

RESEARCH ARTICLE

Open Access

Genome based analysis of type-I polyketide synthase and nonribosomal peptide synthetase gene clusters in seven strains of five representative *Nocardia* species

Hisayuki Komaki¹, Natsuko Ichikawa², Akira Hosoyama², Azusa Takahashi-Nakaguchi³, Tetsuhiro Matsuzawa³, Ken-ichiro Suzuki¹, Nobuyuki Fujita² and Tohru Gono^{1,3*}

Abstract

Background: Actinobacteria of the genus *Nocardia* usually live in soil or water and play saprophytic roles, but they also opportunistically infect the respiratory system, skin, and other organs of humans and animals. Primarily because of the clinical importance of the strains, some *Nocardia* genomes have been sequenced, and genome sequences have accumulated. Genome sizes of *Nocardia* strains are similar to those of *Streptomyces* strains, the producers of most antibiotics. In the present work, we compared secondary metabolite biosynthesis gene clusters of type-I polyketide synthase (PKS-I) and nonribosomal peptide synthetase (NRPS) among genomes of representative *Nocardia* species/strains based on domain organization and amino acid sequence homology.

Results: Draft genome sequences of *Nocardia asteroides* NBRC 15531^T, *Nocardia otitidiscaviarum* IFM 11049, *Nocardia brasiliensis* NBRC 14402^T, and *N. brasiliensis* IFM 10847 were read and compared with published complete genome sequences of *Nocardia farcinica* IFM 10152, *Nocardia cyriacigeorgica* GUH-2, and *N. brasiliensis* HUJEG-1. Genome sizes are as follows: *N. farcinica*, 6.0 Mb; *N. cyriacigeorgica*, 6.2 Mb; *N. asteroides*, 7.0 Mb; *N. otitidiscaviarum*, 7.8 Mb; and *N. brasiliensis*, 8.9 - 9.4 Mb. Predicted numbers of PKS-I, NRPS, and PKS-I/NRPS hybrid clusters ranged between 4–11, 7–13, and 1–6, respectively, depending on strains, and tended to increase with increasing genome size. Domain and module structures of representative or unique clusters are discussed in the text.

Conclusion: We conclude the following: 1) genomes of *Nocardia* strains carry as many PKS-I and NRPS gene clusters as those of *Streptomyces* strains, 2) the number of PKS-I and NRPS gene clusters in *Nocardia* strains varies substantially depending on species, and *N. brasiliensis* strains carry the largest numbers of clusters among the species studied, 3) the seven *Nocardia* strains studied in the present work have seven common PKS-I and/or NRPS clusters, some of whose products are yet to be studied, and 4) different *N. brasiliensis* strains have some different gene clusters of PKS-I/NRPS, although the rest of the clusters are common within the *N. brasiliensis* strains. Genome sequencing suggested that *Nocardia* strains are highly promising resources in the search of novel secondary metabolites.

Keywords: *Nocardia asteroides*, *Nocardia otitidiscaviarum*, *Nocardia brasiliensis*, *Nocardia farcinica*, *Nocardia cyriacigeorgica*, Genome sequence, Type-I polyketide synthase, Nonribosomal peptide synthetase

* Correspondence: gono¹@faculty.chiba-u.jp

³Medical Mycology Research Center (MMRC), Chiba University, Chuo-ku, Chiba 260-8673, Japan

Full list of author information is available at the end of the article

Background

Actinomycetous strains of the genus *Nocardia* usually live in soil or water and play saprophytic roles in the environment, but also are opportunistic human pathogens, infecting the respiratory tract, skin, brain, and other organs of both immunocompromised and immunocompetent patients. To date, more than 80 species have been established in the genus *Nocardia*, and approximately one-third to one-half of the species have been reported as human pathogens [1-3]. Because of their medical importance, *Nocardia* strains have accumulated in microbial collections as a resource for clinical and scientific studies in the last few decades (e.g., [4-7]).

Although *Nocardia* strains belong to the Order *Actinomycetales* together with *Streptomyces* strains, the latter being known as a rich resource for discovery of secondary metabolites, few studies have been focused on secondary metabolites and their synthetic genes in *Nocardia* strains.

Type I polyketide synthase (PKS-I) and nonribosomal peptide synthetase (NRPS) gene clusters are two of the major secondary metabolite-producing clusters in bacteria and are involved in the biosynthesis of polyketide chains and nonribosomal peptides, respectively. It has been found that these clusters produce several medically and industrially important compounds, such as pathogenic factors, avermectin, erythromycin, and vancomycin.

In the present paper, we searched for PKS-I and NRPS genes in the genomes of representative *Nocardia* strains and analyzed their sequence similarities and differences in domain/module structures. While we were sequencing and analyzing *Nocardia* draft genomes, two new *Nocardia* genomes of *N. cyriacigeorgica* GUH-2 [8,9] and *N. brasiliensis* HUJEG-1 [10] were published. We included them in the present analysis together with *N. farcinica* genome, which our group has published previously [11].

Methods

Strains

N. otitidiscaviarum IFM 11049 and *N. brasiliensis* IFM 10847 were from the IFM culture collections of MMRC, Chiba University, Japan [12]. *N. asteroides* NBRC 15531^T and *N. brasiliensis* NBRC 14402^T were from the NBRC culture collection [5]. Cells were cultured in brain heart infusion liquid culture medium (Difco) in the conventional manner.

Acquisition of whole-genome sequences

Genomic DNA of *N. otitidiscaviarum* IFM 11049, *N. brasiliensis* (IFM 10847, NBRC 14402^T), and *N. asteroides* NBRC 15531^T was prepared as described previously [13]. Genome sequences were read by the pyrosequencing method using genome sequencer GS FLX Instruments and GS FLX Titanium Kits (Roche Applied Science,

Japan). The read redundancy for the four draft genomes ranged between 55 and 104. We assembled the sequence reads of *N. otitidiscaviarum* IFM 11049, *N. brasiliensis* IFM 10847, *N. brasiliensis* NBRC 14402^T, and *N. asteroides* NBRC 15531^T, and obtained 65, 223, 115, and 39 contigs, which were longer than 500 bp. The estimated genome sizes of *N. otitidiscaviarum* IFM 11049, *N. brasiliensis* IFM 10847, *N. brasiliensis* NBRC 14402^T, and *N. asteroides* NBRC 15531^T were 7.9 Mb, 9.2 Mb, 8.9 Mb, and 7.0 Mb, respectively. The draft genome sequences of *N. otitidiscaviarum* IFM 11049, *N. brasiliensis* (IFM 10847, NBRC 14402^T), and *N. asteroides* NBRC 15531^T are available at GenBank/EMBL/DDBJ under the accession numbers BATZ01000001–BATZ01000065, BAUA01000001–BAUA01000223, BAFT01000001–BAFT01000128, and BAFO01000001–BAFO01000049, respectively. The complete genome sequences of *N. cyriacigeorgica* GUH-2, *N. brasiliensis* HUJEG-1 (=ATCC 700358), and *N. farcinica* IFM 10152 were downloaded from DDBJ [14], with accession numbers FO082843, CP0033876, and AP006618, respectively.

Analysis of PKS-I and NRPS gene clusters

The assembled contig sequences were submitted to the auto-annotation pipeline MiGAP [15,16] at DDBJ as described previously [17]. Assigned ORFs were further searched for signature domains of PKS-I and NRPS genes using the InterPro domain database [18,19]. ORFs having ketosynthase (KS) domain (IPR014030, IPR014031, IPR020841) or condensation (C) domain (IPR001242) were identified, and their adjacent genes were further analyzed as PKS-I and NRPS gene candidates. Module organizations were determined manually based on search results using InterPro database, results using PKS/NRPS analysis website [20], and signature sequences deduced using MOTIF search [21]. We also used antiSMASH [22,23], a website for antibiotics and secondary metabolite analysis, for finding orthologous clusters and predicting substrates for adenylation domains. PKS-I and NRPS gene clusters of *N. farcinica* IFM 10152, *N. cyriacigeorgica* GUH-2, and *N. brasiliensis* HUJEG-1 were also identified using the *N. farcinica* genomic database [11,24]. We assumed that two or more PKS-I and/or NRPS genes that were adjacent to each other constitute one cluster for secondary metabolite production (See Additional file 1: Table S1, for details and exceptions). We also assumed that one multi-domain PKS-I or NRPS gene that was not accompanied by adjacent PKS-I/NRPS genes constitute one independent cluster. However, genes having only a single PKS-I or NRPS domain were excluded from the present analysis because we considered them atypical, and focused on multi-domain clusters. The contig sequences containing PKS-I and NRPS gene clusters are available at GenBank/EMBL/DDBJ under the following accession numbers: [AB700569 - AB700587] (*N. otitidiscaviarum*

IFM 11049), [AB701575 - AB701605] (*N. brasiliensis* IFM 10847), [AB701607 - AB701636] (*N. brasiliensis* NBRC 14402^T), and [AB685274], [AB700124 - AB700133], [AB700557 - AB700568] (*N. asteroides* NBRC 15531^T).

Search for orthologous gene clusters among species and strains

BLASTP search was performed using the NCBI Protein BLAST program against the non-redundant protein sequence database [25,26]. We considered *Nocardia* genes homologous to other genes when they have more than 70% sequence similarity in BLASTP search, and also when their domain organizations have high similarity. We also compared clusters with domain organizations that only partially match each other, as described in the text.

Results and discussion

The two leftmost columns in Table 1 list *Nocardia* strains studied in the present paper and their exact (complete genome) or estimated (draft) genome sizes. The genome sizes ranged between 6.0 and 9.4 Mb, similar to those of representative *Streptomyces* strains (5.0 - 11.9 Mb), the most abundant sources of secondary metabolites [27-29]. The fourth column indicates that the strains are from different clinical origins.

Figure 1 illustrates phylogenetic positions of the five *Nocardia* species (seven strains) studied in the present paper among 78 other established *Nocardia* strains. It also includes *Streptomyces coelicolor* and *Mycobacterium*

tuberculosis for comparison. Four out of the five species are located in different clades of the 16S rRNA phylogenetic tree, indicating that the present analysis is based on information from a wide range of *Nocardia* species. *N. asteroides* and *N. cyriacigeorgica* are in the same clade. We also included three strains from *N. brasiliensis* to elucidate intra-species variations.

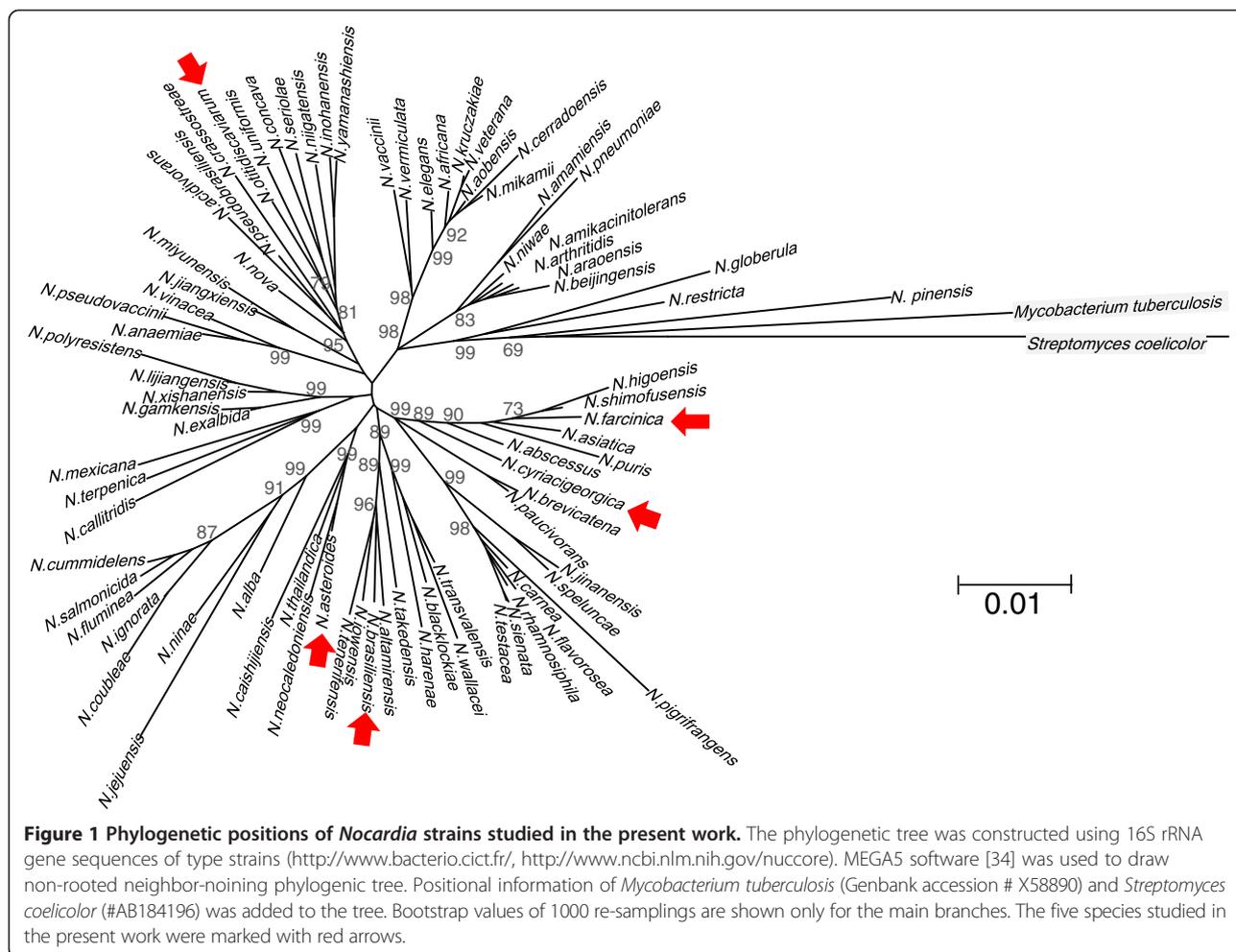
PKS-I, NRPS, and PKS-I/NRPS hybrid gene clusters from the *Nocardia* strains were predicted as described in Methods. Numbers of the three different types of clusters and the total number of clusters in each strain are listed in the four rightmost columns in Table 1. Among the seven strains, the numbers of PKS-I, NRPS, PKS-I/NRPS hybrid clusters, and their total number increased proportionally to the genome size, except for *N. otitidiscaviarum* and *N. asteroides* (Table 1) as reported in other genera [35]. *N. farcinica* had the least while *N. brasiliensis* HUJEG-1 had the highest number of the gene clusters. The total number of clusters within the three *N. brasiliensis* genomes differed (27 to 30), suggesting that different strains of the same species potentially produce their own unique products (see below).

We also counted the numbers of type-II PKS, type-III PKS and terpene synthesis clusters in each genome. The numbers ranged between 0 and 3, except in *N. otitidiscaviarum* and *N. brasiliensis* strains, which have five and eight clusters for terpene synthesis per genome, respectively. In the present paper, however, we focused on PKS-I and NRPS secondary metabolite clusters because their products usually have larger molecular weights with more

Table 1 Genome sizes and numbers of PKS-I, NRPS, and PKS-I/NRPS hybrid gene clusters in *Nocardia* strains

Strain name	Genome size* (Mb)	State of sequence	Source	Number of gene clusters (Average number of genes/modules per cluster)			
				PKS-I	NRPS	PKS-I/NRPS hybrid	Total number of clusters
<i>N. farcinica</i> IFM 10152 [11]	6.0	Complete	Clinical (human sputum)	4 (1.0/1.0) [‡]	7 (1.6/7.7)	1 (6.0/5.0)	12 (1.8/5.6)
<i>N. cyriacigeorgica</i> GUH-2 [8,9]	6.2	Complete	Clinical (human kidney)	5 (1.2/1.0)	9 (1.4/5.8)	1 (8.0/8.0)	15 (1.8/4.3)
<i>N. asteroides</i> NBRC 15531 ^T [1,30]	7.0	Draft	Clinical (fatal brain abscess)	7 (1.4/2.6)	13 (1.2/4.5)	2 (4.0/4.0)	22 (1.5/3.9)
<i>N. otitidiscaviarum</i> IFM 11049 [31]	7.8	Draft	Clinical (human sputum)	4 (1.0/1.0)	12 (1.2/3.3)	2 (4.0/11.0)	18 (1.5/3.6)
<i>N. brasiliensis</i> NBRC 14402 ^T [32,33]	8.9	Draft	Clinical (leg lesion)	10 (1.4/1.1)	13 (1.6/5.0)	4 (5.5/4.3)	27 (2.1/3.4)
<i>N. brasiliensis</i> IFM 10847 [12]	9.2	Draft	Clinical (human pus)	11 (1.5/1.1)	13 (1.5/5.1)	6 (4.8/4.3)	30 (2.2/3.5)
<i>N. brasiliensis</i> HUJEG-1 [10]	9.4	Complete	Clinical (human mycetoma)	11 (1.4/1.2)	13 (1.3/4.5)	6 (4.8/4.3)	30 (2.0/3.2)

*In draft genomes, genome sizes are estimated values. [‡]Average numbers of genes/modules per cluster are indicated in parenthesis.



complex chemical structures than the others and have unique pharmacological activities.

Figure 2 shows all the clusters found in each genome. Presumptive orthologous clusters, as defined in Methods, are aligned in the same row of the table. The rightmost column shows secondary metabolites referred from the database, e.g., [23], and also those inferred using the tools described in Methods.

Clusters common among the seven strains

Figure 2 suggests that seven presumable products (lines #1, #2, #4, #5, #25, #27, and #35) are common among the seven strains belonging to the five species. It is noteworthy that clusters #1, #2, #4, and #5 reside close to the original points of replication (ori.) in the three species whose completed genome sequences are known, in accordance with a report showing that conserved genes reside in the internal core region of actinomycete genomes [27].

Mycolic acid (pks13)

We predicted that the products of the PKS-I genes in line #1 were mycolic acids, cell wall components in members

belonging to *Corynebacterineae*, because these PKS-IS showed the same domain organization as those of *pks13* in *Mycobacterium tuberculosis* for the synthesis of mycolic acids [36], and also showed over 80% sequence similarities to PKS-IS of *N. farcinica* annotated for mycolic acid synthesis [24].

Poly-lysine (Pls)

NRPS genes of line #2 were predicted to be for poly-lysine synthesis because their module organizations are the same as that of poly-lysine synthetase (Pls) in *Streptomyces albulus* [37]. The corresponding gene, *nfa3790*, is also annotated as a Pls homolog in the *N. farcinica* database [24]. The sequence identity between Pls in *S. albulus* and *nfa3790* in *N. farcinica* was 55%.

Ser/The-rich nonribosomal peptides

NRPS gene clusters in line #4 were present in all strains examined, but only a partial sequence was found in *N. otitidiscaviarum* (Additional file 2: Figure S1A). Zoropogui et al., [9] has suggested that #4 in *N. cyriaci-georgica* was for synthesis of 2-amino-9,10 epoxy-8-

#	synthase type	<i>N. farciica</i>	<i>N. cyriacigeorgica</i>	<i>N. asteroides</i>	<i>N. otitidiscaviarum</i>	<i>N. brasiliensis</i>			putative product§
		IFM 10152	GUH-2	NBRC 15531	IFM 11049	NBRC 14402	IFM 10847	HUJEG-1	
		nfa	NOCYR	NCAST	NOTIT	NBRGN	NBRGI	O3I	
1	pks	1890†	145	25_03140	34_00620	027_03110	042_00590	755	mycolic acid
2	nrps	3790	388	25_01210	11_06890	027_00550	073_00150	2055	poly-lysine
3	nrps	-	-	-	-	027_00030	158_00030	2325	Ser-Pro
4	nrps	7170 - 7200	751 - 753	11_00880 - 11_00850	41_01750	056_00550, 112_01590 - 112_01580	221_00020, 160_00010 - 160_00020	4080 - 4090	Ser/Thr-rich 11aa peptide (or 4-7aa peptide & 7-5aa peptide)
5	hybrid	7630 - 7680	792 - 801	11_00460 - 11_00420, 13_1740	03_00430, 41_00480	112_01110 - 112_01030	009_00460 - 009_00380	4310 - 4350	nocobactin-related siderophore
6	nrps	-	-	-	-	110_00930	170_00010	7355	Thr-x
7	nrps	-	1701	-	-	-	-	-	Ser-Pro
8	pks	-	2012 - 2013	-	-	-	-	-	-
9	hybrid	-	-	-	-	-	008_00110 - 008_00060	15055 - 15030	x-x
10	hybrid	-	-	-	-	-	149_00160 - 149_00170	15610	Val-Pro-pk?(-x)
11	pks	-	-	32_00390	58_01020	-	087_00780	16575	polyunsaturated fatty acid
12	hybrid	-	-	-	-	032_00450 - 032_00420	006_00860 - 006_00830	17000 - 17015	Thr-x containing pk
13	pks	30250	2493	10_00730	-	060_01820	165_00260	18640	mycocerosic acid?
14	pks	-	-	-	20_02870	060_01510	215_00300 - 215_00280	18810	-
15	pks	-	-	-	-	060_01120	113_00110	18990	-
16	pks	-	-	-	-	060_00540 - 060_00550	047_00080 - 047_00090	19285 - 19280	-
17	nrps	-	-	-	-	061_00070 - 061_00010	016_00070 - 016_00060	21160 - 21155	(Ser-Ser)-x-x
18	nrps	-	-	-	-	061_00440 - 061_00430	107_00450 - 107_00440	21345 - 20340	Thr-x-x-x-Leu-x-x-Val-Leu
19	pks	-	-	-	-	-	-	22445 - 22425	-
20	hybrid	-	-	-	47_07110 - 47_07270	024_00900 - 024_00930	040_00090 - 040_00140	22780 - 22800	-
21	nrps	-	-	-	-	002_00020 - 002_00170, 080_00510	098_00010	22700	(Thr-x)-x-x-Leu-x-x-x
22	pks	-	-	-	-	024_00320	208_00160	23120	enediylne (see Text)
23	hybrid	-	-	-	-	072_00110 - 072_00190	075_00280 - 075_00370	25165 - 25205	x-Thr(x)
24	nrps	-	2584 - 2585	-	-	-	-	-	Cys-Cys-Cys-Cys
25	nrps	31170	3282	32_02950	21_03020	068_01480	081_00200	26465	Ser-x-x-x
26	pks	-	-	-	-	037_00370 - 037_00390	185_00520 - 185_00510	29065 - 29060	-
27	nrps	11050	4062	05_01440	07_00710	016_00570	090_00180	32155	x-x-x
28	nrps	-	-	05_01330	-	-	-	-	x-Thr
29	nrps	27950	-	-	-	-	-	-	Tyr-x-x-x-Ser
30	pks	43240	3144	-	20_01100	-	-	32485	mycocerosate
31	nrps	-	-	37_01260	-	083_00060	062_00950	33045	Dhb-Gly-Om
32	nrps	-	-	-	-	023_00700 - 023_00710	080_00030 - 080_00020	34385	x
33	nrps	-	-	-	11_01800	058_00040	077_00430	34810	x
34	nrps	-	4098	-	-	-	-	-	x
35	nrps	50330	4800	20_04250	28_00280	057_00530	082_00560	37910	Ser-rich 12aa peptide
36	nrps	50630 - 50620	4844 - 4843	20_03980	64_00560 - 64_00570	-	-	-	Ser-rich 13aa peptide
37	pks	55930	5485	-	-	065_00920 - 065_00930	209_00490 - 209_00460	41665	-
38	nrps	-	-	32_06720	48_00420	-	-	-	Pro-x-x-Ser
39	nrps	-	-	37_01040	47_10160	-	-	-	x-x-x
40	nrps	-	-	-	47_08460	-	-	-	Dhb-x
41	nrps	-	-	-	03_00010	-	-	-	x
42	nrps	-	-	-	24_01940 - 24_01930	-	-	-	Cys-Cys-Cys
43	pks	-	-	24_01760	-	-	-	-	-
44	pks	-	-	21_01670	-	-	-	-	-
45	pks	-	-	08_00240	-	-	-	-	-
46	pks	-	-	33_02270 - 33_02300	-	-	-	-	see Fig. 2
47	nrps	-	-	05_01800	-	-	-	-	x-x
48	nrps	-	-	05_04970	-	-	-	-	Ser-Trp-x-x-Ser-Ser-x
49	nrps	-	-	32_04540	-	-	-	-	x
50	hybrid	-	-	30_00340 - 30_00330	-	-	-	-	-
51	pks	-	-	-	-	066_00620	069_00380	-	-
52	pks	-	-	-	-	054_01060	-	-	-
							015_00750		

† nfa, NOCYR, NOTIT, NCAST, NBRGN, NBRGL, and O3I under the strain names are prefix of genes for each strain, and the numbers beneath them are gene numbers.

Figure 2 PKS-I, NRPS, and PKS-I/NRPS hybrid gene clusters identified in genome sequences of *Nocardia* strains.

oxodecanoic acid, a component of HC-toxin, which is produced by a plant pathogen causing corn leaf spots (reviewed in [38]). However, we suggest two other possibilities based on the domain organization of the cluster. One is that the intact #4 cluster is for synthesis of a serine/threonine-rich peptide composed of 11 amino acids. The second possibility is that the same sequence consists of two different NRPS clusters, since two thioesterase domains are present within the sequence, and accordingly, produces two peptide chains. In the latter case, the products of cluster #4 in *N. asteroides* would be two peptides: one composed of four amino acids (NCAST_11_00880 & NCAST_11_00870) and the other composed of seven

amino acids (NCAST_11_00860 & NCAST_11_00850). Likewise, in *N. brasiliensis*, the products of cluster #4 would also be two peptides: one composed of six amino acids (O3I_004080 & O3I_004085) and the other composed of five amino acids (O3I_004090) (Figure 2, #4, rightmost column).

A possible reason why *N. otitidiscaviarum*, unlike the other species, has only a partial #4 cluster is that this species is phylogenetically distant from the other strains, as shown in Figure 1. The relationships among the products of these gene clusters, the phylogenetic positions of the strains, and their pathogenicity to plants and animals are interesting issues to clarify.

NRPS genes in line #35 were also present in all the strains examined, although those of *N. otitidiscaviarum* and *N. brasiliensis* IFM 10847 were partial compared with those of the other strains (Additional file 2: Figure S1B). The genes in five strains, except for *N. otitidiscaviarum* and *N. brasiliensis* IFM 10847, have 12 modules, and many of their adenylation domains were predicted to select Ser as the substrate. Hence, we assumed the products would be Ser-rich 12 aa peptides. Besides the clusters common among the seven strains in line #35, similar NRPS clusters were also present in the adjacent clusters of line #36, but only in four species excluding *N. brasiliensis* strains. Interestingly, the predicted products are Ser-rich 13 aa peptides.

Although the conservation of Ser- (and Thr-) rich large peptides synthesized by clusters of #4, #35, and #36 in the *Nocardia* strains suggests that they have important roles, physiological roles of the products remain to be investigated.

Nocobactin (*nbt*-like) siderophore

The PKS-I/NRPS hybrid cluster #5 in *N. farcinica* produces nocobactin, a siderophore and a pathogenic factor, as proven by Hoshino *et al.*, [39]. Figure 3 compares clusters in line #5, which are candidates of nocobactin-like siderophore-producing genes. *N. asteroides* has a full set of genes required for siderophore synthesis (NCAST_11_00460 through NCAST_11_00420 and NCAST_13_1740 in #5). NCAST_13_1740 contains an *nbtF*-like gene [39], but is found separated from the rest of *nbt* genes in contig 11 by more than 180 Mb based on

the contig sequences. In *N. brasiliensis*, the corresponding cluster structure is different from that in *N. farcinica* in terms of module number and domain organization, suggesting that a functionally similar but structurally different molecule is synthesized in *N. brasiliensis*. In *N. otitidiscaviarum*, cluster #5 lacks genes corresponding to *nbtA-C* and *nbtE*, which are required for nocobactin synthesis in *N. farcinica* [39], suggesting that cluster #5 in *N. otitidiscaviarum* may not be able to produce a nocobactin-like siderophore. Interestingly, however, *nbtD*- and *nbtF*-like genes, which both have 75% similarities to *nbtD* and *nbtF* genes in *N. farcinica*, respectively, are found in the middle of two different contigs in the *N. otitidiscaviarum* genome (contigs 3 and 41, respectively; listed in cluster #5 in Figure 2; see also Figure 3), suggesting gene loss/gain and recombination during evolution within the genus *Nocardia*. Such gene loss/gain is not only observed in the *nbt*-like gene cluster #5, but also found in *nfa*7170-7200 homologs (#4), *nfa*50330 homologs (cluster #35), and *nfa*50630-50620 homologs (#36), as shown in Additional file 2 (Figure S1). It is further noteworthy that *N. otitidiscaviarum* has another candidate gene cluster for siderophore synthesis, which is shown in line #40 of Figure 2. The domain and module organization of this cluster are shown in Figure 3F, which shows the differences from the one for nocobactin synthesis.

Other common gene clusters

Two NRPS genes in lines #25 and #27 (Figure 2) were common in all the strains sequenced, having four and three modules, respectively. However, the amino-acid

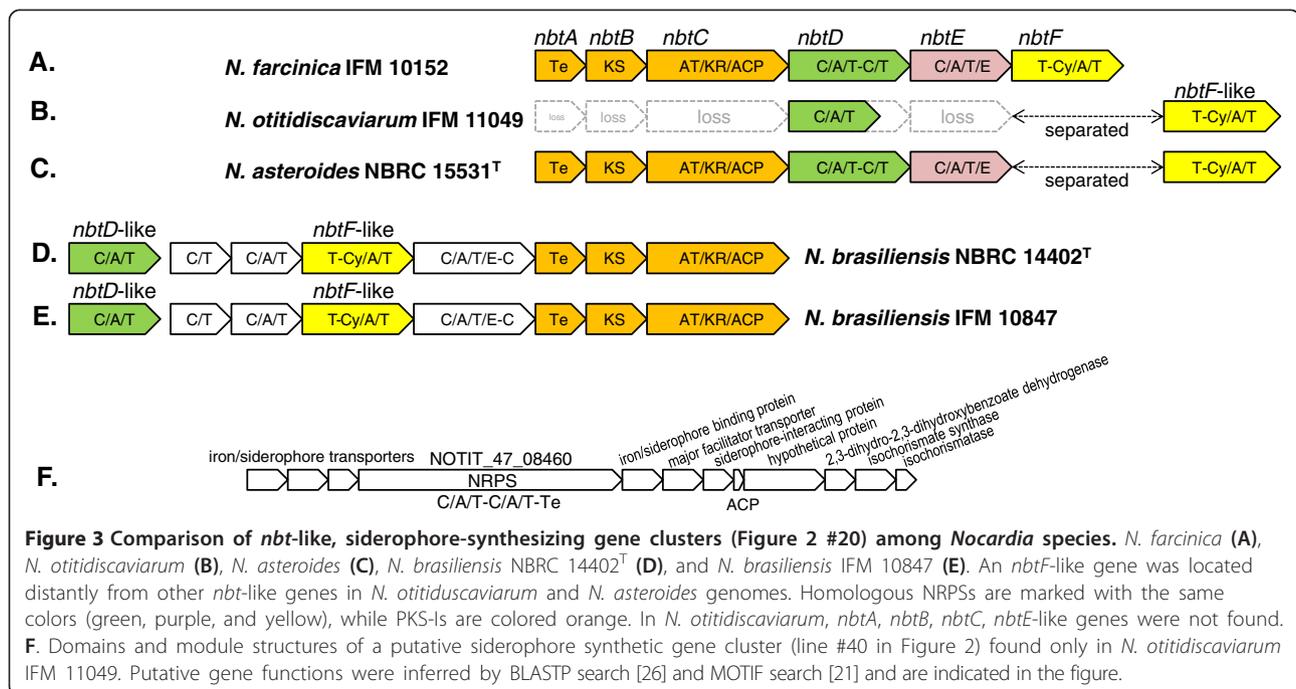


Figure 3 Comparison of *nbt*-like, siderophore-synthesizing gene clusters (Figure 2 #20) among *Nocardia* species. *N. farcinica* (A), *N. otitidiscaviarum* (B), *N. asteroides* (C), *N. brasiliensis* NBRC 14402^T (D), and *N. brasiliensis* IFM 10847 (E). An *nbtF*-like gene was located distantly from other *nbt*-like genes in *N. otitidiscaviarum* and *N. asteroides* genomes. Homologous NRPSs are marked with the same colors (green, purple, and yellow), while PKS-I's are colored orange. In *N. otitidiscaviarum*, *nbtA*, *nbtB*, *nbtC*, *nbtE*-like genes were not found. F. Domains and module structures of a putative siderophore synthetic gene cluster (line #40 in Figure 2) found only in *N. otitidiscaviarum* IFM 11049. Putative gene functions were inferred by BLASTP search [26] and MOTIF search [21] and are indicated in the figure.

composition of the products could not be predicted *in silico* using antiSMASH. Chemical structures and physiological roles of the products remain to be investigated.

Clusters missing in a few strains but found in others

Mycocerosic acid

PKS-I clusters of line #13 (Figure 2) were present in the six strains except *N. otitidiscaviarum*. Other PKS-I clusters in line #30 were present in the four strains but not in *N. asteroides* and in the two *N. brasiliensis* strains (NBRC 14402^T, IFM 10847). Nfa30250 (Figure 2 #13) has been predicted to be involved in mycocerosic acid synthesis in the *N. farcinica* genome project [24]. On the other hand, O3I_032485 in *N. brasiliensis* HUJEG-1 (Figure 2 #30) has been putatively annotated as mycocerosate synthase [10]. Both PKS-Is in *N. farcinica* (#13) and in *N. brasiliensis* (#30) showed approximately 47% amino acid similarities to *Mycobacterium tuberculosis* PKS-I [GenBank/EMBL/DDB] accession number: CCP46654] for the synthesis of mycocerosic acid, a pathogenic factor [40,41]. All the PKS-Is listed in line #13 and #30 showed sequence similarities ranging between 62 and 75%, having almost the same protein length (approximately 2200 aa) and identical domain organization (KS/AT/DH/KR/ER/ACP) (Additional file 1: Table S1).

It is possible that in *N. otitidiscaviarum*, cluster #30 is a substitute for cluster #13. Interestingly, *N. farcinica* and *N. cyriacigeorgica* possess both #13 and #30, which could be related to the strong pathogenicity of the two strains.

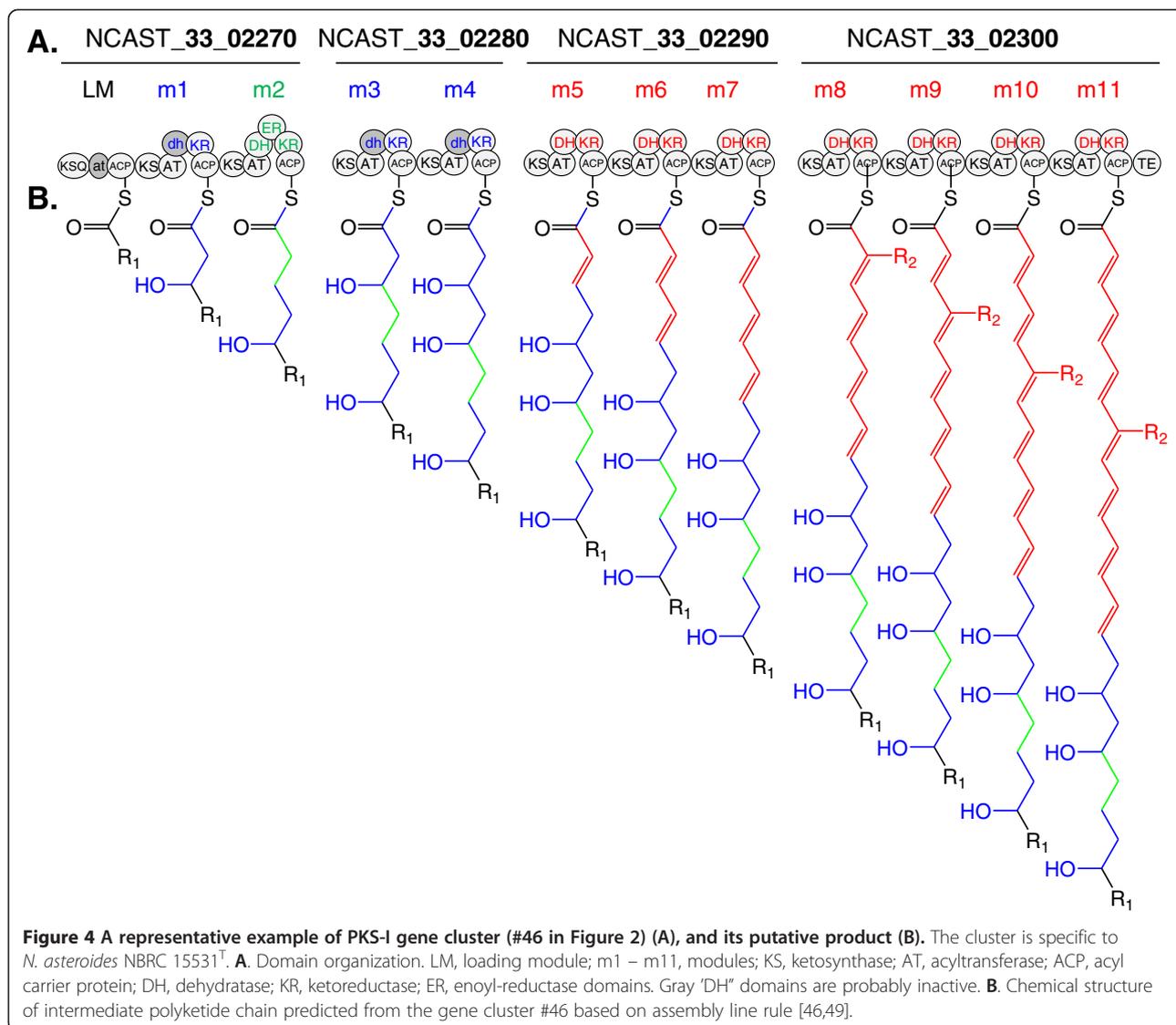
Polyunsaturated fatty acid (pfaA)

Orthologous genes of cluster #11 are found in *N. asteroides*, *N. otitidiscaviarum*, and *N. brasiliensis* IFM 10487 and HUJEG-1, but not in *N. farcinica*, *N. cyriacigeorgica*, and *N. brasiliensis* NBRC 14402 (Figure 2). The PKS-I sequence identities among *N. asteroides*, *N. otitidiscaviarum*, and *N. brasiliensis* ranged from 66 to 78%. The modular structures of PKS-Is in line #11 (KS/AT/ACP/ACP/KR) are unusual because they have two tandem ACP domains, and one KR domain is located after ACP (Additional file 3: Figure S2). These features have been known in polyunsaturated fatty acid (PUFA) synthase, PfaA, of marine bacteria, such as those belonging to the genus *Shewanella* [42,43]. The module organizations are similar between #11 and PfaA (Figure S2), and their amino-acid sequence similarities are over 50%. Hence, we predict that the products of PKS-I in line #11 are polyunsaturated fatty acids, as already reported in *N. brasiliensis* HUJEG-1 genome [10]. However, reports of production of polyunsaturated fatty acids have been limited only in some psychrophilic, piezophilic, or halophilic bacteria in prokaryotes [44,45] and have not been

found in *Nocardia* strains. Thus, future chemical and synthetic analysis are required to explore the potential of *Nocardia* strains as industrial producers of these pharmaceutically and nutraceutically valuable compounds.

Species-specific clusters

Only a few species-specific clusters are found in *N. farcinica* (cluster #29), *N. cyriacigeorgica* (#7, #8, #24, #34), and *N. otitidiscaviarum* (#40, #41, #42), suggesting that the evolution of their polyketides and nonribosomal metabolites are not as dynamic as in other species (Figure 2). On the other hand, *N. asteroides* has nine species-specific clusters (PKS-I, #43 - #46; NRPS, #28, #47 - #49; PKS-I/NRPS hybrid, #50). Among them, we selected #46 as a representative unique cluster structure in *N. asteroides* (Figure 4A). The cluster consists of four adjacent genes, namely, NCAST33_02270, _02280, _02290, and _02300, each of which has the top BLAST-hit to a gene sequence in different *Streptomyces* strains (see Additional file 1: Table S1, for more details). The cluster consisted of twelve modules and was larger than any other cluster in the *Nocardia* strains studied in this paper. The orders of domains within the modules follow the accepted theory for PKS-I gene cluster structures, i.e., having repeats of “KS-AT-optional domains-ACP” [46,47]. The same module organization, however, could not be found in available public databases, and the cluster has an amino-acid sequence similarity of less than 53% to any other known PKS-I clusters. This suggests that the gene cluster may be involved in production of a novel secondary metabolite. We propose a chemical structure of the polyketide chain synthesized by this cluster as shown in Figure 4B based on the PKS-I assembly line rule [46] as follows. The presence of a ketosynthaseQ (KSQ) domain at the N terminus of NCAST_33_02270 and a thioesterase (Te) domain at the C terminus of NCAST_33_02300 indicates that NCAST_33_02270 and NCAST_33_02300 contain the modules that initiate and terminate PKS-I assembly line, respectively. Among the eleven AT domains of module 1 (m1) to 11 (m11), ten AT domains, except that in m8, had the HAFHS signature amino-acid sequence, which is specific for malonyl-CoA in substrate recognition [48,49]. The substrate of the AT domain in m8, which has IASHS amino-acid sequence, and the starter molecules loaded on LM could not be predicted using bioinformatic approach. Hence, this polyketide backbone is predicted to be C_x-C₂-C₂-C₂-C₂-C₂-C₂-C₂-C_y-C₂-C₂-C₂, where C₂ is a unit derived from malonyl-CoA, and C_x and C_y are carbon backbones derived from presently unknown substrates. An inactive DH (dh)-KR pair, which is responsible for formation of a hydroxyl group, is present as an optional domain in m1, m3, and m4 modules, while the optional domain present in m2 is a DH-ER-KR trio, which completely reduces the ketone



residue formed by the m2 module. Optional domains of m5 through m11 are DH-KR pairs, which form double bonds from the ketone residues produced by these modules. The resulting molecule has a polyketide backbone consisting of more than 23 carbons.

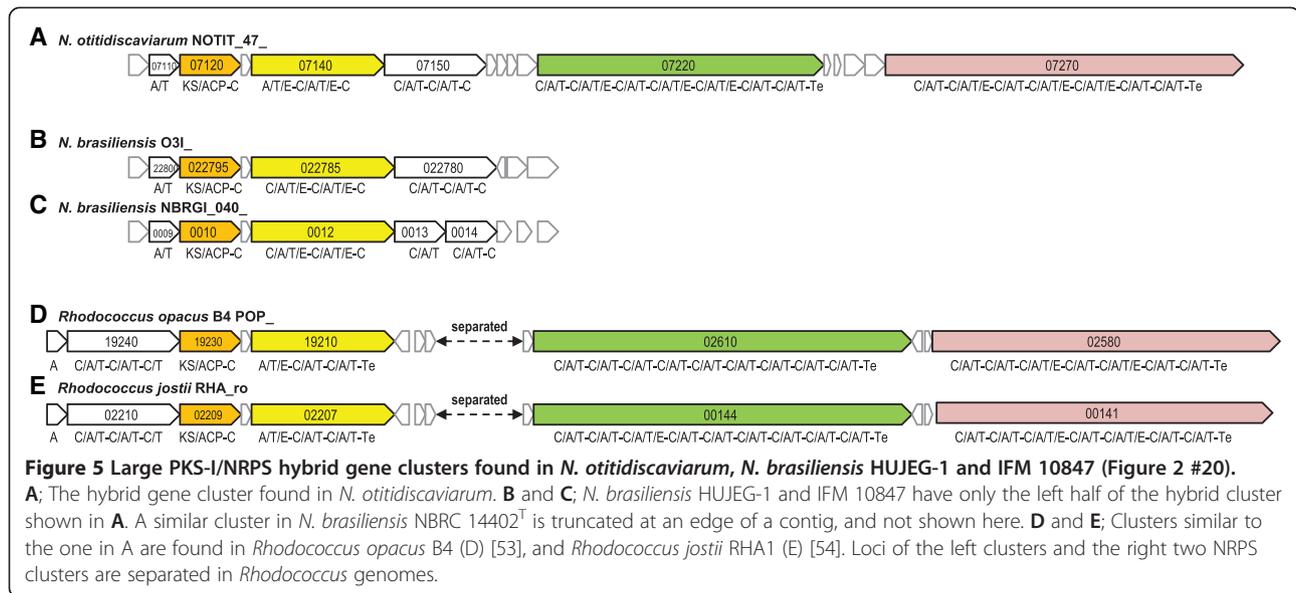
The *N. brasiliensis* strains had the largest number of species-specific clusters among the strains studied in the present work: five to six PKS-I clusters (#15, #16, #19, #22, #26, #51, #52), six NRPS clusters (#3, #6, #17, #18, #21, #32), and two to four PKS-I/NRPS hybrid clusters (#9, #10, #12, #23), depending on the strains. Among them, three PKS-I clusters (#22, #51, #52) consisted of a single module, suggesting that their final or intermediate products may be small in accordance with the assembly line rule [46,47], unless the modules are used iteratively, as has been reported in actinomycetes (e.g., [50,51]). Interestingly, PKS-Is in #22 shows 61% sequence similarity to PksE of *Streptomyces griseus* [GenBank/EMBL/

DBJ accession number: AAO25858], whose product is an unusual polyketide compound including a 9-membered enediyne core [52].

The rest of the clusters with multiple modules may possibly produce large species-specific products. In particular, nonribosomal peptides produced by clusters #18 and #21 are, respectively, predicted to consist of nine and six (to eight) amino acids.

Intra-species variations of clusters

Strain-specific as well as species-specific PKS-I clusters were found in *N. brasiliensis* strains (#19, #30, #52), indicating different strains of the same species may potentially produce different products. Because there are large numbers of *Nocardia* strains stored in several bio-resource centers (e.g., [5-7]), and are rapidly accumulating [12], these *Nocardia* strains constitute a highly promising future resource for exploring secondary metabolites.



Other unique examples of clusters

Figure 5A shows a module structure of PKS-I/NRPS hybrid cluster #20 (Figure 2) in *N. otitidiscaviarum*. The left half (approximately 8,300 aa) of the cluster has high similarity to *N. brasiliensis* hybrid cluster #20 (75 - 82% similarities), but the right half (7,746 aa of NOTIT_47_07220 plus 9,157 aa of NOTIT_47_07270) has top hits to two proteins in *Rhodococcus opacus* (7,746 aa) [GenBank/EMBL/DDBJ accession number: YP_002777453] (Figure 5D) and *Gordonia aichiensis* (9,517 aa) [GenBank/EMBL/DDBJ accession number: WP_005170336], with 59% (4,667/7,781) and 54% (4,543/8,322) amino acid similarities, respectively. The right-half consisting of two NRPSs is widely conserved in *Rhodococcus* spp. including *R. jostii* as shown in Figure 5E.

Because the gene cluster of *N. otitidiscaviarum* #20 contains one PKS-I module and 19 NRPS modules, the product has been tentatively predicted to include one polyketide chain and 19 amino acids (Additional file 4: Figure S4). It should also be mentioned, however, that the two genes in cluster #20, NOTIT_47_07270 and NOTIT_47_07220, each contain thioesterase domains (Te) at their C-terminal ends, suggesting another possibility that the product may contain 12 amino acids instead of 19.

All cluster structures we analyzed in the present work are listed in (Additional file 1: Table S1).

Conclusions

We conclude the following: 1) genomes of *Nocardia* strains carry as many PKS-I and NRPS clusters as *Streptomyces* strains, 2) the number of PKS-I and NRPS gene clusters in *Nocardia* strains varies substantially depending on species, and *N. brasiliensis* strains carry the largest number of

clusters among the species studied, 3) the seven *Nocardia* strains studied in the present work have six common PKS-I/NRPS clusters, some of whose products are yet to be studied, and 4) different *N. brasiliensis* strains have a few different clusters for secondary metabolite synthesis. Also, the following are suggested: 1) there is no clear relation between genome size and pathogenicity in *Nocardia* strains, e.g. *N. farcinica* and *N. brasiliensis* are both prevalent pathogens, but their genome sizes are 6.0 Mb and 9.4 Mb, respectively, the minimum and maximum among the strains studied, and 2) some genes (e.g. cluster #17 in Figure 2) are likely to have been horizontally transferred from (or to) other actinomycetous strains, such as *Rhodococcus* spp.

To summarize, in this study, we compared complete and draft genome sequences of seven strains from five representative *Nocardia* species. The sequences we obtained provided useful information for inferring numbers and molecular structures of secondary metabolites potentially produced by *Nocardia* strains. Genome sequencing revealed the possibility that *Nocardia* strains are as attractive resources as *Streptomyces* strains, the largest resource of natural compounds, in the search for new useful secondary metabolites.

Additional files

Additional file 1: Table S1. ORFs and module/domain structures of PKS-I, NRPS, and PKS-I/NRPS hybrid gene clusters in genomes. Data for *N. otitidiscaviarum* IFM 11049, *N. asteroides* NBRC 15531 T, *N. brasiliensis* NBRC 14402 T and IFM 10847 are shown.

Additional file 2: Figure S1. Representative NRPS gene clusters in *N. farcinica* and their homologs in other strains. A. *N. asteroides* has a cluster with an overall similarity to nfa7170-7200; but the third ORF, NCAST_11_00860, is similar to the first ORF nfa7170, rather than the third

ORF nfa7190 of the corresponding position. *N. brasiliensis* NBRC 14402^T and IFM 10847 lack ORFs corresponding to nfa7180. B. *N. brasiliensis* IFM 10847 has only partial sequences of *N. farcinica* nfa50330-homologous gene, while the homolog in *N. otitidiscaeviarum* is not only partial but also distantly located in the genome. C. *N. asteroides* possesses an nfa50630 homolog, but lacks an nfa50620 homolog. *N. brasiliensis* strains have no homologs.

Additional file 3: Figure S2. Comparison of putative polyunsaturated fatty acid synthase (PfaA) genes between the genus *Nocardia* (Figure 2 #11) and the genus *Shewanella*.

Additional file 4: Figure S3. Predicted chemical structure of the product from PKS-I/NRPS hybrid gene cluster #20 in *N. otitidiscaeviarum*.

Abbreviations

PKS: Polyketide synthase; PKS-I: Type-I polyketide synthase; NRPS: Nonribosomal peptide synthetase; NBRC: Biological Resource Center, National Institute of Technology and Evaluation; MMRC: Medical Mycology Research Center; KS: Ketosynthase; AT: Acyltransferase; ACP: Acyl carrier protein; KR: Ketoreductase; Te: Thioesterase; LM: Loading module; DH: Dehydratase; dh: Inactive dehydratase; ER: Enoyl-reductase.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HK and NI: analyzed gene clusters and predicted their functions; AH, AT-N, and TM: performed sequencing; KS and NF: organized the sequencing project and edited the manuscript; TG: wrote the manuscript. All authors read and approved the final manuscript.

Authors' information

HK: Senior Chief, Biological Resource Center, National Institute of Technology and Evaluation (NBRC).

NI: Chief, NBRC.

AH: Senior Chief, NBRC.

AT-N: Postdoctoral Fellow, Medical Mycology Research Center (MMRC), Chiba University.

TM: Research Associate, MMRC, Chiba University.

KS: Senior Director, NBRC.

NF: Senior Director, NBRC.

TG: Professor, MMRC, Chiba University.

Acknowledgement

This work was supported by research grants of the Ministry of Education, Culture, Sports, Science, & Technology in Japan [grants #21406003], by the National BioResource Project (<http://www.nbrp.jp/>), and by the Cooperative Research Grant of NEKKEN, 2012 to TG. We thank to Mr. Syuji Yamazaki and Dr. Atsushi Yamazoe for launching the massive genome sequencing project of the genus *Nocardia* and for providing us with the sequences.

The GenBank/EMBL/DDBJ accession numbers for the sequences containing the type-I PKS and NRPS gene clusters are: AB685274, AB700124-AB700133, AB700557-AB700568 (*Nocardia asteroides* NBRC 15531^T), AB700569-AB700587 (*Nocardia otitidiscaeviarum* IFM 11049), AB701575-AB701605 (*Nocardia brasiliensis* IFM 10847), and AB701607-701636 (*Nocardia brasiliensis* NBRC 14402^T).

Author details

¹Biological Resource Center, National Institute of Technology and Evaluation (NBRC), Kisarazu, Chiba 292-0818, Japan. ²NBRC, Shibuya-ku, Tokyo 151-0066, Japan. ³Medical Mycology Research Center (MMRC), Chiba University, Chuo-ku, Chiba 260-8673, Japan.

Received: 11 October 2013 Accepted: 15 April 2014

Published: 30 April 2014

References

1. Beaman BL, Beaman L: *Nocardia* species: host-parasite relationships. *Clin Microbiol Rev* 1994, **7**(2):213–264.
2. Kageyama A, Yazawa K, Ishikawa J, Hotta K, Nishimura K, Mikami Y: Nocardial infections in Japan from 1992 to 2001, including the first

- report of infection by *Nocardia transvalensis*. *Eur J Epidemiol* 2004, **19**(4):383–389.
3. Brown-Elliott BA, Brown JM, Conville PS, Wallace RJ Jr: **Clinical and Laboratory Features of the *Nocardia* spp. Based on Current Molecular Taxonomy.** *Clin Microbiol Rev* 2006, **19**(2):259–282.
4. Medical Mycology Research Center: Japan: Chiba University. <http://www.pf.chiba-u.ac.jp/eng/index.html>.
5. NBRC culture collection: www.nbrc.nite.go.jp/e/index.html.
6. American Type Culture Collection. <http://www.atcc.org/>.
7. Leibniz Institute DSMZ-German Collection of Microorganisms and cell culture: <https://www.dsmz.de/>.
8. Zoropogui A, Pujic P, Normand P, Barbe V, Beaman B, Beaman L, Boiron P, Colinon C, Deredjian A, Graindorge A, Mangelot S, Nazaret S, Neto M, Petit S, Roche D, Vallenet D, Rodriguez-Nava V, Richard Y, Courmoyer B, Blaha D: **Genome sequence of the human- and animal-pathogenic strain *Nocardia cyriacigeorgica* GUH-2.** *J Bacteriol* 2012, **194**(8):2098–2099.
9. Zoropogui A, Pujic P, Normand P, Barbe V, Belli P, Graindorge A, Roche D, Vallenet D, Mangelot S, Boiron P, Rodriguez-Nava V, Ribun S, Richard Y, Courmoyer B, Blaha D: **The *Nocardia cyriacigeorgica* GUH-2 genome shows ongoing adaptation of an environmental Actinobacteria to a pathogen's lifestyle.** *BMC Genomics* 2013, **14**(1):286.
10. Vera-Cabrera L, Ortiz-Lopez R, Elizondo-Gonzalez R, Perez-Maya AA, Ocampo-Candiani J: **Complete genome sequence of *Nocardia brasiliensis* HUJEG-1.** *J Bacteriol* 2012, **194**(10):2761–2762.
11. Ishikawa J, Yamashita A, Mikami Y, Hoshino Y, Kurita H, Hotta K, Shiba T, Hattori M: **The complete genomic sequence of *Nocardia farcinica* IFM 10152.** *Proc Natl Acad Sci U S A* 2004, **101**(41):14925–14930.
12. IFM culture collections of MMRC: Japan: Chiba University. <http://www.pf.chiba-u.ac.jp/eng/bioresources/index.html>.
13. Ichikawa N, Oguchi A, Ikeda H, Ishikawa J, Kitani S, Watanabe Y, Nakamura S, Katano Y, Kishi E, Sasagawa M, Ankaï A, Fukui S, Hashimoto Y, Kamata S, Otoguro M, Tanikawa S, Nihira T, Horinouchi S, Ohnishi Y, Hayakawa M, Kuzuyama T, Arisawa A, Nomoto F, Miura H, Takahashi Y, Fujita N: **Genome sequence of *Kitasatospora setae* NBRC 14216 T: an evolutionary snapshot of the family Streptomycetaceae.** *DNA Res* 2010, **17**(6):393–406.
14. DNA Data Bank of Japan: <http://www.ddbj.nig.ac.jp/index-e.html>.
15. Microbial Genome Annotation Pipeline: <http://www.migap.org/index.php/en/aboutpipeline>.
16. Sugawara H, Ohyama A, Mori H, Kurokawa K: **Microbial Genome Annotation Pipeline (MiGAP) for diverse users [abstract].** *The 20th International Conference on Genome Informatics (GIW2009)* 2009, **20**(S001):1–2.
17. Komaki H, Ichikawa N, Oguchi A, Hanamaki T, Fujita N: **Genome-wide survey of polyketide synthase and nonribosomal peptide synthetase gene clusters in Streptomyces turgidiscabies NBRC 16081.** *J Gen Appl Microbiol* 2012, **58**(5):363–372.
18. InterPro domain database: <http://www.ebi.ac.uk/interpro/>.
19. Mulder NJ, Kersey P, Pruess M, Apweiler R: **In silico characterization of proteins: UniProt, InterPro and Integr8.** *Mol Biotechnol* 2008, **38**(2):165–177.
20. PKS/NRPS analysis Web-site: <http://nrps.igs.umaryland.edu/nrps>.
21. KEGG MOTIF Search: <http://www.genome.jp/tools/motif/>.
22. Antibiotics and Secondary Metabolite Analysis Shell (antiSMASH): <http://www.secondarymetabolites.org/>.
23. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R: **antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.** *Nucleic Acids Res* 2011, **39**(Web Server issue):W339–W346.
24. *Nocardia farcinica* genome project page. <http://nocardia.nih.gov/>.
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403–410.
26. NCBI/BLAST Home: <http://blast.ncbi.nlm.nih.gov/>.
27. Nett M, Ikeda H, Moore BS: **Genomic basis for natural product biosynthetic diversity in the actinomycetes.** *Nat Prod Rep* 2009, **26**(11):1362–1384.
28. Berdy J: **Bioactive microbial metabolites.** *J Antibiot (Tokyo)* 2005, **58**(1):1–26.
29. Berdy J: **Thoughts and facts about antibiotics: where we are now and where we are heading.** *J Antibiot (Tokyo)* 2012, **65**(8):385–395.

30. Eppinger H: Über eine neue pathogene *Cladotrix* und eine durch sie hervorgerufene Pseudotuberculosis (*Cladotrichica*). *Beitrage zur pathologischen Anatomie* 1891, **9**:287–328.
31. Snijders EP: Cavia-scheefkopperij, een nocardiose. *Geneeskundig Tijdschrift voor Nederlandsch-Indie* 1924, **64**:85–87.
32. Lindenberg A: Un nouveau mycétome. *Archives de Parasitologie* 1909, **13**:265–282.
33. Gordon RE, Mihm JM: A Comparison of *Nocardia asteroides* and *Nocardia brasiliensis*. *J Gen Microbiol* 1959, **20**:129–135.
34. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011, **28**:2731–2739.
35. Jenke-Kodama H, Sandmann A, Müller R, Dittmann E: Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* 2005, **22**(10):2027–2039.
36. Portevin D, De Sousa-D'Auria C, Houssin C, Grimaldi C, Chami M, Daffé M, Guilhot C: A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc Natl Acad Sci U S A* 2004, **101**(1):314–31937.
37. Yamanaka K, Murayama C, Takagi H, Hamano Y: Epsilon-poly-L-lysine dispersity is controlled by a highly unusual nonribosomal peptide synthetase. *Nat Chem Biol* 2008, **4**(12):766–772.
38. Walton JD: HC-toxin. *Phytochem* 2006, **67**(14):1406–1413.
39. Hoshino Y, Chiba K, Ishino K, Fukai T, Igarashi Y, Yazawa K, Mikami Y, Ishikawa J: Identification of nocardin NA biosynthetic gene clusters in *Nocardia farcinica*. *J Bacteriol* 2011, **193**(2):441–448.
40. Mathur M, Kolattukudy PE: Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from *Mycobacterium tuberculosis* var. *bovis* *Bacillus Calmette-Guerin*. *J Biol Chem* 1992, **267**(27):19388–19395.
41. Brennan PJ, Nikaido H: The envelope of mycobacteria. *Annu Rev Biochem* 1995, **64**:29–63.
42. Kaulmann U, Hertweck C: Biosynthesis of polyunsaturated fatty acids by polyketide synthases. *Angew Chem Int Ed Engl* 2002, **41**(11):1866–1869.
43. Jiang H, Zirkle R, Metz JG, Braun L, Richter L, Van Lanen SG, Shen B: The role of tandem acyl carrier protein domains in polyunsaturated fatty acid biosynthesis. *J Am Chem Soc* 2008, **130**(20):6336–6337.
44. Wallis JG, Watts JL, Browne J: Polyunsaturated fatty acid synthesis: what will they think of next? *Trends Biochem Sci* 2002, **27**(9):467.
45. Okuyama H, Orikasa Y, Nishida T: Significance of antioxidative functions of eicosapentaenoic and docosahexaenoic acids in marine microorganisms. *Appl Environ Microbiol* 2008, **74**(3):570–574.
46. Fischbach MA, Walsh CT: Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* 2006, **106**(8):3468–3496.
47. Weissman KJ: Introduction to polyketide biosynthesis. *Methods Enzymol* 2009, **459**:3–16.
48. Del Vecchio F, Petkovic H, Kendrew SG, Low L, Wilkinson B, Lill R, Cortes J, Rudd BA, Staunton J, Leadlay PF: Active-site residue, domain and module swaps in modular polyketide synthases. *J Ind Microbiol Biotechnol* 2003, **30**:489–494.
49. Kakavas SJ, Katz L, Stassi D: Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*. *J Bacteriol* 1997, **179**:7515–7522.
50. Gaisser S, Trefzer A, Stockert S, Kirschning A, Bechthold A: Cloning of an avilamycin biosynthetic gene cluster from *Streptomyces viridochromogenes* Tu57. *J Bacteriol* 1997, **178**(20):6271–7278.
51. Van Lanen SG, Oh TJ, Liu W, Wendt-Pienkowski E, Shen B: Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne. *Biosynthesis* 2007, **129**(43):13082–13094.
52. Zazopoulos E, Huang K, Staffa A, Liu W, Bachmann BO, Nonaka K, Ahlert J, Thorson JS, Shen B, Farnet CM: A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nat Biotechnol* 2003, **21**(2):187–190.
53. A genome database of microorganisms sequenced at NITE: www.bio.nite.go.jp/dogan/project/view/OPACUS.
54. McLeod MP, Warren RL, Hsiao WW, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, Dosanjh M, Hara H, Petrescu A, Morin RD, Yang G, Stott JM, Schein JE, Shin H, Smailus D, Siddiqui AS, Marra MA, Jones SJ, Holt R, Brinkman FS, Miyauchi K, Fukuda M, Davies JE, Mohn WW, Eltis LD: The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A* 2006, **103**(42):15582–15587.

doi:10.1186/1471-2164-15-323

Cite this article as: Komaki et al.: Genome based analysis of type-I polyketide synthase and nonribosomal peptide synthetase gene clusters in seven strains of five representative *Nocardia* species. *BMC Genomics* 2014 **15**:323.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

