

Received: 25 January 2018 Accepted: 13 March 2018 Published online: 02 April 2018

# **OPEN** Functional metagenomic approach to identify overlooked antibiotic resistance mutations in bacterial rRNA

Kentaro Miyazaki<sup>1,2</sup> & Kei Kitahara<sup>3,4</sup>

Our knowledge as to how bacteria acquire antibiotic resistance is still fragmented, especially for the ribosome-targeting drugs. In this study, with the aim of finding novel mechanisms that render bacteria resistant to the ribosome-targeting antibiotics, we developed a general method to systematically screen for antibiotic resistant 16 S ribosomal RNAs (rRNAs), which are the major target for multiple antibiotics (e.g. spectinomycin, tetracycline, and aminoglycosides), and identify point mutations therein. We used  $\textit{Escherichia coli}\ \Delta 7$ , a null mutant of the rrn (ribosomal RNA) operons, as a surrogate host organism to construct a metagenomic library of 16 S rRNA genes from the natural (non-clinical) environment. The library was screened for spectinomycin resistance to obtain four resistant 16 S rRNA genes from non-E. coli bacterial species. Bioinformatic analysis and site-directed mutagenesis identified three novel mutations - U1183C (the first mutation discovered in a region other than helix 34), and C1063U and U1189C in helix 34 - as well as three well-described mutations (C1066U, C1192G, and G1193A). These results strongly suggest that uncharacterized antibiotic resistance mutations still exist, even for traditional antibiotics.

Antibiotic resistance is a serious problem for human beings because pathogenic microorganisms that acquire such resistance void antibiotic treatments. Therefore, a tremendous effort has been made by researchers to identify specific resistance mechanisms and mutations that render bacteria resistant to antibiotics. These studies are beneficial for the timely detection and early diagnosis of resistant bacteria, which is key to prevent the spread of unwanted infectious diseases.

There are three main mechanisms for microorganisms to acquire antibiotic resistance: (i) enzymatic inactivation or modification of antibiotics (e.g. β-lactamases inactivate penicillin antibiotics)<sup>1</sup>; (ii) acquisition of mutation(s) in target sites of the antibiotics; and (iii) decreasing the net drug concentration in the cell by reducing drug permeability via cell wall or by increasing the activity of efflux pumps (e.g. tetracycline resistance)<sup>2</sup>. Among these, the second mechanism is often observed for ribosome-targeting drugs such as spectinomycin (Spc), aminoglycosides (e.g. kanamycin, streptomycin, neomycin), tetracycline, chloramphenicol, macrolides, lincomycins, streptogramin A, and oxazolidinones; the former three are known to target the 30 S subunit that contains the 16 S rRNA as its main component, whereas the others are known to attack the 50 S subunit that contains the 23 S rRNA as its

As described above, there are a large number of antibiotics that target the ribosome. This is because ribosomes play an essential role in protein biosynthesis, translating messenger RNA-encoded genetic information into proteins, which consists of sequential multistep reactions - initiation, elongation, termination, and recycling. Owing to these extremely elaborate reaction dynamics, there are different kinds of inhibitors targeting each step of the translation process<sup>3-5</sup>. As the ribosome is RNA-rich, and functionally critical sites exist mainly on RNAs (the

<sup>1</sup>Department of Life Science and Biotechnology, Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki, 305-8566, Japan. <sup>2</sup>Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561, Japan. <sup>3</sup> Ambitious Leader's Program, Graduate School of Science, Hokkaido University, N10W8, Kita-ku, Sapporo, 060-0810, Hokkaido, Japan. Department of Chemistry, Graduate School of Science, Hokkaido University, N10W8, Kita-ku, Sapporo, 060-0810, Hokkaido, Japan. Correspondence and requests for materials should be addressed to K.M. (email: miyazaki-kentaro@aist.go.jp)

decoding centre in 16 S rRNA and peptidyl transferase centre in 23 S rRNA), many antibiotic target sites exist on rRNAs, as do several resistant point mutations, accordingly<sup>3,6,7</sup>.

Researchers have long tried to list as many resistant point mutations in rRNAs as possible, by means of classical genetic experiments using organisms such as *Escherichia coli*<sup>8</sup>, *Halobacterium halobium*<sup>9</sup>, and *Mycoplasma smegmatis* (*rrn*<sup>-</sup>)<sup>10</sup>. There are, however, unavoidable drawbacks in these systems. In the *E. coli* system, the organism has as many as seven rRNA operons (*rrn* operons) in its genome. Owing to the high background derived from the endogenous (wild-type) rRNA genes, it is necessary to use a high copy number vector to characterize the function (i.e. antibiotic susceptibility) of mutant 16 S rRNA genes *in vivo*<sup>8,11,12</sup>. However, the handling of such non-simple genetic systems suffers from various technical difficulties, as pointed out previously<sup>13</sup>. Although *H. halobium* and *M. smegmatis* only have one *rrn* operon in their genome, and thus can partly solve the underlying problem in the *E. coli* system, they only show slow growth phenotypes and therefore, make it difficult to conduct reliable genetic experiments. It is thus uncertain whether all possible resistant mutations to an antibiotic have successfully and correctly been listed using these systems. It should be noted that *Thermus thermophilus*, a thermophilic strain with a single *rrn* operon, has been used for similar purposes i.e. to generate interesting insights on antibiotic resistance mutations<sup>14,15</sup>.

Here, we propose a new approach to circumvent these methodological problems in a simple way; we use E. coli  $\Delta 7$ , a null mutant of the rrn operon 16, as a surrogate host organism and screen for the antibiotic resistance of various species' rRNA genes, which are retrieved from environmental metagenomes. Admittedly, the use of the null mutant E. coli strain is not a novel approach and there are in fact some studies that use the strain to determine antibiotic resistance mutations in rRNAs<sup>17,18</sup>. The originality in our study resides in the methodology i.e. we use metagenomic rRNA genes that are directly extracted from the environment, while E. coli  $\Delta 7$  is simply used as a surrogate organism that can be handled easily as a model microorganism. The rationale of our approach using the combination of non-E. coli rRNA genes and the E. coli host is based on our recent finding that various 16 S rRNAs, including those from a different class<sup>19</sup> or a phylum<sup>20</sup>, are functionally compatible with the *E. coli* ribosome. We can genetically characterize the functions of a diverse array of heterologous 16 S rRNAs using E. coli  $\Delta 7$  as a common platform<sup>20</sup>. In this study, we applied this technique, named Comparative RNA Function Analysis<sup>20</sup>, to test whether we can find novel and biologically relevant antibiotic resistance mutations. Specifically, we used a traditional antibiotic, Spc, mutations to which are supposed to have been thoroughly investigated, as a model antibiotic. A metagenomic library of non-clinical environments, which are considered to be a reservoir of antibiotic resistance  $^{21-23}$ , was constructed using *E. coli*  $\Delta 7$  as a host, then functionally screened for Spc resistance. As the result, we successfully obtained four 16 S rRNA genes from non-E. coli bacterial species carrying Spc resistance mutations. Further analysis of these genes revealed that three point mutations (C1063U, U1183C, and U1189C, E. coli numbering), which have not been, to our best knowledge, reported in any literature thus far, render bacteria resistant to Spc. Our results strongly suggest that there are many unfound and therefore uncharacterized antibiotic resistance point mutations in 16 S rRNA genes in natural environments, and also possibly in clinical environments.

#### Results

Metagenomic library screening for Spc resistant 16 S rRNA genes. To construct a metagenomic library of 16 S rRNA genes, we prepared a mixture of metagenomic samples directly extracted from natural environments (fermentation products, activated sludge, and wood compost). The full-length 16 S rRNA genes were PCR-amplified from the metagenomic mixture using a set of universal primers<sup>24</sup> and the amplicon was cloned in the expression vector pMY205mPAG2 (encoding trimethoprim [Tmp] resistance gene, p15A ori, and *E. coli rrnB* operon containing a 16 S rRNA gene) by replacing the pre-existing *E. coli* 16 S rRNA gene in the vector with the amplicon<sup>24</sup>. An *E. coli* Δ7 strain (MY101), a null mutant of the *rrn* operons in the genome, was then transformed with the plasmid library and screened for functionally compatible heterologous 16 S rRNA genes based on the ability to support the growth of the host (in the absence of Spc). MY205 colonies that appeared on the LB/Tmp agar plates containing sucrose (approximately 2,000) were then secondarily screened on LB/Tmp agar plates containing 40 μg/mL Spc. Four Spc resistant clones were selected, which were named NHMcSpc1, mgSpc1, mgSpc2, and mgSpc5.

Table 1 summarizes the sequence properties of the resistant 16S rRNA genes. All the genes were derived from Gammaproteobacteria, which showed 82–96% sequence identities to the *E. coli* 16S rRNA (*rrnB*) gene. Each gene sequence was used as a query to BLAST-search for the closest homologues in the NCBI database; sequence comparison between the queries and the homologues, as well as *E. coli* sequence (Spc susceptible) would be effective for predicting resistance mutations. Figure 1A shows the sequence alignment of the two hot spots (G1058 to G1071 and G1174 to U1199, *E. coli* numbering) where Spc resistance mutations have been frequently discovered (G1064, C1066, A1191, C1192, and G1193, underlined in the *E. coli* sequence) $^{25-34}$ . The secondary structure of 16S rRNA, including the hot spots, is shown in Fig. 1B. Putative resistance mutations identified in the metagenomically retrieved resistance genes are highlighted in red in Fig. 1A,B and summarized in Table 2. U1183C and U1189C mutations have not been described as Spc resistance mutations, but our sequence analysis (Fig. 1A) suggested that these mutations might be involved in resistance. Our approach to confirm the effect of these putative resistance mutations included: (i) changing the putative resistance mutations in the metagenomically-retrieved resistant genes to the non-resistant type nucleotides (*E. coli* sequence as a reference), either individually or in combination with other nucleotides and (ii) introducing the putative resistance mutations into the *E. coli* 16S rRNA gene. The constructed mutant 16S rRNA genes were then introduced into *E. coli*  $\Delta$ 7 and the resistance was tested in the presence of various concentrations of Spc.

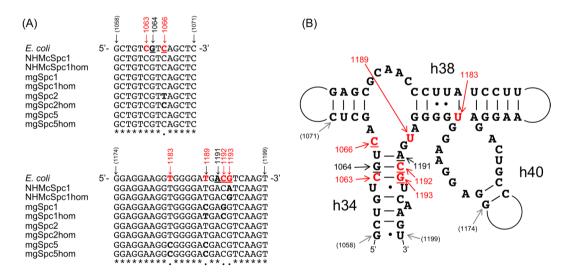
**Mutation study of the putative Spc resistance mutations in the metagenomically retrieved 16 S rRNA genes.** *E. coli* MY205 derivatives were grown in LB broth containing various concentrations (0–1,024 µg/mL) of Spc and the growth phenotypes (growth curves drawn from OD<sub>600</sub> values) were monitored for 10 h. The

Clone (Accession ID)	Closest homologues				
	Accession ID	Strain	Phylogeny (phylum; class; order; family)	Identity	Identity to E. coli 16 S rRNA (rrnB)
NHMcSpc1 (LC306682)	CP017802.1	Raoultella ornithinolytica strain MG	Proteobacteria; Gammaproteobacteria; Enterobacterales; Enterobacteriaceae	99% (1536/1540)	96% (1484/1542)
mgSpc1 (LC306679)	NR_074692.1	Thioalkalivibrio sulfidophilus strain HL-EbGR7	Proteobacteria; Gammaproteobacteria; Chromatiales; Ectothiorhodospiraceae	92% (1418/1549)	83% (1302/1558)
mgSpc2 (LC306680)	NR_145539.1	Chujaibacter soli strain KIS55–21	Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae	99% (1466/1486)	82% (1286/1550)
mgSpc5 (LC306681)	NR_108606.1	Thalassolituus marinus strain IMCC1826	Proteobacteria; Gammaproteobacteria; Oceanospirillales; Oceanospirillaceae	99% (1464/1472)	86% (1330/1542)

**Table 1.** Spectinomycin (Spc)-resistant 16 S rRNA genes retrieved from the metagenome and their closest homologues.

Clone	Putative resistance mutations (helix number)	
NHMcSpc1	G1193A (helix 34)	
mgSpc1	U1189C (helix 34), C1192G (helix 34)	
mgSpc2	C1066U (helix 34)	
mgSpc5	U1183C (helix 38), U1189C (helix 34)	

**Table 2.** Putative resistance mutations in the Spectinomycin (Spc)-resistant 16 S rRNA genes retrieved from the metagenome.



**Figure 1.** Putative spectinomycin (Spc) resistance point mutations in 16 S rRNA gene as inferred by Spc resistant 16 S rRNA genes from the metagenome. (**A**) Multiple sequence alignment of 16 S rRNA genes from *E. coli*, metagenomically retrieved Spc resistant clones (NHMcSpc1, mgSpc1, mgSpc2, and mgSpc5), and their closest homologues (suffixed "hom" to its parent's name) from the NCBI database. Known resistance mutation sites are underlined in the *E. coli* sequence and putative resistance nucleotides to Spc are highlighted in red. Sites that share identical nucleotides among all sequences are shown with asterisks and those not completely conserved are shown with a dot below the alignment. (**B**) Secondary structure of *E. coli* 16 S rRNA around the Spc resistance mutation sites aligned in (**A**). Putative resistance nucleotides to Spc are coloured in red.

wild-type *E. coli* (MY205 complemented by self  $16\,S$  rRNA gene from rrnB) had a minimum inhibitory concentration (MIC) of  $32\,\mu\text{g/mL}$  in this system (Table 3, Supplementary Fig. S1a). Similar analysis was conducted for each metagenomic  $16\,S$  rRNA clone (Supplementary Fig. S1b–S1e) and the MICs were determined as summarized in Table 4.

MY205 with NHMcSpc1 had high-level Spc resistance with MIC of  $>1,024\,\mu\text{g/mL}$  (Table 4, Supplementary Fig. S1b). Within this gene, a putative resistance mutation, G1193A, was inferred (Fig. 1A, Table 2), which is known to confer a low-level resistance in plant chloroplasts<sup>25,35</sup>. As G1193 forms a Watson-Crick base pair with C1063 (C1063–G1193) in its homologue (NHMcSpc1hom) and also in the *E. coli* 16 S rRNA (Fig. 1), we considered that incorrect pairing of these bases might be related to the expression of Spc resistance, as well as their base-identities. Interestingly, by introducing a single point mutation, C1063U, in NHMcSpc1 to introduce a U-A Watson-Crick base pair (U1063-A1193), we found that Spc resistance was slightly (but significantly) reduced

Mutations MIC, μg/mL		Notes	
None	32	Wild-type 16 S rRNA gene of E. coli	
C1063U	256	Novel resistance mutation (helix 34) forming a stable (non-Watson-Crick) U–G base pair (U1063–G1193)	
C1066U	>1,024	A mutation known to confer high level resistance in E. coli <sup>23</sup>	
U1183C	1024	Novel resistance mutation (helix 38)	
U1189C	256	Novel resistance mutation (helix 34)	
U1183C/U1189C	256	Double mutant	
C1192G	>1,024	A mutation known to confer high-level resistance in E. coli <sup>22</sup>	
U1189C/C1192G	>1,024	Double mutant	
G1193A	>1,024	A mutation known to confer low-level resistance in the chloroplast <sup>21,31</sup> forming an unstable base pair (C1063–A1193)	
C1063U/G1193A	1024	Double mutant forming a stable Watson–Crick base pair (U1063–A1193)	

**Table 3.** Minimal inhibitory concentrations (MICs) of Spectinomycin (Spc) for *E. coli* MY205 harbouring wild-type or mutated *E coli* 16 S rRNA gene.

(MIC =  $512 \mu g/mL$ ) (Table 4). Additional introduction of G in the position 1093, which is expected to form a stable U-G base pair (U1063-G1193)<sup>36</sup> was also effective in significantly reducing resistance (MIC =  $256 \mu g/mL$ ) (Table 4). When A1193 was reverted to G (C1063-G1193), the resistance was completely eradicated (MIC =  $32 \mu g/mL$ ) (Table 4). These results suggest that the mode of base-pairing between nucleotides 1063 and 1193 as well as the base identity at position 1193 determines whether a given 16 S rRNA sequence expresses Spc resistance.

MY205 with mgSpc1 showed a high-level resistance (MIC  $>1,024\,\mu g/mL$ ) (Table 4, Supplementary Fig. 1C) in which two putative resistance mutations, U1189C and C1192G, were inferred (Fig. 1A, Table 2). These bases in mgSpc1 were singly or doubly reverted to the nucleotides used in *E. coli* or mgSpc1hom. C1189U alone did not significantly alter the resistance (MIC  $>1,024\,\mu g/mL$ ) (Table 4), whereas G1192C largely reduced the resistance (MIC  $=256\,\mu g/mL$ ) (Table 4). Their combination further reduced the resistance (MIC  $=64\,\mu g/mL$ ) (Table 4), suggesting that G1192 is the major determinant and C1189 is a minor determinant for Spc resistance.

MY205 with mgSpc2 showed a high-level resistance (MIC = 1,024  $\mu$ g/mL) (Table 4, Supplementary Fig. S1d), in which a single putative resistance mutation C1066U was inferred (Fig. 1A, Table 2). Reverting the U1066 nucleotide to the nucleotide used both in *E. coli* 16 S rRNA and mgSpc2hom (C) largely reduced the resistance (MIC = 256  $\mu$ g/mL) (Table 4), verifying that the C1066U mutation was the key determinant in mgSpc2 for Spc resistance, as reported in many Spc resistant bacterial 16 S rRNAs<sup>27</sup>.

MY205 with mgSpc5 showed a mild but significant resistance (MIC =  $256\,\mu\text{g/mL}$ ) (Table 4, Supplementary Fig. S1e). One unique characteristic in this gene is that the closest homologue mgSpc5hom (i.e. *T. marinus* strain IMCC1826) also shared the putative resistance mutations, U1183C and U1189C (Fig. 1A, Table 2). Both the C1183U and C1189U reverting mutations did not significantly reduce the resistance of mgSpc5, respectively (Table 4), whereas the double mutations (C1183U/C1189U) in the same gene slightly but significantly reduced the resistance (MIC =  $128\,\mu\text{g/mL}$ ) (Table 4), suggesting that the simultaneous mutation of both nucleotides is the determinant of the Spc resistance in mgSpc5.

Introducing the putative Spc resistance mutations into the *E. coli* 16 S rRNA gene. In the above subheading, we introduced reverting point mutations in metagenomically-retrieved 16 S rRNA genes to confirm that the putative resistance mutations we predicted in Table 2 were involved in Spc resistance (according to the first approach). During this process, we also addressed the significance of the mode of base-pairing between nucleotides 1063 and 1193 with respect to the expression of Spc resistance. In this subheading, we report the results of our second approach, in which the putative resistance mutations were introduced into the *E. coli* 16 S rRNA gene (*rrsB*). The effect of the mode of base-pairing between nucleotides 1063 and 1193 was also elucidated using the same *E. coli* 16 S rRNA gene.

In MY205, introducing G1193A mutation (found in NHMcSpc1) in *E. coli* 16 S rRNA gene conferred high-level resistance (MIC >1,024 µg/mL, Table 3, Supplementary Fig. S2a), confirming that this mutation renders bacterial 16 S rRNA highly resistant to Spc. The effect of G1193A, however, was slightly diminished when combined with C1063U mutation (U1063-A1193) (MIC = 1,024 µg/mL, Table 3, Supplementary Fig. S2a). C1063U single mutation (U1063-G1193) conferred a modest level of resistance (MIC = 256 µg/mL, Supplementary Fig. S2a). Collectively, strong base-pairing between nucleotides 1063 and 1193 seems to be an important factor in determining Spc susceptibility of the bacterial ribosome; C-A unstable base pair makes 16 S rRNA highly resistant to Spc while U-A and U-G stable base pairs at least partially reduce the Spc resistance compared to the C-A pair, although the U-A and U-G pairs certainly make 16 S rRNA less susceptible to Spc, compared to the cognate C-G pair (the strongest Watson-Crick base pair), respectively.

When U1189C (found in mgSpc1) was introduced into the *E. coli* 16 S rRNA gene, it rendered MY205 modestly resistant to Spc (MIC =  $256 \,\mu\text{g/mL}$ ) (Table 3, Supplementary Fig. S2b), whereas very high-level resistance (MIC >1,024  $\,\mu\text{g/mL}$ ) was conferred by C1192G (also found in mgSpc1) mutation (Table 3, Supplementary Fig. S2b), confirming the previous report showing that C1192G is a high resistance mutation<sup>26</sup> and our new finding that U1189C mutation is involved in Spc resistance. Growth phenotype of the double mutant (U1189C/C1192G) (Table 3, Supplementary Fig. S2b) was similar to that of the C1192G alone. The high resistance was

16 S rRNA genes	Mutations	MIC, μg/mL	Notes	
	None (Wild-type)	>1,024	Harbouring the putative resistance mutation G1193A (in helix 34) forming an unstable base pair (C1063–A1193)	
NHMcSpc1	C1063U	512	Introduces a stable Watson-Crick base pair (U1063-A1193)	
_	A1193G	32	Introduces a E. coli-type Watson-Crick base pair (C1063-G1193)	
	C1063U/A1193G	256	Introduces a stable (non-Watson-Crick) U-G base pair (U1063-G1193)	
	None (Wild-type)	>1,024	Contains the putative resistance mutations U1189C (helix 34) and C1192G (helix 34)	
mgSpc1	C1189U	>1,024	Reverts to the putative resistance nucleotide C1189 in <i>E. coli</i> type (U)	
	G1192C	256	Reverts to the putative resistance nucleotide G1192 in E. coli type (C)	
	C1189U/G1192C	64	Double mutant	
	None (Wild-type)	1,024	Contains the putative resistance mutation C1066U (helix 34)	
mgSpc2	U1066C	256	Reverts to the putative resistance nucleotide U1066 in E. coli type (C)	
	None (Wild-type)	256	Contains the putative resistance mutations U1183C (helix 38) and U1189C (helix 34)	
mgSpc5	C1183U	512	Reverts to the putative resistance nucleotide C1183 in <i>E. coli</i> type (U)	
	C1189U	512	Reverts to the putative resistance nucleotide C1189 in E. coli type (U)	
	C1183U/C1189U	128	Double mutant	

**Table 4.** Minimal inhibitory concentrations (MICs) of Spectinomycin (Spc) for *E. coli* MY205 derivatives harbouring the 16 S rRNA genes retrieved from the metagenome.

also reproduced in *E. coli* 16 S rRNA gene with the mutation found in mgSpc2 (C1066U) (MIC >1,024  $\mu$ g/mL) (Table 3, Supplementary Fig. S2c). As for putative resistance mutations found in mgSpc5, U1183C and U1189C mutations rendered *E. coli* 16 S rRNA highly (MIC = 1024  $\mu$ g/mL) and modestly (MIC = 256  $\mu$ g/mL) resistant to Spc, respectively (Table 3, Supplementary Fig. S2c). The MIC of the mutant carrying the double mutations (U1183C/U1189C) was similar to that observed in U1189C individual mutant (MIC = 256  $\mu$ g/mL) (Table 3, Supplementary Fig. S2c). Some of the resistance mutations (U1189C, C1192G, and G1193A) in the 16 S rRNA gene were moderately disadvantageous for the host *E. coli* MY205, consistent with the previous finding that antibiotic resistance mutations often occur in functionally important sites<sup>7</sup>. (Supplementary Fig. S4).

# Discussion

Investigating the mechanisms of antibiotic resistance has been one of the central issues in the field of ribosomal studies from both clinical and biochemical points of view. However, despite decades of effort, we still know only a little about the diversity of resistance mutations in rRNAs. More precisely, we may not recognize if we know much or little about them. In this study, we developed a general approach to systematically survey resistant rRNA genes/mutations using E.  $coli~\Delta 7$ , a null mutant of the rrn operons. The idea to use the null mutant as a surrogate host organism is based on our recent finding on the functional compatibility of  $16 \, S \, rRNAs$  between phylogenetically distant species i.e.  $16 \, S \, rRNA$  genes from the Acidobacterial lineage, which were different from E. coli~ at the phylum level, supported the growth of proteobacterial E.  $coli~\Delta 7^{20}$ . We took advantage of this unexpectedly high functional compatibility of  $16 \, S \, rRNAs$  in bacteria to characterize the functionality of metagenomic (non-E. coli~  $16 \, S \, rRNA$  genes in an E. coli~ genetic background, successfully demonstrating the utility of our method by finding novel resistance mutations to Spc, which are supposed to have been thoroughly investigated.

Historically, soon after the clinical use of Spc, lots of studies were undertaken to investigate resistance to the drug. Knowledge obtained in these five decades pointed to Spc inhibiting the translocation step in protein synthesis<sup>21</sup> by tightly binding to the helix 34 of 16 S rRNAs. Genetic studies revealed that resistance mutations exclusively localized in upper stem of helix 34 in 16 S rRNA<sup>25-34,37</sup>. It is also suggested that Spc indirectly interacts with the ribosomal protein S5<sup>38</sup>; various mutations including point mutations and deletions in S5 confer resistance<sup>39-43</sup>.

As the result of screening a mixed metagenomic library containing 16 S rRNA genes from various environments, we identified four 16 S rRNA genes that rendered the host E. coli resistant to Spc, from which five putative resistance mutations (C1066U, U1183C, U1189C, C1192G, and G1193A) were identified (Table 2, Table 4). First, we conducted systematic reverting mutation experiments for these nucleotides, and these experiments generally rendered the metagenomic 16S rRNA genes susceptible to Spc (Table 4), suggesting that these putative resistance mutations play important roles in rendering these 16 S rRNA genes resistant to Spc. As rRNAs are known to be highly conserved across kingdoms, both in structure and function, particularly for the decoding centre (including Spc binding site) and peptidyl transferase centre, we could study the mechanism of Spc resistance using E. coli as the common genetic platform without paying much attention to the species-specific uniqueness of the metagenomic 16 S rRNA genes. In fact, the five putative resistance mutations also rendered E. coli 16 S rRNA gene resistant to Spc when each mutation was individually or doubly introduced (Table 3), confirming that they are species-nonspecific resistance mutations. Among them, C1066U, C1192G and G1193A are the known mutations reported in the literature<sup>25–34</sup>. Notably, U1183C and U1189C have not been reported as Spc-resistance mutations; the former is especially interesting because it is the first mutation found in a region other than helix 34 (found in helix 38). Identification of such mutations was unexpected, but ideally coincides with a structural study of the Spc-bound form of the 30 S subunit, which suggested the occurrence of structural rearrangement of the connections between helix 34 and helices 35 and 3838. The U1189C mutation had not been described before, but was repetitively identified during our screening (mgSpc1 and mgSpc5) (Table 1). Structurally, both U1183 and U1189 do not interact directly with Spc (Supplementary Fig. S4), which makes it difficult to clearly explain the mechanism behind why mutations in these positions render 16 S rRNA resistant to Spc. The ability to discover such resistance mutations, which are usually difficult to identify from structural data, is certainly one of the merits of our new approach. In addition to the two novel mutations identified in the metagenomic 16 S rRNA genes, we also found that the existence of U in the position 1063 provided non-negligible Spc resistance both on NHMcSpc1 (MIC >1,024  $\mu$ g/mL) (Table 4) and *E. coli* 16 S rRNA (MIC = 256  $\mu$ g/mL) (Table 3). C1163 and its base-pairing partner G1193 directly interact with Spc in the crystal structure (Supplementary Fig. S4), which suggests that loss of direct interactions on their mutations renders 16 S rRNA resistant to Spc. Therefore, this C1063U mutation, as well as the other two novel resistance mutations (U1183C and U1189C), have to be added to the list of Spc resistance mutations to monitor the emergence of Spc resistant bacteria.

It should be mentioned, however, that there is one possible insufficiency in this study. Although it is generally true as mentioned above that we do not need to pay much attention about the species-specificities of 16 S rRNA genes (e.g. a specific point mutation can render both E. coli and non-E. coli 16 S rRNAs Spc resistant equally) in the common genetic background of E. coli, there were some cases in which the outcomes of the resistant mutations might be context-dependent. For example, the nucleotide 1189 locating at the upper entrance of h34 conferred a weak resistance to the host E. coli upon U to C mutation in E. coli 16 S rRNA (MIC = 256 µg/mL) (Table 3). However, reverting mutation in mgSpc1 (C1189U) did not alter the susceptibility and retained high resistance (MIC >1,024 µg/mL) (Table 4). For mgSpc5, however, slight decrease in susceptibility was observed with the same mutation (Table 4), suggesting the existence of sequence (or structure) dependence for a specific resistance mutation toward the expression of Spc resistance. In addition, mgSpc5 showed very high sequence identity with 16S rRNA from T. marinus (1464 of 1472 nucleotides identical) and the strain shares the same "mutations" in the sequence. We tested the Spc susceptibility of the strain but did not find significant resistance (MIC < 32 ug/mL). Therefore, although we have effectively identified resistance mutations by genetically reconstituting hybrid 30 S subunits (consist of non-E. coli 16 S rRNAs and E. coli ribosomal proteins), some resistance of non-E. coli 16S rRNAs may have appeared because of structural perturbation in the "artificial" hybrid ribosome, which effect could be different from "pure" ribosome in native non-E. coli bacteria. It is, however, worth noting that this situation could take place after horizontal gene transfer of 16 S rRNA between species<sup>44,45</sup>, suggesting the possibility of a non-canonical scenario for the acquisition of antibiotic resistances.

In this study, we found five resistance mutations to Spc (of which three are novel) in metagenomically-retrieved 16 S rRNA sequences, a traditional and well-studied antibiotic, demonstrating the validity of our experimental approach. The same methodology should be readily applicable to investigate the rRNA-based resistome in clinical samples and other antibiotics. For example, researchers would be able to systematically screen for aminoglycoside resistant 16 S rRNA genes from a faecal sample to estimate the percentage and/or phylogenetic origins of Spc-resistant bacteria in the large intestine at the same time. Another application could be to check the antibiotic susceptibility of a 16 S rRNA gene in a specific pathogenic bacterium that exclusively contains no known specific antibiotic resistance mutations. In such cases, if the 16 S rRNA shows resistance to the antibiotics in *E. coli*, it will strongly suggest the presence of uncharacterized and novel resistance mutation(s) in 16 S rRNA. It would also be interesting to expose environmental or medical samples to a selective pressure of specific antibiotics and select rRNA genes with resistance mutations, from which novel resistance mutations can be found more efficiently. Such experiments would help to better investigate and monitor resistance mutations in rRNAs that have yet to be studied properly.

### **Materials and Methods**

**Reagents.** Ampicillin (Amp), Tmp, and sucrose were purchased from Wako Pure Chemicals. In-Fusion Cloning Kit was purchased from Takara Bio. Lennox LB powder was purchased from Merck. Oligonucleotides were purchased from Sigma Genosys.

**Bacterial strains and growth conditions.** *E. coli* MY101 (ΔrrnG ΔrrnA ΔrrnD ΔrrnE ΔrrnB ΔrrnB ΔrrnC, pMY101,  $rna::Km^R$ ) is a derivative of SQ171 (Δ7 prrn strain)<sup>16,46</sup>, a null mutant of the rRNA (rrn) operons in the chromosome. The plasmid pMY101 (E. coli rrnB, tRNA<sup>Glu</sup>, tRNA<sup>Asp</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Ala</sup>, tRNA<sup>Trp</sup>, sacB, Amp<sup>R</sup>, pSC101 ori) was constructed by transferring the tRNA gene cluster encoded by pTRNA67<sup>16</sup> into pRB101<sup>47</sup> at the site between the 16S and 23 S rRNA genes. The resultant pMY101 plasmid contains the entire E. coli rrnB operon, and complements the growth of MY101. The strain was cultured at 37 °C in LB (1% [w/v] tryptone, 0.5% [w/v] yeast extract, 0.5% [w/v] NaCl) medium containing 100 μg/mL Amp. E. coli MY205 is a derivative of MY101, in which pMY101 was completely replaced with pMY205mPAG2 (E. coli rrnB, tRNA<sup>Glu</sup>, tRNA<sup>Asp</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Ala</sup>, tRNA<sup>Trp</sup>, Tmp<sup>R</sup>, p15a ori) using sucrose-induced counter-selection. The pMY205mPAG2 plasmid was used as a vector to introduce foreign 16S rRNA genes. The MY205 strain was cultured in LB medium containing 10 μg/mL Tmp at 37 °C. *Thalassolituus marinus* (NBRC 107590) was obtained from NITE Biological Resource Centre (Japan). It was grown in LB medium in the presence or absence of 40 μg/mL Spc at 25 °C.

Constructing and screening a metagenomic library of 16 S rRNA genes for Spc resistance. Environmental metagenomes were purified from various sources including fermentation products<sup>48</sup>, wood composts<sup>49</sup>, and activated sludge<sup>50</sup>, as described previously<sup>24</sup>. These metagenomic samples were mixed and used as the source material of the various 16 S rRNAs. The 16 S rRNA gene fragments were PCR-amplified from the metagenomic DNA as described<sup>24</sup>. Briefly, a set of primers Bac1f (5'-AAATTGAAGAGTTTGATC-3') and UN1542r (5'-TAAGGAGGTGATCCA-3') were used to amplify the full-length of the 16 S rRNA genes, which was replaced with the *E. coli* 16 S rRNA gene in pMY205mPAG2. To this effect, the vector was inversely amplified using another set of primers Bac1r (5'-GATCAAACTCTTCAATTTAAAAGTTTGACGCTCAAAG-3') and UN1542f (5'-TGGATCACCTCCTTAACGTTAAAGAAGCGT-3')<sup>24</sup>. Equimolar vector and insert fragments were combined

and ligated by incubation at 50 °C for 1 h using the In-Fusion Cloning Kit. The reaction mixture was introduced into *E. coli* JM109 and the colonies were grown on LB/Tmp agar plates at 37 °C. Plasmids were extracted from the pooled colonies (approximately 10,000) and used to transform the *E. coli*  $\Delta$ 7 strain MY101. We then recovered the colonies from the LB/Tmp agar plates (approximately 10,000), resuspended them in LB broth, appropriately diluted and spread the broth on LB/Tmp agar plates containing 5% (w/v) sucrose. After counter-selection on sucrose-containing plates to eliminate pMY101, we obtained approximately 2,000 colonies. The MY205 library was then screened on LB/Tmp agar plates containing 40 µg/mL Spc. MY205 carrying *E. coli* 16 S rRNA (as the Spc sensitive control) was constructed using pMY205mPAG2 (containing a *rrnB* operon) without mutations.

**Antibiotic resistance test.** MY205 derivatives were grown overnight in LB/Tmp at 37 °C. The saturated culture was then diluted at a concentration of 1/1,000 in LB/Tmp. The diluted culture (1  $\mu$ L) was then inoculated into 200  $\mu$ L of LB/Tmp containing varied concentrations (0–1,024  $\mu$ g/mL) of Spc in a flat-bottomed 96-well plate. The plate was then incubated with vigorous agitation (9.2 Hz) on a Sunrise Thermo RC-R plate reader (Tecan) at 37 °C and the OD<sub>600</sub> was continuously monitored every 15 min, without reducing the baseline value of the negative control in which bacteria was not inoculated into the medium. The minimum inhibitory concentration (MIC) for each MY205 derivative was defined as the minimal concentration of Spc which OD<sub>600</sub> value after 10 h-cultivation did not exceed 0.15, close to the baseline value (0.13).

**DNA sequencing and analyses.** DNA sequencing was carried out using the Sanger method with an Applied Biosystems automatic DNA sequencer (ABI PRISM 3130xl Genetic Analyzer) and an Applied Biosystems BigDye (ver. 3.1) Kit. BLAST search<sup>51</sup> was carried out on 15 August, 2017 using the NCBI nucleotide database "16 S rRNA sequences (Bacteria and Archaea)" with the program selection optimized for "Highly similar sequences (MegaBLAST)". Multiple sequence alignment was performed using the MAFFT v7 program<sup>52</sup>.

**Data and materials availability.** DNA sequence data reported in this study have been deposited under the accession numbers, LC306679 (mgSpc1), LC306680 (mgSpc2), LC306681 (mgSpc5), and LC306682 (NHMcSpc1), respectively.

## References

- Li, X. Z. & Nikaido, H. Efflux-mediated drug resistance in bacteria: an update. Drugs 69, 1555-1623, https://doi. org/10.2165/11317030-000000000-00000 (2009).
- 2. Bassetti, M. et al. Antimicrobial resistance in the next 30 years, humankind, bugs and drugs: a visionary approach. Intensive Care Med 43, 1464–1475, https://doi.org/10.1007/s00134-017-4878-x (2017).
- 3. Wilson, D. N. In *Protein* synthesis and ribosome structure (eds K. H. Nierhaus & D. N. Wilson) Ch. 12, 449–527, https://doi.org/10.1002/3527603433.ch12 (Wiley-VCH, 2004).
- 4. Lambert, T. Antibiotics that affect the ribosome. Rev Sci Tech 31, 57-64, https://doi.org/10.20506/rst.31.1.2095 (2012).
- Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. Nat Rev Microbiol 12, 35–48, https://doi. org/10.1038/nrmicro3155 (2014).
- Moazed, D. & Noller, H. F. Interaction of antibiotics with functional sites in 16 S ribosomal RNA. Nature 327, 389–394, https://doi. org/10.1038/327389a0 (1987).
- 7. Yassin, A., Fredrick, K. & Mankin, A. S. Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. *Proc Natl Acad Sci USA* **102**, 16620–16625, https://doi.org/10.1073/pnas.0508444102 (2005).
- 8. Xiong, L. et al. Oxazolidinone resistance mutations in 23S rRNA of Escherichia coli reveal the central region of domain V as the primary site of drug action. J Bacteriol 182, 5325–5331, https://doi.org/10.1128/JB.182.19.5325-5331.2000 (2000).
- 9. Kloss, P., Xiong, L., Shinabarger, D. L. & Mankin, A. S. Resistance mutations in 23 S rRNA identify the site of action of the protein synthesis inhibitor linezolid in the ribosomal peptidyl transferase center. *J Mol Biol* 294, 93–101, https://doi.org/10.1006/jmbi.1999.3247 (1999).
- Sander, P. et al. Ribosomal and non-ribosomal resistance to oxazolidinones: species-specific idiosyncrasy of ribosomal alterations. Mol Microbiol 46, 1295–1304, https://doi.org/10.1046/j.1365-2958.2002.03242.x (2002).
- 11. Brosius, J., Dull, T. J., Sleeter, D. D. & Noller, H. F. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli. J Mol Biol* 148, 107–127, https://doi.org/10.1016/0022-2836(81)90508-8 (1981).
- 12. Brosius, J. et al. Construction and fine mapping of recombinant plasmids containing the rrnB ribosomal RNA operon of E. coli. Plasmid 6, 112–118, https://doi.org/10.1016/0147-619X(81)90058-5 (1981).
- 13. Nomura, M. Engineering of bacterial ribosomes: replacement of all seven *Escherichia coli* rRNA operons by a single plasmid-encoded operon. *Proc Natl Acad Sci USA* 96, 1820–1822, https://doi.org/10.1073/pnas.96.5.1820 (1999).
- Gregory, S. T., Carr, J. F., Rodriguez-Correa, D. & Dahlberg, A. E. Mutational analysis of 16 S and 23S rRNA genes of Thermus thermophilus. J Bacteriol 187, 4804–4812, https://doi.org/10.1128/jb.187.14.4804-4812.2005 (2005).
- Gregory, S. T., Connetti, J. L., Carr, J. F., Jogl, G. & Dahlberg, A. E. Phenotypic interactions among mutations in a *Thermus thermophilus* 16S rRNA gene detected with genetic selections and experimental evolution. *J Bacteriol* 196, 3776–3783, https://doi.org/10.1128/jb.02104-14 (2014).
- Asai, T., Zaporojets, D., Squires, C. & Squires, C. L. An Escherichia coli strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. Proc Natl Acad Sci USA 96, 1971–1976, https://doi.org/10.1073/pnas.96.5.1971 (1999).
- 17. Orelle, C. et al. Tools for characterizing bacterial protein synthesis inhibitors. Antimicrob Agents Chemother 57, 5994–6004, https://doi.org/10.1128/aac.01673-13 (2013).
- 18. Polikanov, Y. S. et al. Negamycin interferes with decoding and translocation by simultaneous interaction with rRNA and tRNA. Mol Cell 56, 541–550, https://doi.org/10.1016/j.molcel.2014.09.021 (2014).
- 19. Kitahara, K., Yasutake, Y. & Miyazaki, K. Mutational robustness of 16 S ribosomal RNA, shown by experimental horizontal gene transfer in Escherichia coli. *Proc Natl Acad Sci USA* 109, 19220–19225, https://doi.org/10.1073/pnas.1213609109 (2012).
- Tsukuda, M., Kitahara, K. & Miyazaki, K. Comparative RNA function analysis reveals high functional similarity between distantly related bacterial 16 S rRNAs. Sci Rep 7, 9993, https://doi.org/10.1038/s41598-017-10214-3 (2017).
- 21. Davies, J., Anderson, P. & Davis, B. D. Inhibition of protein synthesis by spectinomycin. Science 149, 1096–1098, https://doi.org/10.1126/science.149.3688.1096 (1965).
- 22. Riesenfeld, C. S., Goodman, R. M. & Handelsman, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol* 6, 981–989, https://doi.org/10.1111/j.1462-2920.2004.00664.x (2004).

- D'Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. Sampling the antibiotic resistome. Science 311, 374–377, https://doi. org/10.1126/science.1120800 (2006).
- 24. Miyazaki, K., Sato, M. & Tsukuda, M. PCR primer design for 16 S rRNAs for experimental horizontal gene transfer test in *Escherichia coli. Front Bioeng Biotechnol* 5, 14, https://doi.org/10.3389/fbioe.2017.00014 (2017).
- 25. Fromm, H., Edelman, M., Aviv, D. & Galun, E. The molecular basis for rRNA-dependent spectinomycin resistance in *Nicotiana chloroplasts*. *EMBO J* 6, 3233–3237 (1987).
- 26. Brink, M. F., Brink, G., Verbeet, M. P. & de Boer, H. A. Spectinomycin interacts specifically with the residues G1064 and C1192 in 16 S rRNA, thereby potentially freezing this molecule into an inactive conformation. *Nucleic Acids Res* 22, 325–331, https://doi.org/10.1093/nar/22.3.325 (1994).
- 27. Johanson, U. & Hughes, D. A new mutation in 16 S rRNA of *Escherichia coli* conferring spectinomycin resistance. *Nucleic Acids Res* 23, 464–466, https://doi.org/10.1093/nar/23.3.464 (1995).
- 28. Bilgin, N., Richter, A. A., Ehrenberg, M., Dahlberg, A. E. & Kurland, C. G. Ribosomal RNA and protein mutants resistant to spectinomycin. EMBO J 9, 735–739, https://doi.org/10.1002/j.1460-2075.1990.tb08167.x (1990).
- 29. Criswell, D., Tobiason, V. L., Lodmell, J. S. & Samuels, D. S. Mutations conferring aminoglycoside and spectinomycin resistance in *Borrelia burgdorferi*. *Antimicrob Agents Chemother* **50**, 445–452, https://doi.org/10.1128/aac.50.2.445-452.2006 (2006).
- 30. Galimand, M., Gerbaud, G. & Courvalin, P. Spectinomycin resistance in *Neisseria* spp. due to mutations in 16 S rRNA. *Antimicrob Agents Chemother* 44, 1365–1366, https://doi.org/10.1128/AAC.44.5.1365-1366.2000 (2000).
- 31. Makosky, P. C. & Dahlberg, A. E. Spectinomycin resistance at site 1192 in 16 S ribosomal RNA of *E. coli*: an analysis of three mutants. *Biochimie* 69, 885–889, https://doi.org/10.1016/0300-9084(87)90216-1 (1987).
- Mark, L. G., Sigmund, C. D. & Morgan, E. A. Spectinomycin resistance due to a mutation in an rRNA operon of Escherichia coli. J Bacteriol 155, 989–994 (1983).
- O'Connor, M. & Dahlberg, A. E. Isolation of spectinomycin resistance mutations in the 16S rRNA of Salmonella enterica serovar Typhimurium and expression in Escherichia coli and Salmonella. Curr Microbiol 45, 429–433, https://doi.org/10.1007/s00284-002-3684-y (2002).
- 34. Sigmund, C. D., Ettayebi, M. & Morgan, E. A. Antibiotic resistance mutations in 16 S and 23S ribosomal RNA genes of *Escherichia coli*. *Nucleic Acids Res* 12, 4653–4663, https://doi.org/10.1093/nar/12.11.4653 (1984).
- Harris, E. H., Burkhart, B. D., Gillham, N. W. & Boynton, J. E. Antibiotic resistance mutations in the chloroplast 16 S and 23S rRNA genes of *Chlamydomonas reinhardtii*: correlation of genetic and physical maps of the chloroplast genome. *Genetics* 123, 281–292 (1989).
- 36. Varani, G. & McClain, W. H. The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep* 1, 18–23, https://doi.org/10.1093/embo-reports/kyd001 (2000).
- 37. Galimand, M., Courvalin, P. & Lambert, T. Plasmid-mediated high-level resistance to aminoglycosides in Enterobacteriaceae due to 16S rRNA methylation. *Antimicrob Agents Chemother* 47, 2565–2571, https://doi.org/10.1128/AAC.47.8.2565-2571.2003 (2003).
- 38. Carter, A. P. et al. Functional insights from the structure of the 30 S ribosomal subunit and its interactions with antibiotics. *Nature* 407, 340–348, https://doi.org/10.1038/35030019 (2000).
- 39. Piepersberg, W., Bock, A., Yaguchi, M. & Wittmann, H. G. Genetic position and amino acid replacements of several mutations in ribosomal protein S5 from Escherichia coli. Mol Gen Genet 143, 43–52, https://doi.org/10.1007/BF00269419 (1975).
- 40. Itoh, T. Amino acid replacement in the protein S5 from a spectinomycin resistant mutant of *Bacillus subtilis. Mol Gen Genet* 144, 39–42, https://doi.org/10.1007/BF00277301 (1976).
- 41. Unemo, M. et al. Neisseria gonorrhoeae strain with high-level resistance to spectinomycin due to a novel resistance mechanism (mutated ribosomal protein S5) verified in Norway. Antimicrob Agents Chemother 57, 1057–1061, https://doi.org/10.1128/aac.01775-12 (2013).
- 42. Kehrenberg, C. & Schwarz, S. Mutations in 16 S rRNA and ribosomal protein S5 associated with high-level spectinomycin resistance in *Pasteurella multocida*. Antimicrob Agents Chemother 51, 2244–2246, https://doi.org/10.1128/aac.00229-07 (2007).
- Ilina, E. N. et al. Mutation in ribosomal protein S5 leads to spectinomycin resistance in Neisseria gonorrhoeae. Front Microbiol 4, 186, https://doi.org/10.3389/fmicb.2013.00186 (2013).
- 44. Kitahara, K. & Miyazaki, K. Revisiting bacterial phylogeny: Natural and experimental evidence for horizontal gene transfer of 16S rRNA. *Mob Genet Elements* 3, e24210, https://doi.org/10.4161/mge.24210 (2013).
- 45. Sato, M. & Miyazaki, K. Phylogenetic Network analysis revealed the occurrence of horizontal gene transfer of 16 S rRNA in the genus *Enterobacter. Front Microbiol* 8, 2225, https://doi.org/10.3389/fmicb.2017.02225 (2017).
- 46. Quan, S., Skovgaard, O., McLaughlin, R. E., Buurman, E. T. & Squires, C. L. Markerless Escherichia coli rrn deletion strains for genetic determination of ribosomal binding sites. G3(5), 2555–2557, https://doi.org/10.1534/g3.115.022301 (2015).
- 47. Kimura, S. et al. Base methylations in the double-stranded RNA by a fused methyltransferase bearing unwinding activity. Nucleic Acids Res 40, 4071–4085, https://doi.org/10.1093/nar/gkr1287 (2012).
- 48. Uchiyama, T., Yaoi, K. & Miyazaki, K. Glucose-tolerant beta-glucosidase retrieved from a Kusaya gravy metagenome. Front Microbiol 6, 548, https://doi.org/10.3389/fmicb.2015.00548 (2015).
- 49. Uchiyama, T., Miyazaki, K. & Yaoi, K. Characterization of a novel beta-glucosidase from a compost microbial metagenome with strong transglycosylation activity. *J Biol Chem* 288, 18325–18334, https://doi.org/10.1074/jbc.M113.471342 (2013).
- 50. Suenaga, H., Ohnuki, T. & Miyazaki, K. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ Microbiol* 9, 2289–2297, https://doi.org/10.1111/j.1462-2920.2007.01342.x (2007).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. J Mol Biol 215, 403–410, https://doi.org/10.1016/s0022-2836(05)80360-2 (1990).
- 52. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780, https://doi.org/10.1093/molbev/mst010 (2013).

#### Acknowledgements

This work was partly supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Challenging Exploratory Research 26670219 awarded to K.M. and JSPS Grant-in-Aid for Young Scientists (B) 25830132 awarded to K.K.

# **Author Contributions**

K.M. and K.K. designed the study, conducted the data analysis, and wrote the manuscript.

# **Additional Information**

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-23474-4.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>.

© The Author(s) 2018