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, 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 ISDge2-1 region resulted in a combination of the disrupted proteins Dgeo_2586 and Dgeo_2588 into the full 448 aa-long IS66 family Tpase. Thus, the Dgeo_2586 + 2588 region as an ISDge4 type IS element was integrated into the C-terminal region of IS3/IS911 Tpase and IS66 Tpase of a genomic DNA locus and followed further integration by the ISDge2-1 type of Dgeo_2587 region. Therefore, the integration order is as follows: IS3/IS911 (Dgeo_2585)-IS66 (Dgeo_2589) > IS66 (ISDge4 of Dgeo_2586 + 2588 region) > IS1 (ISDge2-1 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.
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 pDGE001 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. The SynMap data analysis of fully assembled genomes showed the relationship of gene arrangement similarity between D. geothermalis and D. desertii (Figure S2A). The “X”-like pattern represents genome inversion, including insertion, deletion, and recombination.

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 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 DSM11300T 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
Figure 6. Synteny analysis of genomic DNA sequences between plasmid_01 of *D. geothermalis* and contig_15 of *Deinococcus* sp. strain S9 by YASS analysis (A) and comparison of detail gene arrangements (B). Contig_15 of *Deinococcus* strain S9 contains a yellow marked box area of Figure S2B Spot 7 on *D. geothermalis* genomes. There are 4 gaps which were contributed by the integration of IS elements and additional genes. Black arrows indicate the conserved ORFs and their transcriptional direction; blue color indicates *Deinococcus* strain S9 unique genes; orange color indicates *D. geothermalis* unique genes; red filled and red marked gene ID numbers indicate IS elements. The white–gray tone scales indicated amino acid sequence identity between 85% and 100%. Especially, * marked 100% identical proteins.

IS indicates insertion sequences.
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. These TEs 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.

Case studies on the effects of TE on bacterial genomes have been published for a variety of bacteria, including Escherichia coli, Acinetobacter baumannii, Bacillus, Klebsiella pneumoniae, Mycobacterium, Lactococcus garvieae, Burkholderia multivorans, Pseudomonas aeruginosa, Pseudomonas stutzeri, Thermus spp., D. radiodurans, and D. geothermalis.

Taking L. garvieae as an example, 77 published genomes from 12 strains reveal 15 types of IS elements and 4 families of ISs (IS3, IS4, IS6, and IS21). Interestingly, 2 strains (8831 and TRF1) contain no ISs within their genomic DNA. 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. A mobile 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. In particular, ISTb7 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. For this reason, we can realize genome variation and environmental adaptation repertoires much better (e.g., 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. 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 Tpase 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. 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). 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. 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. 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. 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 Tpases 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
1. 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.

2. Mahillon J, Chandler M. Insertion sequences. Microbiol Mol Biol Rev. 1998;62:725-774.

3. Sugier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014;38:865-891.

4. Hickman AB, Dyda F. Mechanisms of DNA transposition. Microbial Spectr. 2015;3:MDNA3-0014-2014.

5. Guerillot R, Siguier P, Gourbeyre 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.

6. Sugier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M. Everyman’s guide to bacterial insertion sequences. Microbial Spectr. 2015;3:MDNA3-0030-2014.

7. Chandler M, Fayet O, Rousseau P, Ton Hoang B, Duval-Valentin G. Copy-out “paste-in” transposition of IS911: a major transposition pathway. Microbial Spectr. 2015;3:MDNA3-0031-2014.

8. Sugier P, Perezbon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34:D32-D36.

9. Robinson DG, Lee M-C, Marx CJ. An automatic program for global investigation of bacterial and archaeal insertion sequences. Nucleic Acid Res. 2012;40,e74.

10. Biswas A, Gautheir DT, Ranjan D, Zubiak M. ISQuest: finding insertion sequences in prokaryotics sequence fragment data. Bioinformatics. 2015;31:3406-3412.

11. Al-Nayyer H, Gueyec 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.

12. Durrant MG, Li MM, Siranossian BA, Montgomery SB, Bhatt AS. A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. Cell Host Microbe. 2020;27:140-151.

13. Sugier P, Gourbeyre E, Chandler M. Known knowns, known unknowns and unknown unknowns in prokaryotic transposition. Curr Opin Microbiol. 2017;38:171-180.

14. Narumi I, Cherdchu K, Kitayama S, Watanabe H. The Deinococcus radiodurans uveo gene: identification of mutation sites in two mitomycin-sensitive strains and the first discovery of selection insertion element from deinobacteria. Gene. 1997;198:115-126.

15. Menneceier S, Servant P, Coste G, Bailone A, Sommer S. Mutagenesis via IS transposition in Deinococcus radiodurans. Mol Microbiol. 2006;59:317-325.

16. Pasternak C, Dulemro R, Ton-Hoang B, et al. ISDns2 transposition in Deinococcus radiodurans is downregulated by TnpB. Mol Microbiol. 2013;88:443-455.

17. 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 geothermalis. Microorganisms. 2019;7:E446.

18. Lee C, Choo K, Lee S-J. Active transposition of insertion sequences by oxidative stress in Deinococcus geothermalis. Front Microbiol. 2020;11:558747.

19. Oberto J. SynTxs: a web server linking synteny to prokaryotic taxonomy. BMC Bioinformatics. 2013;14:44.

20. Haug-Baltzell A, Stephens SA, Davey S, Scheidegger CE, Lyons E. SyncMap2 and SyncMap3D: web-based whole-genome synteny browsers. Bioinformatics. 2017;33:2197-2198.

21. Noe L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Res. 2005;33:w540-w543.

22. Talwar C, Singh AK, Singh DN, et al. Draft genome sequence of Deinococcus sp. IS6110 isolated from microbial mat deposits of hot springs located atop the Himalaya ranges at Manikaran, India. Microbiol Res Annu. 2019;5:e00316-19.

23. Harmer CJ, Hall RM. An analysis of the IS66/IS26 family of insertion sequences: is it a single family? Microbiol. 2019;5:e00291.

24. Varani A, He S, Sugier P, Ross K, Chandler M. The IS family, a scientifically complex single-strand transposition mechanism. Microbial Spectr. 2015;3:MDNA3-0039-2014.

25. Blesa A, Sanchez M, Sacristan-Horcasida 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.

26. Eisen JA, Heidelberg JF, White O, Salzberg SL. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. 2000;1:RESEARCH0101.

27. Troupcon M, Rocha EPC. Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol. 2007;24:969-981.

28. Sawyer SA, Dykhuisen DE, DuBois RP, et al. Distribution and abundance of insertion sequences among natural isolates of Escherichia coli. Genetica. 1987;115:51-63.

29. Sousa A, Bourgard C, Wahl LM, Gordo I. Rates of transposition in Escherichia coli. Mol Biol Lett. 2018;29:20130838.

30. Wright MS, Mountain S, Beerli K, Adams MD. Assessment of insertion sequence mobilization as an adaptive response to oxidative stress in Azotobacter chroococcum using IS-seq. J Bacteriol. 2017;199:e00833-16.

31. Oberto J. SyntTax: a web server linking synteny to prokaryotic taxonomy. BMC Genomics. 2019;20:436.

32. 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.

33. Gonzalo-Asensio J, Perez I, Aguilo N, et al. IS903: a transposition mechanism of IS610 and its dynamic distribution between Mycobacterium tuberculosis complex lineages. PLoS Genet. 2018;14:e1007282.

34. Erazco G, Racci G, Fortina MG. Insertion sequence elements in Lactobacillus garvieae. Gen. 2015;555:291-296.

35. Ohtsubo Y, Genka H, Komatsu H, Nagata Y, Tsuda M. High-temperature induced transposition of insertion elements in Baeckhildia multivorans ATCC17616. Appl Environ Microbiol. 2005;71:1822-1828.

36. de Sousa LP. Mobile genetic elements in Peudomonas stutzeri. Curr Microbiol. 2020;77:179-184.

37. 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.

38. Beim H, Venkateswaran A, Kostandarithes HM, Fredrickson JK, Daly MJ. Engineering Deinococcus geothermalis for bioremediation of high-temperature radioactive waste environments. Appl Environ Microbiol. 2003;69:4575-4582.

39. Makarova KS, Omelchenko MV, Gaidamakova EK, et al. Deinococcus geothermalis: the pool of extreme radiation resistance genes shrinks. PLoS ONE. 2010;5:e975.

40. 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:yya008. 