ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 880-890 (2021)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

OPEN ACCESS GOLD

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 17(10)

Research Article

Insights from the comparative genome analysis of natural rubber degrading *Nocardia* species

Biraj Sarkar[1], Aayatti Mallick Gupta[2] & Sukhendu Mandal[1*]

[1]Laboratory of Molecular Bacteriology, Department of Microbiology, University of Calcutta, 35, Ballygunge Circular Road, Kolkata, 700019, India; [2]Department of Chemical, Biological & Macro-Molecular Sciences, S. N. Bose National Centre for Basic Sciences, Block-JD, Sector-III, Salt Lake, Kolkata-700 106; Sukhendu Mandal; ORCID ID:0000-0002-7752-0982 - E-mail: sukhendu1@hotmail.com; *Corresponding author:

Received October 5, 2021; Revised October 16, 2021; Accepted October 16, 2021, Published October 31, 2021

DOI: 10.6026/97320630017880

Declaration on official E-mail:
The corresponding author declares that official e-mail from their institution is not available for all authors

Declaration on Publication Ethics:
The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Abstract:
Nocardia are known to be a facultative human pathogen and can cause infection in immune compromised patients. Though the details research on the virulence factors of *Nocardia* are scanty but numerous genes that code such factors were reported from different species of *Nocardia*. Despite of the presence of several virulence factors, species of this genus have been shown to have role in remediation of many toxic and hazardous materials from the environment. In this study, genome sequences of rubber degrading *Nocardia* sp. BSTN01 and *N. nova* SH22a have been analyzed to locate the potential virulence genes. Also, the genomes of facultative pathogenic *Nocardia* like, *N. africana*, *N. brasiliensis*, *N. kruczakiae*, *N. transvalensis* and *N. veterana* have been analyzed to find the gene encoding latex clearing protein (Lcp), a rubber oxygenase enzyme of Gram-positive action bacteria. The study provides an insight about the potentiality of rubber-degrading *Nocardia* species to emerge as future human pathogens and also the probability of a serious concern if the studied facultative pathogens of *Nocardia* like *N. africana*, *N. brasiliensis*, *N. kruczakiae*, *N. transvalensis* and *N. veterana*are capable of degrading rubber, a regularly used material in clinics. Moreover, use of such possible pathogenic strains for their known role in bioremediation of rubber waste from the environment might be deleterious.

Keywords: poly-cis-isoprene; rubber degradation; *Nocardia*; whole genome analysis; phylogenomics

Background:
The genus *Nocardia* consists of aerobic, filamentous, Gram-positive, non-motile actinomycetes which have been reported to be isolated from a diverse range of environments around the globe [1-3]. Various *Nocardia* spp. are now known to be facultative intracellular human pathogens which were reported to be capable of causing localized or disseminated infection in both immune competent and immune compromised humans and in animals [3-5]. Although infection causing *Nocardia* are generally known to be opportunistic and infection occurs in immune-compromised patients, around 15% of such infections do not even exhibit any definable predisposing condition [6, 7]. According to a report published in 2017, out of 92 recognized species of *Nocardia*, listed in LPSN (the List of Prokaryotic names with Standing in the literature), 54 species have

been found to be clinically relevant [8]. Moreover, the number of reported cases of *Nocardia* infections have been increasing since last few decades [9, 10], which is indeed a matter of great concern in clinical relevance. Numerous cases have been reported where infections caused by *Nocardia* were found to affect lungs, central nervous system, skin and othervital organs. Nonetheless, *Nocardia* can also possess serious life-threatening disseminated infections like osteomyelitis and nocardial sepsis [10-12]. Studies involving genomics helps in understanding the diversity, taxonomy and evolutionary relatedness of species belonging to the genus *Nocardia* [13]. Multiple genome sequences of different species of *Nocardia* have been reported till date. These sequences provide insights about different virulence factors which are known to play crucial role in cell invasion, modulation of phagocyte functions and survival inside the host macrophages resulting successful pathogenesis [14-19]. Not just the genomes of pathogenic species of *Nocardia*, but species of *Nocardia* that are known to be endophytes, has also been studied to explore their biological roles [20, 21]. Reports involving studies on the whole genome sequence (WGS) also provided insights into the molecular basis of resistance towards different antibiotics exhibited by different species of *Nocardia* [22-25]. Studies on the WGSs of *N. cyriacigeorgica* strains helped in understanding the physiology, evolution and adaptation of the environmental bacteria for being converted into a pathogenic one [26, 27].Genome based analysis of different species of *Nocardia* revealed some of them to harbour genes responsible for the production of important secondary metabolites like polyketide synthase (PKS-I) and non-ribosomal peptide synthetases (NRPS) [28, 29]. Species belonging to the genus *Nocardia* were also reported to produce valuable metabolites like antibacterial compounds, UV-protectant molecules, immune suppressant, nocavionin and nocobactin derivatives with anti muscarinic activities [30-33].

On the other hand, some species belonging to the genus *Nocardia* have been reported to be capable of degrading varied kinds of environmental toxic and hazardous hydrocarbon-based materials, including rubber [34-37]. Such behaviour exhibited by these Gram-positive actinomycetes is attributed to their ability to use the polyisoprene backbone of rubber as a source of carbon for their growth and proliferation in scarcity of other easily accessible nutrients. Several studies reported that these group of bacteria are capable of extracellular cleaving the high-molecular-weight polyisoprene backbone of rubber by virtue of their extracellularly secreted latex clearing protein, Lcp (EC 1.13.11.87), which in turn allows the bacteria to easily uptake the resultant low-molecular-weight oligoisoprenoids into the cell for further metabolic processes required in the quest of energy and survival[36, 38-40]. All rubber degrading actino bacteria are known to harbour one or more *lcp* gene in their genome or plasmid which encodes the latex clearing protein [38]. The functionality of *lcp* gene and its enzyme product were analysed using by means of gene deletion, cloning and protein expression studies [40-43].

Numerous polymer materials which are chiefly made of rubber are extensively used in the field of clinical science, prosthetics and hygiene products. As examples, urinary catheters and gloves used clinically are generally made up of rubber [44-46], gutta-percha which is basically a trans- 1,4-polyisoprene form of rubber has been widely used for years in endodontics [47] and products like menstrual cups are also primarily composed of rubber which comes under the category of feminine hygiene products [48]. Interestingly, *Nocardia* species can colonize and produce biofilm on clinical or hospital resources that are primarily composed of rubber or poly-isoprene. Such materials can put the treated patients under serious conditions causing *Nocardia*-bacteremia. Such an instance is stated in case of*N. nova* complex, as it was capable of forming biofilm in polyisoprene made catheters, like central venous catheters which are normally needed by cancer patients [49]. According to a study performed at the University of Texas M.D. Anderson Cancer Center, around 7% of all invasive infections caused by *Nocardia* had central venous catheters in common [50]. Therefore, the comparative genome analysis of an unreported rubber degrading *Nocardia* (i) *Nocardia* sp BSTN01 and (ii) a reported efficient rubber degrading *N. nova* SH22a have been analysed in to find potential virulence factor in the genome. We have also analysed the complete genome sequences of other related bacteria for *lcp* gene in their genome. This data will help us to know the aspect of emerging infections caused by rubber degrading species of *Nocardia,* particularly for environmental remediation without knowing their pathogenicity attributes.

Table 1: Genome assembly accession numbers and average nucleotide identity (ANI) calculated between BSTN01 and other *Nocardia* genomes of different strains considered in this study.

Strains	Accession	ANI values
Nocardia sp. BSTN01	NZ_JADKYP000000000	
N. kruczakiae NBRC 101016	NZ_BDBL00000000	94.56
N. africana NCTC13184	NZ_UGRU01000000	91.93
N. veterana NBRC 100344	NZ_BAGM000000000	84.60
N. nova SH22a	NZ_CP006850	83.21
N. transvalensis NBRC 15921	NZ_BAGL00000000	78.80
N. brasiliensis ATCC 700358	NZ_018681	76.53

Materials and Methods:

Retrieve of genomic data of *Nocardia* spp.

In this study the WGS sequence of BSTN01 was compared with the complete and WGS sequences of 6 other different species of *Nocardia*which have been obtained from NCBI database. A list of organisms with their accession number has been tabulated in Table 1.

Phylogenetic analysis

Phylogenetic analysis of the WGS and complete genomesequences used in this study is inferred with REALPHY 1.12 [51]. It employs Maximum likelihood algorithm with the general time reversible (GTR) of the nucleotide substitution model with gamma distribution (G) to construct the tree. The resultant tree is visualized and annotated in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).The average nucleotide identity (ANI) was calculated in EzGenome (https://www.ezbiocloud.net/taxonomy) [52].

Annotation of the genomes

In order to locate the *lcp* gene and different virulent factors (VFs) from the WGS of the selected *Nocardia* spp., the genome sequences of respective strains were annotated in the RAST (Rapid Annotations Using Subsystems Technology) server[53] and

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 880-890 (2021)

PATRIC bioinformatics resource centre (https://patricbrc.org/app/Annotation), an online platform [54].

Comparison of Lcp homologues

LCP homologues are identified and multiple sequence alignment (MSA) using clustal omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) of all the sequences were carried out.

Pathogenomic analysis

To identify the virulence factor within various species of *Nocardia*, standard virulence factors of *Mycobacterium tuberculosis* has been used as reference. Genes homologous to those of *M. tuberculosis* with functions related to stress adaptation, phagosome arresting, nitrate reductases, secreted protein, effector delivery system and mammalian cell entry (*mce*) have been harnessed using VFDB database [55]. *M. tuberculosis* is a well-studied known species of pathogenic actinobacteria. As both *Nocardia* and *M. tuberculosis* are from the same phylum, in this study the genes responsible for encoding the virulence factors in *M. tuberculosis* were searched in the genomes of selected species of *Nocardia*. All the above-mentioned virulence factors from the genome sequences of *Nocardia* have been retrieved further fromRAST (Rapid Annotations Using Subsystems Technology) server [53] and PATRIC bioinformatics resource centre (https://patricbrc.org/app/Annotation) [54].BLAST of the amino acid sequences of the VF genes from the different *Nocardia* species used in this study were also performed against the *M. tuberculosis* (NCBI taxonomy ID: 1773) using Basic Local Alignment Tool (BLASTp).

Results and Discussion:

Nocardia sp. strain BSTN01 is an isolate capable of degrading both synthetic and natural rubber, and was isolated from the water stored in latex collecting cup from the wastes of a rubber processing unit. The whole genome shotgun sequence of the strain is submitted in NCBI under the accession number NZ_JADKYP010000000. Another well-established rubber degrader *N. nova* SH22a is known to be capable of degrading both cis-isoprene rubber (NR) and trans-isoprene rubber (gutta-percha) [36]. Studies has already confirmed the presence of *lcp* gene encoding the latex clearing protein (Lcp) which is the primary and most vital enzyme involved in the oxidative cleavage of the polyisoprene backbone in the process of biodegradation in both BSTN01 and SH22a [36]. These strains were isolated from different geographical locations and yet had not been observed or reported for exhibiting virulence in humans. There are numerous pathogenic and facultative pathogenic *Nocardia*known to cause health issues in humans. For instance, *N. veterana* NBRC 100344 is an actinomycete which was isolated from the bronchoscopic lavage of a 78-year-old patient with a past history of tuberculosis pleurisy exhibited bilateral upper lobe lesions [56]. *N. africana* NCTC13184 is such a pathogen which was isolated from patients with pulmonary infections [57]. *N. kruczakiae* was found to be responsible for causing traumatic endo phthalmitis in a patient with eye injury. Strains of *N. kruczakiae*are are known to be opportunistic pathogens infecting immune compromised patients [58, 59]. Similarly, there are reports of mycetoma caused by *N. transvalensis* and *N. Brasiliensis* [60, 61].

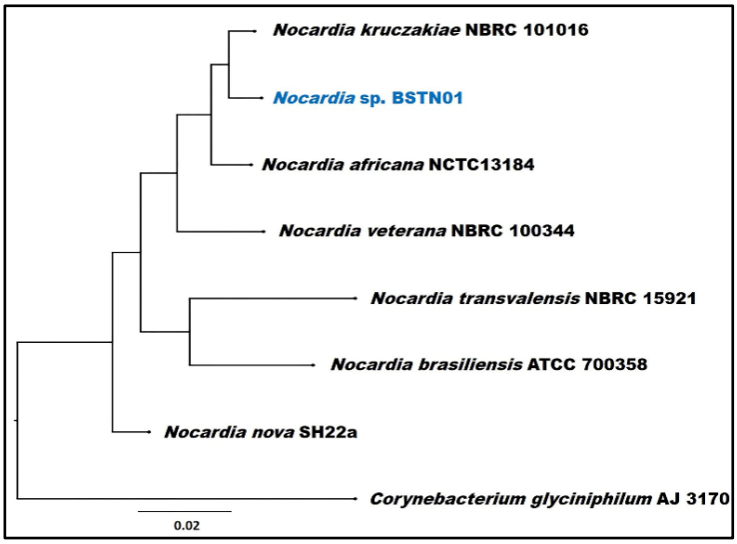

Figure 1: Maximum likelihood phylogenetic analysis of the WGS and complete genome sequences used in this study.

Table 2: Showing the identity (%) of amino acid sequences of virulence factors from the two strains of *Nocardia* with respect to that of *M. tuberculosis*

	Nocardia sp. BSTN01			*N. nova* SH22a		
Virulence factors	**Locus tag (IRT45_RS)**	**Identity (%)**	**E-value**	**Locus tag (NONO_RS)**	**Identity (%)**	**E-value**
sodA (Superoxide dismutase [Mn/Fe])	12770	85.59	1e-65	00420	87.44	7e-134
katG(catalase-peroxidase KatG)	09560	71.95	0.0	17245	75.98	0.0
	28740	78.00	0.0	18055	65.44	0.0
ahpC(alkyl hydroperoxide reductase protein C)	15570	86.60	5e-126	26420	84.54	9e-125
ptpA (low molecular weight protein-tyrosine-phosphatase)	15300	59.49	3e-59	10640	59.49	2e-58
ndk (nucleoside diphosphate kinase)	04255	81.20	1e-74	29270	82.22	6e-78
eis(enhanced intracellular survival protein)	00090	45.75	2e-95	03520	47.25	6e-100
narG (Respiratory nitrate reductase alpha chain)	02535	77.57	0.0	07200	72.06	0.0
	10940	71.58	0.0	19820	77.25	0.0
narH (Respiratory nitrate reductase beta chain)	02530	73.92	0.0	07195	62.67	0.0
	10945	64.98	0.0	19815	73.74	0.0
narI (Respiratory nitrate reductase gamma chain)	02520	67.37	5e-114	07185	60.63	4e-97
	10955	57.87	1e-97	19805	65.27	2e-111
narJ (Respiratory nitrate reductase delta chain)	02525	58.45	3e-66	07190	50.68	6e-63
	10950	52.49	2e-63	19810	57.79	4e-65
eccA(AAA+ family protein ATPase EccA1, component of Type VII secretion system ESX-1)	01005	62.77	0.0	04675	62.27	0.0
				07995	44.33	5e-148
eccB(Membrane protein EccB-like, component of Type VII secretion system in Actinobacteria)	01000	39.43	5e-103	04670	39.02	5e-100
	01135	38.49	1e-98	04790	38.49	4e-96
eccD (Integral membrane protein EccD-like, component of Type VII secretion system in *Actinobacteria*	01010	30.82	2e-43	04680	30.16	4e-42
	01155	31.47	2e-40	04820	31.06	2e-39
mycP (Serine protease, putative component of Type VII secretion system in Actinobacteria)	01015	53.46	2e-107	04685	52.88	1e-106
	01150	50.96	5e-96	04815	50.62	2e-105
eccCa(FtsK/SpoIIIE family protein, putative EssC/YukB component of Type VII secretion system)	01020	61.52	0.0	08005	40.40	0.0

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 880-890 (2021)

01160	53.02	0.0
2785	41.19	0.0

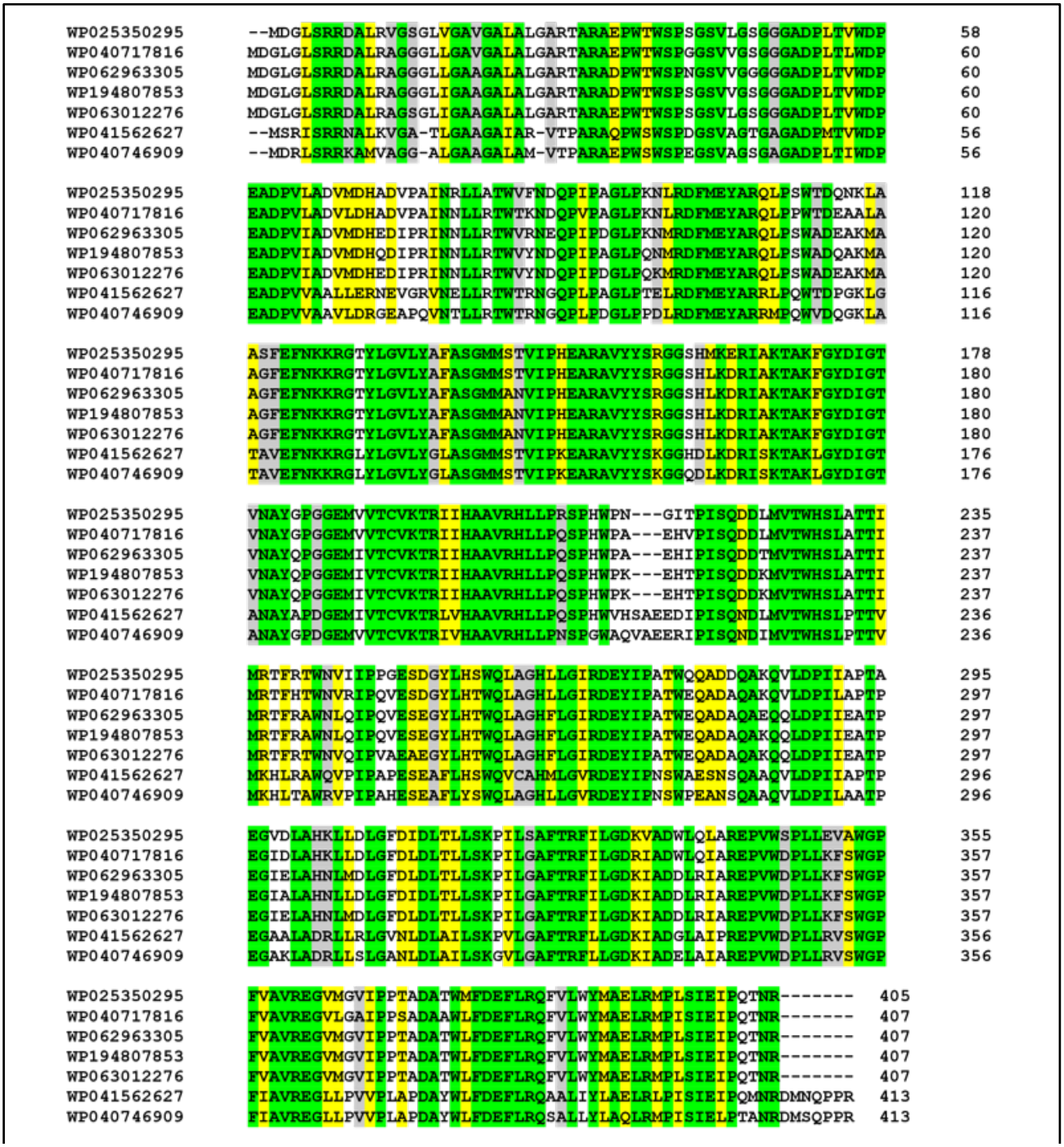

Figure 2: Multiple sequence alignment (MSA) analysis of the DUF2236 amino acid sequences from the considered *Nocardia* species, i.e., *Nocardia* sp. BSTN01 (WP194807853), *N. nova* SH22a (WP025350295), *N. veterana* NBRC 100344 (WP040717816), *N. africana* NCTC13184 (WP062963305), *N. kruczakiae* NBRC 101016 (WP063012276), *N. brasiliensis* ATCC 700358 (WP041562627) and *N. transvalensis* NBRC 15921 (WP040746909). The green highlighted regions represent identical amino acid residues, the yellow highlighted regions represent conserved substitution and the grey highlighted regions represent semi-conserved substitutions in the multiple sequence alignment.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 880-890 (2021)

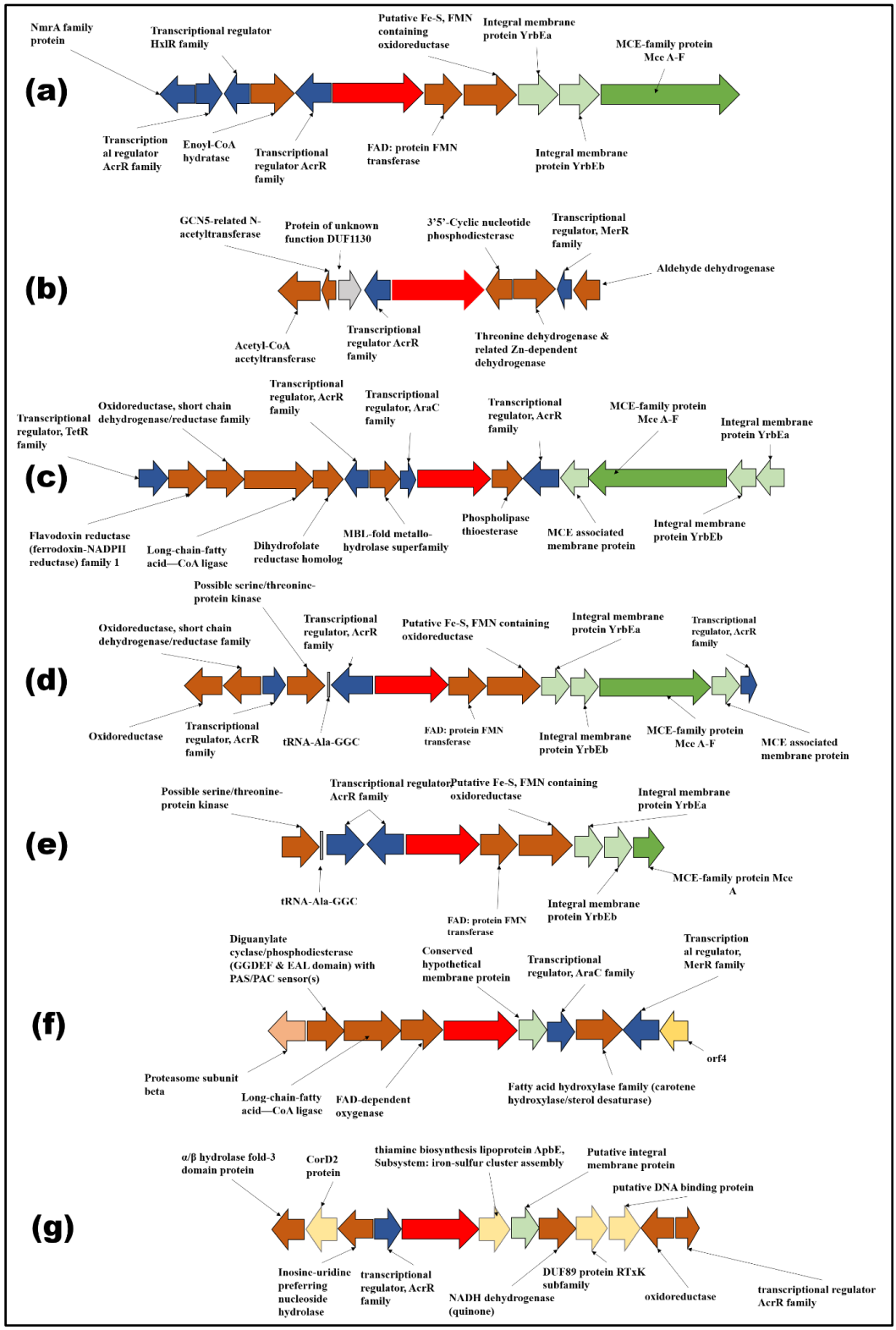

Figure 3: Showing the physical maps of *Nocardia* sp. BSTN01 (a) *N. veterana* NBRC 100344; (b) *N.transvalensis*NBRC 15921; (c) *N. africana* NCTC13181; (d) *N. kruczakiae* NBRC 101016; (e) *N. brasiliensis* ATCC 700358 and (f) *N. nova* SH22a (g) *lcp*genes or DUF2236 containing domains and their respective adjacent genes.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 880-890 (2021)

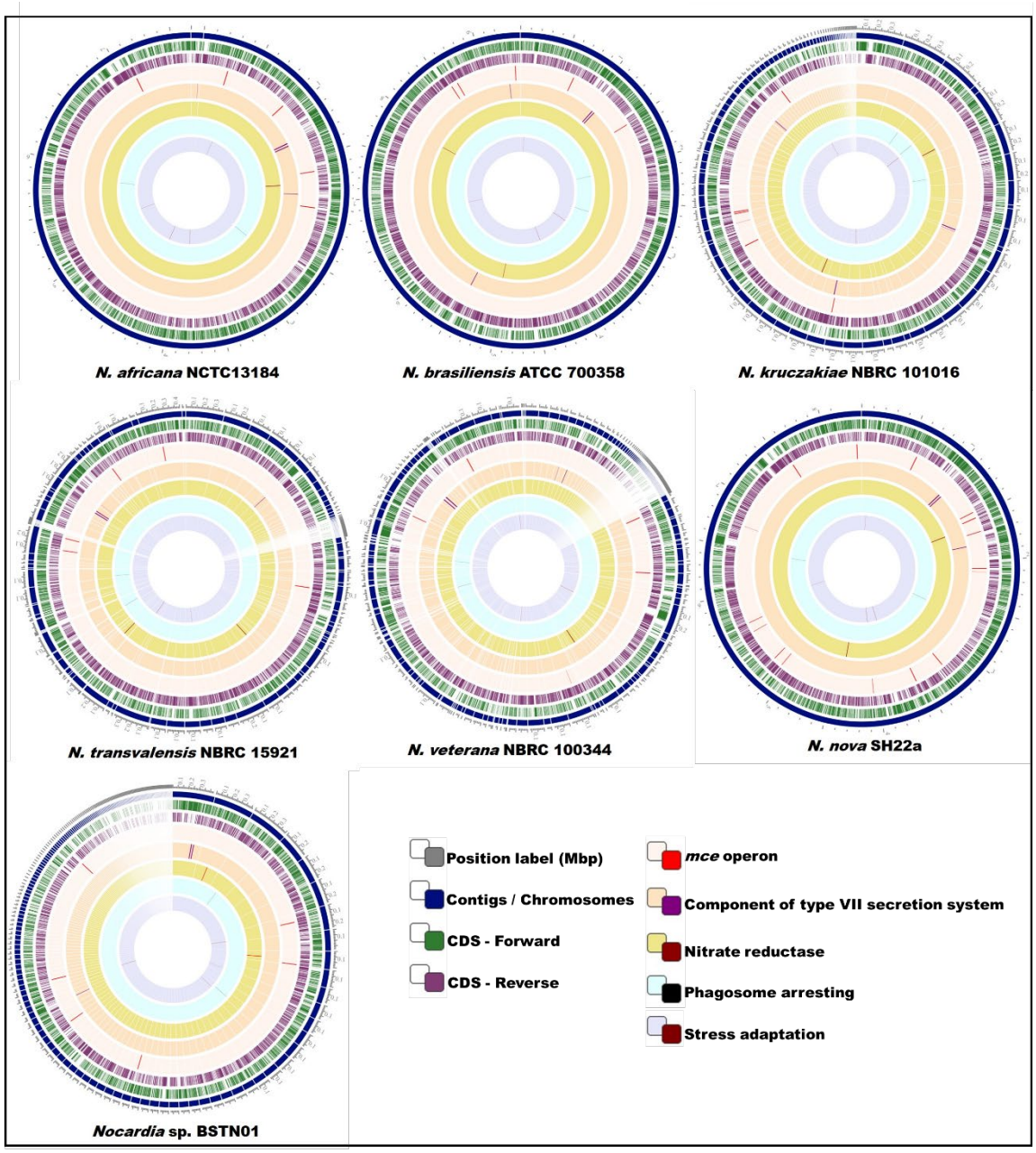

Figure 4: Showing the presence of virulence factors in the genomes of the studied *Nocardia* species. From the outermost circle to innermost circle is an illustration of the circular genomes of studied species of *Nocardia* highlighting their virulence factors. From the outermost circle to innermost (a) Grey, Position label (Mbp); (b) Blue, contigs/chromosomes; (c) green, forward CDS; (d) purple, reverse CDS; (e) red marks in seashell circle represents *mce* operons; (d) purple marks in bisque circle represents, components of type VII secretion system (T7SS); (e) dark red marks in khaki circle represents nitrate reductases; (f) black marks in light cyan circle represents phagosome arresting factors; (g) stress adaptation factors are marked in maroon in lavender colour inner most circle.

Segregation of rubber degrading *Nocardia*

From the phylogenetic analysis of the selected species of *Nocardia*, it was found that *N. kruczakiae* NBRC 101016 and strain BSTN01 share a common clade (Figure 1), indicating the strains to be descended from a common ancestor. But as the average nucleotide identity between the two strain is 94.56% (Table 1), which is well below the cut-off level (95-96%) recommended as the ANI criterion for interspecies identity [62], the strain BSTN01 does not belong to the same species as strain NBRC 101016. In fact, the strain BSTN01 should be a distinct species with respect to the rest of the considered species of *Nocardia* in this study according to the ANI criterion for interspecies identity [62]. Moreover, even *N. nova*

SH22a is placed in a distinct clade in the phylogenetic tree which indicates both the strains BSTN01 and SH22a to be distantly related.

The Rubber degrading genes of *Nocardia*are conserved

Rubber degrading actinobacteria produces latex clearing protein (Lcp) encoded by the *lcp*gene. This latex clearing protein is in fact a rubber oxygenase which initiates the degradation of rubber by cleaving the polyisoprene backbone extracellularly, which in turn allows the easy uptake of low molecular weight oligo-isoprenoids [38-40]. The amino acid sequences of latex clearing proteins (Lcp) from different species are known to be related and all are found to share a common domain of unknown functioni.e., DUF2236 together with other hypothetical proteins with diverse functional annotations [40, 63, 64]. From the translated CDs' of WGS sequences of the *Nocardia* spp. considered in this study, it was found that, *Nocardia* sp. BSTN01 harbours 10 DUF2236 sequences; whereas, *N. nova* SH22a, *N. africana* NCTC13184, *N. brasiliensis* ATCC 700358, *N. kruczakiae* NBRC 101016, *N. transvalensis* NBRC 15921 and *N. veterana* NBRC 100344 harbours 12, 8, 11, 13, 6, and 5 DUF2236 sequences, respectively. The DUF2236 amino acid sequences of BSTN01 showing highest similarity with the already proven *Lcp*SH22a (WP025350295) from *N. nova* SH22a were considered for MSA analysis (Figure 2). According to the percentage identity matrix performed for the considered DUF2236 sequences employing the online platform of https://www.ebi.ac.uk/Tools/msa/clustalo/, WP194807853 (from *Nocardia* sp. BSTN01) showed 85.43% identity with respect to the WP025350295 (from *N. nova* SH22a); whereas, the WP040717816 (from *N. veterana* NBRC 100344), WP062963305 (from *N. africana* NCTC13184), WP063012276 (from *N. kruczakiae* NBRC 101016), WP041562627 (from *N. brasiliensis* ATCC 700358) and WP040746909 (from *N. transvalensis* NBRC 15921) exhibited 87.41%, 83.46%, 85.19%,66%, and67% of identity, respectively with WP025350295. This high percentage of identity amongst the candidate sequences and the Lcp protein of *N. nova* SH22a indicates that, all the considered strains might be capable of degrading polyisoprene, provided the environment they are growing on are otherwise favourable.

From the physical maps (Figure 3) of the studied species of *Nocardia*, it was found that the rubber degrading *Nocardia* sp. BSTN01 and three others studied *Nocardia* strains, i.e., NBRC 15921, NCTC1314 and NBRC 101016 harbours similar genes downstream to their Lcp encoding gene. The downstream of *lcp* were followed by genes encoding a FAD:protein FMN transferase, a putative Fe-S, FMN containing oxidoreductase, further followed by two *yrbE* genes (*yrbEA* and *yrbEB*) and mammalian cell entry (*mce*) cluster (Figure3a, c, d and e). Interestingly, it has already been proposed that the transport of low-molecular-weight oligoisoprenoids formed due to the extracellular oxidative cleavage of polyisoprene by latex clearing protein (Lcp) into the cytoplasm via an ATP-dependent, Mce protein-driven mechanism [38]. Generally, the *mce* clusters have similar structures comprising of two *yrbE* genes (*yrbEA* and *yrbEB*), encoding integral membrane protein, followed by mammalian cell entry (*mce*) genes *mceA, mceB, mceC, mceD, mceE*and*mceF.* The *mce* clusters encode exported or membrane-tethered proteins are known to play roles in physiological functions (like, acting as substrate-binding proteins and acting as novel ABC transporters) as well as exhibiting virulence factors [36, 38].In almost all the studied strains in this work, the gene encoding Lcp is located downstream of transcriptional regulator of AcrR family. Such regulators are located proximal to a*lcp* gene in all the genomes harbouring *lcp* or DUF2236-containing domain and they might play role in regulating the expression of *lcp* genes while bacterial growth on polyisoprene or rubber (Figure 3). The physical map of *N. nova* SH22a *lcp* and its adjacent genes displays an ORF encoding a α/β-hydrolase like protein upstream to the *lcp* and an ORF encoding lipoprotein of the ApbE family. This is quite similar to that as observed in the physical map of *lcp* harboured in *Gordoniapolyisoprenivorans* VH2 plasmid, p174 [38]. It should be noted that their role in microbial rubber degradation is not clear.

Presence of virulence factors in the genome sequences

Genes homologous to those of *M. tuberculosis* with functions related to stress adaptation (*sodA*, *katG* and *ahpC*), phagosome arresting (*ndk* and *ptpA*), nitrate reductases (*narG*, *narH*, *narI* and *narJ*), secreted protein (*eis*), components of type VII secretion system (*eccA*, *eccB*, *eccD*, *mycP* and*eccCa*)and mammalian cell entry (*mce*) were detected in all the strains (Figure 4; Table S1). Almost all of the studied *Nocardia* strainshave genes related with enhanced intracellular survival protein (*eis*) whose encoded protein is associated to the intracellular survival of *M. tuberculosis* in macrophage cell lines [65]. The studied genomes from the different *Nocardia* harboured nucleoside diphosphate kinase gene (*ndk*) which is responsible for the protection of bacterial cells against reactive oxygen species. Basically, nucleoside diphosphate kinase gene (*ndk*) and protein tyrosine phosphatase (*ptpA*) have the ability to arrest macrophage phagosomal maturation for the sake of survival and persistence [66, 67]. Also, all the considered species of *Nocardia* are found to harbour catalase peroxidase (*KatG*), superoxide dismutase (*SodA*) and alkyl hydroperoxide reductase protein C (*ahpC*) which are known to play role in stress adaptation. The alkyl hydroperoxide reductase protein C helps to resist the peroxynitrite produced by macrophages as a host defensemechanism [68]. Interestingly, as it is known that rubber or polyisoprene materials exhibit the tendency to automatically get oxidized in the presence of atmospheric oxygen and ozone, giving rise to the formation of reactive oxygen species, the presence of SodA and KatG can be supposed to help the *Nocardia* strains in such conditions. A similar condition has been proposed for *Gordoniapolyisoprenivorans* VH2 [38]. The studied species of *Nocardia* were found to harbour genes encoding respiratory nitrate reductase alpha chain (*narG*), nitrate reductase beta chain (*narH*), nitrate reductase gamma chain (*narI*) and nitrate reductase delta chain (*narJ*). These are actually believed to be involved in the nitrate reduction pathway and its encoded products are involved in the survival of the opportunistic facultative pathogens during low-oxygen-levels-mediated dormancy in host infections [69, 70]. The mammalian cell entry (*mce*) genes are responsible for coding the MCE-family proteins which are known to have the ability to invade into mammalian cells and survive inside the macrophages [71].Similarly, the presence of components of type seven secretion

system (T7SS)(*ccA*, *eccB*,*mycP*,etc.) in the selected strains of *Nocardia*, exhibit virulence in *M. tuberculosis* indicates that, indicates that they might also be capable of exhibiting similar characteristics[72].Interestingly, amino acid sequences of virulence factors with functions related to stress adaptation (*sodA*, *katG* and *ahpC*), phagosome arresting (*ndk* and *ptpA*), nitrate reductases (*narG*, *narH*, *narI* and *narJ*), secreted protein (*eis*), and components of type VII secretion system (*eccA*, *eccB*, *eccD*, *mycP* and *eccCa*) from both the rubber degrading strains of *Nocardia* i.e., BSTN01 and SH22a exhibited similar range of percentage identity (Table 2) of the virulence factors exhibited by the other studied pathogenic species of *Nocardia*with respect to that of *M. tuberculosis* (Table S2). There is a report of application of genome editing technology employing CRISPR/Cas9 technique to desirably knockout secondary alcohol dehydrogenase in *N. cholesterolicum* [73], if such technology is employed in case of rubber degrading microbes, we can expect positive outputs. To cope up with the threat of *Nocardia* contaminated wastes, lytic bacteriophages can be used, as a similar case has been reported for biocontrol of Nocardia-stabilized foams in activated sludge plants by lytic bacteriophage GTE2 [74].

Conclusion:

From the WGS and complete genome sequences of the studied *Nocardia* genomes, it is evident that all the genomes from different *Nocardia* sp. harbour DUF2236 containing domain. This is very apparent that DUF2236 containing domain is conserved in the diverse studied *Nocardia* genomes and is also responsible for encoding the latex clearing protein which is responsible for the primary steps in the process of biodegradation of rubber. Nevertheless, all the studied species of *Nocardia* are also found to harbour virulence factors which are homologous to those of *M. tuberculosis* with functions related to stress adaptation, phagosome arresting, nitrate reductases, secreted protein, effector delivery system and mammalian cell entry (*mce*). Interestingly, as already discussed above, the *mce* clusters which were originally identified because of their critical function of invading and surviving inside mammalian cells are now also known to play roles as substrate-binding proteins and mediate the movement of intermediate products formed due to the oxidative cleavage of polyisoprene across the bacterial cell wall. Moreover, the presence of SodA and KatG in rubber degrading *Nocardia* is supposed to help in survival from the oxidative stress due to the formation of activated oxygen species as a result of auto-oxidizing effect of rubber. Such relevant virulent genes in rubber degrading *Nocardia* and the presence of conserved *lcp* gene or DUF2236 containing domain in facultative human pathogens of *Nocardia* raises a concern in terms of clinical science. A number of clinical products like renal tubes, catheters and central vascular tubes are made up of rubber or latex. Thus, in patients who are immune compromised and are required to undergo treatment employing such rubber/latex products *Nocardia* infection might pose a threat. Such conditions make the studied rubber degrading a potential emerging human pathogen. Further studies would definitely help to better understand any relatedness between the rubber degradation genes and virulence factors if present.

Funding: Not applicable

Conflict of interest: The authors declare that they have no competing interests

Authors' contributions: BS designed and performed the experiment and wrote the main manuscript text; AGM helped acquisition and analysing the in-silico data; SM conceptualized the research, designed and critically revised the manuscript.

Acknowledgements: BS is thankful to University of Calcutta for his fellowship. We are especially thankful to Dr. Prithidipa Sahoo, Department of Chemistry, Visva-Bharati University for her generous help to proofreading the manuscript.

References:

[1] Dhakal D *et al. CurrProtocMicrobiol*. 2016 **40** [PMID: 26855280]
[2] Shariff M & Gunasekaran J,*Can Respir J*. 20167494202 [PMID:27445562]
[3] Bafghi MF*et al.Malays J Med Sci*. 2014**21:**75-76. [PMID: 24639618]
[4] Hashemi-Shahraki Aet al. *Sci Rep*. 2015**5:**17862. [PMID: 26638771]
[5] Fellag M *et al. Genome Announc*. 2018 **6**:e00212-18 [PMID: 29674534]
[6] Beaman BL*et al.J Infect Dis*. 1976**134:**286-289[PMID: 789786]
[7] AlaviDarazam I*et al. Iran Red Crescent Med J*. 2013**15:**436-439. [PMID: 24349735]
[8] Conville PS*et al. J Clin Microbiol*. 2017**56:**e01419-17.[PMID:29118169]
[9] Matulionyte R*et al.J Clin Pathol*. 2004**57:**807-812. [PMID: 15280400]
[10] Lu SH*et al. Bosn J Basic Med Sci*. 2020**20:**531-538. [PMID: 32415818]
[11] Vanegas S*et al.Int J Infect Dis*. 2014 **20:**63–65. [PMID: 24362018]
[12] McGuinness SL*et al. Open Forum Infect Dis*. 2016**3:**208. [PMID: 27942539]
[13] Xu S *et al. PLoSNegl Trop Dis*. 2021 **15**:e0009665 [PMID: 32816977]
[14] Zoropogui A *et al. J Bacteriol*. 2012 194:2098–2099 [PMID: 22461543]
[15] Imajoh M *et al. Genome Announc.*2015**3**:e00082-15 [PMID: 25767227]
[16] Xia L *et al. Genome Announc*. 2015 **3**:e01223-14[PMID: 25555736]
[17] Chakrabortti A *et al. Genome Announc.* 2016 **4**:e00593-16 [PMID: 27445366]
[18] Vera-Cabrera L *et al. PLoS One*. 2013 **8**:e65425 [PMID: 23755230]
[19] Yasuike M *et al. PLoS One*. 2017 **12**:e0173198 [PMID: 28257489]
[20] Ghodhbane-Gtari F *et al. Genome Announc.* 2016 **4**:e00770-16 [PMID: 27491997]

[21] Ghodhbane-Gtari F *et al. Genome Announc.* 2016 **4**:e00799-16 [PMID: 27491980]
[22] Perez-Maya A *et al. J Bacteriol*. 2012**194**:2761-2762 [PMID: 22535940]
[23] Imajoh M *et al. Genome Announc*. 2016 **4**:e01606-15. [PMID: 26798107]
[24] Valdezate S *et al.Genome Announc*. 2017 **5**:e01212-17 [PMID: 29146850]
[25] Toyokawa M *et al. MicrobiolResourAnnounc.* 2020 **9**:e01022-20 [PMID: 33214300]
[26] Zoropogui A *et al. BMC Genomics*. 2013 **14**:286 [PMID: 23622346]
[27] Vautrin*et al. MicrobiolResourAnnounc*. 2019 **8**:e00600-19 [PMID: 31416867]
[28] Komaki H *et al. BMC Genomics.* 2014 **15**:323 [PMID: 24884595]
[29] Carrasco G *et al. Genome Announc*. 2017 **5**:e00551-17 [PMID: 28963198]
[30] Dhakal D *et al. MicrobiolResourAnnounc.* 2019 **8**:e01098-19 [PMID: 31806741]
[31] Buchmann A *et al. Genome Announc.* 2016**4**:e01391-16 [PMID: 27979943]
[32] Buchmann A *et al. MicrobiolResourAnnounc.* 2020 **9**:e00689-20 [PMID: 32816977]
[33] Chen J *et al. Chembiochem*. 2020**21**:2205–2213 [PMID: 32196864]
[34] Rodrigues EM *et al. Braz J Microbiol.* 2017**3**:391-392 [PMID: 28017289]
[35] Tsuchii A *et al.Biodegradation*. 1996 **7:**405–413. [PMID: 9144970]
[36] Luo Q *et al. Appl Environ Microbiol*. 2014**80:**3895-3907. [PMID: 24747905]
[37] Linh DV*et al. J BiosciBioeng*. 2017**123:**412-418. [PMID: 28065456]
[38] Hiessl S *et al.Appl Environ Microbiol*. 2012**78:**2874-2887. [PMID: 22327575]
[39] Sharma V *et al.PLoS One*. 2018**13:**e0197448. [PMID: 29851965]
[40] Basik AA*et al. Polymers (Basel)*. 2021**13:**1989. [PMID: 34204568]
[41] Rose K*et al. Biomacromolecules*. 2005**6:**180-188. [PMID: 15638519]
[42] Yikmis M&Steinbüchel A, *Appl Environ Microbiol*. 2012**78:**4543-4551. [PMID: 22504822]
[43] Yikmis M *et al.Appl Environ Microbiol*. 2008**74:**5373-5382. [PMID: 18606806]
[44] Lawrence EL& Turner IG, *Med Eng Phys*. 2005**27:**443-453. [PMID: 15990061]
[45] Polat*et al.Int J Clin Exp Med*. 2015**8:**13421-13423. [PMID: 26550278]
[46] Kumar RP*et al. Indian J Dermatol*. 2012**57:**66-70. [PMID: 22470217]
[47] Vishwanath V& Rao HM,*J Conserv Dent*. 2019**22:**216-222. [PMID: 31367101]
[48] VanEijk AM *et al.The Lancet Public Health.*2019**4** (8):e376-e393 [PMID: 31324419]
[49] Al Akhrass F *et al.Emerg Infect Dis*. 2011**17:**1651-1658. [PMID: 21888790]
[50] Torres HA *et al.Medicine (Baltimore)*. 2002**81:**388-397. [PMID: 12352633]
[51] Guindon S *et al.Syst Biol*. 2010**59:** 307–321. [PMID: 20525638]
[52] Goris J *et al.Int J SystEvolMicrobiol*. 2007**57:**81-91.[PMID:17220447]
[53] Aziz RK*et al. BMC Genomics*. 2008**9:**75.[PMID:18261238]
[54] Davis JJ*et al. Nucleic Acids Res*. 2020, **48:**D606–D612. [PMID: 31667520]
[55] Chen L *et al.Nucleic Acids Res*. 2005**33**(Database issue):D325-D328. [PMID: 15608208]
[56] Gürtler V *et al.Int J SystEvolMicrobiol*. 2001**51:**933-936. [PMID: 11411717]
[57] **[57]**Hamid ME*et al. J Clin Microbiol*. 2001**39:**625-630. [PMID: 11158119]
[58] Compte RB*et al.J Ophthalmic Inflamm Infect*. 2015**5:**36. [PMID: 26604206]
[59] Conville PS*et al.J Clin Microbiol*. 2004**42:**5139-5145. [PMID: 15528707]
[60] Mirza SH& Campbell C,*J Clin Pathol*. 1994**47:**85-86. [PMID: 8132817]
[61] Verma P& Jha A, *Clin Exp Dermatol*. 2019**44:**123-129. [PMID: 29808607]
[62] Richter M&Rosselló-Móra R, *Proc Natl Acad Sci U S A*. 2009**106:**19126-19131. [PMID: 19855009]
[63] Röther W *et al.Appl Environ Microbiol*. 2016**82:**6593-6602. [PMID: 27590810]
[64] Hiessl S *et al.Appl Environ Microbiol*. 2014**80:**5231-5240. [PMID: 24928880]
[65] Lella RK& Sharma C, *J Biol Chem*. 2007**282**(26):18671-18675. [PMID: 17449476]
[66] Yu H *et al.Microbiol Res*. 2017**205:**125-134. [PMID: 28942838]
[67] Forrellad MA *et al.Virulence*. 2013**4:**3-66. [PMID: 23076359]
[68] Master SS *Microbiology (Reading)*. 2002**148:** 3139–3144. [PMID: 12368447]
[69] Khan A& Sarkar D, *Microbiology*. 2012**158:**301-307. [PMID: 22174380]
[70] Akhtar S *et al.J Bacteriol*. 2013**195:**4592-4599. [PMID: 23935045]
[71] Arruda S *et al.Science*. 1993**261:**1454-1457. [PMID: 8367727]
[72] Roy S*et al.Eur J Microbiol Immunol (Bp)*. 2020**10:**202-209. [PMID: 33174865]
[73] Huang JK *et al. PLoS One.* 2020**15**:e0230915 [PMID: 32218601]
[74] **[74]**Petrovski S *et al. Appl Environ Microbiol.*2011 **77**:3923-3929 [PMID: 21498753]

Edited by P Kangueane
Citation: Sarkar ***et al.*** Bioinformation 17(10): 880-890 (2021)

ISSN 0973-2063 (online) 0973-8894 (print)

Bioinformation 17(10): 880-890 (2021)

Table S1: Showing the presence of virulence factors of studied *Nocardia* genomes with their locus tags.

Virulence factor encoding gene (and product)	Strains (locus tag)						
	***Nocardia* sp. BSTN01 (IRT45_RS)**	***N. nova* SH22a (NONO_RS)**	***N. veterana* NBRC 100344 (ON37_RS)**	***N. transvalensis* NBRC 15921 (ON41_RS)**	***N. kruczakiae* NBRC 101016 (NK1_RS)**	***N. brasiliensis* ATCC 700358 (O3I_RS)**	***N. africana* NCTC13184 (DX133_RS)**
Stress adaptation							
sodA (Superoxide dismutase [Mn/Fe])	12770	00420	00270	01605	17330	00385	01765
katG (catalase-peroxidase KatG)	09560, 28740	17245, 18055	22015	15720	04770	14720, 18760	17710, 19915
ahpC (alkyl hydroperoxide reductase protein C)	15570	26420	15680	20665	31640	29880	24315
Phagosome arresting							
ptpA (low molecular weight protein-tyrosine-phosphatase/protein-tyrosine-phosphatase)	15300	10640	08195	11040	04610	25845	10720
ndk (nucleoside diphosphate kinase)	04255	29270	07325	09970	03135, 03070	09150, 09215	27005
Secreted protein							
eis (enhanced intracellular survival protein)	00090	03520	03250	04785	28025	03710	
Nitrate reductases							
narG (Respiratory nitrate reductase alpha chain)	02535, 10940	07200, 19820	19305	14925, 24620	05755, 19400	34885, 22335	08045
narH(Respiratory nitrate reductase beta chain)	02530, 10945	07195, 19815	19310	14930, 24625	05750, 19405	34890, 22325	08040
narI (Respiratory nitrate reductase gamma chain)	S02520, 10955	07185, 19805	19320	14940, 24635	05740, 19415	34900, 22315	08030
narJ (Respiratory nitrate reductase delta chain)	02525, 10950	07190, 19810	19315	14935, 24630	05745, 19410	34895, 22320	08035
Component of Type VII secretion system							
eccA(AAA+ family protein ATPase EccA1, component of Type VII secretion system ESX-1)	01005	04675, 07995	04240, 28365	05995	18475, 29710	04760, 23950	05660, 08530
eccB(Membrane protein EccB-like, component of Type VII secretion system in *Actinobacteria*)	01000, 01135	04670, 04790	28360, 04370, 04235	06135, 05990	29705, 18480, 10580	23945, 04895, 04755	05655, 08535, 05795
eccD(Integral membrane protein EccD-like, component of Type VII secretion system in *Actinobacteria*)	01010, 01155	04680, 04820	28320, 04390, 04245	06000, 06150	18470, 10560, 29665	23905, 04920, 04765	05665, 05815
mycP(Serine protease, putative component of Type VII secretion system in *Actinobacteria*)	01015, 01150	04685, 04815	29060, 28315, 04385, 04250	06145, 06005, 33095	10565, 18465, 29660	04770, 04915, 23900	05670, 35435, 05810, 08580
eccCa(FtsK/SpoIIIE family protein, putative EssC/YukB component of Type VII secretion system)	01020, 01160, 27835	08005	28355	33055	29700, 18460, 10555	23940	05675, 08540, 05820
Mammalian cell entry (mce) protein							
mceA	09045, 11270, 22775, 26760, 28990, 34710	06545, 06855, 07550, 09395, 14710, 02745, 16230, 18125, 23140, 23575, 25115, 25615, 34175, 37465	02620, 06080, 23285, 26175	01075, 04080, 08300, 09005, 28250	13515, 23385, 25030, 30800	03000, 06745, 38180, 38580, 42025	01205, 04205, 07460, 09140, 32410
mceB	09040, 11265, 22780, 26765, 26770, 28995, 34715	06550, 06860, 07545, 09400, 14705, 16225, 18120, 23145, 23580, 25110, 02750, 34170, 37460, 25610	02625, 06085, 23280, 26170	01070, 01065, 04085, 08305, 09000, 28245	13520, 13525, 23380, 25025, 30805	38575, 38175, 06750, 03005, 42015, 42020	01200, 01195, 04210, 07465, 09145, 32405
mceC	09035, 11260, 22785, 29000, 34720	06555, 06865, 07540, 09405, 14700, 16220, 18115, 23150, 23585, 25105, 25605, 02755, 37455	02630, 06090, 23275, 26165	04090, 08310, 08995, 28240	23375, 25020, 30810	38170, 06755, 03010, 38570	04215, 07470, 09150, 32400
mceD	09030, 11255, 22790, 26775, 29005, 34725	06560, 06870, 07535, 09410, 14695, 16215, 18110, 23155, 23590, 25100, 25600, 02760, 34160, 37450	02635, 06095, 23270, 26160	01060, 04095, 08315, 08990, 28235	13530, 23370, 25015, 30815	38165, 06760, 03015, 42010, 38565	01190, 04220, 07475, 09155, 32395
mceE	09025, 11250, 22795, 26780, 29010, 34730	06565, 06875, 07530, 09415, 14690, 16210, 18105, 23160, 23595, 25095, 25595, 02765, 34155, 37445	02640, 06100, 23265, 26155	01055, 04100, 08320, 08985, 28230	13535, 23365, 25010, 30820	38160, 06765, 03020, 42005, 38560	01185, 04225, 07480, 09160, 32390
mceF	09020, 11245, 22800, 26785, 29015, 34735	06570, 06880, 07525, 09420, 14685, 16205, 23165, 23600, 25090, 25590, 02770, 34150, 37440	02645, 06105, 23260, 26150	01050, 04105, 08325, 08980, 28225	13540, 23360, 25005, 30825	38155, 06770, 03025, 42000, 38555	01180, 04230, 07485, 09165, 32385

Table S2: Showing the range of identity (%) of the amino acid sequences of virulence factors from the studied pathogenic *Nocardia* species with respect to that of *M. tuberculosis*

Virulence factors	Range of Identity (%)
sodA(Superoxide dismutase [Mn/Fe])	84.54-87.44
katG(catalase-peroxidase KatG)	71.24-79.33
ahpC(alkyl hydroperoxide reductase protein C)	86.60-87.63
ptpA(low molecular weight protein-tyrosine-phosphatase)	28.87-61.54
ndk(nucleoside diphosphate kinase)	23.91-81.95
eis(enhanced intracellular survival protein)	46.25-48
narG(Respiratory nitrate reductase alpha chain)	68.18-77.65
narH(Respiratory nitrate reductase beta chain)	63.57-77.30
narI(Respiratory nitrate reductase gamma chain)	57.45-67.36
narJ(Respiratory nitrate reductase delta chain)	51.38-60.47
eccA(AAA+ family protein ATPase EccA1, component of Type VII secretion system ESX-1)	43.67-62.93
eccB(Membrane protein EccB-like, component of Type VII secretion system in *Actinobacteria*)	36.76-41.70
eccD(Integral membrane protein EccD-like, component of Type VII secretion system in *Actinobacteria*)	28.46-34.76
mycP(Serine protease, putative component of Type VII secretion system in *Actinobacteria*)	43.86-56.38
eccCa(FtsK/SpoIIIE family protein, putative EssC/YukB component of Type VII secretion system)	40.40-66.57

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 880-890 (2021)

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.