



# Genomic characterization of *Nocardia seriolae* strains isolated from diseased fish

Hyun-Ja Han<sup>1\*</sup> | Min-Jung Kwak<sup>2\*</sup> | Sung-min Ha<sup>2,3</sup> | Seung-Jo Yang<sup>2</sup> |  
Jin Do Kim<sup>1</sup> | Kyoung-hee Cho<sup>2</sup> | Tae-Wook Kim<sup>2</sup> | Mi Young Cho<sup>1</sup> |  
Byung-Yong Kim<sup>2</sup> | Sung-Hee Jung<sup>1</sup> | Jongsik Chun<sup>2,3</sup>

<sup>1</sup>Pathology Research Division, National Institute of Fisheries Science, Busan, Korea

<sup>2</sup>ChunLab Inc., Seoul, Korea

<sup>3</sup>Laboratory of evolutionary bioinformatics, Seoul National University, Seoul, Korea

## Correspondence

Byung-Yong Kim, ChunLab Inc., JW TOWER, Nambusunhwan-ro 2477, Seocho-gu, Seoul 06725, Korea.

Email: bykim@chunlab.com  
and

Sung-Hee Jung, Pathology Research Division, National Institute of Fisheries Science, Gijanghaean-ro 216, Gijang-eup, Busan, 46083, Korea.  
Email: immu@korea.kr

## Funding information

National Institute of Fisheries Science, Republic of Korea, Grant/Award Number: R2018062

## Abstract

Members of the genus *Nocardia* are widespread in diverse environments; a wide range of *Nocardia* species are known to cause nocardiosis in several animals, including cat, dog, fish, and humans. Of the pathogenic *Nocardia* species, *N. seriolae* is known to cause disease in cultured fish, resulting in major economic loss. We isolated two *N. seriolae* strains, CK-14008 and EM15050, from diseased fish and sequenced their genomes using the PacBio sequencing platform. To identify their genomic features, we compared their genomes with those of other *Nocardia* species. Phylogenetic analysis showed that *N. seriolae* shares a common ancestor with a putative human pathogenic *Nocardia* species. Moreover, *N. seriolae* strains were phylogenetically divided into four clusters according to host fish families. Through genome comparison, we observed that the putative pathogenic *Nocardia* strains had additional genes for iron acquisition. Dozens of antibiotic resistance genes were detected in the genomes of *N. seriolae* strains; most of the antibiotics were involved in the inhibition of the biosynthesis of proteins or cell walls. Our results demonstrated the virulence features and antibiotic resistance of fish pathogenic *N. seriolae* strains at the genomic level. These results may be useful to develop strategies for the prevention of fish nocardiosis.

## KEYWORDS

*Anguilla japonica*, *Channa argus*, comparative genomics, fish pathogen, *Nocardia seriolae*

## 1 | INTRODUCTION

Bacteria of the genus *Nocardia* belong to the class *Actinobacteria*, and are well known as the cause of nocardiosis, an infectious disease that occurs mainly in immunocompromised patients (Brown-Elliott, Brown, Conville, & Wallace, 2006). Currently, 113 *Nocardia* species have been isolated from diverse environments including soil, water,

the rhizosphere, insects, and fish, as well as from human patients and medical samples. Most *Nocardia* species have been isolated from natural environments, especially soils, but more than 30 species are known to cause disease in humans (Yu, Wang, Fang, Zhang, & Yan, 2016). Moreover, some species have been confirmed as pathogenic in animals including dogs (Eroksuz et al., 2017), cats (Harada, Endo, Sekiguchi, Setoguchi, & Momoi, 2009), and goats (Ellwood, 1973) as

\*Co-first authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *MicrobiologyOpen* published by John Wiley & Sons Ltd.

well as marine animals such as oysters (Friedman et al., 1998) and fish (Kudo, Hatai, & Seino, 1988). Nocardiosis has been identified across a diverse range of fish species, and causes mass mortality rates in economically important farming fish such as *Oncorhynchus tshawytscha* (Chinook salmon) (Brosnahan et al., 2017), *Micropterus salmoides* (Largemouth bass) (Ho et al., 2016), and *Seriola dumerili* (Greater amberjack) (Matsumoto, Hayashi, Suetake, Yamamoto, & Araki, 2016). The infected fish exhibited skin ulcers and tubercles on the gills, and in the liver, kidney, and spleen as the clinical signs (Chen et al., 2000). In particular, typical granulomas in the liver, kidney, and spleen were observed from histopathological examination (Shimahara et al., 2008).

Currently, several *Nocardia* species including *N. crassostreae* (Friedman et al., 1998), *N. salmonicida* (Isik, Chun, Hah, & Goodfellow, 1999), *N. seriolae* (Kudo et al., 1988), and *N. xestospongiae* (Thawai, Rungjindamai, Klanbut, & Tanasupawat, 2017) have been isolated from marine organisms. Of these species, *N. seriolae* was the most frequently detected species in diseased fish (Nayak & Nakanishi, 2016; Wang et al., 2009). For these reasons, many studies on the virulence mechanisms of the bacteria and strategies for preventing infection have been performed (Byadgi, Chen, Wang, Tsai, & Chen, 2016; Ho et al., 2016; Huang, Lou, Wu, & Chen, 2008; Itano, Kawakami, Kono, & Sakai, 2006; Nayak & Nakanishi, 2016; Yasuike et al., 2017). In addition, five genomes of *N. seriolae* strains are publicly available, all of which were isolated from the diseased fish (Blotched snakehead), *Seriola quinqueradiata* (Japanese amberjack), and *Trachinotus ovatus* (Pompano) (Imajoh et al., 2015, 2016; Xia et al., 2015; Yasuike et al., 2017). Furthermore, Yasuike et al. (2017) have reported putative virulence factors from *N. seriolae* strain UTF1 through comparative analysis with the genomes of other *Nocardia* species.

Recently, we isolated two *N. seriolae* strains CK-14008 and EM15056 from diseased *Channa argus* (Northern snakehead) and *Anguilla japonica* (Japanese eel), respectively. To reveal the putative virulence factors and the genomic features of *N. seriolae* strains, we determined genome sequences of them using PacBio sequencing platform and compared their genomes with the genomes of phylogenetically close *Nocardia* species and other *N. seriolae* strains.

## 2 | MATERIALS AND METHODS

### 2.1 | Bacterial isolation, cultivation, and identification

Diseased snakehead (*C. argus*) and Japanese eel (*A. japonica*), which exhibited lethargy and skin ulcers were reported in 2014 (Busan, Korea) and 2015 (Gimcheon Gyeongsangbuk-do, Korea), respectively. Diseased fish samples were collected in ice-cooled boxes and transported directly to the laboratory of Korean National Institute of Fisheries Science for further diagnosis. Several swabs from the kidney, spleen, and liver of diseased fish were streaked on tryptic soy agar (TSA) and brain heart infusion agar (BHIA) plates and incubated at 25°C for 2 weeks. Two isolates, CK-14008

and EM-150506 were selected and cultured in tryptic soy broth medium at 25°C for 7 days under constant shaking with 100 rpm to obtain cell mass for DNA extraction experiment. The preparation of genomic DNA and PCR amplification of the 16S RNA gene were carried out following Chun and Goodfellow (1995) for the identification of two isolates.

### 2.2 | Genome sequencing, assembly and annotation

Genomic DNA from *N. seriolae* CK-14008 and EM150506 were extracted using the Qiagen DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. For genome sequencing, a 20-kb PacBio SMRTbell library was prepared for each genome and PacBio RS II (Pacific Biosciences, Menlo Park, CA USA) was used for genome sequencing with P6 polymerase and C4 chemistry onto a single-molecular real-time (SMRT) cell for each genome.

De novo assembly of the sequences and circularization of the assembled sequences were conducted using HGAP2 in the SMRTpipe of the PacBio portal and circulator program (Hunt et al., 2015), respectively. Structural gene prediction was conducted using Glimmer 3.02 (Delcher, Bratke, Powers, & Salzberg, 2007) and functional annotation of the predicted genes was conducted by a homology search using the nonredundant RefSeq protein (NR; NCBI), COG (Tatusov et al., 2001), EggNOG, SEED (Overbeek et al., 2014), Swiss-Prot (Watanabe & Harayama, 2001), and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa, Goto, Kawashima, Okuno, & Hattori, 2004) databases. rRNA and tRNA were predicted using RNAmmer 1.2 (Lagesen et al., 2007) and tRNAscan-SE (Lowe & Eddy, 1997), respectively.

### 2.3 | Comparative genomics

For comparative genomics, the genome sequences of closely related *Nocardia* strains were obtained from the EzBioCloud genome database (Yoon et al., 2017) and NCBI genome database. For the comparison between *Nocardia* species, the genomes of *N. acidivorans* NBRC 108247 (Kampfer et al., 2007), *N. concava* NBRC 100430 (Hirayama et al., 2016), *N. crassostreae* NBRC 100342 (Friedman et al., 1998), *N. inohanensis* NBRC 100128 (Kageyama, Yazawa, Nishimura, & Mikami, 2004), *N. jejuensis* NBRC 103114 (Lee, 2006), *N. niigatensis* NBRC 100131 (Kageyama et al., 2004), and *N. yamanashiensis* NBRC 100130 (Kageyama et al., 2004) were obtained from EzBioCloud. For comparison between *N. seriolae* strains, the genomes of the strains N-2927 (Imajoh et al., 2015), U-1 (Imajoh et al., 2016), and ZJ0503 (Xia et al., 2015) were also downloaded from EzBioCloud, and the nucleotide sequences from the strains SY-24 (unpublished) and UTF1 (Yasuike et al., 2017) were downloaded from the NCBI genome database. For consistency of gene prediction and functional annotation of the analyzed genomes, the genome sequences of the strains SY-24 and UTF1 were uploaded to the Whole Genome (WG) pipeline of BIOiPLUG (<https://www.bioiplug.com>, ChunLab Inc., Seoul, Republic of Korea) and quality-controlled genome information was obtained.

To understand the phylogenetic relationships of the analyzed strains, OrthoANI values were calculated and UPGMA dendrograms were generated using the Orthologous ANI Tool (OAT) of ChunLab (Lee, Kim, Park, & Chun, 2015). For synteny analysis of the genomes at the nucleotide level, Blast Ring Image Generator (BRIG) was used with the default parameters (Alikhan, Petty, Ben Zakour, & Beatson, 2011).

Analysis of the pan-genome and core genome was conducted using the Comparative Genomics (CG) pipeline of BIOiPLUG Apps (<https://www.bioiplug.com/apps>, ChunLab Inc.). Pan-genome orthologous groups (POGs) were determined by a combined reciprocal best hit (RBH) method using uBLAST with an *E*-value threshold of  $1 \times 10^{-6}$  (Ward & Moreno-Hagelsieb, 2014) and an open reading frame (ORF)-independent method using nucleotide sequences with cutoff values of at least 70% of gene coverage (Chun et al., 2009). A plot for the pan-genome and core genome sizes, Venn diagrams for the numbers of orthologous genes, and a heat-map for the presence and absence of the genes were generated using the CG pipeline of BIOiPLUG in ChunLab.

Analysis of the secondary metabolite biosynthetic gene clusters and antibiotic resistance genes were conducted using the antiSMASH webserver (Blin et al., 2017) and Antibiotic Resistance Genes Database (ARDB) webserver (Liu & Pop, 2009), respectively. Investigation of the putative virulence genes were conducted by keyword searches using annotated genes or homology searches using amino acid sequences against the putative virulence proteins of *N. farcinica* (Ishikawa et al., 2004; Yasuike et al., 2017).

### 3 | RESULTS

#### 3.1 | General genomic features of *N. seriolae* isolates

To analyze the genomic features of *Nocardia* strains isolated from the diseased fish, we sequenced the genomes of *N. seriolae* CK-14008 and EM150506, isolated from diseased *C. argus* (Northern snakehead) and *A. japonica* (Japanese eel), respectively. The genome of strain EM150506 consisted of a complete circular chromosome, and the genome of strain CK-14008 consisted of a complete circular chromosome with two incomplete plasmids. Both strains had an 8.3-Mb chromosome with 68.1% G/C content. The general genomic characteristics of the two strains are described in Table 1.

#### 3.2 | Phylogenetic relationships and genomic features of *Nocardia* bacteria

The similarity of the 16S rRNA gene and the average nucleotide identity (ANI) of the genomes of the two *N. seriolae* strains showed the highest similarity to *N. concava*. Moreover, an unweighted pair group method with arithmetic mean (UPGMA) dendrogram based on OrthoANI values of 101 type strains in the genus *Nocardia* showed that *N. seriolae* shares a common ancestor with *N. acidivorans*, *N. concava*, *N. crassostreae*, *N. inohanensis*, *N. jejuensis*, *N. niigatensis*, and

**TABLE 1** General genomic features of *N. seriolae* CK-14008 and EM150506

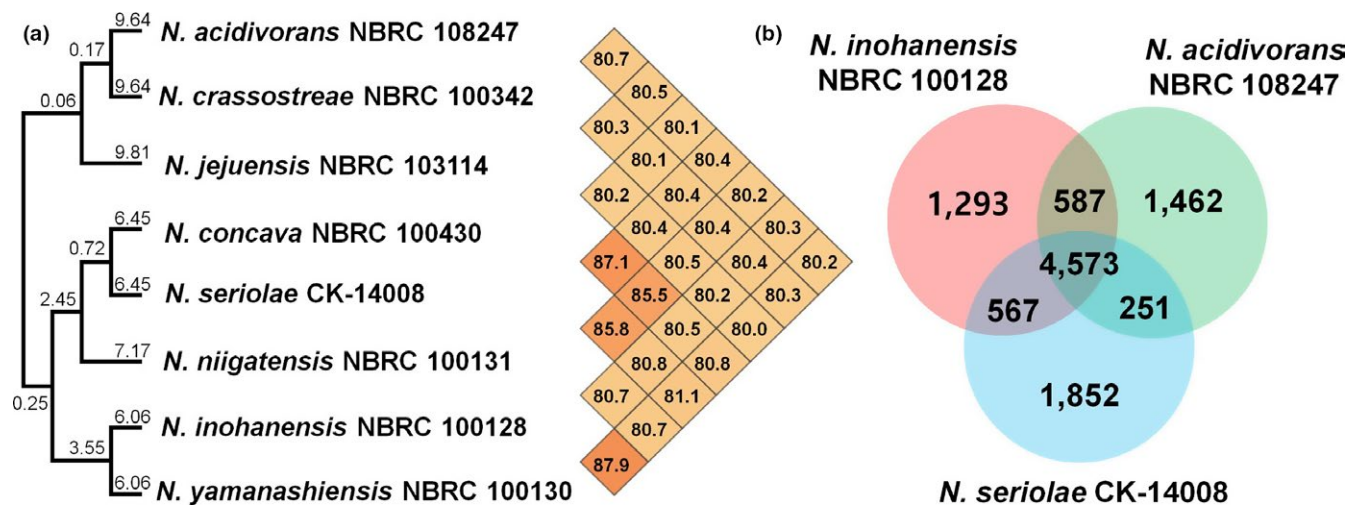
Strain	CK-14008	EM150506
Isolation origin	Diseased <i>Channa argus</i>	Diseased <i>Anguilla japonica</i>
No. of chromosomes	1	1
No. of plasmids	2	0
Total bases	8,370,754	8,304,518
No. of CDSs	7,903	7,794
No. of rRNAs	12	12
No. of tRNAs	66	65
Accession no.	NZ_MOYO000000000.1 <sup>a</sup>	CP017839.1

<sup>a</sup>NZ\_MOYO000000000.1" is a draft genome accession number that includes a complete chromosome and two incomplete plasmids.

*N. yamanashiensis* (<https://www.ezbiocloud.net/>) (Yoon et al., 2017). Out of these species, *N. concava*, *N. niigatensis*, *N. inohanensis*, and *N. yamanashiensis* were isolated from medical samples of patients; *N. acidivorans* and *N. jejuensis* were isolated from soil; and *N. crassostreae* was isolated from diseased oysters (Supporting Information Table S1).

To identify the genomic features of the *N. seriolae* strains isolated from diseased fish, we compared their genomes with the genomes of *Nocardia* species isolated from human medical samples and natural sources. The general features of the analyzed *Nocardia* genomes are described in Supporting Information Table S1. A phylogenetic tree of eight *Nocardia* species based on the OrthoANI algorithm (Yoon et al., 2017) showed that the analyzed *Nocardia* species could be divided into two clades (Figure 1a). One clade contains the species isolated from the diseased fish (CK-14008) and human medical samples (NBRC 100128, NBRC 100130, NBRC 100131, and NBRC 100430), while the other clade contains the species isolated from soil (NBRC 103114 and NBRC 108247) and diseased oysters (NBRC 100342). These phylogenomic relationships were identical with the results in the previous studies (Tamura et al., 2012, 2018).

The distributions of the genes assigned to clusters of orthologous groups (COG) showed that the genes associated with the COG categories "transcription" and "amino acid transport and metabolism" were the most abundant in the genomes of *Nocardia* species, except for the genes assigned to the COG category "general function prediction only" (Supporting Information Table S2). The relative abundances of the genes assigned to each COG category were highly similar among *Nocardia* species. However, in the genomes of the *N. seriolae* strains, the relative abundances of the genes assigned to the COG categories "transcription" and "secondary metabolites biosynthesis, transport, and catabolism" were approximately 1% lower than in other species. Of the COG assigned genes, the genes assigned to the COG category "replication, recombination, and repair" were significantly higher in the genomes of *N. seriolae* CK-14008, EM150506, and UTF1. Many of the genes in this



**FIGURE 1** Genomic comparison of bacterial species in the genus *Nocardia*. (a) Unweighted pair group method with arithmetic mean (UPGMA) dendrogram based on the OrthoANI values of eight species in the genus *Nocardia*. The numbers on the branches indicate the branch length and the numbers in the heat-map indicate the OrthoANI values between two genomes. (b) Venn diagram of the number of orthologous genes in three *Nocardia* species. *N. inohanensis* NBRC 100128, *N. seriolae* CK-14008, and *N. acidivorans* NBRC 108247 represent the species isolated from human medical samples, diseased fish, and natural environments, respectively

category are involved in mobile elements such as transposase and phage, which have highly repeatable sequences. Of the analyzed genomes, only three genomes from *N. seriolae* CK-14008, EM150506, and UTF1 were completely sequenced by the PacBio sequencing platform; therefore, repeat sequences were fully accounted for in these three genomes. For this reason, the relative abundances of the genes assigned to the COG category “replication, recombination, and repair” were significantly higher in the genomes of *N. seriolae* CK-14008, EM150506, and UTF1.

Analysis of the orthologous genes of three *Nocardia* species isolated from soil (*N. acidivorans* NBRC 108247), human medical samples (*N. inohanensis* NBRC 100128), and diseased fish (*N. seriolae* CK-14008) demonstrated that the *Nocardia* species have a core genome of 4,573 genes that contains many genes encoding antibiotic resistance proteins as well as genes for basic cell metabolism (Figure 1b and Supporting Information Table S3). In the core genome of *Nocardia* species, genes encoding diphtheria toxin repressors, lysostaphin, ESX-1 secretion system proteins, and ESX-3 secretion-associated proteins were also identified. *N. seriolae* CK-14008 shared twofold more genes with *N. inohanensis* (567 genes) than with *N. acidivorans* (251 genes). Furthermore, the number of shared genes between *N. inohanensis* and *N. acidivorans* (587 genes) was similar to the number of shared genes between *N. inohanensis* and *N. seriolae*. Interestingly, four genes encoding gas vesicle biosynthetic proteins were detected only in the genomes of *N. concava*, *N. inohanensis*, *N. niigatensis*, and *N. yamanashiensis*, which were isolated from human medical samples (Supporting Information Table S3).

Through these results, we observed that *Nocardia* species might be phylogenetically clustered according to their isolation origins, and that putative pathogenic species form a single clade. Moreover, *Nocardia* species showed different genomic features according to their isolation origins.

### 3.3 | Phylogenetic relationships between strains of *N. seriolae*

*N. seriolae* strains CK-14008 and EM150506 were isolated from diseased *C. argus* and *A. japonica*, respectively. *C. argus* belongs to the family Channidae and *A. japonica* belongs to the family Anguillidae. To investigate the genomic features based on the hosts of each isolate, we compared the genomes of CK-14008 and EM150506 with the genomes of other *N. seriolae* strains. Currently, seven *N. seriolae* genomes, including CK-14008 and EM150506, are publicly available, and all of them were isolated from diseased fish (Supporting Information Table S1).

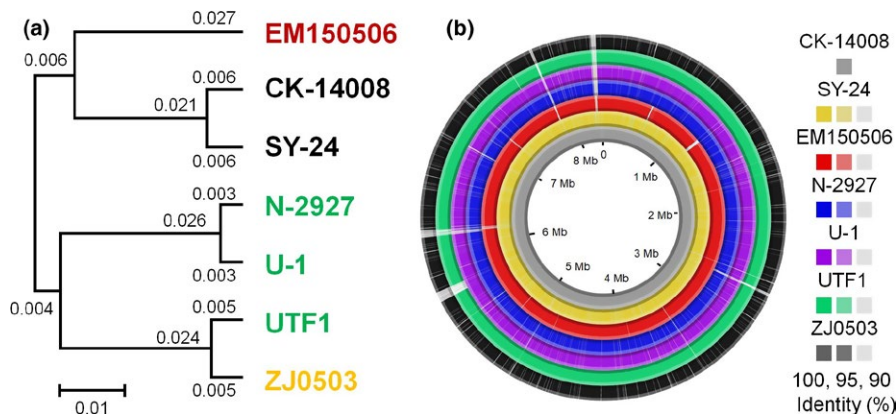
The OrthoANI values calculated from the genomes of seven *N. seriolae* strains were over 99.9%. However, a UPGMA dendrogram based on the OrthoANI values showed that the seven *N. seriolae* strains could be divided into four phylogenetic clusters (Cluster 1: EM150506; Cluster 2: CK-14008 and SY-24; Cluster 3: N-2927, and U-1; Cluster 4: UTF1 and ZJ0503) (Figure 2a). Interestingly, these clusters have similar grouping pattern according to the host family from which the strain was isolated. The strains in clusters 1–3 were isolated from fish belonging to the families Anguillidae, Channidae, and Carangidae, respectively. Meanwhile, the strains in cluster 4 were isolated from Carangidae and Stromateidae. These classifications were also observed in the results of the comparison of genome synteny based on the percent identity in nucleotide BLAST (Figure 2b).

These results indicate that there is a high possibility that the putative pathogenic *N. seriolae* strains might display host specificity according to their genomic features.

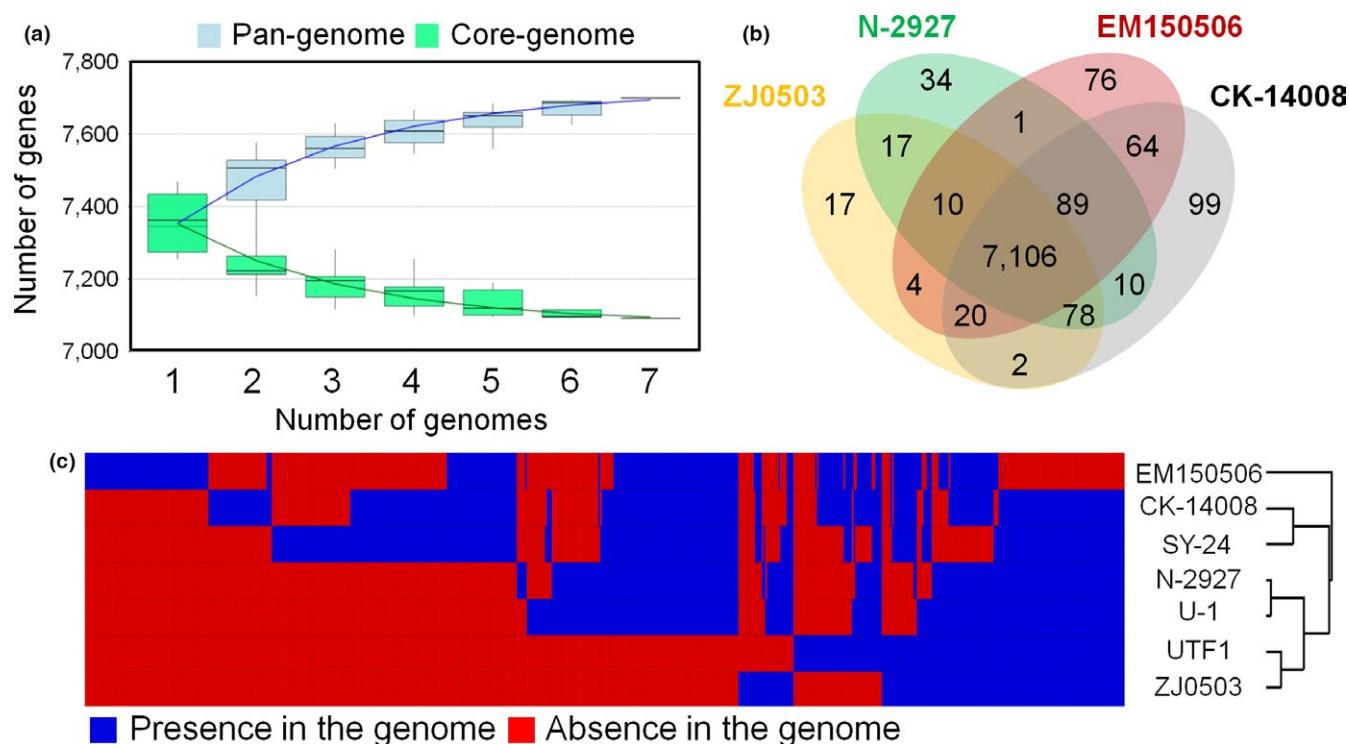
### 3.4 | Comparison of gene content among *N. seriolae* strains

Seven *N. seriolae* strains contained approximately 7,700 and 7,000 genes in their pan-genomic and core genome sizes, respectively





**FIGURE 2** Genomic comparison of different strains of *Nocardia seriolae*. (a) UPGMA dendrogram based on the OrthoANI values of seven *N. seriolae* strains (*N. seriolae* CK-14008, *N. seriolae* EM150506, *N. seriolae* N-2927, *N. seriolae* SY-24, *N. seriolae* U-1, *N. seriolae* UTF1, and *N. seriolae* ZJ0503). Colors of the strains indicate the isolation host of each strain. Red, black, green, and yellow colored strains indicate strains isolated from fish of the Anguillidae family, Channidae family, Carangidae family, and Stromateidae family, respectively. (b) Circular representation of genome synteny across seven *N. seriolae* strains. For synteny analysis, the genome sequence of CK-14008 was used as the reference, and sequence identities calculated using BLASTn between CK-14008 and each strain were used for colored representation of the genomic regions. Darkest colors: sequence identity of 95%–100%. Middle colors: sequence identity of 90%–95%. Brightest colors: sequence identity of <90%



**FIGURE 3** Pan-genome and core genome of *Nocardia seriolae* strains. (a) Pan-genome and core genome sizes of seven *N. seriolae* strains. Pan-genome:  $y = 7720.969 - 567.657e^{-x/2.300}$  ( $R^2 = 0.990$ ); core genome:  $y = 7078.227 + 437.414e^{-x/2.144}$  ( $R^2 = 0.973$ ). (b) Number of orthologous genes across four *N. seriolae* strains. The four strains were selected based on their isolation host and phylogenetic location in Figure 2a. (c) Heat-map illustrating the presence and absence of the genes in each genome of *N. seriolae*. The core genes of the seven strains are not displayed in the heat-map

(Figure 3a). In the core genome of the seven *N. seriolae* strains, genes encoding several kinds of antibiotic resistance proteins, cholesterol oxidase, filamentous hemagglutinin, the putative toxin HlgB2, and several antitoxins were detected, as well as proteins

involved in general cell metabolism (Figure 3b and Supporting Information Table S4).

To investigate the metabolic features associated with host specificity, we compared the orthologous genes from four *N. seriolae* strains

isolated from different fish families (Figure 3b), and analyzed the strain-specific genes in the seven *N. seriolae* strains (Figure 3c). In the genomes of the strain EM150506, CK-14008, SY-24, N-2927, U-1, UTF1, and ZJ0503, total of 74, 35, 47, 5, 11, 13, and 9 genes were detected as strain-specific genes, respectively. However, as most encoded hypothetical proteins, it was difficult to identify the host-specific metabolic features of the strains through the comparison of orthologous genes.

### 3.5 | Putative virulence-associated genes

Dozens of *Nocardia* species have been isolated from diseased hosts, and *N. seriolae* strains were mainly isolated from diseased fish. In the genomes of the analyzed strains, dozens of genes known to be candidate virulence factors in *N. farcinica* (Ishikawa et al., 2004; Yasuike et al., 2017) were detected using the parameters  $E \leq 1 \times 10^{-5}$  and  $\geq 50\%$  sequence identity (Table 2). The detected genes encoded proteins involved in penetration into mammalian cells (invasion), oxidative/nitrosative stresses (catalase and superoxide dismutase), and metal transporter (*nbt* genes for nocobactin biosynthesis) (Yasuike et al., 2017). In addition to these genes, several other genes encoding putative virulence factors, such as vibriolysin and toxin proteins, were detected in the genomes of the analyzed *Nocardia* strains.

Interestingly, eight to 12 *mce* operons, which encode mammalian cell entry proteins and are known as virulence factors in *Mycobacterium tuberculosis*, were detected in the genomes of all analyzed *Nocardia* strains (eight operons in *N. acidivorans*, *N. concava*, *N. crassostreae*, *N. niigatensis*, and *N. seriolae*, 10 operons in *N. yamanashiensis*, 11 operons in *N. jejuensis*, and 12 operons in *N. inohanensis*).

Moreover, a gene encoding vibriolysin was only detected in the genomes of all *N. seriolae* strains, *N. crassostreae* NBRC 100342, and *N. niigatensis* NBRC 100131. Although *N. crassostreae* phylogenetically clustered with the strains isolated from natural environments, the strain NBRC 100342 was isolated from diseased oysters, and *N. niigatensis* phylogenetically clustered with *N. seriolae*.

In the genomes of *Nocardia* species, several genes encoding toxin and antitoxin proteins were detected. Particularly, *N. seriolae* and *N. concava* showed approximately twofold more genes than the other species (Supporting Information Table S5). The genes encoding putative toxin HigB2, antitoxin MazE3, and putative antitoxin HigA3 were only detected in the genomes of *N. seriolae* strains and *N. concava* NBRC 100430. The gene encoding HC-toxin synthetase was also only detected in the genomes of *N. seriolae* strains and *N. yamanashiensis* NBRC 100130. In addition, the genes encoding antitoxin HipB and putative antitoxin Rv0268c were only detected in the genomes of *N. seriolae* N-2927, U-1, and ZJ0503.

These distributions of virulence-associated genes demonstrated that the virulence features of *Nocardia* strains were species-specific and highly related to their phylogenetic position and isolation origin.

### 3.6 | Secondary metabolite biosynthetic genes

Dozens of secondary metabolite biosynthetic genes were detected in the genomes of *Nocardia* strains, and were identified in

94 biosynthetic gene clusters (Supporting Information Table S6). *N. concava* NBRC 100430 had the most secondary metabolite biosynthetic genes, while *N. seriolae* had a relatively low number of genes compared to other *Nocardia* species.

The nocobactin biosynthetic gene cluster was detected in all analyzed genomes and showed a high similarity (at least 62%) with the gene cluster of *N. farcinica* IFM 1015 (accession no. AP006618) (Figure 4a). The biosynthetic gene cluster for nocobactin includes five core biosynthetic genes (OJF78308 to OJF78312); the biosynthetic modules for adenylation, acyltransferase, condensation, ketoreductase, ketoacyl synthase, peptidyl carrier protein, and thioesterase were detected in these core biosynthetic genes (Figure 4b). Approximately 30-kb upstream of the nocobactin biosynthetic gene cluster, the biosynthetic gene cluster for laspartomycin biosynthesis was detected (OJF78260 to OJF83928) with 37% similarity to the gene cluster of *Streptomyces viridochromogenes* ATCC 29814 (accession no. HM756254). The gene structures of the laspartomycin and nocobactin biosynthetic genes were highly conserved in the genomes of *N. seriolae* CK-14008, EM150506, and UTF1, which are currently the only complete genomes analyzed in this study.

Of the detected secondary metabolite biosynthetic gene clusters, the gene cluster for mycobactin biosynthesis was detected in the genomes of *Nocardia* species isolated from diseased fish or human medical samples (Supporting Information Table S6). The gene cluster showed 30%–40% gene similarity to the gene cluster of *M. tuberculosis* H37Rv (accession no. AL123456). Of the mycobactin biosynthetic genes in *M. tuberculosis*, only three genes shared homology with the genes from *Nocardia* species (Figure 4c,d). Approximately 10-kb downstream of the mycobactin biosynthetic gene cluster, two large nonribosomal peptide synthetase genes containing adenylation, condensation, and peptidyl carrier modules were detected (OJF82997 and OJF82998). However, they showed no homology with previously known secondary metabolite biosynthetic genes.

### 3.7 | Antibiotic resistance genes

In the genomes of *Nocardia* strains, resistance genes against several kinds of antibiotics were detected with more than 30% similarity to previously known antibiotic resistance genes (Figure 5). According to Yasuike et al., the *N. seriolae* strains can be divided into two groups according to their  $\alpha$ -glucosidase activity and susceptibility to erythromycin or oxytetracycline (Ismail, Takeshita, Umeda, Itami, & Yoshida, 2011; Yasuike et al., 2017). All *N. seriolae* strains analyzed in this study contained the gene encoding  $\alpha$ -glucosidase, but did not have resistance genes against erythromycin or oxytetracycline. Of the detected antibiotic resistance genes, genes involved in the resistance to macrolide were the most abundant in the genomes of *Nocardia* species, and the diversity of antibiotic resistance genes was the highest in the genomes of *N. seriolae* strains. In particular, the number of genes involved in resistance to vancomycin was approximately twofold higher in the genomes of *N. seriolae* strains than in those of other *Nocardia* species. Furthermore, the resistance genes against amikacin, cephalosporin, dibekacin, fluoroquinolone,

**TABLE 2** Number of the virulence gene candidates in the genomes of the analyzed *Nocardia* strains

<i>N. farcinica</i> IFM 10152			Analyzed <i>Nocardia</i> species													
Gene ID	Gene	Product	A	B	C	D	E	F	G	H	I	J	K	L	M	N
nfa37890	<i>ahpC</i>	Putative alkylhydroperoxide reductase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa37900	<i>ahpD</i>	Putative alkylhydroperoxidase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa1810	<i>fbpA</i>	Putative mycolyltransferase	6	7	6	6	7	6	6	6	6	6	6	5	4	6
nfa1820	<i>fbpB</i>	Putative mycolyltransferase	6	7	6	6	7	6	6	6	6	6	7	5	5	7
nfa1830	<i>fbpC</i>	Putative mycolyltransferase	4	5	4	4	5	4	4	5	5	6	6	4	3	5
nfa37790	<i>ideR</i>	Putative transcriptional regulator	1	1	1	1	1	1	1	1	1	2	1	1	1	1
nfa34810	<i>inv</i>	Putative invasins	1	1	1	1	1	1	1	1	1	2	1	1	1	1
nfa29500	<i>katG</i>	Putative catalase-peroxidase	1	1	1	1	1	1	1	1	1	1	1	1	3	1
nfa45490	<i>narG</i>	Putative nitrate reductase alpha subunit	2	2	2	2	2	2	2	1	2	2	2	2	2	2
nfa45500	<i>narH</i>	Putative nitrate reductase beta subunit	2	2	2	2	2	2	2	1	2	1	2	2	2	2
nfa45520	<i>narI</i>	Putative nitrate reductase gamma subunit	2	2	2	2	2	2	2	1	2	1	1	2	2	2
nfa45510	<i>narJ</i>	Putative nitrate reductase delta subunit	2	2	2	2	2	2	2	1	1	1	2	1	3	1
nfa45610	<i>nirB</i>	Putative nitrite reductase (NAD(P)H) subunit	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa45600	<i>nirD</i>	Putative nitrite reductase (NAD(P)H) subunit	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa7630	<i>nbtA</i>	Thioesterase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa7640	<i>nbtB</i>	Polyketide synthase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa7650	<i>nbtC</i>	Polyketide synthase	1	1	1	1	1	1	1	1	1	2	1	2	2	1
nfa7660	<i>nbtD</i>	Nonribosomal peptide synthetase	3	3	2	3	2	3	2	2	2	1	3	1	1	1
nfa7670	<i>nbtE</i>	Nonribosomal peptide synthetase	1	1	1	1	1	1	1	0	0	1	1	1	1	1
nfa7680	<i>nbtF</i>	Nonribosomal peptide synthetase	3	3	3	3	3	3	3	0	0	3	1	2	1	4
nfa7610	<i>nbtG</i>	Lysine-N-oxygenase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa6190	<i>nbtS</i>	Salicylate synthase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa6200	<i>nbtT</i>	Salicylate-AMP ligase	1	1	1	1	1	1	1	1	1	2	1	1	1	2
nfa13510	<i>ndk</i>	Putative nucleoside diphosphate kinase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa37880	<i>oxyR</i>	Putative hydrogen peroxide sensing transcriptional regulator	3	3	3	3	3	3	3	2	2	1	1	2	1	1
nfa16310	<i>ptpA</i>	Putative protein-tyrosine phosphatase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa18680	<i>ptpB</i>	Putative protein-tyrosine phosphatase	1	1	1	1	1	1	1	1	1	1	1	1	1	1

(Continues)

TABLE 2 (Continued)

<i>N. farcinica</i> IFM 10152			Analyzed <i>Nocardia</i> species													
Gene ID	Gene	Product	A	B	C	D	E	F	G	H	I	J	K	L	M	N
nfa52980	<i>sodC</i>	Putative superoxide dismutase	1	1	1	1	1	1	1	1	1	1	2	1	1	1
nfa1210	<i>sodF</i>	Putative superoxide dismutase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
nfa19960	<i>tlyA</i>	Putative cytotoxin/hemolysin	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Total number of CDSs			53	56	52	53	55	53	52	44	47	52	52	47	47	52

Virulence gene candidates were identified based on the genomic features of *N. seriolae* UTF1 (Yasuike et al., 2017). The amino acid sequences of the genes were obtained from the *N. farcinica* Genome Project Page (<http://nocardia.nih.go.jp/>). Coding DNA sequences (CDSs) with more than 50% sequence identity to the *N. farcinica* virulence proteins are shown. A, *N. seriolae* CK-14008; B, *N. seriolae* EM150506; C, *N. seriolae* N-2927; D, *N. seriolae* SY-24; E, *N. seriolae* U-1; F, *N. seriolae* UTF1; G, *N. seriolae* ZJ0503; H, *N. acidivorans* NBRC 108247; I, *N. jejuensis* NBRC 103114; J, *N. crassostreae* NBRC 100342; K, *N. concava* NBRC 100430; L, *N. inohanensis* NBRC 100128; M, *N. niigatensis* NBRC 100131; N, *N. yamanashiensis* NBRC 100130.

iseipamicin, netilmicin, sisomicin, streptomycin, tobramycin, and tobramycinilmicin were mainly detected in the genomes of *N. seriolae* strains.

## 4 | DISCUSSION

*Nocardia* strains are widespread in diverse habitats such as soil and water (Luo, Hiessl, & Steinbuchel, 2014). Some of pathogenic *Nocardia* species causing nocardiosis have been detected in human and animal such as cat, dog, and fish (Eroksuz et al., 2017; Harada et al., 2009; Kudo et al., 1988). Recently, the occurrence of nocardiosis in farming fish has been increasing gradually because of the high-environmental stresses caused by the dense cultivation of fish and environmental conditions that favor pathogens, such as like warming seawater (Le Roux et al., 2015; Pulkkinen et al., 2010). However, there are currently no treatment options to cure nocardiosis in fish; suppressing the growth of the pathogens by antibiotic treatment is the only viable method to prevent disease occurrence (Nayak & Nakanishi, 2016).

For several decades, the virulence factors of some pathogenic *Nocardia* species have been actively studied, and several virulence features were identified including invasion into the host cells, survival in the cells, and bacterial lytic activity (Beaman & Beaman, 1994).

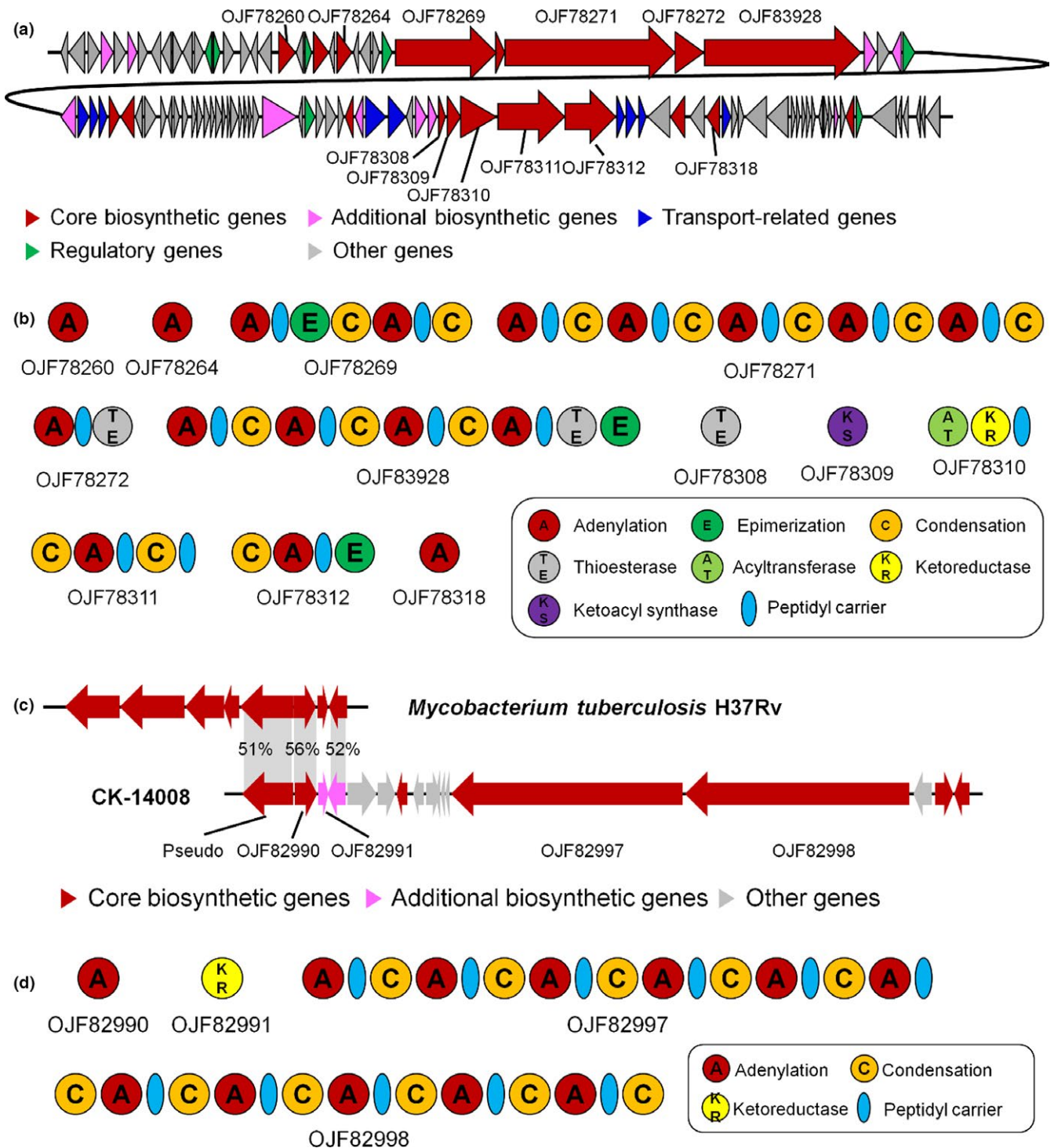
The major virulence feature of pathogenic *Nocardia* species is the invasion of the bacterium into host cells including macrophages (Beaman & Beaman, 1994). For attachment and invasion into the host cells, the most well-known virulence factor in the genus *Nocardia* is the mammalian cell entry (Mce) family of proteins (Arruda, Bomfim, Knights, Huima-Byron, & Riley, 1993; Yasuike et al., 2017). In our study, eight to 12 copies of *mce* operons were detected in the genomes of *Nocardia* species. Particularly, eight and 11 copies of *mce* operons were detected in the genomes of soil-derived *N. acidivorans* NBRC 108247 and *N. jejuensis* NBRC 103114, respectively. This indicates that *Nocardia* species isolated from natural environments such as soil and water might invade animal cells under certain conditions.

In host cells, especially macrophages, invading pathogens have to defend against reactive oxygen species produced by the defense responses of the host cells (Fang, 2004). To defend against this oxidative attack by host cells, *Nocardia* species produce antioxidant proteins such as catalase and superoxide dismutase. Like the *mce* operons, the genes encoding catalase and superoxide dismutase were detected in all analyzed genomes (Table 2). For survival in the host cells, invading pathogens must actively acquire small concentrations of metal compounds such as iron; *Nocardia* species were confirmed to have genes encoding biosynthetic proteins for several kinds of siderophores (Table 2 and Supporting Information Table S6). However, in the genomes of the two soil isolates *N. acidivorans* NBRC 108247 and *N. jejuensis* NBRC 103114, the *nbtE* and *nbtF* genes for the biosynthesis of nocobactin were not detected, (Table 2) although the gene cluster was detected by the prediction program for the secondary metabolite biosynthetic gene cluster (Supporting Information Table S6). In addition, the gene cluster for mycobactin biosynthesis was only detected in the genomes of strains isolated from diseased fish and human medical samples. Moreover, the gene cluster for griseobactin biosynthesis was only detected in the genomes of *N. seriolae* strains and *N. concava* NBRC 100430. These results demonstrated that iron acquisition might be an important factor for transition from the natural *Nocardia* strains to the pathogenic strains.

Interestingly, the gene encoding vibriolysin was only detected in the genomes of *N. seriolae*, *N. crassostreae* NBRC 100342, and *N. niigatensis* NBRC 100131. *N. seriolae* and *N. crassostreae* were isolated from diseased marine organisms. However, the gene encoding cytotoxin/hemolysin was detected in all of the analyzed *Nocardia* genomes. This indicates that vibriolysin, which was previously reported as a virulence factor in fish pathogens (Bjornsdottir et al., 2009), can be a distinguishing virulence factor for pathogenic *Nocardia* strains in aquatic environments.

Resistance against several kinds of antibiotics is a major feature of the genus *Nocardia* (Hashemi-Shahraki et al., 2015; Ismail et al., 2011). Pathogens possessing diverse antibiotic resistance genes can cause severe problems, such as the transfer





**FIGURE 4** Nocobactin and mycobactin biosynthetic gene clusters in the genome of *Nocardia seriolae* CK-14008. (a) Genomic region around the nocobactin biosynthetic gene cluster. Red colored genes—from OJF78308 to OJF78312—indicate the core nocobactin biosynthetic genes. The core biosynthetic genes—from OJF78260 to OJF83928—located upstream of the nocobactin biosynthetic genes showed high homology with laspartomycin biosynthetic genes. This gene structure was highly conserved in the genomes of *N. seriolae* EM150506 and UTF1, which are fully sequenced *N. seriolae* genomes. The coding DNA sequence (CDS) numbers follow the protein ID from GenBank. (b) Structure of the biosynthetic modules detected in the core biosynthetic genes of nocobactin (OJF78308 to OJF78312) and laspartomycin (OJF78260 to OJF83928). (c) Comparison of the mycobactin biosynthetic gene cluster between *N. seriolae* and *Mycobacterium tuberculosis* H37Rv. Of the mycobactin biosynthetic genes in *M. tuberculosis*, only three genes (gray shadows) showed homology with the *N. seriolae* genes. Percentages indicate the sequence identity between two genes. Downstream of the mycobactin biosynthetic gene cluster of *N. seriolae*, additionally large secondary metabolite biosynthetic genes were detected (OJF82997 and OJF82998), but showed no homology with previously known secondary metabolites. This gene structure was highly conserved in the genomes of *N. seriolae* EM150506 and UTF1. The CDS numbers follow the protein ID from GenBank. (d) Structure of the biosynthetic modules detected in the core biosynthetic genes

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Amikacin, dibekacin, isepamicin, netilmicin, sisomicin, tobramycin	1	1	1	1	1	1	1	-	-	-	1	-	-	-
Aminoglycoside	-	-	-	-	-	-	-	1	1	-	-	1	-	1
Bacitracin	2	2	2	2	2	2	2	3	1	5	5	2	2	3
Carbapenem, cephalosporin,cephamycin, penicillin	-	-	-	-	-	-	-	1	-	2	-	-	1	-
Cephalosporin	1	1	1	1	1	1	1	-	-	-	-	-	-	1
Chloramphenicol	1	1	1	1	1	1	1	-	-	2	3	-	-	-
Fluoroquinolone	1	1	1	1	1	1	1	-	-	-	-	-	1	-
Hygromycin B	-	-	-	-	-	-	-	-	-	-	-	1	-	-
Lincomycin	-	-	-	-	-	-	-	-	-	-	-	1	-	-
Lincosamide, macrolide, streptogramin B	2	2	2	2	2	2	2	1	4	2	1	2	1	4
Macrolide	4	4	4	4	4	4	4	5	6	6	6	4	7	3
Neomycin, paromomycin, ribostamycin	-	-	-	-	-	-	-	-	1	-	-	-	-	-
Penicillin	1	1	1	1	1	1	1	-	1	-	1	1	-	2
Qa_compound	-	-	-	-	-	-	-	-	-	1	-	-	-	-
Streptogramin A	2	2	3	2	3	2	2	3	3	1	1	1	1	1
Streptomycin	1	1	1	1	1	1	1	-	-	-	-	-	1	-
Streptothricin	-	-	-	-	-	-	-	1	1	1	1	-	1	-
Teicoplanin	-	-	-	-	-	-	-	-	-	-	1	-	-	1
Tetracenomycin C	1	1	-	1	-	-	-	-	-	1	1	-	-	1
Tetracycline	2	2	2	2	2	2	2	2	1	4	1	1	2	2
Trimethoprim	1	1	1	1	1	1	1	1	1	1	-	1	1	1
Tunicamycin	-	-	-	-	-	-	-	-	-	1	1	-	-	-
Vancomycin	4	4	4	4	4	4	4	2	2	2	3	1	1	1
Unidentified	-	-	-	-	-	-	-	-	-	-	-	3	-	1
Total	24	24	24	24	24	23	23	20	22	29	26	19	19	22

**FIGURE 5** Detected genes involved in antibiotic resistance. The color gradient indicates the abundance of genes in each genome resistant to each antibiotic. A, *N. seriolae* CK-14008; B, *N. seriolae* EM150506; C, *N. seriolae* N-2927; D, *N. seriolae* SY-24; E, *N. seriolae* U-1; F, *N. seriolae* UTF1; G, *N. seriolae* ZJ0503; H, *N. acidivorans* NBRC 108247 2; I, *N. jejuensis* NBRC 103114; J, *N. crassostreae* NBRC 100342; K, *N. concava* NBRC 100430; L, *N. inohanensis* NBRC 100128; M, *N. niigatensis* NBRC 100131; N, *N. yamanashiensis* NBRC 100130

of antibiotic resistance genes to other bacteria and an increase in multi-drug resistant pathogens. In the genomes of the analyzed *Nocardia* strains, many genes encoding antibiotic resistance were detected, (Figure 5) and they were classified into three classes: inhibition of protein synthesis, inhibition of cell wall synthesis, and quinolones. Of them, eight genes were detected in all of the analyzed *Nocardia* genomes with a role in the resistance against antibiotics for the inhibition of protein synthesis (lincosamide,

macrolide, streptogramin\_a, streptogramin\_b, and tetracycline) and inhibition of cell wall synthesis (bacitracin, penicillin, and vancomycin). Particularly, the *N. seriolae* strains had a more diverse range of antibiotic resistance genes involved in the inhibition of protein synthesis. These results indicate that in the fish farming industries, the antibiotics involved in the inhibition of protein and cell wall synthesis would not be suitable for the prevention of the nocardiosis.

## ACKNOWLEDGMENTS

This work was financially supported by the National Institute of Fisheries Science, Republic of Korea (R2018062).

## CONFLICT OF INTEREST

The authors declare there are no conflicts of interest.

## ORCID

Byung-Yong Kim  <http://orcid.org/0000-0002-4229-8859>

## REFERENCES

- Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., & Beatson, S. A. (2011). BLAST ring image generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics*, 12, 402. <https://doi.org/10.1186/1471-2164-12-402>
- Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T., & Riley, L. W. (1993). Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science*, 261, 1454–1457. <https://doi.org/10.1126/science.8367727>
- Beaman, B. L., & Beaman, L. (1994). *Nocardia* species: Host-parasite relationships. *Clinical Microbiology Reviews*, 7, 213–264. <https://doi.org/10.1128/CMR.7.2.213>
- Bjornsdottir, B., Fridjonsson, O. H., Magnúsdóttir, S., Andrésdóttir, V., Hreggvidsson, G. O., & Gudmundsdóttir, B. K. (2009). Characterisation of an extracellular vibriolysin of the fish pathogen *Moritella viscosa*. *Veterinary Microbiology*, 136, 326–334. <https://doi.org/10.1016/j.vetmic.2008.11.020>
- Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., ... Dickschat, J. S. (2017). antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research*, 45, W36–W41. <https://doi.org/10.1093/nar/gkx319>
- Brosnahan, C. L., Humphrey, S., Knowles, G., Ha, H. J., Pande, A., & Jones, J. B. (2017). Nocardiosis in freshwater reared Chinook salmon (*Oncorhynchus tshawytscha*). *New Zealand Veterinary Journal*, 65, 214–218. <https://doi.org/10.1080/00480169.2017.1314794>
- Brown-Elliott, B. A., Brown, J. M., Conville, P. S., & Wallace, R. J. Jr (2006). Clinical and laboratory features of the *Nocardia* spp. based on current molecular taxonomy. *Clinical Microbiology Reviews*, 19, 259–282. <https://doi.org/10.1128/CMR.19.2.259-282.2006>
- Byadgi, O., Chen, C. W., Wang, P. C., Tsai, M. A., & Chen, S. C. (2016). De Novo transcriptome analysis of differential functional gene expression in largemouth bass (*Micropterus salmoides*) after challenge with *Nocardia seriolae*. *International Journal of Molecular Sciences*, 17, E1315. <https://doi.org/10.3390/ijms17081315>
- Chen, S. C., Lee, J. L., Lai, C. C., Gu, Y. W., Wang, C. T., Chang, H. Y., & Tsai, K. H. (2000). Nocardiosis in sea bass, *Lateolabrax japonicus*, in Taiwan. *Journal of Fish Diseases*, 23, 299–307. <https://doi.org/10.1046/j.1365-2761.2000.00217.x>
- Chun, J., & Goodfellow, M. (1995). A phylogenetic analysis of the genus *Nocardia* with 16S rRNA gene sequences. *International Journal of Systematic Bacteriology*, 45, 240–245. <https://doi.org/10.1099/00207713-45-2-240>
- Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., ... Brettin, T. S. (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 15442–15447. <https://doi.org/10.1073/pnas.0907787106>
- Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, 23, 673–679. <https://doi.org/10.1093/bioinformatics/btm009>
- Ellwood, D.C. (1973). Pulmonary nocardiosis in a goat in Malawi. *The British Veterinary Journal*, 129, iv–vii. [https://doi.org/10.1016/S0007-1935\(17\)36561-2](https://doi.org/10.1016/S0007-1935(17)36561-2)
- Eroksuz, Y., Gursoy, N. C., Karapinar, T., Karabulut, B., Incili, C. A., Yerlikaya, Z., ... Eroksuz, H. (2017). Systemic nocardiosis in a dog caused by *Nocardia cyriacigeorgica*. *BMC Veterinary Research*, 13, 30–35.
- Fang, F. C. (2004). Antimicrobial reactive oxygen and nitrogen species: Concepts and controversies. *Nature Reviews Microbiology*, 2, 820–832. <https://doi.org/10.1038/nrmicro1004>
- Friedman, C.S., Beaman, B. L., Chun, J., Goodfellow, M., Gee, A., & Hedrick, R. P. (1998). *Nocardia crassostreae* sp. nov., the causal agent of nocardiosis in Pacific oysters. *International Journal of Systematic Bacteriology*, 48 Pt 1, 237–246. <https://doi.org/10.1099/00207713-48-1-237>
- Harada, H., Endo, Y., Sekiguchi, M., Setoguchi, A., & Momoi, Y. (2009). Cutaneous nocardiosis in a cat. *The Journal of Veterinary Medical Science*, 71, 785–787. <https://doi.org/10.1292/jvms.71.785>
- Hashemi-Shahraki, A., Heidarieh, P., Bostanabad, S. Z., Hashemzadeh, M., Feizabadi, M. M., Schraufnagel, D., & Mirsaedi, M. (2015). Genetic diversity and antimicrobial susceptibility of *Nocardia* species among patients with nocardiosis. *Scientific Reports*, 5, 17862.
- Hirayama, T., Takazono, T., Horai, Y., Tashiro, M., Saijo, T., Kosai, K., ... Miyazaki, T. (2016). Pulmonary nocardiosis caused by *Nocardia concava* with a literature review. *Internal Medicine*, 55, 1213–1217. <https://doi.org/10.2169/internalmedicine.55.6228>
- Ho, P. Y., Byadgi, O., Wang, P. C., Tsai, M. A., Liaw, L. L., & Chen, S. C. (2016). Identification, molecular cloning of IL-1beta and its expression profile during *Nocardia seriolae* infection in largemouth bass, *Micropterus salmoides*. *International Journal of Molecular Sciences*, 17, E1670. <https://doi.org/10.3390/ijms17101670>
- Huang, Y., Lou, H., Wu, X., & Chen, Y. (2008). Characterization of the BPI-like gene from a subtracted cDNA library of large yellow croaker (*Pseudosciaena crocea*) and induced expression by formalin-inactivated *Vibrio alginolyticus* and *Nocardia seriolae* vaccine challenges. *Fish & Shellfish Immunology*, 25, 740–750. <https://doi.org/10.1016/j.fsi.2008.02.012>
- Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., & Harris, S. R. (2015). Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, 16, 294. <https://doi.org/10.1186/s13059-015-0849-0>
- Imajoh, M., Fukumoto, Y., Yamane, J., Sukeda, M., Shimizu, M., Ohnishi, K., & Oshima, S. (2015). Draft genome sequence of *Nocardia seriolae* strain N-2927 (NBRC 110360), isolated as the causal agent of nocardiosis of yellowtail (*Seriola quinqueradiata*) in Kochi Prefecture, Japan. *Genome Announcements*, 3, e00082–00015.
- Imajoh, M., Sukeda, M., Shimizu, M., Yamane, J., Ohnishi, K., & Oshima, S. (2016). Draft genome sequence of erythromycin- and oxytetracycline-sensitive *Nocardia seriolae* strain U-1 (NBRC 110359). *Genome Announcements*, 4, e01606–e01615.
- Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., ... Hattori, M. (2004). The complete genomic sequence of *Nocardia farcinica* IFM 10152. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 14925–14930. <https://doi.org/10.1073/pnas.0406410101>
- Isik, K., Chun, J., Hah, Y. C., & Goodfellow, M. (1999). *Nocardia salmonicida* nom. rev., a fish pathogen. *International Journal of Systematic Bacteriology*, 49 Pt 2, 833–837. <https://doi.org/10.1099/00207713-49-2-833>
- Ismail, T. F., Takeshita, A., Umeda, N., Itami, T., & Yoshida, T. (2011). Application of  $\alpha$ -glucosidase activity and drug susceptibility tests to epidemiological studies on the fish pathogen *Nocardia*

- seriolae*. *Fisheries Science*, 77, 113–118. <https://doi.org/10.1007/s12562-010-0302-y>
- Itano, T., Kawakami, H., Kono, T., & Sakai, M. (2006). Experimental induction of nocardiosis in yellowtail, *Seriola quinqueradiata* Temminck and Schlegel by artificial challenge. *Journal of Fish Diseases*, 29, 529–534. <https://doi.org/10.1111/j.1365-2761.2006.00748.x>
- Kageyama, A., Yazawa, K., Nishimura, K., & Mikami, Y. (2004). *Nocardia inohanensis* sp. nov., *Nocardia yamanashiensis* sp. nov. and *Nocardia niigatensis* sp. nov., isolated from clinical specimens. *International Journal of Systematic and Evolutionary Microbiology*, 54, 563–569. <https://doi.org/10.1099/ijs.0.02794-0>
- Kampfer, P., Huber, B., Buczolits, S., Thummes, K., Grun-Wollny, I., & Busse, H. J. (2007). *Nocardia acidivorans* sp. nov., isolated from soil of the island of Stromboli. *International Journal of Systematic and Evolutionary Microbiology*, 57, 1183–1187. <https://doi.org/10.1099/ijs.0.64813-0>
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Research*, 32, D277–D280. <https://doi.org/10.1093/nar/gkh063>
- Kudo, T., Hatai, K., & Seino, A. (1988). *Nocardia seriolae* sp. nov. causing nocardiosis of cultured fish. *International Journal of Systematic and Evolutionary Microbiology*, 38, 173–178.
- Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAMmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35, 3100–3108. <https://doi.org/10.1093/nar/gkm160>
- Le Roux, F., Wegner, K. M., Baker-Austin, C., Vezzulli, L., Osorio, C. R., Amaro, C., ... Mazel, D. (2015). The emergence of *Vibrio* pathogens in Europe: Ecology, evolution, and pathogenesis (Paris, 11–12th March 2015). *Frontiers in Microbiology*, 6, 830.
- Lee, S. D. (2006). *Nocardia jejuensis* sp. nov., a novel actinomycete isolated from a natural cave on Jeju Island, Republic of Korea. *International Journal of Systematic and Evolutionary Microbiology*, 56, 559–562. <https://doi.org/10.1099/ijs.0.63866-0>
- Lee, I., Kim, Y. O., Park, S. C., & Chun, J. (2015). OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology*, 66, 1100–1103.
- Liu, B., & Pop, M. (2009). ARDB—antibiotic resistance genes database. *Nucleic Acids Research*, 37, D443–D447. <https://doi.org/10.1093/nar/gkn656>
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25, 955–964. <https://doi.org/10.1093/nar/25.5.0955>
- Luo, Q., Hiessl, S., & Steinbuchel, A. (2014). Functional diversity of *Nocardia* in metabolism. *Environmental Microbiology*, 16, 29–48. <https://doi.org/10.1111/1462-2920.12221>
- Matsumoto, M., Hayashi, K., Suetake, H., Yamamoto, A., & Araki, K. (2016). Identification and functional characterization of multiple interleukin 12 in amberjack (*Seriola dumerili*). *Fish & Shellfish Immunology*, 55, 281–292. <https://doi.org/10.1016/j.fsi.2016.05.025>
- Nayak, S. K., & Nakanishi, T. (2016). Development of vaccines against nocardiosis in fishes. *Methods in Molecular Biology*, 1404, 193–201. <https://doi.org/10.1007/978-1-4939-3389-1>
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., ... Vonstein, V. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Research*, 42, D206–D214. <https://doi.org/10.1093/nar/gkt1226>
- Pulkkinen, K., Suomalainen, L. R., Read, A. F., Ebert, D., Rintamaki, P., & Valtonen, E. T. (2010). Intensive fish farming and the evolution of pathogen virulence: The case of columnaris disease in Finland. *Proceedings Biological Sciences*, 277, 593–600. <https://doi.org/10.1098/rspb.2009.1659>
- Shimahara, Y., Nakamura, A., Nomoto, R., Itami, T., Chen, S. C., & Yoshida, T. (2008). Genetic and phenotypic comparison of *Nocardia seriolae* isolated from fish in Japan. *Journal of Fish Diseases*, 31, 481–488. <https://doi.org/10.1111/j.1365-2761.2008.00920.x>
- Tamura, T., Matsuzawa, T., Oji, S., Ichikawa, N., Hosoyama, A., Katsumata, H., ... Fujita, N. (2012). A genome sequence-based approach to taxonomy of the genus *Nocardia*. *Antonie van Leeuwenhoek*, 102, 481–491. <https://doi.org/10.1007/s10482-012-9780-5>
- Tamura, T., Ohji, S., Ichikawa, N., Hosoyama, A., Yamazoe, A., Hamada, M., ... Suzuki, K. I. (2018). Reclassification of *Nocardia* species based on whole genome sequence and associated phenotypic data. *The Journal of Antibiotics*, <https://doi.org/10.1038/s41429-018-0043-1>
- Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., ... Koonin, E. V. (2001). The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*, 29, 22–28. <https://doi.org/10.1093/nar/29.1.22>
- Thawai, C., Rungjindamai, N., Klanbut, K., & Tanasupawat, S. (2017). *Nocardia xestospongiae* sp. nov., isolated from a marine sponge in the Andaman Sea. *International Journal of Systematic and Evolutionary Microbiology*, 67, 1451–1456.
- Wang, P. C., Chen, S. D., Tsai, M. A., Weng, Y. J., Chu, S. Y., Chern, R. S., & Chen, S. C. (2009). *Nocardia seriolae* infection in the three striped tigerfish, *Terapon jarbua* (Forsskal). *Journal of Fish Diseases*, 32, 301–310. <https://doi.org/10.1111/j.1365-2761.2008.00991.x>
- Ward, N., & Moreno-Hagelsieb, G. (2014). Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PLoS ONE*, 9, e101850. <https://doi.org/10.1371/journal.pone.0101850>
- Watanabe, K., & Harayama, S. (2001). SWISS-PROT: The curated protein sequence database on Internet. *Tanpakushitsu Kakusan Koso Protein, Nucleic Acid, Enzyme*, 46, 80–86.
- Xia, L., Cai, J., Wang, B., Huang, Y., Jian, J., & Lu, Y. (2015). Draft genome sequence of *Nocardia seriolae* ZJ0503, a fish pathogen isolated from *Trachinotus ovatus* in China. *Genome Announcements*, 3, e01223–01214.
- Yasuike, M., Nishiki, I., Iwasaki, Y., Nakamura, Y., Fujiwara, A., Shimahara, Y., ... Katoh, M. (2017). Analysis of the complete genome sequence of *Nocardia seriolae* UTF1, the causative agent of fish nocardiosis: The first reference genome sequence of the fish pathogenic *Nocardia* species. *PLoS ONE*, 12, e0173198. <https://doi.org/10.1371/journal.pone.0173198>
- Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., & Chun, J. (2017). Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International Journal of Systematic and Evolutionary Microbiology*, 67, 1613–1617.
- Yu, S., Wang, J., Fang, Q., Zhang, J., & Yan, F. (2016). Specific clinical manifestations of *Nocardia*: A case report and literature review. *Experimental and Therapeutic Medicine*, 12, 2021–2026. <https://doi.org/10.3892/etm.2016.3571>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Han H-J, Kwak M-J, Ha S-M, et al. Genomic characterization of *Nocardia seriolae* strains isolated from diseased fish. *MicrobiologyOpen*. 2019;8:e656. <https://doi.org/10.1002/mbo3.656>