

# SCIENTIFIC REPORTS



OPEN

## Integrative metagenomic and biochemical studies on rifamycin ADP-ribosyltransferases discovered in the sediment microbiome

Jae Hong Shin<sup>1</sup>, Hyunuk Eom<sup>2</sup>, Woon Ju Song<sup>2</sup> & Mina Rho<sup>1,3</sup>

Antibiotic resistance is a serious and growing threat to human health. The environmental microbiome is a rich reservoir of resistomes, offering opportunities to discover new antibiotic resistance genes. Here we demonstrate an integrative approach of utilizing gene sequence and protein structural information to characterize unidentified genes that are responsible for the resistance to the action of rifamycin antibiotic rifampin, a first-line antimicrobial agent to treat tuberculosis. Biochemical characterization of four environmental metagenomic proteins indicates that they are adenosine diphosphate (ADP)-ribosyltransferases and effective in the development of resistance to FDA-approved rifamycins. Our analysis suggests that even a single residue with low sequence conservation plays an important role in regulating the degrees of antibiotic resistance. In addition to advancing our understanding of antibiotic resistomes, this work demonstrates the importance of an integrative approach to discover new metagenomic genes and decipher their biochemical functions.

Antibiotics are indispensable therapeutic agents for human health against infection from pathogenic bacteria. The emergence of resistance against antibiotics, however, has outpaced the discovery of new antibiotics. The Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) declared that we are in the post-antibiotic era and are facing an antibiotic resistance crisis. This critical situation urges for immediate actions in politics, research, and health care systems<sup>1</sup>.

Tuberculosis is a widespread infectious disease, which is caused mostly by *Mycobacterium tuberculosis*<sup>2,3</sup>. Currently, nearly one-third of the global population is infected by the pathogenic bacteria. Rifampin, also known as rifampicin, belongs to the rifamycin family, and is a first-line antimicrobial agent against tuberculosis. Rifampin binds to the  $\beta$  subunit of DNA-dependent RNA polymerase, and inhibits the RNA synthesis of bacteria through steric occlusion<sup>4,5</sup>.

Since rifampin is an effective anti-tuberculosis agent, molecular mechanisms underlying the resistance to rifampin have been studied extensively<sup>6–9</sup>. A few mutations in bacterial RNA polymerase attenuate the binding affinity of rifampin<sup>10,11</sup>. Alternatively, rifampin can be modified by glycosylation<sup>12</sup>, adenosine diphosphate (ADP)-ribosylation<sup>13</sup>, monooxygenation<sup>14</sup>, or phosphorylation<sup>7</sup>, which also results in decreased binding affinity to RNA polymerase. The glycosyltransferase Rgt1438 confers rifampin resistance by converting rifampin to 23-O-glu-rifampin<sup>12</sup>. Rifampin ADP-ribosyltransferases (Arr) attach an ADP-ribosyl group to the hydroxyl group at the C21 position of rifampin to form ADP-ribosylrifampin (Additional file 1: Scheme 1a)<sup>13,15–17</sup>. Rifampin phosphotransferase (RPH), recently identified from *Listeria monocytogenes*<sup>7</sup>, is a kinase that modifies rifampin to phospho-rifampin at the C21 position<sup>7,18</sup>. The diversity in the chemical origins implies that antibiotic resistance can readily be acquired by multiple organisms. Therefore, antibiotic resistance genes and related proteins should be explored as immediate targets of research to understand comprehensive antibiotic resistomes<sup>19,20</sup>.

With the advances in sequencing technology, metagenomic approaches have been applied to identify antibiotic resistance genes from environmental samples<sup>21–24</sup>. Since many antibiotics originate from secondary metabolites of bacteria<sup>25,26</sup>, resistance determinants are ubiquitously found in microorganisms, even in ancient

<sup>1</sup>Department of Computer Science and Engineering, Hanyang University, Seoul, Korea. <sup>2</sup>Department of Chemistry, Seoul National University, Seoul, 08826, Korea. <sup>3</sup>Department of Biomedical Informatics, Hanyang University, Seoul, Korea. Jae Hong Shin and Hyunuk Eom contributed equally to this work. Correspondence and requests for materials should be addressed to W.J.S. (email: [woonjusong@snu.ac.kr](mailto:woonjusong@snu.ac.kr)) or M.R. (email: [minarho@hanyang.ac.kr](mailto:minarho@hanyang.ac.kr))

environments<sup>27–31</sup>. As such, reservoirs of resistance determinants have been investigated to profile abundant resistance genes and their effects on microbiome composition<sup>32</sup>. Nesme *et al.* analyzed antibiotic resistance gene determinants from 71 environmental shotgun metagenomic samples from a global environment survey<sup>33</sup>. Li *et al.* applied network analysis to identify 18 types and 260 subtypes of antibiotic resistance genes<sup>34</sup>.

In order to deduce the functions of particular genes and proteins, sequence similarity has been widely applied as an important measure. For example, resistome profiling<sup>35</sup> or functional metagenomics<sup>36,37</sup> have been applied to reveal the genes related to the antibiotic resistances. Studies of conserved residues alone, however, might not be sufficient for predicting the function of the genes<sup>38</sup>. An inherent limitation of sequence similarity-based approaches becomes more pronounced in determining the functions of newly discovered genes<sup>39</sup>. Recently, Balskus *et al.* have developed a profiling method by combining protein sequence similarity network analysis with quantitative metagenomics to discover an unidentified enzyme, suggesting that additional information can greatly improve the quality of metagenomic data<sup>40</sup>.

Within this context, we have developed an integrative approach of combining sequence analysis and molecular interaction analysis of the protein structure with highest sequence similarity. Potential *arr* genes (*arr-wd*), obtained from the sediment microbiome, were analyzed to locate conserved functional sites. As an experimental validation of our computational prediction, *in vitro* and *in vivo* resistance activity assays were carried out with rifamycins to assess the ADP-ribosyltransferase activity of expressed proteins. This work represents an example of using microbiome sequence data to discover proteins related to the antibiotic resistance, to construct detailed three-dimensional models to extract functionally important structural features, and to validate computational predictions by biochemical experimental studies.

## Results and Discussion

**Novel rifamycin ADP-ribosyltransferase genes in the sediment microbiome.** A total of 45 rifamycin ADP-ribosyl transferase (*arr*) homologous genes (*arr-wd*) were obtained from the sediment microbiome with the query of seven *arr* genes annotated in the comprehensive antibiotic resistance database (CARD) (Additional file 1: Table S1)<sup>41,42</sup>. The similarity between the seven known *arr* genes and 45 *arr-wd* genes ranges from 48.23 to 71.01%. These 45 putative *arr* genes were aligned with the seven known *arr* genes to find the conserved functional regions in the amino acid sequences (Fig. 1a). ADP-ribosyltransferases exhibit multiple conserved motifs that are presumably related to rifampin binding and nicotinamide adenine dinucleotide (NAD<sup>+</sup>) binding through hydrogen bonding, hydrophobic interactions, and cation- $\pi$  interactions (Fig. 1b). Our results include the rifampin and NAD<sup>+</sup> binding sites reported in the previous studies of *arr-ms* from *Mycobacterium smegmatis*, *arr-sc* from *S. coelicolor*, and *arr-2* and homologous genes (*arr-3*, *arr-4*, *arr-5*, *arr-7*, and *arr-8*) from several gram-negative pathogenic bacteria<sup>13,43–46</sup>. In particular, H19 and Y49, which were identified for NAD<sup>+</sup> binding<sup>13</sup>, were highly conserved in our alignment.

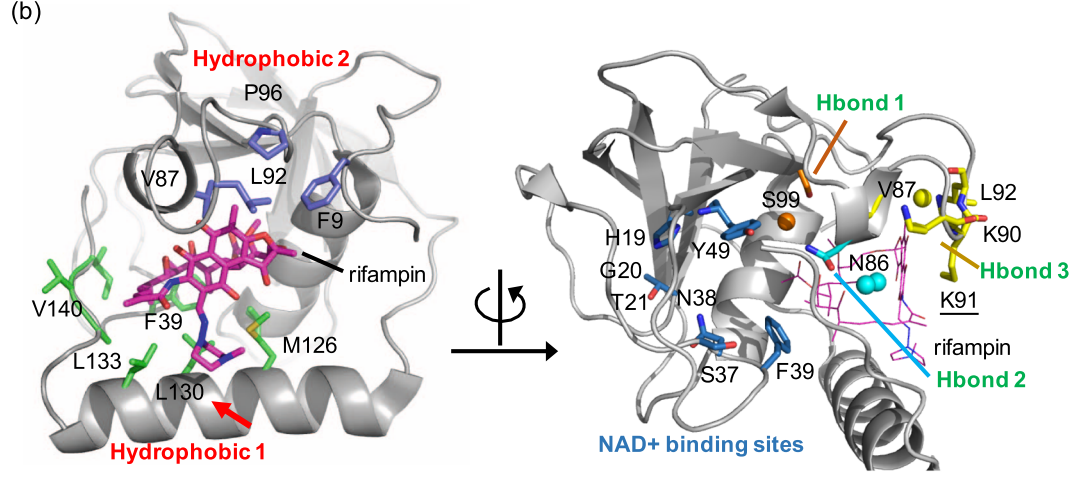
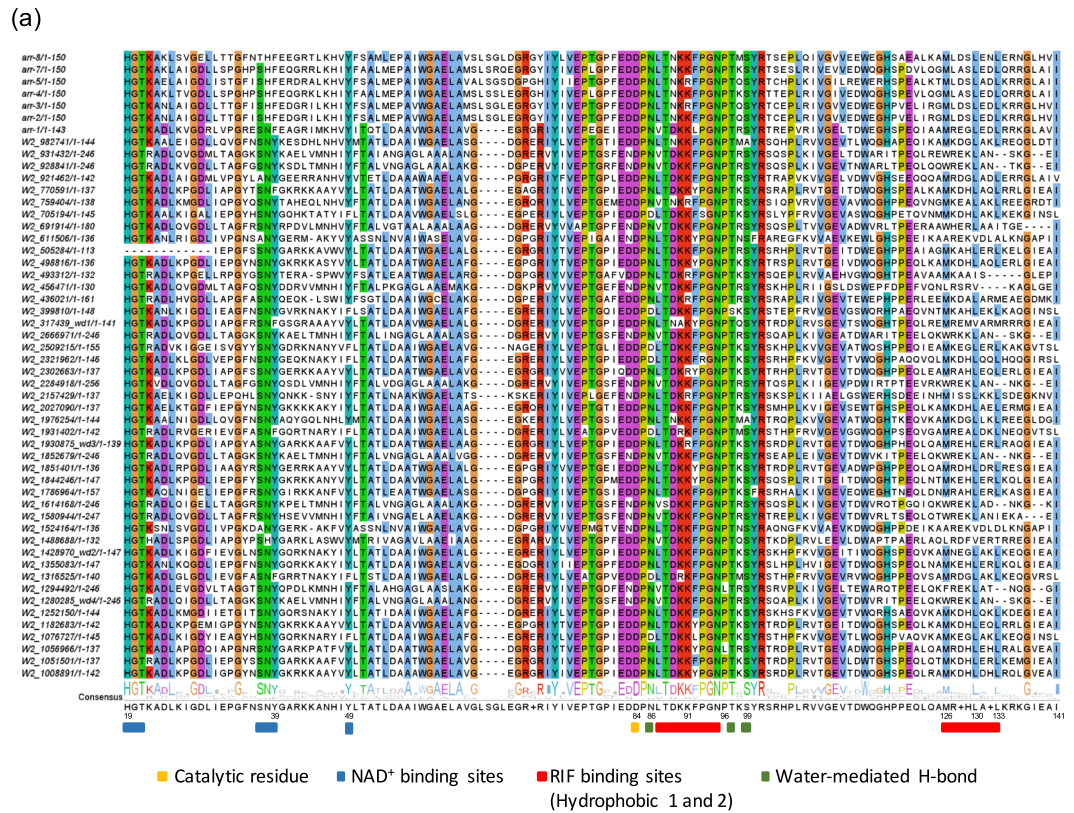
**Structure-guided conservation analysis.** It has been proposed that NAD<sup>+</sup> binding sites in ADP-ribosyltransferases are formed by a long flexible loop between  $\beta$ 1 and  $\beta$ 3 strands (Fig. 1b; Additional file 1: Figure S1), and several residues constituting the loop are essential for the function of Arr<sup>13</sup>. A HGT motif (H19, G20, and T21 in Arr-*ms*) at the end of  $\beta$ 1 strand and a tyrosine (Y49 in Arr-*ms*) in  $\beta$ 3 strand are spatially adjacent to an aspartate (D84). Site-directed mutagenesis studies have demonstrated that these amino acids, H19, Y49, and D84 in Arr-*ms*, are critical for the catalysis through hydrogen-bonding interactions with NAD<sup>+</sup><sup>13</sup>. Indeed, these residues are highly conserved in the putative rifampin ADP-ribosyltransferase genes of the current study: 98.0% for H19, 87.0% for Y49, and 100% for D84 (Fig. 1a).

In addition to the HGT motif, we found that SN[Y|F] motif in the flexible loop region between  $\beta$ 2 and  $\beta$ 3 strands (Figs 1a and 2b; Additional file 1: Figure S1) is highly conserved (92.0% for S, 87.0% for N, 79% for Y, 21% for F). Although the X-ray crystal structure of Arr complexed with NAD<sup>+</sup> is not available, a docking simulation indicates that the SN[Y|F] motif is likely to be located at the entrance to the active site, providing the shape and/or charge complementarity to the sugar-phosphate group of NAD<sup>+</sup> (Additional file 1: Figure S2). Presumably, such non-covalent interaction with HGT and SN[Y|F] motifs polarizes the nicotinamide portion of NAD<sup>+</sup> to help generate the oxocarbenium intermediate.

The X-ray crystal structure of Arr-*ms* (PDB ID 2HW2)<sup>13</sup> displays a large binding pocket to accommodate rifampin (solvent accessible surface area = 621.95 Å<sup>2</sup>), comprised of amino acid residues from  $\alpha$ 1 and  $\alpha$ 2 helices and a loop between  $\beta$ 5 and  $\beta$ 6 strands. We categorized the residues into four types, depending on their molecular interactions, *hydrophobic 1*, *hydrophobic 2*, *cation- $\pi$* , and *water-mediated hydrogen bond* (Fig. 1b). High conservation was indeed observed in the *arr* genes that we predicted from the metagenomic data. All key amino acid residues comprising the rifampin-binding site were well-conserved (Fig. 1a).

The *hydrophobic 1* region is a main hydrophobic pocket for rifampin-binding. It is comprised of hydrophobic bulky residues, F39, M126, L130, L133, A139, and I141 in Arr-*ms* (Fig. 1b). The three residues M126, L130, and L133, surrounding the piperazine ring portion of rifampin, are located in helix  $\alpha$ 2. The piperazine ring of rifampin is likely to be inserted between L133 and M126 by a hydrophobic interaction. L130 also possesses hydrophobic characteristics and is highly conserved in our metagenome sequences [L(88.0%)|V(10.0%)|I(2.0%)], thus contributing to the formation of hydrophobic pocket that is positioned close to the piperazine ring. In helix  $\alpha$ 2, M[R|K]|D[E]XL residues seem to be essential for rifampin-binding; they are highly conserved for both the reference and metagenome sample sequences. Since the methionine, the fourth amino acid (labeled as X), and two leucine residues are positioned toward the pocket in Arr-*ms*, a variation of these key residues would regulate the binding affinity and/or activities. In the aliphatic ansa chain of rifampin (Fig. 2a), two methyl groups (C31 and C32) and diene units are surrounded by F39, A139, and I141, which allow hydrophobic interactions.

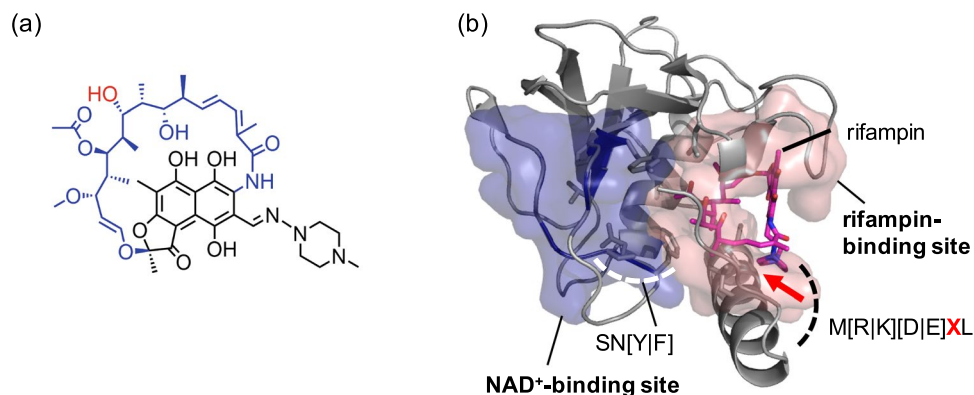
The *hydrophobic 2* region is a shallow and rather surface-exposed pocket. This region consists of F9, V87, L92, and P96 (Fig. 1b). While the residues F9, L92, and P96 are positioned in the loops, and directly exposed to the



**Figure 1.** Integrated analysis of rifampicin ADP-ribosyltransferase (Arr). (a) The sequence conservation of rifampicin ADP-ribosyltransferase (Arr). The residues that are conserved more than 78% of the sequences are highlighted based on the clustalw color scheme. The genes selected for experimental validation are shown as wd1–wd4. (b) Proposed key residues and their molecular interactions in the rifampin-binding sites of Arr-*ms* structure (PDB 2HW2). In Arr-*ms*, the variable X residue in M[R]K[D]E]XL motif is Gly129, and its position is highlighted with a red arrow. Water molecules are depicted with spheres.

aqueous environment, they are spatially close to engage in hydrophobic interaction with a methyl group (C14) of rifampin. The residues L92 and P96 are located in the loop structure between  $\beta 5$  and  $\beta 6$  strands, and are part of the rifampin binding loop<sup>13</sup>. The residue V87 is positioned between  $\beta 5$  and  $\beta 6$  strands, generating a small  $\alpha$ -helix structure.

K91 is located in the flexible loop structure between  $\beta 5$  and  $\beta 6$  strands (Fig. 1b). The residue is positioned above the naphthoquinone ring portion of rifampin, and the positively charged  $\epsilon$ -ammonium ( $\text{NH}_3^+$ ) of K91 makes a cation- $\pi$  contact. The residue is highly conserved in most of the Arr protein, constituting 85% of lysine and 15% of arginine. This spatial arrangement effectively shields the aromatic rings from exposure to the aqueous environment. The adjacent amino acids are also highly conserved as [L]V]TK[K]R]P: 73.0% for L; 98.0% for T; 94.0% for K; 85% for K; 96.0% for P.



**Figure 2.** Structural analysis of rifampin ADP-ribosyltransferase (Arr). (a) A structure of rifampin. The aliphatic ansa chain and the site of modification by ADP-ribosyltransferase is colored in blue and red, respectively. (b) Two substrate-binding pockets depicted in the structure of rifampin ADP-ribosyltransferase from *arr-ms* (PDB 2HW2). The variable X residue in M[R|K][D|E]XL motif is highlighted with a red arrow.

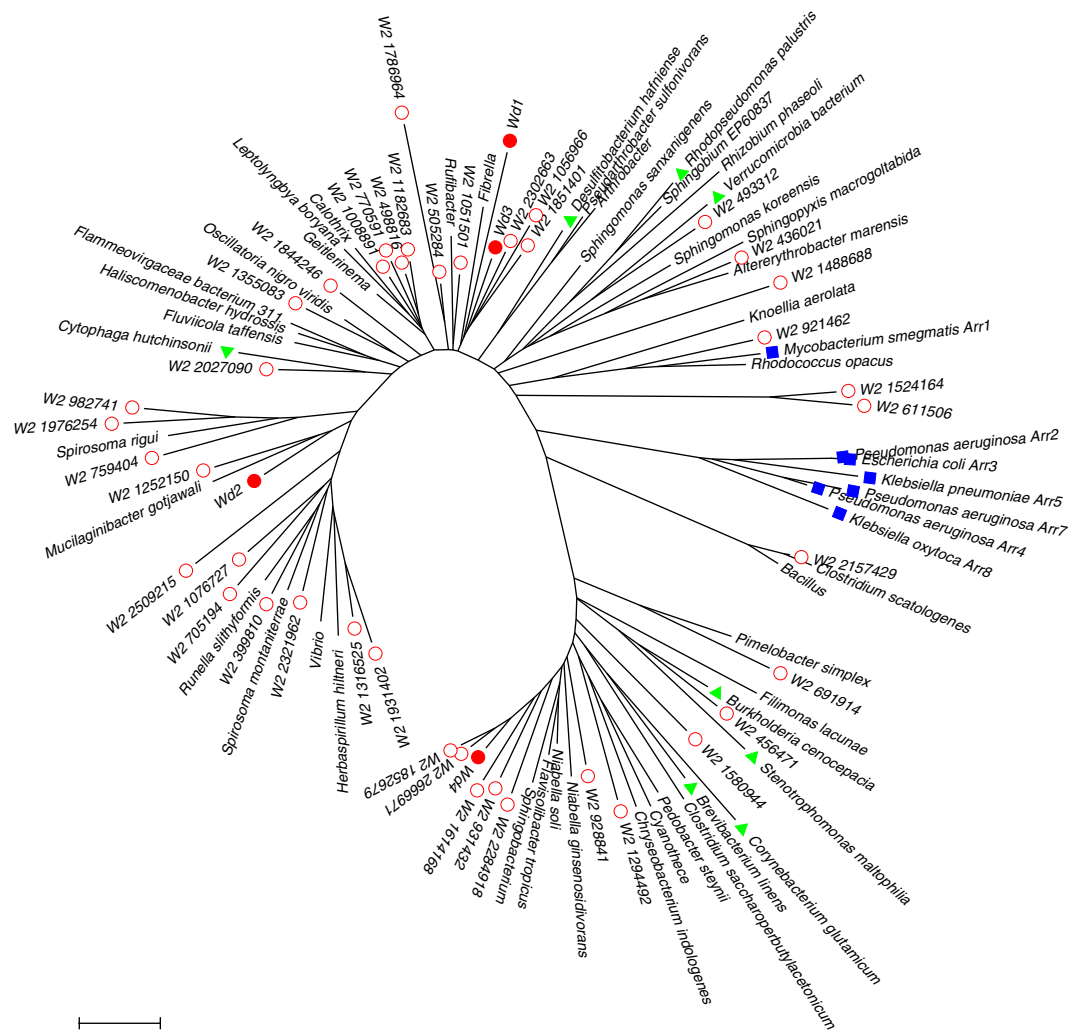
In addition to the rifampin–Arr-*ms* interactions discussed above, we also observe three *water-mediated hydrogen bonds*, which might be necessary for rifampin binding (Fig. 1b, Additional file 1: Figure S1). In the rifampin binding loop, a water molecule (numbered as 1226) forms hydrogen bonding networks between a hydroxyl group (O2, 3.07 Å) of the rifampin naphthoquinone ring and backbone amide groups from V87, T88, K90, and L92 residues (2.89–3.04 Å). As discussed above, these residues are also involved in the hydrophobic interaction with rifampin, which might further contribute to the binding affinity. Another water molecule (numbered as 1335) mediates hydrogen bonding between the carbonyl oxygen atom of the O-acetyl (O8, 3.26 Å) group of rifampin and the hydroxyl group of S99 (OG, 2.77 Å). In addition, two water molecules (numbered as 1230 and 1381) form hydrogen bonding networks between N86, V140, and the N–H group of rifampin. These conserved residues are represented as DPN[V|L]. Here, D84 denotes the catalytic residue for the ADP-ribosylation function in Arr-*ms*<sup>13</sup>.

**Phylogenetic relationship of *arr* genes.** In order to find the phylogenetic relationship of the newly found *arr* genes, *arr-wd*, with the known genes, a neighbor-joining tree was built. In the phylogenetic tree, four *arr-wd* genes (*arr-wd1-4*) experimentally validated in this study were analyzed with the known *arr* genes (Fig. 3). In addition to *arr-ms* and known *arr* genes from gram-negative bacteria (*arr-2*, *arr-3*, *arr-4*, *arr-5*, *arr-7*, and *arr-8*; denoted as blue filled rectangles in Fig. 3), *arr-wd1*, *arr-wd2*, *arr-wd3*, and *arr-wd4* (denoted as red filled circles in Fig. 3) are placed with the homologous *arr* genes identified from the complete genomes. In particular, *arr-wd1*, *arr-wd2*, and *arr-wd3* are clustered with the *arr* genes from *Rhodopseudomonas palustris*, *Desulfitobacterium hafniense*, *Verrucomicrobium spinosum*, and *Cytophaga hutchinsonii*, which were used in the phylogenetic tree of a previous study<sup>13</sup>. On the other hand, *arr-wd4* is clustered with the genes from *Brevibacterium linens*, *Corynebacterium glutamicum*, *Burkholderia cenocepacia*, and *Stenotrophomonas maltophilia*. While the previous study showed a wide distribution of *arr* genes in the phylogenetic tree<sup>13</sup>, our study expanded the phylogenetic tree by incorporating diverse *arr* genes from an environmental sample and validating them experimentally.

To find the origins of the microbial genomes, the *arr-wd* genes were investigated by using homology search. We found that the *arr* gene W2\_2157429 shows the highest homology (98.54% protein similarity) to *Clostridium ljungdahlii*. A total of 13 out of 45 *arr-wd* genes show high homology (>80% protein similarity) to the known species (Additional file 1: Table S2), implying that *arr-wd* like genes might be prevalent in various organisms, although they are yet to be annotated.

**Biochemical analysis of representative *arr* genes.** To evaluate whether the newly discovered genes display ADP-ribosyltransferase activity, we selected four metagenomic genes (*arr-wd 1-4*) and expressed the corresponding proteins (Additional file 1: Table S3 and Figure S3). All four putative *arr* genes possess the conservation of the essential functional residues (Fig. 1a), comprised of NAD<sup>+</sup> and rifampin-binding pockets (Fig. 2), and exhibit rifampin ADP-ribosyltransferase activities (Fig. 4; Additional file 1: Tables S5–S6 and Figure S4). The activities are relatively lower than the previously characterized *arr* genes<sup>13</sup> by about 1–2 orders of magnitude in both turnover number ( $k_{\text{cat}}$ ) and catalytic efficiency ( $k_{\text{cat}}/K_{\text{M}}$ ). The data, however, definitively indicate that all Arr-*wd* proteins discovered herein are ADP-ribosyltransferases that modify rifampin upon the reaction with NAD<sup>+</sup>, and the metagenomic sequences can be denoted as *arr* genes that induce the resistance of rifampin. Interestingly, Michaelis constant ( $K_{\text{M}}$ ) and the dissociation constant ( $K_{\text{D}}$ ) of the Arr-*wd* proteins for rifampin are significantly lower than those of the previously reported Arr proteins (Fig. 4d,e; Additional file 1: Table S7). These results indicate that Arr-*wd* proteins exhibit tighter binding to the rifampin than other Arr proteins, and the *arr-wd* genes could start to develop resistance even at low concentration levels of antibiotics when sufficient  $k_{\text{cat}}$  is provided.

The binding constants for NAD<sup>+</sup> and rifampin were also measured with Arr-*wd3* protein without the His-tag (Arr-*wd3* +  $\Delta 6\text{His}$ ) (Additional file 1: Table S7). The protein yields only slightly perturbed values, indicating that Arr-*wd* proteins expressed with the His-tag are valid for evaluating the biochemical activities.



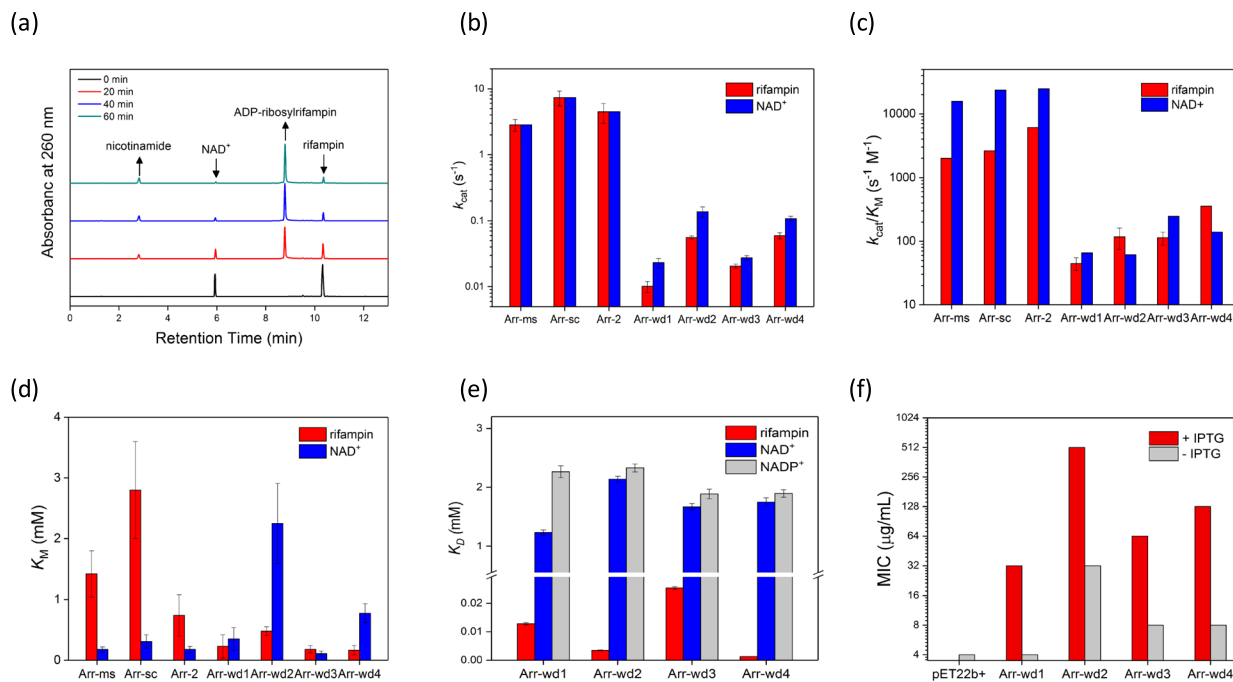
**Figure 3.** A phylogenetic tree of the metagenomic *arr* genes and well-known *arr* genes. Red circles denote *arr* genes collected from the metagenomic sample in this work. Red filled circles denote *arr* genes validated experimentally in this work. Blue rectangles denote previously known and experimentally validated *arr* genes. Green triangles denote *arr* genes referenced previously but not validated experimentally.

The *in vivo* activities of the proteins were determined by the minimum inhibitory concentration (MIC) values (Fig. 4f; Additional file 1: Table S9). Their MIC values are slightly lower than other *arr* proteins, presumably due to their lower *in vitro* activities of Arr-wd proteins. Their MIC values, however, are considerably higher than the background level, again suggesting that the *arr-wd* genes can inactivate the antibiotics in considerable rates. The ADP-ribosyltransferase activities measured both *in vitro* and *in vivo* conditions, therefore, convincingly demonstrate that we have discovered a novel set of *arr* genes from environmental metagenomes, and have predicted their biochemical function.

**Biochemical analysis of NAD<sup>+</sup> binding sites.** Our structure-guided conservation analysis and *in vitro* activity assay described above suggest that putative Arr-wd proteins are likely to have structurally well-defined NAD<sup>+</sup> and rifampin-binding sites, which are essential for the dedicated chemical reaction. When NAD<sup>+</sup> was replaced with nicotinamide adenine dinucleotide phosphate (NADP<sup>+</sup>), NADH, and NADPH, no ADP-ribosyltransferase activity was observed, indicating that Arr-wd proteins exhibit selective interaction with NAD<sup>+</sup> over other analogous cofactors.

As described above, the SN[Y|F] motif is highly conserved, composing the putative NAD<sup>+</sup>-binding site: SNF in Arr-wd1 and SNY in Arr-wd2-4. The variation of the third residue between Phe and Tyr seems to be insignificant because there is no clear correlation between the sequence and binding affinity ( $K_D$ ) or ADP-ribosyltransferase activity (Additional file 1: Tables S6–S7 and Figures S4–S5). Therefore, only the hydrophobic interactions with the benzene ring of the side chain seems to be sufficient for retaining the interaction with NAD<sup>+</sup>. Since the Tyr or Phe residue is located at the intersection of NAD<sup>+</sup> and rifampin-binding sites, it is possible that this hydrophobic residue also protects the oxocarbenium ion from other side reactions (Additional file 1: Scheme S1b and Figure S2).

Our sequence analysis also demonstrate that a large portion of the previously reported genes have SH[F], instead of the SN[Y|F] motif, with histidine (H) rather than asparagine (N) in the second position. To



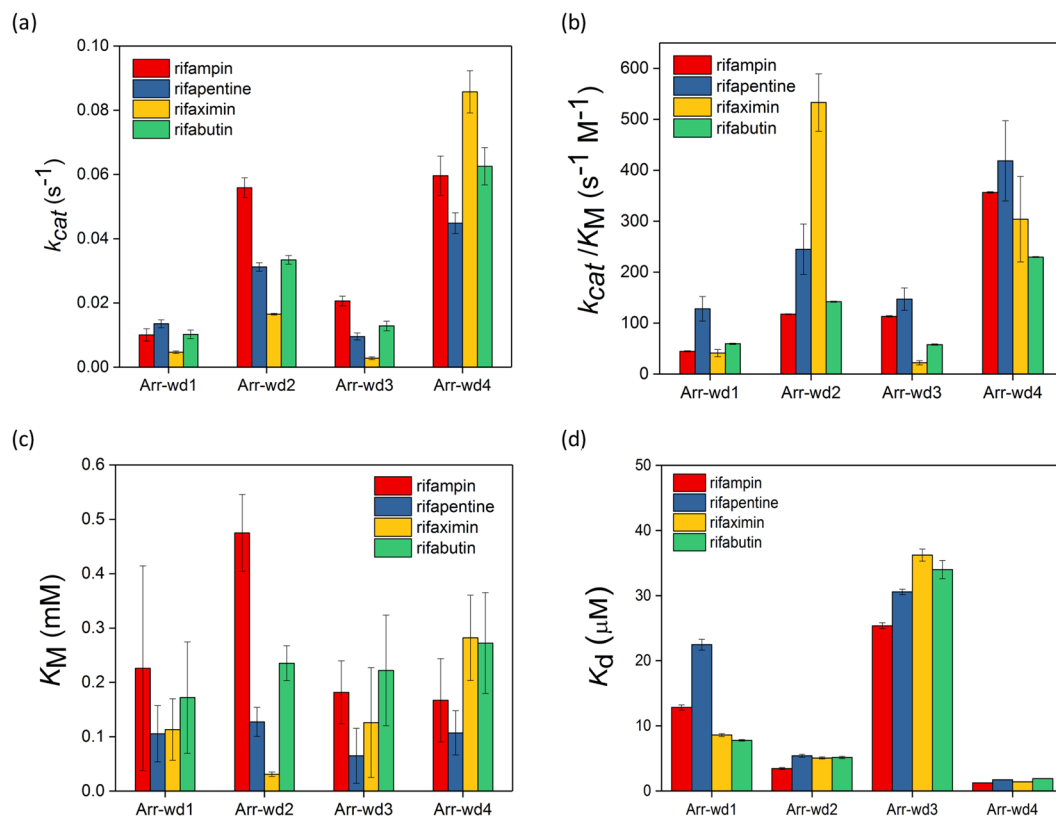
**Figure 4.** Biochemical analysis of Arr-wd 1–4 proteins. **(a)** A representative HPLC trace of rifampin ADP-ribosylation by Arr-wd proteins. **(b–d)** Steady-state ADP-ribosyltransferase activities of Arr-wd proteins together with the reported values of Arr proteins. **(b)** Turnover numbers,  $k_{\text{cat}}$  ( $\text{s}^{-1}$ ) **(c)** Catalytic efficiencies,  $k_{\text{cat}}/K_M$  ( $\text{s}^{-1} \text{M}^{-1}$ ) **(d)** Michaelis constants,  $K_M$  (mM) **(e)** Dissociation constants,  $K_D$  (mM) **(f)** MIC values ( $\mu\text{g}/\text{mL}$ ). Values reported in **(b–d)** and **(e)** are derived from the fit to the Michaelis-Menten equation and Stern-Volmer equation, respectively.

evaluate whether the sequence variation is related to the action of rifampin ADP-ribosyltransferase, we carried out site-directed mutagenesis of one of the metagenomes, arr-wd3. The N34H mutant of Arr-wd3 yields 6-fold lower  $k_{\text{cat}}/K_M$  and slightly elevated  $K_D$  values than Arr-wd3 (Additional file 1: Tables S7–S8). The results imply that the residue is involved in  $\text{NAD}^+$ -binding and/or positioning, and the N34H mutation perturbs the interactions with  $\text{NAD}^+$ . Interestingly, the  $K_D$  value for rifampin has been also altered by the N34H mutation, suggesting that the mutation might have allosterically influenced the thermodynamic parameters associated with the rifampin-binding event.

**Biochemical analysis of rifampin-binding sites.** Steady-state activity data demonstrate that the Arr-wd proteins are kinetically competent in rifampin ADP-ribosylation, although their activities are lower than those of the previously reported Arr proteins. Notably, however, the Arr-wd proteins exhibit considerably lower  $K_M$  and  $K_D$  values for rifampin, implicating that Arr-wd proteins can interact with rifampin in a thermodynamically favorable manner.

We evaluated whether the proteins can react with other semisynthetic rifampin analogues that are clinically approved and prescribed as drugs, such as rifapentine, rifaximin, and rifabutin (Additional file 1: Table S6 and Figures S6 and S7)<sup>47</sup>. All Arr-wd proteins react with these analogues to afford the corresponding ADP-ribosylated products (Table S5). However, their kinetic parameters are markedly altered depending on the substrates (Fig. 5), implying that modified chemical structure at the C8 position of the naphthoquinone core significantly changes the degrees of interaction with Arr-wd proteins. In general, Arr-wd2 and Arr-wd4 exhibit higher steady-state activity than Arr-wd1 and Arr-wd3. But Arr-wd2 and Arr-wd3 exhibit noticeably high and low  $k_{\text{cat}}/K_M$  for rifaximin, respectively, implying that specific interaction with the protein environment is operative in the ADP-ribosylation of the rifamycins. In contrast, rifapentine seems to be easily modified by all Arr-wd 1–4 proteins, indicating that the substrate is more promiscuous than other rifampin analogues.

To evaluate whether the altered catalytic activities are originated from the perturbation in the initial step of the reaction, i.e. association of the substrate to the protein, we measured the rifampin binding constants by monitoring intrinsic tryptophan fluorescence. The  $K_D$  values are in the order of Arr-wd4  $\approx$  Arr-wd2  $<$  Arr-wd1  $<$  Arr-wd3. This trend is similar to the inverse of the  $k_{\text{cat}}/K_M$ , implying that substrate binding is likely to be the part of the rate-determining steps in catalysis. However, no significant difference was observed between the  $K_D$  values of the rifampin analogues. Presumably, fluorescence changes of the tryptophan residue might not be sufficiently sensitive to differentiate the structural changes of the proteins that entail ADP-ribosyl transfer during steady-state reactions.



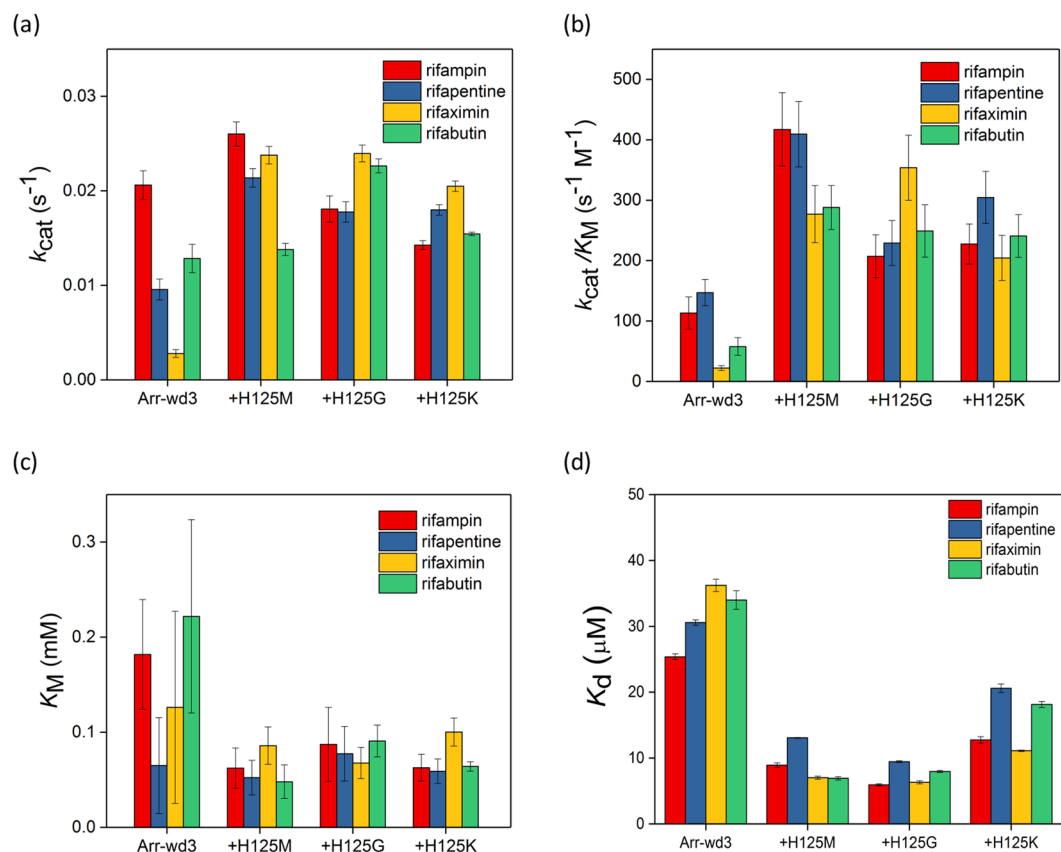
**Figure 5.** Kinetic and thermodynamic parameters of Arr-wd proteins from the steady-state reactions with rifampin derivatives, rifampin, rifapentine, rifaximin, and rifabutin. **(a)** Turnover numbers,  $k_{cat}$  ( $s^{-1}$ ) **(b)** Catalytic efficiencies,  $k_{cat}/K_M$  ( $s^{-1} M^{-1}$ ) **(c)** Michaelis constants,  $K_M$  (mM) **(d)** Dissociation constants,  $K_D$  ( $\mu M$ ). Values reported in **(a–c)** and **(d)** are derived from the fit to the Michaelis-Menten equation and Stern-Volmer equation, respectively.

The chemical origin of the different activities across rifampin analogues is yet to be delineated. We tentatively conclude that amino acid residues comprising the rifampin-binding sites could be responsible for this differentiation. As described above, the  $\alpha 2$ -helix composed of M[R|K][D|E]XL residues at the C-terminus makes direct contact with rifampin (Figs 1b and 2b), and is conserved throughout the metagenomic sequences (Fig. 1a). Although the fourth residue X is one of the least conserved residues in our metagenomic sequences as G(15.56%), K(17.78%), H(40%), A(8.89%), R(4.44%), K(6.67%), M(2.22%), S(2.22%), and F(2.22%), our structural analysis suggests that it is located in proximity to the complexed rifampin, directly pointing towards the substituent at the C8 position. Therefore, we suppose that it may play an important role in determining the ADP-ribosyltransferase activity.

**Site-directed mutagenesis of rifampin-binding sites.** In Arr-wd 1–4 proteins, the specific residue X is Met, Gly, His, and Lys, respectively. To evaluate whether the least conserved residue in the M[R|K][D|E]XL motif plays a role during the enzyme catalysis, we prepared three Arr-wd3 variants, H125M, H125G, and H125K (Additional file 1: Table S4); Met, Gly, and Lys are the amino acids that other Arr-wd 1, 2, and 4 proteins natively have at the analogous position, respectively. Although turnover rates ( $k_{cat}$ ) (Fig. 6a; Additional file 1: Table S8) and MIC values (Additional file 1: Table S9 and Figure S8) are marginally perturbed by the H125 mutations, the mutants H125M, H125G, and H125K display enhanced  $k_{cat}/K_M$  for all rifampin analogues. The data demonstrate that this residue is involved in the enzyme activities, and the relatively low activities of Arr-wd3 are at least partially attributed to the steric hindrance with little flexibility of the histidine residue at this position.

It is also supported by the  $K_D$  values of the H125 variants (Fig. 6d; Additional file 1: Table S7). H125G mutation yields the largest decrease in  $K_D$  value, followed by H125M and H125K, for all rifampin analogues, implicating that the size and the flexibility of the side chain is correlated with the degrees of favorable interactions with the rifamycins.

The most dramatic enhancement of the  $k_{cat}/K_M$  is observed from the reaction with rifaximin. When Arr-wd3 protein is compared with the H125G variant, the  $k_{cat}/K_M$  is increased by 16-fold, indicating that the removal of the imidazole ring presumably alleviates the steric clash with the bulky and strained 5 and 6-membered rings fused to the naphthoquinone core of rifaximin. In contrast, all Arr-wd3 mutants exhibit the least degree of enhancement upon the reaction with rifapentine or rifampin. The reactivity pattern with the rifampin analogues are consistent with our observation described above that rifaximin and rifapentine are the least and the most promiscuous substrates, respectively. These data suggest that the combined effects of the chemical structures of



**Figure 6.** Kinetic and thermodynamic parameters of Arr-wd3 H125 variants from the steady-state reactions with rifampin derivatives. **(a)** Turnover numbers,  $k_{cat}$  ( $s^{-1}$ ) **(b)** Catalytic efficiencies,  $k_{cat}/K_M$  ( $s^{-1} M^{-1}$ ) **(c)** Michaelis constants,  $K_M$  (mM) **(d)** Dissociation constants,  $K_D$  ( $\mu M$ ). Values reported in **(a–c)** and **(d)** are derived from the fit to the Michaelis-Menten equation and Stern-Volmer equation, respectively.

the antibiotics and their dynamic interactions with the enzyme active site may determine the enzyme catalysis, and the X residue in M[R]K[D]E]XL motif plays an important role in the reaction.

Even with the low sequence conservation in the X residue of the M[R]K[D]E]XL motif (Fig. 1a), the role in the reactivity and biochemical function was revealed only after the examination of protein structure. Therefore, the current study suggests that integrative analysis of 1-dimensional sequence and 3-dimensional structural information can provide us comprehensive information to elucidate the function of unidentified genes.

## Conclusions

Using metagenomic analysis on the sediment microbiome, we have identified new ADP-ribosyltransferase genes that are involved in the development of antibiotic resistance. These proteins share low sequence similarity with known enzymes, but have highly conserved amino acid residues around the enzyme active site, which catalyzes the transfer of  $NAD^+$ -derived ADP-ribosyl group to the antibiotic rifampin.

To gain a detailed understanding of this process, we carried out extensive analysis of gene sequence and protein structure to locate functionally relevant regions. For experimental validation of our finding, we expressed four representative genes into the proteins: Arr-wd1, Arr-wd2, Arr-wd3, and Arr-wd4. These proteins are kinetically competent in the ADP-ribosylation of rifampin using  $NAD^+$ . Steady-state kinetic analysis and binding assays revealed that the Arr-wd proteins exhibit much tighter binding interaction with rifampin and its analogues than previously reported Arr proteins. The steady-state activities of the Arr-wd proteins are altered depending on the substructures of the rifamycin substrates, and site-directed mutagenesis studies provided important chemical insights into how the steric interaction between the protein and substrate might affect the enzyme activity.

This work nicely demonstrates the advantage of an integrated bioinformatics approach to discover new enzyme functions from environmental microbiome data. Taking a step further from simple sequence analysis and classification, we have constructed three-dimensional models to identify structural features important for enzyme activity, and validated computational predictions with biochemical experiments. Our studies indicate that chemically guided integrative approach can improve the quality of the metagenomic data and can provide greater knowledge to predict structure-function relationships, thereby facilitating the annotation of protein function with high accuracy. Sequence alignment and site-directed mutagenesis also suggest that even a single amino acid residue with low sequence conservation can alter the *in vitro* and *in vivo* enzymatic activities, which further highlights the importance of identifying key enzyme residues to understand and combat antibiotic resistance.



## Methods

**Data collection.** A microbiome sample was collected from the soil sediment (latitude 33°49'65"N longitude 126°96'84"E), and sequenced using HiSeq™ 2000 platform (Illumina, San Diego, USA). The sample was prepared according to the Illumina protocols. For a 350 bp insert size, one microgram of genomic DNA was fragmented using Covaris. The raw sequencing data described in this study is available at European Nucleotide Archive (ENA) with the accession number PRJEB25358. The 151 bp paired-end sequencing reads were assembled by using MegaHit<sup>48</sup> with the default option. The genes predicted by FragGeneScan<sup>49</sup> were analyzed to find the rifampin resistance functions. Seven *arr* genes (i.e. *arr-1*, *arr-2*, *arr-3*, *arr-4*, *arr-5*, *arr-7*, and *arr-8*) were obtained from the comprehensive antibiotic resistance database (CARD)<sup>41,42</sup>. These genes were used as query to search homologous genes in the microbiome sample by blastp with the threshold of 50% sequence similarity and e-value < 1.0e-10. As a result, 45 homologous genes were obtained (Fig. 1a; Additional file 1: Table S1).

**Phylogenetic analysis.** In order to delineate the phylogenetic relationship between the putative *arr* genes and the annotated genes from the sequenced genomes, a phylogenetic tree was constructed (Fig. 3). The known *arr* genes were obtained from the CARD database and NCBI repository. In order to find homologous *arr* genes in the microbial genomes, 45 *arr* genes were searched against a set of 29,574,723 protein sequences of 8,369 bacterial genomes (downloaded from NCBI ftp server on 12/05/2017). In the phylogenetic tree, we included 45 *arr-wd* genes, 7 known *arr* genes from CARD database, 8 *arr* genes from the previous study<sup>13</sup>, and 37 putative *arr* genes that were searched from the microbial genomes. For each *arr* gene, five most homologous sequences (evalue < 1.0e-10) were obtained, and a set of non-redundant sequences were retained to build the phylogenetic tree. Neighbor-joining method was applied to build the tree with the bootstrapping value of 1000. MUSCLE<sup>50</sup> and MEGA<sup>51</sup> were used for multiple sequence alignment and phylogenetic analysis, respectively.

**Integrated analysis of protein structure and sequence.** Our integrated analysis of ADP-ribosyltransferase exploits rifampin and NAD<sup>+</sup> binding sites. For this purpose, both structure-guided and sequence-based approaches were taken. The structure-guided conservation analysis reveals the amino acids that are sterically and/or spatially conserved and physico-chemically similar, but relatively low in sequence conservations. The analysis consists of three tasks: (1) multiple sequence alignment (MSA) of the reference *arr* genes; (2) ligand-protein complex binding analysis between rifampin and Arr-*ms*; (3) mapping conserved residues from MSA with featured amino acid residues interacting with rifampin. A previously reported X-ray structure of Arr-*ms* complexed with rifampin (PDB ID: 2HW2) was used to find amino acid residues involved in the ADP-ribosylation activity. Both intermolecular (i.e. between rifampin and Arr-*ms*), and intramolecular interactions of Arr-*ms* are considered in the rifampin binding sites. The molecular interactions are determined from the distances between the atoms that participate in hydrogen bonding or hydrophobic interactions. The results were incorporated with sequence conservation data obtained from metagenome analysis.

No crystal structure is currently available for Arr proteins complexed with NAD<sup>+</sup>. We thus performed Autodock<sup>52</sup> simulation to visualize possible orientations of NAD<sup>+</sup> when bound to the protein. The size of grid box was set to 22 × 24 × 28 Å<sup>3</sup>. Structural models of Arr-*wd* were generated from the Swiss modeling<sup>53</sup> to simulate their possible three-dimensional structures (Additional file 1: Figure S2). PyMol<sup>54</sup> was used to visualize the X-ray crystal structure of Arr-*ms* (Figs 1–2; Additional file 1: Figures S1–S2 and S5). Discovery Studio Visualizer<sup>55</sup> was used to generate the 2D diagram of molecular interactions between rifampin and Arr-*ms*. The sequence alignment figure was generated by JalView<sup>56</sup>.

**Protein expression and purification.** Among 45 homologous genes from the metagenomic analysis, four genes denominated as *arr-wd1*, *arr-wd2*, *arr-wd3*, and *arr-wd4* were selected as the targets for biochemical characterizations (Fig. 1a; Additional file 1: Table S3). The codons of the DNA fragments were optimized for *E. coli* expression prior to the gene synthesis (General Biosystems), and were inserted into pET22b(+)/amp<sup>R</sup> vector using NdeI and XhoI cut-sites. Each plasmid was transformed to either DH5α or BL21(DE3) competent cells for sequencing or protein expression, respectively. All proteins were expressed with a His-tag at the C-terminus, and Arr-*wd3* without the His-tag (Arr-*wd3* + Δ6His) was additionally prepared to ensure that the presence of His-tag yields no significant perturbation in their activities. Single variants such as N34H, H125M, H125G, and H125K were prepared by site-directed mutagenesis. The PCR was carried out with KOD DNA polymerase (Toyobo) and primers (Additional file 1: Table S4). After PCR, the reaction products were digested with DpnI and transformed to DH5α competent cells. The plasmids were extracted by using miniprep kits, and sequenced at Macrogen, Inc.

For protein overexpression, a single colony of BL21(DE3) was picked and inoculated in 10 mL autoclaved LB media containing 100 mg/L ampicillin. The culture was grown in 200 rpm orbital shaker at 37 °C overnight, and inoculated in 1 L autoclaved LB media containing 100 mg/L ampicillin. When the optical density at 600 nm (OD<sub>600</sub>) reached ca 0.7, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at the final concentration was added. After 3 h of induction at 37 °C, the cells were harvested by centrifugation at 5000 rpm (4715 × g) for 15 min at 4 °C. The cell pastes were frozen in liquid nitrogen and stored at –80 °C for further usage.

The protein purification was initiated by thawing the cell pastes in lysate buffer (20 mM sodium phosphate buffer, pH 7.4 with 500 mM NaCl). The cells were lysed by sonication for 15 min in an ice bath (on/off = 3 s each). The lysates were centrifuged at 13000 rpm (18800 × g) for 30 min at 4 °C, and the supernatants were loaded to Ni affinity column (HisTrap HP column, GE Healthcare Life Sciences). The column was pre-equilibrated with the lysate buffer at 4 °C using ÄKTA Protein Purification Systems. After loading the cell lysates, elution buffer (20 mM sodium phosphate buffer, pH 7.4 with 500 mM NaCl and 500 mM imidazole) was applied in a linear gradient (5–50%) (Additional file 1: Figure S3). All proteins were eluted when ca 100 mM imidazole was applied. The protein expressed without a His-tag (Arr-*wd3* + Δ6His) was purified by Q anion exchange chromatography by varying the concentration of NaCl (0–1 M) in 20 mM sodium phosphate buffer, pH 8.0 at 4 °C. Relatively pure

fractions (ca 80–90%) were concentrated using Amicon stirred cells (EMD Millipore) or centrifugal concentration devices with 10 kDa cutoff membrane filters; the protein was further purified by size exclusion chromatography (S75 column), eluted with 20 mM Tris/HCl buffer, pH 7.0 with 150 mM NaCl. The purity of the protein was determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The mass of the protein was confirmed by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometer (MALDI-TOF; Bruker Ultraflex extreme TOF/TOF) with sinapinic acid as matrix solution. The protein was concentrated up to ca 0.5–5 mM and stored at  $-80^{\circ}\text{C}$  until further usage. The protein concentration was determined by UV-vis spectrophotometer (Agilent Cary 8454). The extinction coefficients at 280 nm were estimated from the sequence (Additional file 1: Table S3)<sup>57</sup>.

***In vitro* activity assay.** The assays were carried out in dark due to the light-sensitivity of rifampin and its analogues. Rifampin was dissolved in methanol, and diluted with the buffer used for the assay (Fig. 4). Initially, we followed the previously reported procedure (Additional file 1: Figure S4a)<sup>13</sup>. In short, 10  $\mu\text{M}$  protein in 100  $\mu\text{L}$  50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer, pH 7.5 was mixed with various concentrations of the substrates,  $\text{NAD}^+$  and rifampin. The solution was left for 5 min at  $22^{\circ}\text{C}$ , and the reaction was quenched by the addition of 100  $\mu\text{L}$  methanol. The reaction mixture was analyzed by HPLC (Agilent 1260 Infinity II, C18 column with 2.7  $\mu\text{m}$ , 120  $\text{\AA}$ ,  $4.6 \times 100$  mm). We later modified the quenching method by increasing the volume of methanol to 400  $\mu\text{L}$  because 100  $\mu\text{L}$  methanol was not sufficient to completely quench the reaction (Additional file 1: Figure S4b). In addition, we modified the HPLC method to accurately detect the formation of nicotinamide from the reaction mixture as follows: 100%  $\text{H}_2\text{O}$  and 0.05% trifluoroacetic acid (TFA) for 3.5 min, followed by a linear gradient to 90%  $\text{CH}_3\text{CN}$  with 0.1% TFA over 9.5 min (Fig. 4a). The ADP-ribosylated products were confirmed by electrospray ionization mass spectrometry (ESI-MS; Thermo Finnigan LTQ) (Additional file 1: Table S5).

Steady-state kinetic parameters of Arr-wd enzymes were obtained by varying the concentrations of one of the substrates, either  $\text{NAD}^+$  or rifampin, when the other was fixed to be 1 mM. The product yields were measured by detecting the concentrations of nicotinamide by HPLC. The standard curve of nicotinamide was measured independently (Additional file 1: Figure S4c). The kinetic parameters,  $k_{\text{cat}}$ ,  $K_{\text{M}}$ , and  $k_{\text{cat}}/K_{\text{M}}$ , were determined from non-linear iteration curve fits to the Michaelis-Menten equation (Figs 4 and 5; Additional file 1: Table S6 and Figure S4). Activities of Arr-wd3 variants were also measured by the same procedure as described above (Fig. 6; Additional file 1: Table S8).

*In vitro* activity assay was also carried out with  $\text{NADP}^+$ , NADH, or NADPH, instead of  $\text{NAD}^+$ , and with rifampin analogues, such as rifapentine, rifaximin, and rifabutin (Fig. 5; Additional file 1: Figure S6). Representative HPLC traces, Michaelis-Menten fits, and kinetic parameters are included in the Additional file 1: Table S6 and Figure S7. Rifampin analogues and the corresponding ADP-ribosylated products were eluted at 10.6 min and 9.1 min for rifapentine, 10.3 min and 8.6 min for rifaximin, and 9.9 min and 8.4 min for rifabutin, respectively. The reaction products were further analyzed by ESI-MS (Additional file 1: Table S5).

**Determination of binding constants by intrinsic fluorescence changes.** Intrinsic tryptophan fluorescence changes were monitored with microplate reader (Biotek Synergy H1m) at  $22\text{--}25^{\circ}\text{C}$ . Arr-wd1-3 and Arr-wd4 proteins possess two and four tryptophan residues, respectively. Sequence alignment together with structure homology modeling suggests that two tryptophan residues are located at similar positions in Arr-wd1-4 (Additional file 1: Figure S5). Conformational changes induced by the addition of substrates were monitored by using 295 nm excitation to avoid any possible interference from tyrosine and phenylalanine residues. The fluorescence emission was observed at 325–450 nm, and the fluorescence emission at 325 nm was used to obtain the dissociation constants ( $K_{\text{D}}$ ). Because rifamycins exhibits strong absorption at both excitation and emission wavelengths (Additional File 1: Figure S5b), the concentrations of rifamycins were kept at 1–4  $\mu\text{M}$  so that the absorption at both excitation and emission wavelengths remains less than 0.05, thereby minimizing the inner filter effect<sup>58</sup>. The plot of  $F_0/F$  versus the concentrations of rifamycins yielded a linear correlation (Additional File 1: Figure S5c), which is fitted to the Stern-Volmer equation (Additional File 1: Figure S5d) to determine the dissociation constants ( $K_{\text{D}}$ ) for rifampin and its derivatives (Figs 4e, 5d, 6d, and Additional file 1: Table S7). The protein samples (5–20  $\mu\text{M}$ ) were prepared in 100  $\mu\text{L}$  of 50 mM HEPES buffer, pH 7.5.

**Determination of MIC values.** Minimal inhibitory concentration (MIC) values for BL21(DE3) cells containing arr-wd genes were measured by the procedures reported previously<sup>13</sup>. In short, the plasmids containing arr-wd genes were transformed into BL21(DE3) cells, and were grown on the LB/agar plate containing 100 mg/L ampicillin overnight at  $37^{\circ}\text{C}$ . The cells were mixed with 0.85% saline to adjust the  $\text{OD}_{625}$  to be 0.1. The aliquots of 100  $\mu\text{L}$  cell culture in saline solution were mixed with 18.9 mL Mueller Hinton broth media and 0–512  $\mu\text{g}/\text{mL}$  rifamycin dissolved in methanol. Either 0 or 100  $\mu\text{M}$  IPTG at final concentration was added to the solution. After the solutions were incubated at  $37^{\circ}\text{C}$  overnight, optical cell density ( $\text{OD}_{625}$ ) was measured with the microplate reader to determine the minimum inhibitory concentrations of the antibiotics (Fig. 3f; Additional file 1: Table S9 and Figure S8).

**Availability of data and materials.** The raw sequencing data described in this study is available at European Nucleotide Archive (ENA) with the accession number PRJEB25358.

## References

- Ventola, C. L. The antibiotic resistance crisis: Part 1: Causes and threats. *Pharm. Ther.* **40**, 277–283 (2015).
- Chakraborty, S. & Rhee, K. Y. Tuberculosis drug development: History and evolution of the mechanism-based paradigm. *Cold Spring Harbor perspectives in medicine* **5**, a021147, <https://doi.org/10.1101/cshperspect.a021147> (2015).

3. Sensi, P. History of the development of rifampin. *Rev. Infect. Dis.* **5**(Suppl 3), S402–406 (1983).
4. Marriner, G. A. *et al.* In *Third World Diseases* Vol. 7 (ed. Elliott, R. L.) 47–124 (Springer-Verlag Berlin Heidelberg, 2011).
5. Campbell, E. A. *et al.* Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **104**, 901–912 (2001).
6. Alifano, P., Palumbo, C., Pasanisi, D. & Talà, A. Rifampicin-resistance, rpoB polymorphism and RNA polymerase genetic engineering. *J. Biotechnol.* **202**, 60–77, <https://doi.org/10.1016/j.jbiotec.2014.11.024> (2015).
7. Stogios, P. J. *et al.* Rifampin phosphotransferase is an unusual antibiotic resistance kinase. *Nature Comm.* **7**, 11343, <https://doi.org/10.1038/ncomms11343>, <https://www.nature.com/articles/ncomms11343supplementary-information> (2016).
8. Goldstein, B. P. Resistance to rifampicin: a review. *J. Antibiot.* **67**, 625–630, <https://doi.org/10.1038/ja.2014.107> (2014).
9. Floss, H. G. & Yu, T.-W. Rifamycin mode of action, resistance, and biosynthesis. *Chem. Rev.* **105**, 621–632, <https://doi.org/10.1021/cr030112j> (2005).
10. Ahmad, S., Mokaddas, E. & Fares, E. Characterization of rpoB mutations in rifampin-resistant clinical Mycobacterium tuberculosis isolates from Kuwait and Dubai. *Diagn. Microbiol. Infect. Dis.* **44**, 245–252 (2002).
11. Jamieson, F. B. *et al.* Profiling of rpoB mutations and MICs for rifampin and rifabutin in Mycobacterium tuberculosis. *J. Clin. Microbiol.* **52**, 2157–2162, <https://doi.org/10.1128/JCM.00691-14> (2014).
12. Spanogiannopoulos, P., Thaker, M., Koteva, K., Waglechner, N. & Wright, G. D. Characterization of a rifampin-inactivating glycosyltransferase from a screen of environmental actinomycetes. *Antimicrob. Agents. Chemother.* **56**, 5061–5069, <https://doi.org/10.1128/aac.01166-12> (2012).
13. Baysarowich, J. *et al.* Rifamycin antibiotic resistance by ADP-ribosylation: Structure and diversity of Arr. *Proc. Natl. Acad. Sci. USA* **105**, 4886–4891, <https://doi.org/10.1073/pnas.0711939105> (2008).
14. Liu, L. K. *et al.* The structure of the antibiotic deactivating, N-hydroxylating rifampicin monooxygenase. *J. Biol. Chem.* **291**, 21553–21562, <https://doi.org/10.1074/jbc.M116.745315> (2016).
15. Dabbs, E. R. *et al.* Ribosylation by mycobacterial strains as a new mechanism of rifampin inactivation. *Antimicrob. Agents. Chemother.* **39**, 1007–1009 (1995).
16. Agrawal, P., Miryala, S. & Varshney, U. Use of Mycobacterium smegmatis deficient in ADP-ribosyltransferase as surrogate for Mycobacterium tuberculosis in drug testing and mutation analysis. *PLOS ONE* **10**, e0122076, <https://doi.org/10.1371/journal.pone.0122076> (2015).
17. Morisaki, N. *et al.* Structure determination of ribosylated rifampicin and its derivative: new inactivated metabolites of rifampicin by mycobacterial strains. *J. Antibiot.* **48**, 1299–1303 (1995).
18. Qi, X. *et al.* Structural basis of rifampin inactivation by rifampin phosphotransferase. *Proc. Natl. Acad. Sci. USA* **113**, 3803–3808, <https://doi.org/10.1073/pnas.1523614113> (2016).
19. Martínez, J. L. Antibiotics and antibiotic resistance genes in natural environments. *Science* **321**, 365–367, <https://doi.org/10.1126/science.1159483> (2008).
20. 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).
21. Subirats, J., Sánchez-Melsió, A., Borrego, C. M., Balcázar, J. L. & Simonet, P. Metagenomic analysis reveals that bacteriophages are reservoirs of antibiotic resistance genes. *Int. J. Antimicrob. Agents.* **48**, 163–167, <https://doi.org/10.1016/j.ijantimicag.2016.04.028> (2016).
22. Flórez, A. B., Vázquez, L. & Mayo, B. A functional metagenomic analysis of tetracycline resistance in cheese bacteria. *Front. Microbiol.* **8**, 907, <https://doi.org/10.3389/fmicb.2017.00907> (2017).
23. Guo, J., Li, J., Chen, H., Bond, P. L. & Yuan, Z. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Research* **123**, 468–478, <https://doi.org/10.1016/j.watres.2017.07.002> (2017).
24. Chen, B. *et al.* Metagenomic analysis revealing antibiotic resistance genes (ARGs) and their genetic compartments in the tibetan environment. *Environ. Sci. Technol.* **50**, 6670–6679, <https://doi.org/10.1021/acs.est.6b00619> (2016).
25. Crofts, T. S., Gasparrini, A. J. & Dantas, G. Next-generation approaches to understand and combat the antibiotic resistome. *Nat. Rev. Microbiol.* **15**, 422, <https://doi.org/10.1038/nrmicro.2017.28>, <https://www.nature.com/articles/nrmicro.2017.28supplementary-information> (2017).
26. Wright, P. M., Seiple, I. B. & Myers, A. G. The evolving role of chemical synthesis in antibacterial drug discovery. *Angew. Chem. Int. Ed.* **53**, 8840–8869, <https://doi.org/10.1002/anie.201310843> (2014).
27. Pawlowski, A. C. *et al.* A diverse intrinsic antibiotic resistome from a cave bacterium. *Nature Comm.* **7**, 13803, <https://doi.org/10.1038/ncomms13803>, <https://www.nature.com/articles/ncomms13803supplementary-information> (2016).
28. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457, <https://doi.org/10.1038/nature10388>, <https://www.nature.com/articles/nature10388supplementary-information> (2011).
29. Bhullar, K. *et al.* Antibiotic resistance is prevalent in an isolated cave microbiome. *PLOS ONE* **7**, e34953, <https://doi.org/10.1371/journal.pone.0034953> (2012).
30. Segawa, T. *et al.* Distribution of antibiotic resistance genes in glacier environments. *Environ. Microbiol. Rep.* **5**, 127–134, <https://doi.org/10.1111/1758-2229.12011> (2013).
31. Forsberg, K. J. *et al.* Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612, <https://doi.org/10.1038/nature13377>, <https://www.nature.com/articles/nature13377supplementary-information> (2014).
32. Allen, H. K. *et al.* Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* **8**, 251–259, <https://doi.org/10.1038/nrmicro2312> (2010).
33. Nesme, J. *et al.* Large-scale metagenomic-based study of antibiotic resistance in the environment. *Curr. Biol.* **24**, 1096–1100, <https://doi.org/10.1016/j.cub.2014.03.036> (2014).
34. Li, B. *et al.* Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *ISME J.* **9**, 2490–2502, <https://doi.org/10.1038/ismej.2015.59> (2015).
35. Hu, Y. *et al.* Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Comm.* **4**, 2151, <https://doi.org/10.1038/ncomms3151>, <https://www.nature.com/articles/ncomms3151supplementary-information> (2013).
36. Forsberg, K. J. *et al.* The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111, <https://doi.org/10.1126/science.1220761> (2012).
37. Sommer, M. O. A., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131, <https://doi.org/10.1126/science.1176950> (2009).
38. Wass, M. N. & Sternberg, M. J. E. Prediction of ligand binding sites using homologous structures and conservation at CASP8. *Proteins* **77**, 147–151, <https://doi.org/10.1002/prot.22513> (2009).
39. Rost, B. Enzyme function less conserved than anticipated. *Journal of molecular biology* **318**, 595–608, [https://doi.org/10.1016/s0022-2836\(02\)00016-5](https://doi.org/10.1016/s0022-2836(02)00016-5) (2002).
40. Levin, B. J. *et al.* A prominent glycol radical enzyme in human gut microbiomes metabolizes trans-4-hydroxy-L-proline. *Science* **355**, <https://doi.org/10.1126/science.aai8386> (2017).
41. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573, <https://doi.org/10.1093/nar/gkw1004> (2017).
42. McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob. Agents. Chemother.* **57**, 3348–3357, <https://doi.org/10.1128/aac.00419-13> (2013).

43. Houang, E. T. S., Chu, Y.-W., Lo, W.-S., Chu, K.-Y. & Cheng, A. F. B. Epidemiology of rifampin ADP-ribosyltransferase (arr-2) and metallo- $\beta$ -Lactamase (bla(IMP-4)) gene cassettes in class 1 integrons in *Acinetobacter* strains isolated from blood cultures in 1997 to 2000. *Antimicrob. Agents. Chemother.* **47**, 1382–1390, <https://doi.org/10.1128/AAC.47.4.1382-1390.2003> (2003).
44. Tribuddharat, C. & Fennewald, M. Integron-mediated rifampin resistance in *Pseudomonas aeruginosa*. *Antimicrob. Agents. Chemother.* **43**, 960–962 (1999).
45. Chowdhury, G., Pazhani, G. P., Nair, G. B., Ghosh, A. & Ramamurthy, T. Transferable plasmid-mediated quinolone resistance in association with extended-spectrum beta-lactamases and fluoroquinolone-acetylating aminoglycoside-6'-N-acetyltransferase in clinical isolates of *Vibrio fluvialis*. *Int. J. Antimicrob. Agents.* **38**, 169–173, <https://doi.org/10.1016/j.ijantimicag.2011.04.013> (2011).
46. Fonseca, E. Ld, Freitas, Fd. S. & Amorim, J. C. d. & Vicente, A. C. P. Detection of new arr-4 and arr-5 gene cassettes in clinical *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* strains from Brazil. *Antimicrob. Agents. Chemother.* **52**, 1865–1867, <https://doi.org/10.1128/AAC.00017-08> (2008).
47. Aristoff, P. A., Garcia, G. A., Kirchhoff, P. D. & Hollis Showalter, H. D. Rifamycins – Obstacles and opportunities. *Tuberculosis* **90**, 94–118, <https://doi.org/10.1016/j.tube.2010.02.001> (2010).
48. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676, <https://doi.org/10.1093/bioinformatics/btv033> (2015).
49. Rho, M., Tang, H. & Ye, Y. FragGeneScan: Predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191–e191, <https://doi.org/10.1093/nar/gkq747> (2010).
50. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797, <https://doi.org/10.1093/nar/gkh340> (2004).
51. Tamura, K., Stecher, G., Peterson, D., Filipinski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729, <https://doi.org/10.1093/molbev/mst197> (2013).
52. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791, <https://doi.org/10.1002/jcc.21256> (2009).
53. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258, <https://doi.org/10.1093/nar/gku340> (2014).
54. PyMOL. The PyMOL molecular graphics system, version 1.8 Schrödinger, LLC.
55. BIOVIA, D. S. Discovery Studio Visualizer, 2017 R2, San Diego: Dassault Systèmes (2017).
56. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191, <https://doi.org/10.1093/bioinformatics/btp033> (2009).
57. Gill, S. C. & von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326, [https://doi.org/10.1016/0003-2697\(89\)90602-7](https://doi.org/10.1016/0003-2697(89)90602-7) (1989).
58. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. 3 edn, (Springer, 2006).

## Acknowledgements

We thank Youna Cho for sequence data processing, and Dr. Dongwhan Lee for helpful discussion. This work was supported by a grant from the Collaborative Genome Program (20140428 to W.J.S. and M.R.) funded by the Ministry of Oceans and Fisheries, Korea, a NRF grant from Korea government (NRF-2016R1C1B2007898 to W.J.S.), and a grant (2017NER54070 to M.R.) by Research of Korea Centers for Disease Control and Prevention.

## Author Contributions

J.H.S. and M.R. conducted the metagenomic sequence analysis. H.E. and W.J.S. conducted biochemical experiments. J.H.S., H.E., W.J.S. and M.R. designed the experiment, analysed the data, and wrote the manuscript. All authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-30547-x>.

**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 <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018