# Mono-homologous linear DNA recombination by the non-homologous end-joining pathway as a novel and simple gene inactivation method: a proof of concept study in *Dietzia* sp. DQ12-45-1b

Running title: a novel and simple gene inactivation method

Shelian Lu[1,#], Yong Nie[1,#], Meng Wang[1], Hong-Xiu Xu[2,1], Dong-Ling Ma[1], Jie-Liang Liang[1], Xiao-Lei Wu[1,*]

[1]College of Engineering, Peking University, Beijing 100871, P. R. China

[2]College of Architecture and Environment, Sichuan University, Chengdu 610065, P. R. China

[#] Contribute equally[1]

* To whom correspondence should be addressed. Tel/Fax: +86-10-62759047; Email: xiaolei_wu@pku.edu.cn; Address: Department of Energy and Resources Engineering, College of Engineering, Peking University, Beijing 100871, P. R. China.

[1] The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

## ABSTRACT

Non-homologous end-joining (NHEJ) is critical for genome stability because of its roles in double-strand break repair. Ku and ligase D (LigD) are the crucial proteins in this process, and strains expressing Ku and LigD can cyclize linear DNA *in vivo*. Herein, we established a proof-of-concept mono-homologous linear DNA recombination for gene inactivation or genome editing by which cyclization of linear DNA *in vivo* by NHEJ could be used to generate non-replicable circular DNA and could allow allelic exchanges between the circular DNA and the chromosome. We achieved this approach in *Dietzia* sp. DQ12-45-1b, which expresses Ku and LigD homologs and presents NHEJ activity. By transforming the strain with a linear DNA mono homolog to the sequence in chromosome, we mutated the genome. This method did not require the screening of suitable plasmids and was easy and time-effective. Bioinformatic analysis showed that more than 20% prokaryotic organisms contain Ku and LigD, suggesting the wide distribution of NHEJ activities. Moreover, the *Escherichia coli* strain also showed NHEJ activity when the Ku and LigD of *Dietzia* sp. DQ12-45-1b were introduced and expressed in it. Therefore, this method may be a widely applicable genome editing tool for diverse prokaryotic organisms, especially for non-model microorganisms.

**KEYWORDS:** mono-homologous linear DNA, non-homologous end-joining, gene inactivation

35 **IMPORTANCE**

36 The non-model gram-positive bacteria lack efficient genetic manipulation systems, but they
37 express genes encoding Ku and LigD. The NHEJ pathway in *Dietzia* sp. DQ12-45-1b was
38 evaluated and was used to successfully knockout eleven genes in the genome. Since bioinformatic
39 studies revealed that the putative genes encoding Ku and LigD ubiquitously exist in
40 phylogenetically diverse bacteria and archaea, the mono-homologous linear DNA recombination
41 by the NHEJ pathway could be a potentially applicable genetic manipulation method for diverse
42 non-model prokaryotic organisms.

43

## INTRODUCTION

Bacterial genome editing such as deletion, mutation, and insertion of genes is an efficient method to understand gene functions and modify metabolic activities. The most often used gene recombination methods in bacteria include the bacteriophage recombination system using the linear double-strand DNA (dsDNA) as recombination substrate (1-3), the single or double crossover homologous recombination (HR) using the circular plasmid DNA as substrate (4, 5), and the transposase-mediated transposition recombination (6, 7). Recently, several genome editing technologies have emerged, which are mediated by the targeted nucleases, including zinc finger nucleases (ZFNs) (8), transcription activator–like effector nucleases (TALENs) (9), and short palindromic repeats (CRISPR)-associated Cas9 endonuclease (10). All these methods generally require suitable plasmids and a laborious process such as cloning of genes into plasmids. Moreover, it is nearly impossible to find compatible plasmids for diverse prokaryotic organisms in nature, especially for the non-model microorganisms. Therefore, developing efficient and suitable genetic manipulation methods is still of great necessity.

Generally, a genetic manipulation method is developed according to a natural biological process. For example, the HR method was developed following a natural HR process to repair DNA double-strand break (DSB) (11), which evolves in all cellular organisms (12). Another DSB repairing process is the non-homologous end joining (NHEJ) process, which was first discovered in mammalian cells (13). During the eukaryotic NHEJ process, the ends of broken DNA are approximated by the DNA-end-binding protein Ku (Yku70/Yku80 in yeast) and then joined by an ATP-dependent DNA ligase IV/XRCC4/XLF (LXX) complex (LigD, Dnl4/Lif1/Nej1 in yeast) (13).

Consequently, a homologous DNA template is not needed (13, 14). Recently, NHEJ was also discovered in prokaryotes with the evidence of the circulation of a linear plasmid DNA in Ku and LigD containing mycobacteria (15, 16) and the circulation of a linear DNA in *Escherichia coli* expressing mycobacterial Ku and LigD (17). In addition, the Ku homodimer and LigD are crucial proteins in the bacterial NHEJ process (18-20), whose presence could indicate NHEJ activities in bacteria (21-23). NHEJ also offer protection to bacteria when only a single copy of the genome is available such as after sporulation or during stationary phase under environmental stresses (24, 25).

Similar to the HR DSB repairing process, the NHEJ process may also be used for genetic manipulation. In this case, a target homologous sequence can be amplified, PCR fused with a selective marker to form a linear DNA fragment and then transformed into the recipient bacterium, which expresses Ku and LigD. If the linear DNA can be cyclized by the NHEJ pathway *in vivo*, the resultant circular DNA may act as a circular incompatible plasmid in bacterial genetic manipulation (26). Because the linear/circular DNA without replication region cannot be replicated, the transformed cells surviving against selective stresses such as antibiotic should be mutants with insertion in the target sequences. Consequently, the target gene can be knocked out by using the NHEJ pathway.

Herein, we proposed a novel and much simpler gene knockout method by using the NHEJ pathway in *Dietzia* sp. DQ12-45-1b. *Dietzia* strains found in diverse environments are powerful *n*-alkane degraders and potential pathogens. They can utilize a wide range of compounds as the sole carbon source such as hydrocarbons and aniline (27-29). Moreover, a number of *Dietzia* strains can be used as potential probiotics to inhibit fecal *Mycobacterium avium* subspecies

86 *paratuberculosis in vitro* (30) and might be useful for the treatment of patients with Crohn's disease
87 (31, 32). They are non-model gram-positive bacteria and lack efficient genetic manipulation
88 systems, but they express genes encoding Ku and LigD. The NHEJ pathway in *Dietzia* sp. DQ12-
89 45-1b was evaluated and was used to successfully knockout eleven genes in the genome. Since
90 bioinformatic studies revealed that the putative genes encoding Ku and LigD ubiquitously exist in
91 phylogenetically diverse bacteria and archaea, the mono-homologous linear DNA recombination
92 by the NHEJ pathway could be a potentially applicable genetic manipulation method for diverse
93 non-model prokaryotic organisms.
94

95 ## MATERIALS AND METHODS

96 ### Bacterial strains, plasmids, media, and primers

97 The bacterial strains and plasmids used in this study are listed in Table 1. All primers are listed in
98 Table S1. *Dietzia* sp. DQ12-45-1b, isolated from oil-production water, has been intensively studied
99 in our laboratory (28, 33-38). *Dietzia* sp. DQ12-45-1b was cultured in GPY medium (1% [wt/vol]
100 glucose, 1% [wt/vol] tryptone, 0.5% [wt/vol] yeast extract) at 30°C on a rotary shaker at 150 rpm.
101 The alkane hydroxylase-rubredoxin fusion gene (*alkW1*) (34) mutant of *Dietzia* sp., DQ12-45-1b
102 (*alkW1*$^-$), was grown in GPY medium with 40 mg/liter kanamycin. To determine their functions in
103 alkane degradation, strains *alkW1*$^-$ and *Dietzia* sp. DQ12-45-1b were grown in minimal medium
104 (4) supplemented with 0.1% (vol/vol) hexadecane as the sole carbon source at 30°C, as described
105 previously (34). *E. coli* DH5α was grown in lysogeny broth (LB) medium at 37°C. *E. coli* strains
106 and *Dietzia* strains harboring plasmids were grown with appropriate antibiotics (ampicillin, 100
107 μg/ml; kanamycin, 30 μg/ml; streptomycin, 30 μg/ml).

108 The vector pNV18-PEK (Table 1) was constructed as follows. Firstly, the promoter p45 DNA
109 fragment was amplified from plasmid pNV18-Dsred (44) with primer p45-EF and p45-BR, *egfp*
110 gene from pK18-egfp (unpublished result) with primer egfp-BF and egfp-HR, and kanamycin-
111 resistant gene (Km) from plasmid pK18 (39) with primer Km-HF and Km-PR. Then the promoter
112 p45 was clone into pXL1801 (40) with EcoRI and BamHI restriction sites, *egfp* with BamHI and
113 HindIII restriction sites and Km with HindIII and PstI restriction sites sequentially, obtaining the
114 plasmid pNV18-PEK. The plasmid pJV53-PA was also constructed. Briefly, a large fragment of
115 vector pJV53 was first amplified with primer 53-HF and 53-ER (Table S1). The resulting 3,013 bp

linear fragment lacking the acetamidase promoter and Chec9c 60-61 was digested with EcoRI and HindIII. The promoter p45 was amplified from pNV18-PEK with primer p45-EF and p45-(amp)-R (Table S1), and the ampicillin resistant gene was amplified from pUC19 with primers amp-(p45)-F and amp-HR (Table S1). The p45 and ampicillin resistant gene was fused by PCR with primer p45-EF and amp-HR, resulting in PA DNA fragment. Then PA was ligated into linear pJV53 by the EcoRI and HindIII restriction sites to obtain the plasmid pJV53-PA.

**Sequence analysis**

Genes encoding the Ku and LigD homologs in the genome of *Dietzia* sp. DQ12-45-1b, designated as Dt-Ku and Dt-LigD, were identified by comparing the proteome of strain DQ12-45-1b (unpublished result) against the non-redundant (NR) database of protein sequences at the National Center for Biotechnology Information (NCBI) using BLASTP (41). The conserved domains of proteins were determined by comparing the protein sequences against the conserved domain database (CDD) at NCBI (42). To identify the Ku and LigD homologous genes in prokaryotic organisms, we searched the KEGG Orthology (KO) (43) for ID K10979 and K01971, which indicated Ku and LigD, respectively, in the Integrated Microbial Genomes (IMG) system (44) against a total of 25,270 bacterial genomes and 528 archaeal genomes (until April 2015). Sequence alignment and phylogenetic analysis were performed in MEGA6 (45).

**NHEJ assay in *Dietzia* sp. DQ12-45-1b**

To identify the circular efficiency of different linear DNA ends, three types linear DNA was generated. The linear DNA fragment with 5′-“A” tails (LA18) was amplified from 4 ng plasmid

136 pNV18-Sm (46) using LA Taq (Takara, Tokyo, Japan) under the conditions suggested by the
137 manufacturer with the primers pNV18-F and pNV18-R, followed by treatment with DpnI digestion
138 (Takara, Tokyo, Japan) for 4 h at 37°C to eliminate the circular plasmid pNV18-Sm template.
139 Secondly, the plasmid pNV18-Sm was digested by HindIII to produce 5′-overhang DSB fragments
140 HD18, and by HindIII and BamHI simultaneously to produce 5′-overhang DSB fragments HB18
141 overnight, respectively (Fig. 1). All three linear DNA fragments contained intact streptomycin-
142 resistant gene (Fig. S1). Gel purification was performed to remove the undigested plasmids and
143 collect the linear plasmid only with DNA purification kit (Tiangen biotech, Beijing, China). Five
144 microliter purified linear DNA fragments, LA18 (108 ng), HD18 (437 ng), and HB18 (188 ng),
145 were then transformed into 100 μL competent cells of strain DQ12-45-1b (~$10^{10}$ CFU/ml and the
146 transformation efficiency was $10^6$ CFU/μg DNA), respectively. After transformation, cells were
147 recovered at 30℃ for 4 h, followed by plating on LB agar containing streptomycin. Positive clones
148 were numbered and picked after 3 days incubation. The method of competent cell preparation and
149 electro-transformation was as previously described (33).

150 The resultant streptomycin-resistant colonies on LB agar were then suspended in 10 μL double
151 distilled water and incubated at 98°C for 30 min to release DNA from the cells, followed by
152 centrifugation at 20,000 ×*g* for 5 min. The DNA-containing supernatant was then subjected to PCR
153 using r*Taq* (Takara, Tokyo, Japan) for 26 cycles with the primers IdenF and IdenR to amplify the
154 joint region of the circular DNA. The PCR product was purified, cloned into pGEM-T (Promega,
155 Madison, WI, USA). Recombinant plasmids were transformed into *E. coli* DH5α for white-blue

156 screening, and the white colonies were sequenced to assess NHEJ activities.

157 To determine the role of *ku* gene in *Dietzia* sp. DQ12-45-1b, we constructed a *ku* inserted
158 mutant with streptomycin resistance gene using a double-homologous-recombination method as
159 described previously (Fig. S2) (47-49) and the primers used were listed in Table S1. The Ku
160 expression vector for complementation studies were constructed as follows: the *ku* gene was
161 amplified from *Dietzia* sp. DQ12-45-1b chromosomal DNA with the primers Ku-BF and Ku-HR
162 (Table S1), and then the 1046 bp DNA fragment was ligated into pNV18-PEK using BamHI and
163 HindIII restriction sites to construct a new plasmid, pNV18-PKK (Table 1). Vector pNV18-PKK
164 was introduced into *ku* mutant cells via electroporation as described previously to obtain the
165 complementary mutant cell (33). The effect of *ku* gene on the circular function of linear DNA in
166 *Dietzia* sp. DQ12-45-1b were identified on the *ku*-complementary mutant cell using the linear
167 plasmid DNA pJV53-PA (L53) with a p45 promoter and an ampicillin resistance gene (Fig. S3),
168 which was amplified from plasmid pJV53-PA with the primers p45-F and 53-ER (Table S1)
169 followed by being processed with DpnI digestion (Takara, Tokyo, Japan) for 4 h at 37°C to
170 eliminate the circular plasmid pNV18-Sm template before being transformed into cells as described
171 above, by taking the treatments operated on the WT cells and *ku* mutant cells transformed with
172 pNV18-PEK as the negative controls (Table 1). After the 3-day incubation, ampicillin resistance
173 colonies were counted as described above.

174 **Expression of Dt-Ku and Dt-LigD and NHEJ assay in *E. coli***

175 The genes coding for Dt-Ku protein and Dt-LigD were amplified from the genome of strain DQ12-

45-1b using PrimeSTAR HS DNA Polymerase (Takara, Tokyo, Japan) according to the manufacturer's instructions with the primers, KuF and KuR for Dt-Ku and LigDF and LigDR for Dt-LigD. The PCR products were purified and digested with the KpnI and EcoRI for Dt-Ku and with HindIII and KpnI for Dt-LigD. The plasmid pUC19-Ku and pUC19-LigD were then constructed by cloning *ku* and *ligD* DNA sequences into the corresponding sites of the plasmid pUC19. Similarly, the plasmid pUC19-Ku-LigD was constructed by cloning these two fragments into the HindIII and EcoRI sites of the plasmid pUC19, using standard methods (50), and then transformed into *E. coli* DH5α to construct the strain expressing single Dt-Ku or Dt-LigD and both of them, respectively.

To assess Dt-LigD and Dt-Ku activities, a heterogeneous expression test was performed. A linear DNA fragment P28 (3,818 bp, containing the pBR322 origin and kanamycin-resistant gene) was amplified from pET-28a(+) (Qiagen, Hilden, Germany) with the primers pET-28F and pET-28R. Specifically, 4 ng pET-28a(+) was used as template for PCR with PrimeSTAR HS DNA Polymerase (Takara, Tokyo, Japan) according to the manufacturer's instructions. Then the PCR product was treated with DpnI digestion (Takara, Tokyo, Japan) for 4 h at 37°C to degrade the circular plasmid DNA and only the 3.8 kb band was collected for further purification with DNA purification kit (Tiangen biotech, Beijing, China). The pUC19, pUC19-Ku, PUC19-LigD and pUC19-Ku-LigD (100 ng) were then transformed by 42 °C heat shock for 90 s into 100 μL *E.coli* DH5α. After transformation, 900 μL of room temperature SOC medium was added to the tubes, and the cells were allowed to recover at 37°C with shaking at 300 rpm for one hour. 50 μL cells were cultured on LB agar containing 100 μg/ml ampicillin. The bacteria expressing no target

197 protein or expressing Ku, LigD and Ku-LigD were prepared competent cells and then electro-
198 transformed fragment P28 (500 ng) according to the standard instructions (50). All the cells were
199 then collected by centrifugation (1,500 ×*g*, 5 min) and cultured on LB agar containing 100 μg/ml
200 ampicillin and 40 μg/ml kanamycin. The transformation of pUC19 and P28 into *E. coli* DH5α was
201 used as a negative control. The kanamycin resistant colonies indicated that the P28 was cyclized,
202 allowing the expression of the kanamycin-resistant gene. Plasmids were then extracted from the
203 positive colonies and digested with NdeI, followed by electrophoresis to determine the lengths of
204 digested fragments.

205 **Gene disruption in *Dietzia* sp. DQ12-45-1b using mono-homologous arm linear DNA**
206 **fragment**

207 We used the *alkW1* gene as an example to determine whether NHEJ could be used for target
208 gene manipulation. First, the homologous DNA fragment (from 143$^{rd}$ to 552$^{nd}$*nt* of *alkW1* gene),
209 was amplified from the genome of strain DQ12-45-1b using primers 143-F and 552-R. For easy
210 identification, a triplet cytidine (CCC) tag was additionally designed at the 5′ end of primer 552-
211 R. Secondly, the kanamycin resistance cassette P45-EGFP-Km with a P45 promoter and an *egfp*
212 gene as the reporter was amplified from the plasmid pNV18-PEK using the primer P45-F and Km-
213 R. Finally, the homologous DNA and P45-EGFP-Km cassette were fused and amplified by fusion
214 PCR (51) using the primer P45-F and 552-R. The generated fragment (alk-Km) with homologous
215 DNA of *alkW1* and kanamycin resistance cassette was purified and 5 μl of the purified product
216 (500 ng) transformed into *Dietzia* sp. DQ12-45-1b as previously described (33). Cells were then

grown on LB plate with kanamycin for 4 days. The kanamycin-resistant colonies were subjected to PCR amplification of gene *alkW1* with the primers alkW1-F and alkW1-R, which was further verified by sequencing. The recombinant efficiency was expressed as the number recombinants per microgram DNA divided by the cell competency which was indicated by the transformation efficiency with plasmid pNV18-Sm. The expression of AlkW1 in the cells cultured in liquid GPY medium or minimal medium with hexadecane as the sole carbon source was detected by western blot analysis as described previously (40).

To verify whether this method could be used for other genes, we selected seven genes encoding histidine kinase of two-component system and two other genes (Table 2) for disruption. The homologous linear DNA fragments were constructed and transformed as described above. The kanamycin-resistant colonies were selected and verified by PCR, fluorescence microscopic analysis and sequencing.

**Nucleotide sequence accession number**

The GenBank accession numbers of Dt-Ku and Dt-LigD in *Dietzia* sp. DQ12-45-1b were KP074897 and KP074898, respectively. The GenBank accession numbers of the seven histidine kinases, putative phosphodiesterase, and cobyrinic acid a,c-diamide synthase were listed in Table 2.

## RESULTS

### NHEJ activity in *Dietzia* sp. DQ12-45-1b and *E. coli* expressing Ku and LigD

Genes encoding Ku and LigD homologs, Dt-Ku and Dt-LigD, were identified by screening the genome of *Dietzia* sp. DQ12-45-1b. These homologs showed 57% and 44% amino acid identities with Mt-Ku and Mt-Lig in *Mycobacterium* (18-20), respectively. To evaluate their NHEJ activities in strain DQ12-45-1b *in vivo*, the linear DNA fragments, LA18 with short overhang ends, HD18 with long overhang complementary ends, and HB18 with long overhang non-complementary ends (Fig. 1), were transformed into the strain DQ12-45-1b. In addition, negative controls were used in which the LA18, HD18, and HB18 fragments were transformed into *E. coli* DH5α that did not express components of the NHEJ pathway. No streptomycin-resistant colonies were detected in the negative controls, while transformation of LA18, HD18, and HB18 fragments into strain DQ12-45-1b generated 2130, 359, and 229 streptomycin-resistant colonies per microgram of linear DNA, respectively, indicating that the streptomycin-resistant gene originally embedded in the linear DNA fragments was functional in strain DQ12-45-1b (Fig. S4). Since the streptomycin gene containing DNA could be replicated only if it was in circular form, these results suggested that the linear DNA fragments were successfully cyclized in the strain, indicating that the NHEJ pathway was active *in vivo* in strain DQ12-45-1b. Among the three linear DNA fragments, it was more difficult for linear DNA with longer 5′-overhang ends to cyclize.

In addition, the three linear DNA fragments were cyclized differently by NHEJ activity. In case of LA18 with short overhang ends, 61 positive colonies were sequenced and 7 different NHEJ joints were detected (Fig. 1A). The overhung “A” at the ends was deleted before ligation in all

256 colonies, suggesting the presence of a 5′-3′ exonuclease activity in strain DQ12-45-1b. Non-
257 templated insertions and sequence deletions were detected in 38 and 23 out of the 61 colonies,
258 respectively. Among the 38 colonies (62% of the 61 colonies) with insertion, three types of non-
259 templated insertion sequences were identified. Two of them were detected in 25 colonies composed
260 of G+C nucleotides (Fig. 1A). Among the 23 colonies with sequence deletions, 12 colonies
261 contained short-scale (9 bp) deletions and the remaining 11 colonies contained 175–308 bp long
262 deletions. It is notable that the deletions, being short or long, happened at the ends of the side
263 without "CCC" tag, suggesting that the repeated G+C sequence might protect the linear DNA from
264 exonuclease activity. In contrast, for HD18 and HB18 with long overhang ends, long unidirectional
265 deletions up to 148 bp, were detected at both ends (Fig. 1B and 1C). The results further confirmed
266 the DNA polymerase and exonuclease activities in the NHEJ process in strain DQ12-45-1b (15).
267 To evaluate the NHEJ activities of Ku and LigD in strain DQ12-45-1b, *ku* mutant strain was
268 constructed. It was notable that *ligD* mutant was failed to be obtained even though two different
269 methods had been tried to disrupt ligD gene from the strain (data not shown), The result suggested
270 that *ligD* might be an essential gene for the strain DQ12-45-1b. The linear DNA L53 with blunt
271 ends was transformed into wild type, *ku* gene mutant, and the mutant with *ku* complementation to
272 observe the NHEJ activities. No ampicillin-resistant colonies were obtained from the *ku* mutant
273 cells, while there were 463 and 351 ampicillin-resistant colonies per microgram of linear DNA in
274 the wild type cells and the mutant with *ku* complementation (Fig. S5), indicating that the linear
275 DNA L53 were cyclized in the cells. In addition, the L53 fragment was also transformed into *E.*
276 *coli* DH5α, and no ampicillin-resistant colonies were detected, suggesting that there was no circular

277 plasmid contaminant.

278 To further confirm the cyclization capability of Dt-Ku and Dt-LigD, we transformed pUC19,
279 pUC19-Ku, pUC19-LigD and pUC19-Ku-LigD firstly into *E. coli* DH5α following electro-
280 transformation P28 and successfully recovered colonies in presence of kanamycin, while no
281 kanamycin-resistant colony was detected in the bacteria including pUC19, pUC19-Ku and pUC19-
282 LigD. From the kanamycin-resistant colonies, plasmids were extracted and digested by NdeI to
283 further verify the circulation results. Three bands with the expected sizes of about 1.2 kb, 5.2 kb,
284 and 3.8 kb were observed, corresponding to the digestion products of pUC19-Ku-LigD and
285 cyclized P28 (Fig. S6), respectively. These results supported that Dt-Ku and Dt-LigD could
286 function in *E. coli* to cyclize the linear DNA and both activities were needed. The electro-
287 transformation of *E.coli* was $2.4 \times 10^4$ CFU/μg DNA. The cyclization efficiency was calculated as
288 228 CFU/μg DNA in *E. coli* DH5α expressing Dt-Ku and Dt-LigD. The ratio of Amp/Kan resistant
289 colonies was about 100:1. All the results were performed in triplicates.

290 **Recombination of linear DNA in *Dietzia* sp. DQ12-45-1b via NHEJ pathway**

291 Since linear DNA could be cyclized in strain DQ12-45-1b by NHEJ activity *in vivo*, we
292 assumed that it could act as a non-replicable plasmid that could further be used to inactivate a target
293 gene by homologous recombination. To verify this hypothesis, we constructed a linear alk-Km
294 DNA fragment containing a mono homologous arm DNA fragment of the *alkW1* gene and the
295 kanamycin resistance cassette. After transforming 200 ng of this linear alk-Km DNA fragment into
296 *Dietzia* sp. DQ12-45-1b, colonies were recovered in the presence of kanamycin. Among them,
297 eight colonies were randomly selected and subjected to PCR amplification with primers alkW1-F

298 and alkW1-R to amplify the *alkW1* gene. DNA fragments with the expected size (about 3.6 kb)
299 were amplified from four out of the eight colonies, indicating that half of the colonies presented
300 the correct insertion into the *alkW1* gene (Fig. 2A). Sequencing of these DNA fragments further
301 confirmed the successful insertion of linear DNA into the *alkW1* gene in the genome (Fig. 2B).
302 The recombinant efficiency was approximately $3\times10^{-6}$.

303 To further verify the gene insertion, we carried out western blotting with the proteomes of
304 *Dietzia* sp. DQ12-45-1b and *alkW1*$^{-}$, which were grown in GPY medium or mineral medium with
305 hexadecane as the sole carbon source. AlkW1 was induced in *Dietzia* sp. DQ12-45-1b by
306 hexadecane, but not detected in the *alkW1*$^{-}$ mutant strain either cultured in GPY or with hexadecane
307 (Fig. 3A). As expected, the growth of the *alkW1*$^{-}$ mutant strain in hexadecane as the sole carbon
308 source was inhibited within eight days of incubation (Fig. 3B). These results proved that *alkW1*
309 gene was successfully inactivated by the mono-homologous arm DNA fragment integration. Using
310 the mono-homologous linear DNA recombination, we also successfully disrupt nine genes of
311 *Dietzia* sp. DQ12-45-1b (Table 2), as confirmed by the PCR analysis (Fig. S7), the presence of
312 green fluorescence (Fig. S8), and sequencing (Supplementary file 1). The recombinant efficiencies
313 were approximately about $1.2 \times 10^{-6}$- $6.5 \times 10^{-5}$ (Table 2). Besides, the successful inactivation of
314 another gene *alkX*, the regulator of *alkW1* using this method (38), suggested the reliability of this
315 method. In addition, the mutant was continuously cultured for 30 re-inoculations in GPY medium
316 without antibiotics with no change in its genotype (data not shown), which demonstrated that the
317 insertion was stable and heritable.

318 **Distribution of Ku and LigD in prokaryotic organisms**

By searching the *ku* and *ligD* homologous genes in the available 25,270 bacterial and 528 archaeal genomes using KO annotation, we identified 6,118 *ku* homologous genes in 5,098 bacterial and 13 archaeal genomes belonging to 14 bacterial phyla and Euryarchaeota. We also identified 18,952 *ligD* homologous genes in 7,631 bacterial and 64 archaeal genomes belonging to 26 bacterial phyla and 6 archaeal phyla (Fig. 4). Among them, 4,783 bacterial and 13 archaeal genomes contained both genes. Strikingly, about 76% and 77% of the total Actinobacteria genomes analyzed contained *ku* and *ligD* homologous genes, respectively, indicating that NHEJ might be common in Actinobacteria. Phylogenetic analysis showed that *ku* genes could be clustered into six clusters (Fig. 5), largely matching their taxonomic classification, suggesting the conservative property of the *ku* genes. In contrast, the *ligD* genes could be clustered into five clusters and their phylogenetic tree topology was much different from their taxonomic classification, suggesting possible vigorous gene transfer of *ligD* genes among microorganisms (Fig. 5).

The *ku* genes were generally juxtaposed with *ligD* genes (Fig. 6A), for which a functional association between Ku and LigD was suggested (21). In contrast, the Dt-Ku and Dt-LigD coding genes were not adjacent in *Dietzia* sp. DQ12-45-1b, but within 9.8 kb to each other (Fig. 6A). Similar gene arrangements were detected in other *Dietzia* species such as *Dietzia alimentaria*72, *Dietzia cinnamea* P4, and *Dietzia* sp. UCD-THP (Fig. S9A), whose Ku shared 72–81% amino acid identity with Dt-Ku. All the Ku protein sequences retrieved had a Ku core functional domain, which could bind DNA ends and transiently bring them together (Fig. 6B), suggesting the Ku functions in these strains (21). All the LigD sequences presented an ATP-dependent DNA ligase domain (Fig. 6C), which was essential in NHEJ (13, 14). The Ku core and ATP-dependent DNA

ligase domain found in all Ku and LigD proteins suggested the intact NHEJ functions in all these strains containing both *ku* and *ligD*. Besides, Dt-LigD and its homologs from other *Dietzia* species contained a polymerase domain and a phosphoesterase domain from the N-terminal to the C-terminal, which was similar to that of the homolog Mt-LigD in *Mycobacterium* (Fig. 6C and Fig. S9B).

## DISCUSSION

As the major DSB repair pathway, HR has been extensively studied in *E. coli* and widely used for gene manipulation. This classical genetic engineering method uses incompatible or suicide plasmids and requires a time-consuming process. A phage-mediated HR, usually referred to as "recombineering", constructs linear DNA with short homologies (52) and was conveniently applied in genomic manipulation of *E. coli*, *Mycobacterium*, *Salmonella*, and *Shigella* (47, 53-55). Unfortunately, the known phage systems are not applicable for recombineering many other bacteria of interest. For example, 666 novel species were published in 2013 (LPSN-list of prokaryotic names with standing in nomenclature, http://www.bacterio.net/), a lot of which are of great importance in bioremediation, chemical production, and human health. However, it is barely possible to find compatible phage systems as well as shuttle vectors for these strains. Although ZFNs, TALENs, and CRISPR have been developed for precise genome editing in eukaryotes (8-10), they also need shuttle vectors to express a nuclease in the host strain. The lack of shuttle vectors expressing these systems is additionally hindering their use for diverse prokaryotes. In this study, we proposed a novel and much simpler gene inactivation method by using the NHEJ pathway mediated by Ku and LigD proteins in *Dietzia* sp. DQ12-45-1b without additional vectors or genome editing systems.

After the linear DNA with mono-homology was transformed into *Dietzia* sp. DQ12-45-1b cells, it performed self-cyclization *in vivo*, recombined with the target homologous sequence in the genome by HR and produced the mutant strain (Fig. 7) (38). Although the efficiency was not very high, around $10^{-5}$ - $10^{-6}$, when targeting the genes in *Dietzia* sp. DQ12-45-1b, it may be acceptable

366 for those newly discovered and important bacteria without compatible genetic manipulation system.
367 Although NHEJ was firstly found in mammalian cells and thought to be specific to eukaryotic
368 organisms, a number of studies showed that NHEJ also exists in prokaryotic organisms (15, 16,
369 18-23, 56). Ku were proved to be essential in NHEJ in the strain DQ12-45-1b. It was interesting
370 that we tried to disrupt the *ku* and *ligD* genes using two methods, the mono-homologous
371 recombination method as described in this paper and the double-homologous recombination
372 method described previously (47), but neither *ku* nor *ligD* mutants was obtained using the mono-
373 homologous linear DNA recombination method. In contrast, *ku* gene could be replaced using
374 double-homologous recombination method which expressed additional recombination enzymes.
375 The result suggested unknown roles of Ku in the recombination process using mono-homologous
376 linear DNA. It was notable that no *ligD* mutant was obtained using both two methods. LigD, as the
377 main ligase of NHEJ, could repair DSBs that had a lethal effect on cell mitosis unless repaired in
378 time (57-60). According to previous reports, *ligD* could be knocked out from *Mycobacterium*
379 *smegmatis*, because there was another ATP-dependent DNA ligase (LigC) that could provide a
380 backup function of LigD-independent error-prone repair of blunt-end DSBs in the strain (61).
381 However, in *Dietzia* sp. DQ12-45-1b, only one *ligD* was identified, and its disruption might be
382 lethal. Furthermore, there was no cyclization of linear plasmid DNA in *ku* mutant of strain DQ12-
383 45-1b, implying that *ku* might be critical in the process of linear DNA cyclization.

384 In addition, 5,111 and 7,695 out of the 25,798 bacterial and archaeal genomes available in
385 the IMG database contain *ku* and *ligD* homologous genes, respectively. The distribution of *ku* and
386 *ligD* genes may suggest the wide distribution of the NHEJ pathway in microorganisms.

Consequently, the introduction of mutagenesis during the cyclization might have played an important role in the genome diversification, which could increase the natural microbial diversity and adaptability. In addition, the wide distribution of *ku* and *ligD* genes may also suggest that the mono-homologous linear DNA recombination by using the NHEJ pathway could be commonly used as a gene inactivation method for diverse prokaryotic organisms, especially for newly discovered, but important microorganisms. Especially, more than 3/4 of the genomes of Actinobacteria, with high GC content and not easy for genetic manipulation, express the *ku* and *ligD* genes, suggesting that the mono-homologous linear DNA recombination is a potentially good method, at least for Actinobacteria. However, more studies are needed to test the versatility of this method.

For organisms lacking Ku and LigD, the mono-homologous linear DNA recombination method requires the construction of a compatible and inducible Ku and LigD system, i.e., one plasmid needs to be constructed to express Ku and LigD. In this case, the expression of Ku and LigD could be controlled by appropriate inducible promoters as described for the expression of Mt-Ku and Mt-LigD in *E. coli* (17). Once the plasmid expressing Ku and LigD is constructed, it can be used for disrupting any gene without constructing new plasmids. Thus, this method is still simpler than the classic approaches requiring incompatible or suicide plasmids.

This method presents two major concerns. One is how to improve the cyclization efficiency and the other is how to reduce the risk of linear DNA degradation by nucleases *in vivo*. In prokaryotic organisms, NHEJ activity was particularly efficient during the stationary phase to

counteract DSBs induced by heat, desiccation, and other factors (62). However, the electroporation efficiency reached its highest value when the recipient cells were in the early exponential phase (33). A balance is therefore necessary to increase the cyclization and recombination activities and to increase the transforming efficiency. Several methods have been developed, including heating the cells, to increase both the transforming efficiencies and the NHEJ process (33), which may promote the efficiency of this method.

Unlike in yeast, linear dsDNA in bacteria such as *E. coli* can be rapidly degraded by nucleases (26). This phenomenon was also observed in strain DQ12-45-1b during sequence deletions when cyclizing the linear DNA LA18, HD18, and HB18 *in vivo* (Fig. 1). In *E. coli*, *recB,* and *recC* are the major nucleases degrading the dsDNA and are essential for the normal activities of the bacterium. Deletion of *recB* and *recC* resulted in poor growth of *E. coli*, producing up to 80% of nonviable cells (12). To inhibit the nuclease activities and promote the recombination efficiency, several phage systems such as RecET system and λ Red system were developed, which again require plasmids (53, 63). For example, in the λ Red system, the *exo*, *bet*, and *gam* genes are under the *lac* promoter control on a multicopy plasmid. Gam function could inhibit RecBCD nuclease to protect the transformed linear DNA from being degraded. Then, with the help of Exo and Bet, gene recombination was conducted (12). Another trial was to recombine linear DNA in wild-type *E.coli* containing the RecBCD nuclease (64). In this case, special sites were engineered in the linear DNA that decreased the RecBCD activities (64). Similarly, when we added a triplet cytidine (CCC) tag at one terminus of the linear DNA (LA18) by PCR, the long deletions were unidirectional and

427 occurred at the opposite terminus with lower "G+C" contents (Fig. 1), suggesting that addition of
428 the CCC tag could protect the target sequence from being digested by nucleases and increased the
429 recombination efficiency. Additionally, electroporation itself has been suggested to reduce DNA
430 degradation by RecBCD nuclease and to allow recombination with linear dsDNA (65).

431 If this method is reliable, it could be used not only to inactivate one gene by inserting the
432 selective cassette into the chromosome (Fig. S10A), but also for gene over-expression and
433 exogenous gene insertion. For example, for over-expression of a gene, the cassette consists of a
434 selectable integrated gene that would be used as the mono-homologous DNA (Fig. S10B). For
435 exogenous DNA insertion, the cassette could be linked to an exogenous DNA before the
436 homologous arm (Fig. S10C).

437 In conclusion, we proposed an easy and fast method for gene inactivation in *Dietzia* sp. DQ12-
438 45-1b, using a linear DNA with mono-homology. This method uses the NHEJ pathway and it might
439 be applicable to prokaryotic organisms that harbor genes encoding Ku and LigD, especially for the
440 newly identified and non-model species. Because the linear DNA with cassettes and homologous
441 DNA could be amplified by fusion PCR conventionally, these procedures should be rapid,
442 convenient, and reliable.

443 **FUNDING**

This work was supported by the National Natural Science Foundation of China (31200099, 31225001 and 31300108), and the National High Technology Research and Development Program ("863"Program: 2012AA02A703 and 2014AA021505).

## ACKNOWLEDGMENTS

We thank Dr. Guang Hu for generously providing the antibody of polyclonal mouse anti-AlkW1, Ms. Hui Fang for her kindly help in bioinformatic analysis, Ms. Xiao-Yu Qin for her help in Ku knockout, and we thank Dr. Yue-Qin Tang for reading the manuscript and discussions.

# REFERENCES


1. Marinelli LJ, Hatfull GF, Piuri M. 2012. Recombineering: A powerful tool for modification of bacteriophage genomes. Bacteriophage 2:5-14.
2. Matsubara K, Malay AD, Curtis FA, Sharples GJ, Heddle JG. 2013. Structural and functional characterization of the Redbeta recombinase from bacteriophage lambda. PLoS One 8:e78869.
3. Nafissi N, Slavcev R. 2014. Bacteriophage recombination systems and biotechnical applications. Appl Microbiol Biotechnol 98:2841-51.
4. Bihari Z, Szvetnik A, Szabo Z, Blastyak A, Zombori Z, Balazs M, Kiss I. 2011. Functional analysis of long-chain n-alkane degradation by Dietzia spp. FEMS Microbiol Lett 316:100-7.
5. Zhao Z, Ding JY, Ma WH, Zhou NY, Liu SJ. 2012. Identification and characterization of gamma-aminobutyric acid uptake system GabPCg (NCgl0464) in Corynebacterium glutamicum. Appl Environ Microbiol 78:2596-601.
6. Sallam KI, Tamura N, Imoto N, Tamura T. 2010. New vector system for random, single-step integration of multiple copies of DNA into the Rhodococcus genome. Appl Environ Microbiol 76:2531-9.
7. Bishop AH, Rachwal PA. 2014. Identification of genes required for soil survival in Burkholderia thailandensis by transposon-directed insertion site sequencing. Curr Microbiol 68:693-701.
8. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636-46.
9. Bogdanove AJ, Voytas DF. 2011. TAL effectors: customizable proteins for DNA targeting. Science 333:1843-6.
10. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31:233-9.
11. Kuzminov A. 1999. Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. Microbiol Mol Biol Rev 63:751-813, table of contents.
12. Cromie GA, Connelly JC, Leach DR. 2001. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Mol Cell 8:1163-74.
13. Pitcher RS, Brissett NC, Doherty AJ. 2007. Nonhomologous end-joining in bacteria: a microbial perspective. Annu Rev Microbiol 61:259-82.
14. Shuman S, Glickman MS. 2007. Bacterial DNA repair by non-homologous end joining. Nat Rev Microbiol 5:852-61.
15. Gong C, Bongiorno P, Martins A, Stephanou NC, Zhu H, Shuman S, Glickman MS. 2005. Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nat Struct Mol Biol 12:304-12.
16. Gong C, Martins A, Bongiorno P, Glickman M, Shuman S. 2004. Biochemical and genetic analysis of the four DNA ligases of mycobacteria. J Biol Chem 279:20594-606.
17. Malyarchuk S, Wright D, Castore R, Klepper E, Weiss B, Doherty AJ, Harrison L. 2007. Expression of Mycobacterium tuberculosis Ku and Ligase D in Escherichia coli results in RecA and RecB-independent DNA end-joining at regions of microhomology. DNA Repair (Amst) 6:1413-24.
18. Wright D, DeBeaux A, Shi R, Doherty AJ, Harrison L. 2010. Characterization of the roles of the catalytic domains of Mycobacterium tuberculosis ligase D in Ku-dependent error-prone DNA end joining.

Mutagenesis 25:473-81.

19. Della M, Palmbos PL, Tseng HM, Tonkin LM, Daley JM, Topper LM, Pitcher RS, Tomkinson AE, Wilson TE, Doherty AJ. 2004. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. Science 306:683-5.

20. Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M, Devine SK, Day JP, Wilkinson A, d'Adda di Fagagna F, Devine KM, Bowater RP, Jeggo PA, Jackson SP, Doherty AJ. 2002. Identification of a DNA nonhomologous end-joining complex in bacteria. Science 297:1686-9.

21. Aravind L, Koonin EV. 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res 11:1365-74.

22. Weller GR, Doherty AJ. 2001. A family of DNA repair ligases in bacteria? FEBS Lett 505:340-2.

23. Doherty AJ, Jackson SP, Weller GR. 2001. Identification of bacterial homologues of the Ku DNA repair proteins. FEBS Lett 500:186-8.

24. Moeller R, Stackebrandt E, Reitz G, Berger T, Rettberg P, Doherty AJ, Horneck G, Nicholson WL. 2007. Role of DNA repair by nonhomologous-end joining in Bacillus subtilis spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation. J Bacteriol 189:3306-11.

25. Wang ST, Setlow B, Conlon EM, Lyon JL, Imamura D, Sato T, Setlow P, Losick R, Eichenberger P. 2006. The forespore line of gene expression in Bacillus subtilis. J Mol Biol 358:16-37.

26. Court DL, Sawitzke JA, Thomason LC. 2002. Genetic engineering using homologous recombination. Annu Rev Genet 36:361-388.

27. Pleshakova EV, Dubrovskaya EV, Turkovskaya OV. 2008. Efficiencies of introduction of an oil-oxidizing Dietzia maris strain and stimulation of natural microbial communities in remediation of polluted soil. Appl Biochem Microbiol 44:389-395.

28. Wang XB, Chi CQ, Nie Y, Tang YQ, Tan Y, Wu G, Wu XL. 2011. Degradation of petroleum hydrocarbons (C6-C40) and crude oil by a novel Dietzia strain. Bioresour Technol 102:7755-61.

29. Jin Q, Hu Z, Jin Z, Qiu L, Zhong W, Pan Z. 2012. Biodegradation of aniline in an alkaline environment by a novel strain of the halophilic bacterium, Dietzia natronolimnaea JQ-AN. Bioresour Technol 117:148-54.

30. Click RE, Van Kampen CL. 2010. Assessment of Dietzia subsp. C79793-74 for treatment of cattle with evidence of paratuberculosis. Virulence 1:145-55.

31. Click RE. 2012. A Potential 'Curative' Modality for Crohn's Disease---Modeled after Prophylaxis of Bovine Johne's Disease. Mycobact Dis 2:117.

32. Click RE. 2015. Crohn's disease therapy with Dietzia: the end of anti-inflammatory drugs. Future Microbiol 10:147-50.

33. Lu S, Nie Y, Tang YQ, Xiong G, Wu XL. 2014. A critical combination of operating parameters can significantly increase the electrotransformation efficiency of a gram-positive Dietzia strain. J Microbiol Methods 103:144-51.

34. Nie Y, Liang J, Fang H, Tang YQ, Wu XL. 2011. Two novel alkane hydroxylase-rubredoxin fusion genes isolated from a Dietzia bacterium and the functions of fused rubredoxin domains in long-chain n-alkane degradation. Appl Environ Microbiol 77:7279-88.

35. Nie Y, Liang JL, Fang H, Tang YQ, Wu XL. 2014. Characterization of a CYP153 alkane hydroxylase gene in a Gram-positive Dietzia sp. DQ12-45-1b and its "team role" with alkW1 in alkane degradation. Appl

Microbiol Biotechnol 98:163-73.

36. Wang XB, Nie Y, Tang YQ, Wu G, Wu XL. 2013. n-Alkane chain length alters Dietzia sp. strain DQ12-45-1b biosurfactant production and cell surface activity. Appl Environ Microbiol 79:400-2.

37. Liang JL, JiangYang JH, Nie Y, Wu XL. 2015. Regulation of the alkane hydroxylase gene CYP153 in a Gram-positive alkane degrading bacterium Dietzia sp. DQ12-45-1b. Appl Environ Microbiol doi:10.1128/AEM.02811-15.

38. Liang JL, Nie Y, Wang M, Xiong G, Wang YP, Maser E, Wu XL. 2015. Regulation of alkane degradation pathway by a TetR family repressor via an autoregulation positive feedback mechanism in a Gram-positive Dietzia bacterium. Mol Microbiol doi:10.1111/mmi.13232.

39. Pridmore RD. 1987. New and versatile cloning vectors with kanamycin-resistance marker. Gene 56:309-12.

40. Liang JL, Nie Y, Wang M, Xiong G, Wang YP, Maser E, Wu XL. 2016. Regulation of alkane degradation pathway by a TetR family repressor via an autoregulation positive feedback mechanism in a Gram-positive Dietzia bacterium. Mol Microbiol 99:338-59.

41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.

42. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43:D222-6.

43. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480-4.

44. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res 42:D560-7.

45. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-9.

46. Szvetnik A, Bihari Z, Szabo Z, Kelemen O, Kiss I. 2010. Genetic manipulation tools for Dietzia spp. J Appl Microbiol 109:1845-52.

47. van Kessel JC, Hatfull GF. 2007. Recombineering in Mycobacterium tuberculosis. Nat Methods 4:147-52.

48. Liang JL, JiangYang JH, Nie Y, Wu XL. 2016. Regulation of the Alkane Hydroxylase CYP153 Gene in a Gram-Positive Alkane-Degrading Bacterium, Dietzia sp. Strain DQ12-45-1b. Appl Environ Microbiol 82:608-19.

49. Fang H QX, Zhang KD, Nie Y, Wu XL. 3 March 2018. Role of the Group 2 Mrp sodium/proton antiporter in rapid response to high alkaline shock in the alkaline- and salt-tolerant Dietzia sp. DQ12-45-1b. Applied Microbiology and Biotechnology 102:3765-3777.

50. Joseph S, David WR. 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 2.

51. Szewczyk E, Nayak T, Oakley CE, Edgerton H, Xiong Y, Taheri-Talesh N, Osmani SA, Oakley BR. 2006. Fusion PCR and gene targeting in Aspergillus nidulans. Nat Protoc 1:3111-20.

52. Sawitzke JA, Thomason LC, Costantino N, Bubunenko M, Datta S, Court DL. 2007. Recombineering: in vivo genetic engineering in E. coli, S. enterica, and beyond. Methods Enzymol 421:171-99.

53. Murphy KC. 1998. Use of bacteriophage lambda recombination functions to promote gene replacement in

Escherichia coli. J Bacteriol 180:2063-71.

54. Datta S, Costantino N, Court DL. 2006. A set of recombineering plasmids for gram-negative bacteria. Gene 379:109-15.

55. Ranallo RT, Barnoy S, Thakkar S, Urick T, Venkatesan MM. 2006. Developing live Shigella vaccines using lambda Red recombineering. FEMS Immunol Med Microbiol 47:462-9.

56. Wilson TE, Topper LM, Palmbos PL. 2003. Non-homologous end-joining: bacteria join the chromosome breakdance. Trends Biochem Sci 28:62-6.

57. Daley JM, Palmbos PL, Wu D, Wilson TE. 2005. Nonhomologous end joining in yeast. Annu Rev Genet 39:431-51.

58. Dudasova Z, Dudas A, Chovanec M. 2004. Non-homologous end-joining factors of Saccharomyces cerevisiae. FEMS Microbiol Rev 28:581-601.

59. Krejci L, Chen L, Van Komen S, Sung P, Tomkinson A. 2003. Mending the break: two DNA double-strand break repair machines in eukaryotes. Prog Nucleic Acid Res Mol Biol 74:159-201.

60. Lees-Miller SP, Meek K. 2003. Repair of DNA double strand breaks by non-homologous end joining. Biochimie 85:1161-73.

61. Gong CL, Bongiorno P, Martins A, Stephanou NC, Zhu H, Shuman S, Glickman MS. 2005. Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nature Structural & Molecular Biology 12:304-312.

62. Karathanasis E, Wilson TE. 2002. Enhancement of Saccharomyces cerevisiae end-joining efficiency by cell growth stage but not by impairment of recombination. Genetics 161:1015-27.

63. Hall SD, Kane MF, Kolodner RD. 1993. Identification and characterization of the Escherichia coli RecT protein, a protein encoded by the recE region that promotes renaturation of homologous single-stranded DNA. J Bacteriol 175:277-87.

64. Dabert P, Smith GR. 1997. Gene replacement with linear DNA fragments in wild-type Escherichia coli: enhancement by Chi sites. Genetics 145:877-89.

65. El Karoui M, Amundsen SK, Dabert P, Gruss A. 1999. Gene replacement with linear DNA in electroporated wild-type Escherichia coli. Nucleic Acids Res 27:1296-9.

602 **FIGURE LEGENDS**

603 **FIG 1 Sequences of DSB junction in LA18 (A), HD18 (B), and HB18 (C**). The 5′-"A" tail is red.
604 HindIII and BamHI digestion sites are shown in green and purple, respectively. "CCC" label in
605 LA18 added by PCR is in blue. The non-templated inserted nucleotides are in pink and the number
606 of deleted nucleotides is shown at the junction site.

607

608 **FIG 2 Identification of *alkW1* mutant in *Dietzia* sp. DQ12-45-1b by mono-homologous linear**
609 **DNA**. A, Ethidium-bromide-stained agarose gel showing typical results of PCR identification of
610 eight colonies. M, DNA marker; 0, blank control. 1–8, different colonies. The bands designated
611 with the black arrow are the *alkW1* mutant colonies with correct size. The bands designated with
612 the red arrow are the unchanged *alkW1* colonies. B, Schematic diagram showing the linear DNA
613 recombination pattern. P45, promoter p45 (119 bp); Hm, mono-homologous arm (410 bp); Km,
614 kanamycine (795 bp).

615

616 **FIG 3 Functional characterization of the strain *alkW1*$^-$**. A, Western blot analysis of the AlkW1
617 protein expression in *Dietzia* sp. DQ12-45-1b and the strain *alkW1*$^-$. Cells grew in GPY as negative
618 control; WT cells grew in MF+ C16 as positive control; WT, wild type *Dietzia* sp. DQ12-45-1b;
619 MT, the strain*alkW1*$^-$; GPY, cells grown in GPY medium; MF+C16, cells grown in minimal
620 medium containing hexadecane. B, Growth curve of strain DQ12-45-1b and strain *alkW1*$^-$ in
621 minimal medium with hexadecane as the sole carbon source.

622 **FIG 4 Distribution of *ku* and *ligD* homologous genes in microbial genomes at the phylum**

623 **level**. A, Distribution of *ligD* in microbial genomes at the phylum level; B, Distribution of *ku* in
624 microbial genomes at the phylum level.
625
626 **FIG 5 Phylogenetic relationships of Ku and LigD based on amino acid sequences in**
627 **prokaryotic organisms.** The phylogenetic trees were constructed using the neighbor-joining
628 method in MEGA6 (45). The trees were bootstrapped with 500 replicates, indicated at the
629 respective nodes.
630
631 **FIG 6 Organization of genes encoding Dt-Ku and Dt-LigD and their protein domain**
632 **architecture.** A, Protein domain architecture of Ku. Ku core, the Ku core domain; HC2, histone
633 H1-like nucleoprotein domain; Rho, helix–extension–helix domain from the bacterial transcription
634 termination factor Rho. B, Protein domain architecture of LigD. C, Gene organization of *ku* and
635 *ligD* operons. The direction of the arrow indicates the direction of the transcription. The numbers
636 above the arrows are the GenBank accession IDs of their corresponding genes. Sc, *Streptomyces*
637 *coelicolor*; Bs, *Bacillus subtilis*; Bh, *Bacillus halodurans*; Mt, *Mycobacterium tuberculosis*; Pa,
638 *Pseudomonas aeruginosa*; Ml, *Mesorhizobium loti*; Af, *Archaeoglobusfulgidus*.
639
640 **FIG 7 Mechanism of mono-homologous linear DNA for target gene mutant.** The linear DNA
641 transformed into bacterial cells is self-cyclized and then recombined with the chromosomal gene.

642 **TABLE 1** List of strains and plasmids used in the study

| Strain or plasmid | Relevant characteristic or genotype | Reference or source |
|---|---|---|
| Strains | | |
| *Dietzia* sp. DQ12-45-1b | Grows on C10–C40 alkanes; have DNA ligase D and Ku protein | This study |
| *alkW1*$^-$ | *alkW1* defective strain of *Dietzia* sp. DQ12-45-1b, Km$^r$ | This study |
| *E. coli* DH5α | Cloning strain | Takara |
| plasmids | | |
| pET-28a(+) | Km$^r$ | Novagen |
| pNV18-Sm | Sm$^r$ | (46) |
| pNV18-PEK | Kanamycin resistant cassettes and *egfp* controlled by P45 in pNV18-Sm, Sm$^r$ | This study |
| pNV18-PKK | *Ku* expression controlled by P45 in pNV18-Sm, Sm$^r$, Km$^r$ | This study |
| pJV53-PA | Ampicillin resistant expression controlled by P45 promoter, Amp$^r$ | This study |
| pUC19-Ku-LigD | pUC19 with insertion fragment containing genes encoding Dt-Ku and Dt-LigD, Amp$^r$ | This study |
| pUC19-Ku | pUC19 with insertion fragment containing genes encoding Dt-Ku, Amp$^r$ | This study |

| | | |
|---|---|---|
| pUC19-LigD | pUC19 with insertion fragment containing genes encoding Dt-Ku and Dt-LigD, $Amp^r$ | This study |

643 $Sm^r$, Streptomycin resistant; $Km^r$, kanamycin resistant; $Amp^r$, ampicillin resistant.

**TABLE 2** List of genes mutant in *Dietzia* sp. DQ12-45-1b

| Gene | GenBank accession numbers | Name | Recombination efficiency [a] |
|---|---|---|---|
| *alkW1* | HQ850582 | alkane hydroxylase | $3.0 \times 10^{-6}$ |
| orf-4898 | KY436357 | histidine kinase | $1.1 \times 10^{-5}$ |
| orf-3472 | KY436358 | histidine kinase | $2.5 \times 10^{-6}$ |
| orf-4338 | KY436359 | histidine kinase SenX3 | $2.5 \times 10^{-6}$ |
| orf-3908 | KY436360 | histidine kinase | $1.2 \times 10^{-6}$ |
| orf-4160 | KY436361 | histidine kinase | $5.0 \times 10^{-6}$ |
| orf-0371 | KY436362 | histidine kinase | $5.0 \times 10^{-6}$ |
| orf-4729 | KY436363 | histidine kinase | $6.5 \times 10^{-5}$ |
| 0rf 2157 | KY436364 | Putative phosphodiesterase | $3.2 \times 10^{-6}$ |
| orf-0039 | KY436365 | cobyrinic acid a,c-diamide synthase | $4.0 \times 10^{-5}$ |

a: The recombinant efficiency was expressed as the number recombinants per microgram DNA divided by the cell competency which was indicated by the transformation efficiency with plasmid pNV18-Sm.

**FIG 1 Sequences of DSB junction in LA18 (A), HD18 (B), and HB18 (C).**

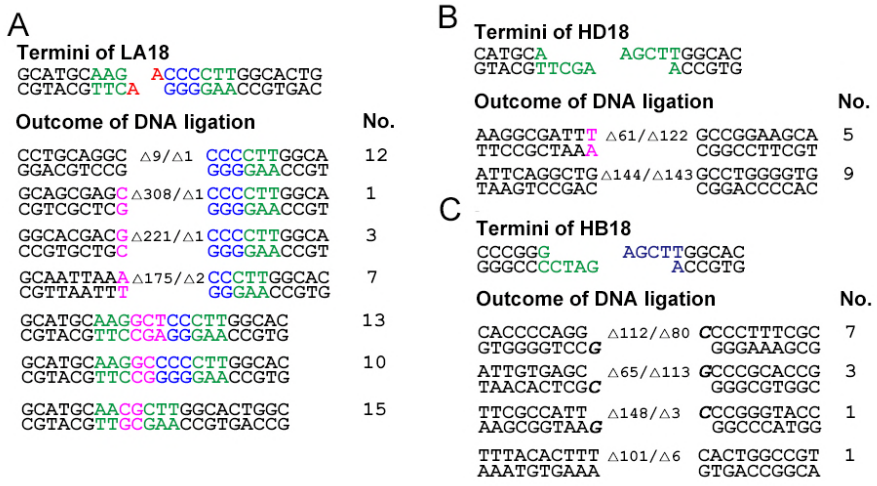

**FIG 2 Identification of *alkW1* mutant in *Dietzia* sp. DQ12-45-1b by mono-homologous linear DNA**.

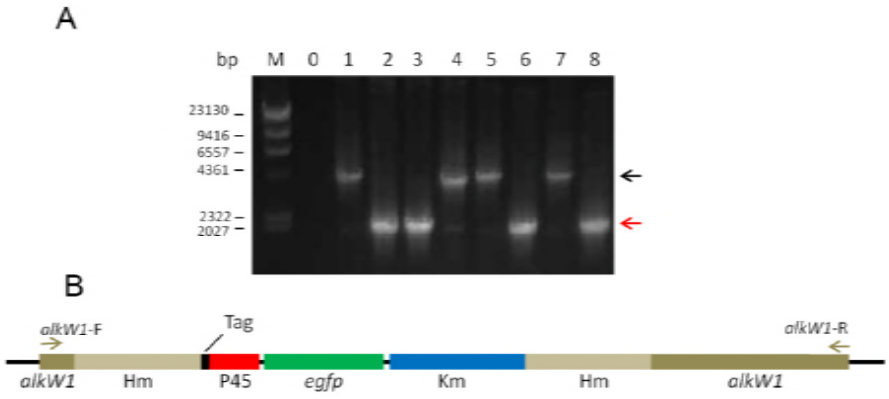

**FIG 3 Functional characterization of the strain *alkW1*$^{-}$.**

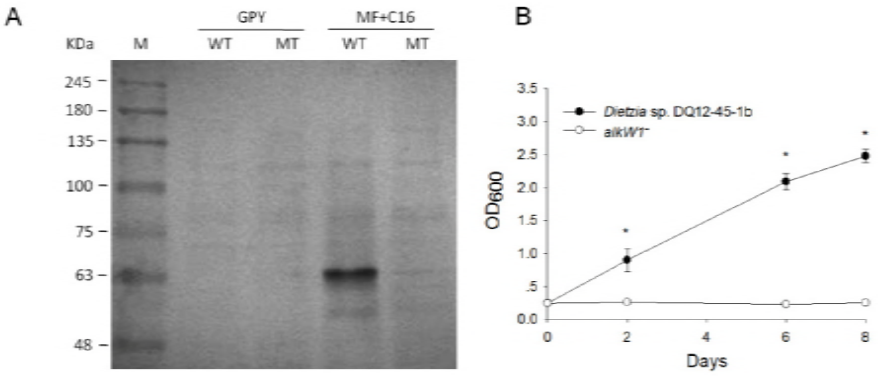

**FIG 4 Distribution of *ku* and *ligD* homologous genes in microbial genomes at the phylum level**.

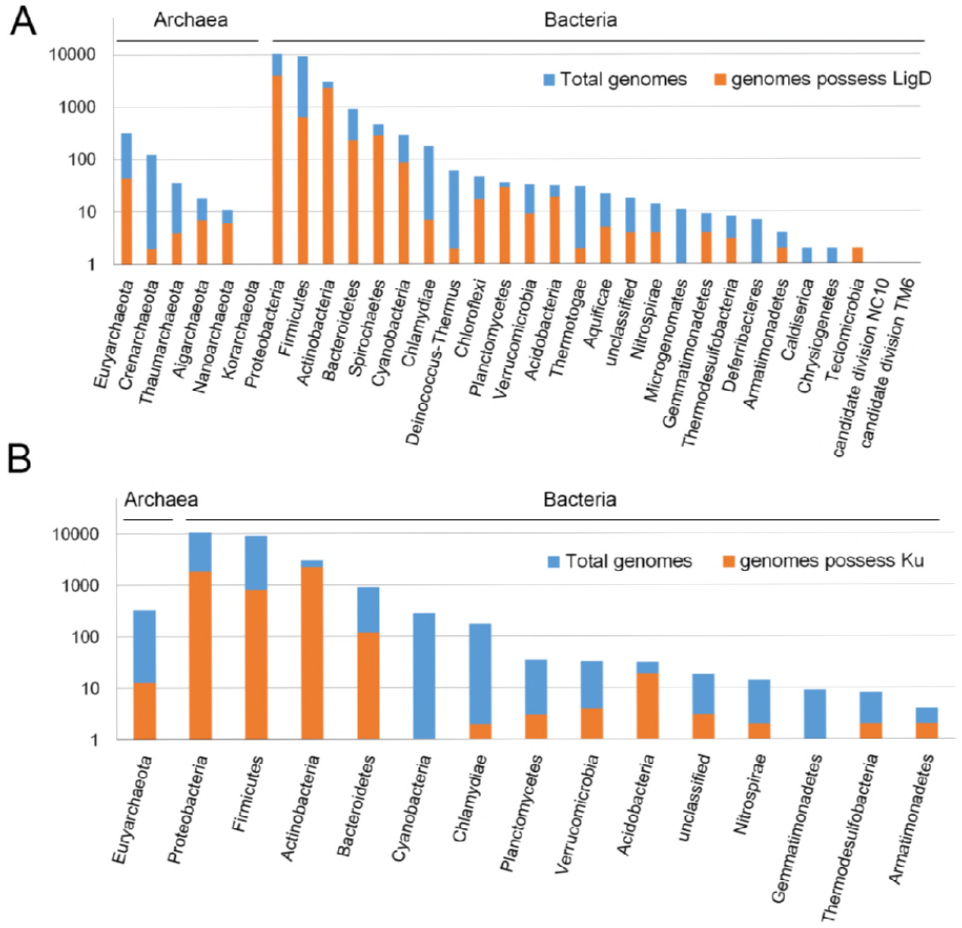

**FIG 5 Phylogenetic relationships of Ku and LigD based on amino acid sequences in prokaryotic organisms.**

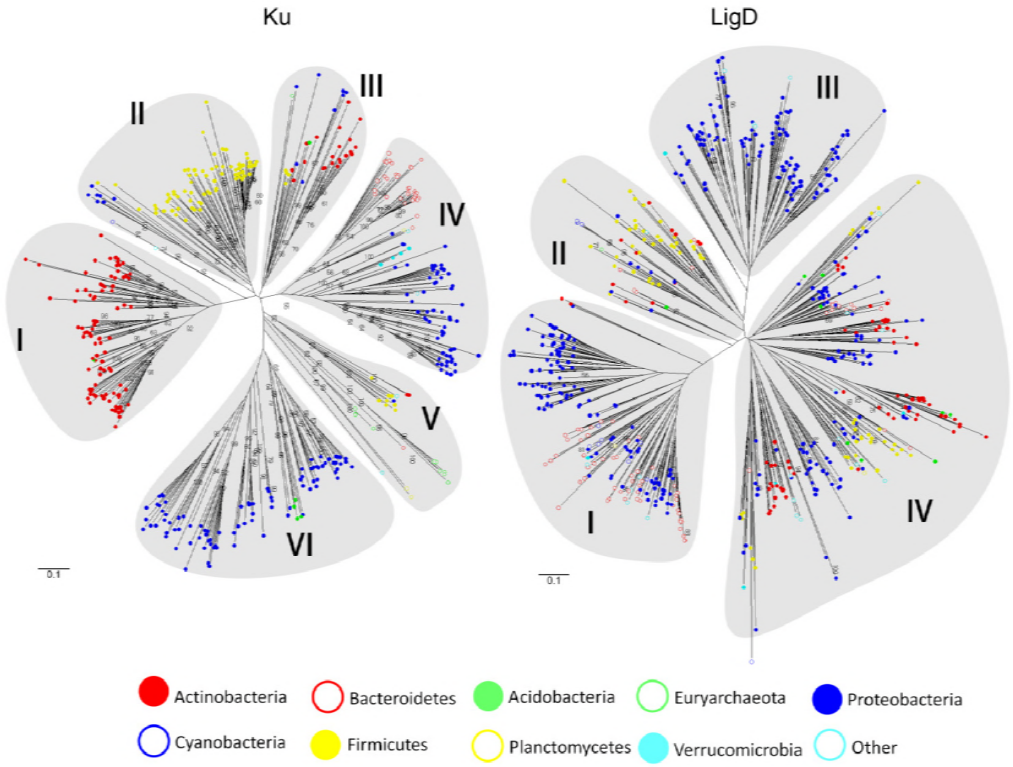

**FIG 6 Organization of genes encoding Dt-Ku and Dt-LigD and their protein domain architecture.**

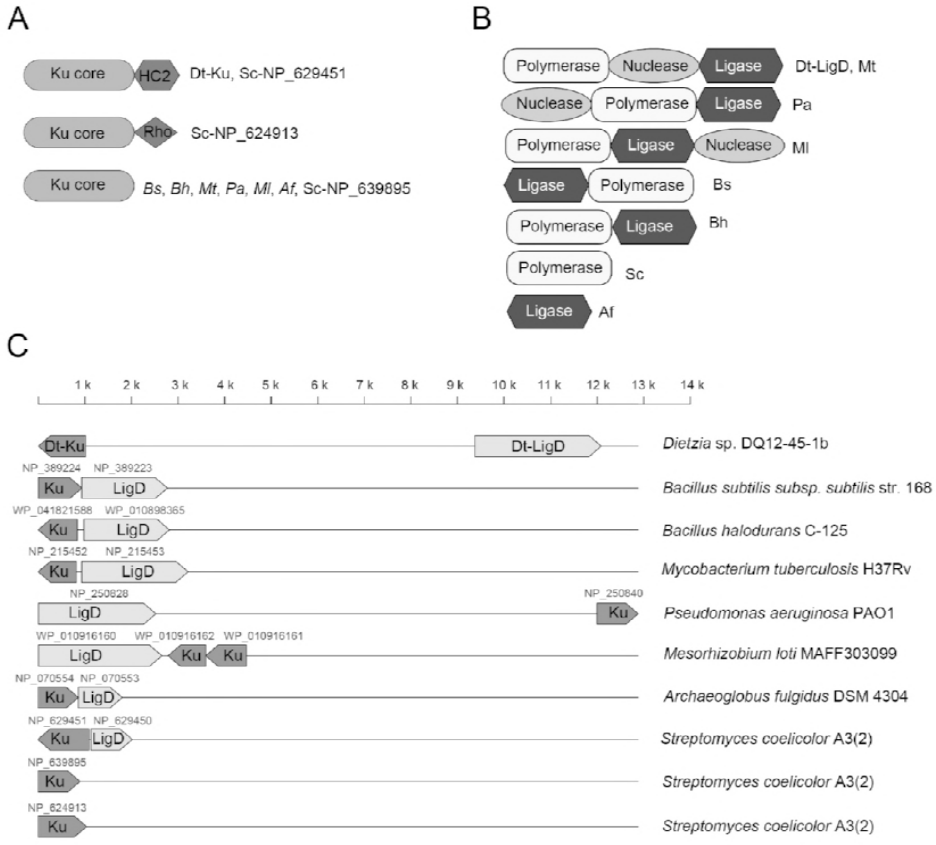

**FIG 7 Mechanism of mono-homologous linear DNA for target gene mutant.**