3 Biotech (2023) 13:47
https://doi.org/10.1007/s13205-023-03459-5

ORIGINAL ARTICLE

Identification and taxonomy of *Streptomyces justiciae* strain RA-WS2: a novel setomimycin producing actinobacterium

Ravi Singh Manhas[1,2] · Devtulya Chander[1,2] · Asha Chaubey[1,2]

Received: 23 July 2022 / Accepted: 2 January 2023 / Published online: 13 January 2023

Abstract

The taxonomic position of novel bianthraquinone antibiotic producer *Streptomyces* strain RA-WS2, a soil isolate from Shivalik region of NW Himalayas, India, has been described. The isolate produces Setomimycin as a major secondary metabolite under defined submerged fermentation conditions. 16S rRNA partial gene sequencing of the isolate indicated its closest similarity (99.4%) with *Streptomyces cyaneochromogenes*, followed by *Streptomyces aquilus.* However, the morphological characteristics i.e. colony colour, mycelium and spore chain arrangement were found to be close to *Streptomyces aquilus.* Therefore, a polyphasic approach was used for taxonomic positioning of the isolate. The Whole genome based similarity with 88.4% dDDH value, 98.65% ANI and 96.99% AAI value indicated its closest identity with *Streptomyces justiciae*. The taxonomic characteristics such as white colony with smooth surface, cylindrical spores arranged in straight chain, diffusible melanin production, high salt tolerance, 16S rRNA gene sequencing and phylogenomic studies, led to the identification of the strain as *Streptomyces justiciae* RA-WS2. The predicted biosynthetic gene clusters further confirmed the presence of the BGC for setomimycin biosynthesis in *Streptomyces justiciae* strain RA-WS2.

Keywords Phylogenomics · *Streptomyces justiciae* · Setomimycin · Whole genome sequence · NW Himalayas

Abbreviations

AAI	Average amino-acid identity
ANI	Average nucleotide identity
antiSMASH	Antibiotics and secondary metabolite analysis shell
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
dDDH	Digital DNA–DNA hybridization
EtBr	Ethidium bromide
ISP	International Streptomyces project
IToL	Interactive Tree of Life
MEGA	Molecular Evolutionary Genetics Analysis
NCBI	National Center for Biotechnology Information
NW Himalayas	North Western Himalayas
RDP	Ribosomal database project
PBS	Phosphate buffer saline
SCA	Starch casein agar
SEM	Scanning electron microscopy
SRA	Sequence read archive
SSY	Starch soyabean meal yeast extract
STAG	Species tree inference all gene method
TAE	Tris acetate EDTA
TYGS	Type strain genome server

Introduction

Members of the genus *Streptomyces*, family Streptomycetaceae, order Actinomycetales are considered to be one of the most important antibiotic producers (Berdy 2005). At the time of writing, the *Streptomyces* genus consists of about 699 validly and correctly named member species based on phylogeny of the 16S rRNA genes and genomes (https://lpsn.dsmz.de/genus/streptomyces). Members of the genus *Streptomyces* are understood to have a complex life cycle.

Ravi Singh Manhas and Devtulya Chander authors contributed equally.

✉ Asha Chaubey
achaubey@iiim.res.in

[1] Fermentation and Microbial Biotechnology Division, CSIR-Indian Institute of Integrative Medicine, Canal Road, Jammu 180001, India

[2] Academy of Scientific and Innovative Research, CSIR- Human Resource Development Centre, (CSIR-HRDC), Ghaziabad, Uttar Pradesh 201 002, India

Their cellular morphologies and colony characteristics such as colony colour, aerial mycelium, formation of branched hyphae, and spore arrangement vary depending on the growth as well as substrate used. Therefore, these traits have been exploited to differentiate between species and strains (Parte et al. 2012).

Cultural studies of *Streptomyces* have provided an insight into the presence of an abundance of bioactive secondary products and the wide range of biosynthetic gene clusters have been confirmed by comparative genome analysis. It is now a common practice to perform genome sequencing, phylogenomics, and comparative genomics to determine the taxonomic placement and bioactive compounds in newly isolated members of genus *Streptomyces*. For example, whole genome sequence studies of *Streptomyces coelicolor* has revealed that it has a potential to produce twenty different secondary metabolites, whereas only three antimicrobials were discovered with cultural studies (Bentley et al. 2002). The whole genome sequencing of *Streptomyces avermitilis* revealed almost thirty different secondary metabolite clusters (Ikeda et al. 2003). These studies have proved that *Streptomyces* genus harbours a large and diverse range of natural bioactives (Stanbury et al. 2016).

We previously reported the isolation of a new *Streptomyces* isolate RA-WS2 which produces setomimycin, a rare 9,9′-bianthraquinone antibiotic (Manhas et al. 2022a, b). Herein, we report the genome sequence of the isolate, its taxonomy, cultural, morphological and genomic traits vis-a-vis its nearest phylogenomic neighbours.

Methods

Growth and characterization

The actinobacterium *Streptomyces* sp. strain RA-WS2 used in the present study was isolated from a soil sample collected from the Shivalik foothills of North-Western Himalayan region (32.7266° N, 74.8570° E), India on starch casein agar medium (Manhas et al. 2022a, b). Growth of the isolate RA-WS2 was observed on SCA and SSY agar media plates. The isolate RA-WS2 was also grown in setomimycin production medium broth (Constituents in g/L: Soluble starch-25, Soyabean meal-15, Calcium carbonate-4 and Yeast extract-2) at different pH (5.0, 6.0, 7.0, 8.0, 9.0); temperatures (10 °C, 20 °C, 30 °C, 40 °C) and NaCl concentrations (0%, 0.25%, 0.5%, 1.0%, 2.5%, 5%, 7.5%, 10%) to evaluate the effect on the growth and morphology. The growth of RA-WS2 was determined according to the methods used by Verma et al. (2009). The utilization of different carbon and nitrogen sources was studied as per the methods of Shirling and Gottlieb (1966), using the basal medium recommended by Pridham and Gottlieb (1948). The sugar utilization ability of strain RA-WS2 was tested using the basal medium supplemented with glucose, adonitol, arabinose, lactose, sorbitol, mannitol, rhamnose, sucrose, xylose, inositol and nitrogen source i.e. arginine to differentiate it with the closely related phylogenetic neighbours.

Additionally, growth of the strain RA-WS2 was tested on International Streptomyces Project (ISP) media (ISP-1, ISP-2, ISP-3, ISP-4, ISP-5, ISP-6, ISP-7) as described by Shirling and Gottlieb (1966) and in various setomimycin production media (PM-1 to PM-12) at 28 °C ± 2 °C for 4 to 6 days, unless stated otherwise. The colour of colonies and soluble pigments were determined according to the colour standard following Ridgway procedure (Ridgway et al. 1912). Biochemical and physiological characteristics such as test for gelatin hydrolysis, starch hydrolysis, lipid hydrolysis were carried out according to the methods described by Li et al. (2020). The production of melanin was screened on specified medium (constituents in g/L: glucose 1.0, peptone 10.0, NaCl 5.0, $CaCl_2$ 0.1 and tyrosine 2.0; pH:7.0) agar plates by the methods described by Guo et al. (2014). Growth was examined by macroscopy as well as microscopy according to methods described by Li et al. (2020) at 40X magnification (Olympus- CH2 Japan) and under a scanning electron microscope (SEM) JEOL JSM-IT300, Tokyo Japan at voltage of 15.0 kV.

Molecular characterization and whole genome sequencing

The isolate RA-WS2 was grown in a 250 mL conical flask containing 50 mL production medium for 4 days at 28 ± 2 °C under shaking conditions. After 4 days of fermentation, broth was centrifuged and the recovered cells were washed with PBS and used for DNA isolation. DNA was isolated by Zymoresearch Bacterial DNA isolation Kit. The 16S rRNA gene from genomic DNA was amplified using 27F and 1492R universal primers by previously reported method (Manhas et al. 2022b). Briefly, reaction mixture contained PCR buffer with $MgCl_2$ (1X) (Takara Cat R500A), 0.5 mM of 1492R and 27F primers each, 0.2 mM of dNTP, 0.05 U/µL Taq polymerase (Takara Cat R500A) and DNA template (2 ng/µL). The amplification was carried out with 95 °C initial denaturing for 5 min followed by denaturing at 95 °C for 30 s, annealing at 60 °C for 30 s, followed by extension at 72 °C for 1 min 30 s. After 35 cycles, final extension was performed at 72 °C for 7 min.

Thus obtained PCR product was gel purified using the Promega gel purification kit (Betz and Strader 2002) and directly used as template for Sanger sequencing on 3130xl Genetic Analyzer from Applied Biosystems. Sequencing PCR reaction mixture (10µL) consisted of 1 µL ready reaction mix supplied by Invitrogen in Big dye terminator 3.1 kit, 1 µL of 5X BigDye Sequencing Buffer, 0.5 µL of primer

(1492R/27F), 50 ng of template and DI water. The sequence PCR conditions were kept same as used for amplification reaction. Following cleanup, the sequencing PCR products were mixed with 10 µL of Hi-Dye formamide. POP-7 was used as matrix for capillary electrophoresis during Sanger sequencing. Fragments of different lengths were analyzed using the 3130xl Genetic analyzer from Applied Biosystems.

The partial consensus sequence of 1419 bp thus obtained was subjected to BLASTn tool of NCBI. Further, 16S rRNA gene sequence-based identification service of EZbiocloud (Yoon et al. 2017a), and sequence match function of RDP database (Cole et al. 2014) were used to find the closely related taxa. The 16S sequences of the closely related taxa, were aligned using the ClustalW and the evolutionary history was inferred using the neighbour-joining method (Saitou and Nei 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The evolutionary distances were computed using the maximum composite likelihood method (Tamura et al. 2004) and are in the units of the number of base substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1). The analysis involved 26 nucleotide sequences and all positions containing gaps and missing data were eliminated using complete deletion option. Evolutionary analyses were conducted in MEGA11 (Tamura et al. 2021) and the obtained tree was visualized in IToL server (Letunic and Bork 2021).

To perform Whole genome sequencing, 100 ng of DNA dissolved in nuclease free water was shredded using the Covaris Ultrasonicator (LE220, Covaris, USA) with 450 W Peak power using following settings: Duty Factor: 20; No. of cycles: 1000; Duration: 150 s. Thereafter, ends of fragmented DNA were enzymatically repaired for dA tailing and ligation of the adapter. Cleanup of adapter ligated fragments was performed using SPRI beads and the cleaned fragments were indexed using limited Cycle PCR to enrich adapter ligated molecules. After determining the quality and quantity, the amplified products were subjected to sequencing in Novaseq 6000 with 2 × 150 reads. The sequencing was performed at MedGenome, India.

The read quality of pair end data was checked using FastQC. Adapter sequences having Phred score below 36 were trimmed using FastP tool (Chen et al. 2018) using Galaxy version 0.23.2 + galaxy 0 (Afgan et al. 2018). The quality of trimmed raw data was re-assessed using the FastQC ver. 0.11.9 (Andrews 2022). The high quality raw data with a Phred score of 36 or above was used for genome assembly using Shovill (Galaxy version 1.1.0 + galaxy 1) and UniCycler (Galaxy ver 0.4.8.0) following procedures described by (Seemann 2022a; Wick et al. 2017). The quality of assembly was assessed using the Quast (version 5.2.0 + plus galaxy 0). The obtained best draft assembly was analyzed on TYGS Server (https://tygs.dsmz.de/) to find the closely related genomes (Meier-Kolthoff and Göker 2019). The complete 16S rRNA sequence was obtained from the Whole genome using Barrnap package ver. 0.9 (Seemann 2022b) and was included in the dataset of closely related taxa, instead of partial 16S rRNA gene sequence for constructing the phylogenetic tree in MEGA 11. The analysis was performed following the method used for partial 16S rRNA gene sequence.

The ANI was calculated on EZbiocloud server by the method of Yoon et al. (2017b) to find the relative similarity against closely related taxa and the AAI was calculated using the method of Rodriguez-R and Konstantinidis (2014) (http://enve-omics.ce.gatech.edu/aai/). Various genes from genome assembly were extracted using Prokka package in Galaxy server (Galaxy ver. 1.14.6 + galaxy1) as described by Seemann (2014). Secondary metabolite gene cluster elucidation was performed in antiSMASH server (http://antismash.secondarymetabolites.org) (Blin et al. 2021).

The orthologs are a set of genes in different species that have evolutionary descended from same gene of common ancestor. Therefore, to construct the phylogenetic tree based on whole genome data, various orthologs were identified in a set of top 24 closely related genomes including the current reference based genome assembly and *Mycobacterium tuberculosis* as an out group genome, using orthofinder package (Emms and Kelly 2019). Various orthologs, thus identified were used to construct a consensus phylogenetic tree based on STAG protocol (Species tree inference all gene method) through orthofinder package Ver 2.5.4 (Emms and Kelly 2018). The orthofinder was installed using Conda, which provided all the dependencies and correct environment in Python version 3.10 and the tree was visualized in IToL server.

To confirm the species level identification, multiphasic approach was used, wherein various macroscopic (colony characteristics, pigment production), microscopy (spore chain morphology) as well as biochemical characteristics (sugar utilization, hydrolytic properties and salt tolerance) of *Streptomyces justiciae* RA-WS2 were compared with the closely related taxa at genome level i.e. *Streptomyces justiciae* 3R004^{T}, *Streptomyces aquilus* GGCR-6 ^{T} and *Streptomyces cyaneochromogenes* MK-45 ^{T} as used by other research groups (Tang et al. 2019; Li et al. 2020 and Phongsopitanun et al. 2021).

Results and discussion

Growth and characterization

The isolate RA-WS2 grows well in starch and soyabean rich media (SCA and SSY) under wide range of temperatures ranging from 20 °C to 40 °C with an optimum temperature

for growth at 25 °C to 30 °C (Fig. 1a, b). Under the submerged conditions, wide range of pH (5.0–9.0) supports growth of RA-WS2 with optimum growth at pH 6.0–7.0 as shown in Fig. 1(c). It was also observed that the isolate has a capability to grow in high NaCl concentrations (upto 10%) indicating its halotolerant nature.

The culture is capable to utilize several sugars such as glucose, adonitol, arabinose, lactose, sorbitol, mannitol, rhamnose, sucrose and xylose as sole carbon source. Additionally, the texture of aerial as well as mature substrate mycelia of the isolate were observed on various ISP culture media series i.e. ISP-1 to ISP-7. The isolate has smooth and powdery aerial mycelium, mostly white in colour on ISP agar plates. Substrate mycelium colour varied from citrine to brown in colour. As shown in Fig. 2(a), the substrate mycelium appears brown on ISP1, ISP2, ISP5, ISP6 and ISP7, whereas on ISP3 and ISP4, it was citrine and light buff in colour, respectively. The strain RA-WS2 was also observed to be positive for diffusible melanin production. The isolate exhibits excellent starch, casein and gelatin hydrolysing

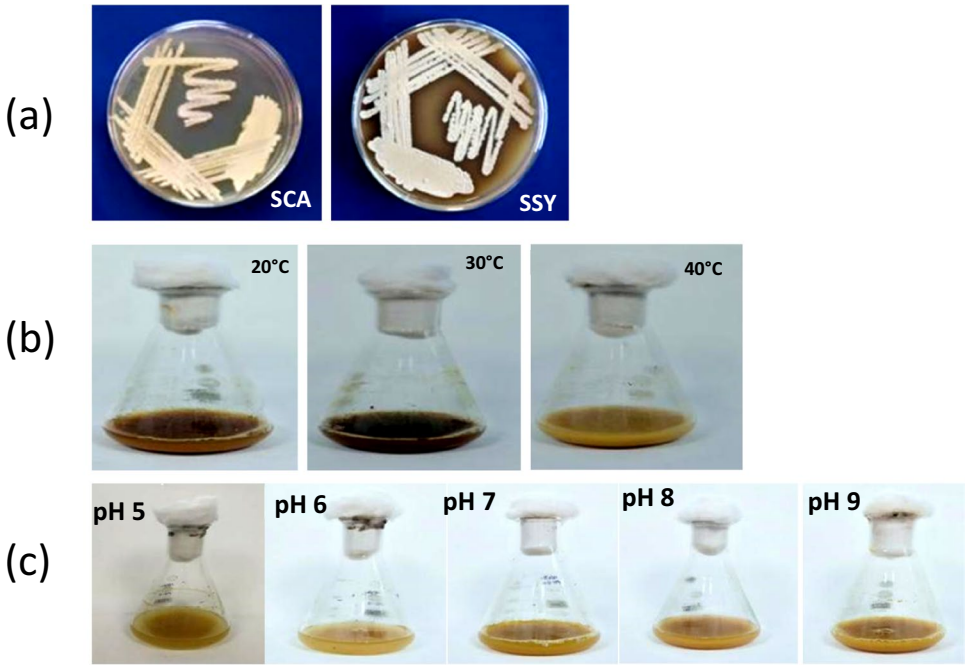

Fig. 1 Growth of isolate RA-WS2: **a** Growth on Starch Casein Agar (SCA) and Starch Soyabean Yeast Extract Agar (SSY) demonstrates that the carbon and nitrogen rich media support the growth of culture **b** Growth of the isolate at 20 °C to 40 oC demonstrates its mesophilic nature **c** pH 5.0 to pH 9.0 support the growth of isolate with an optimum growth at neutral pH

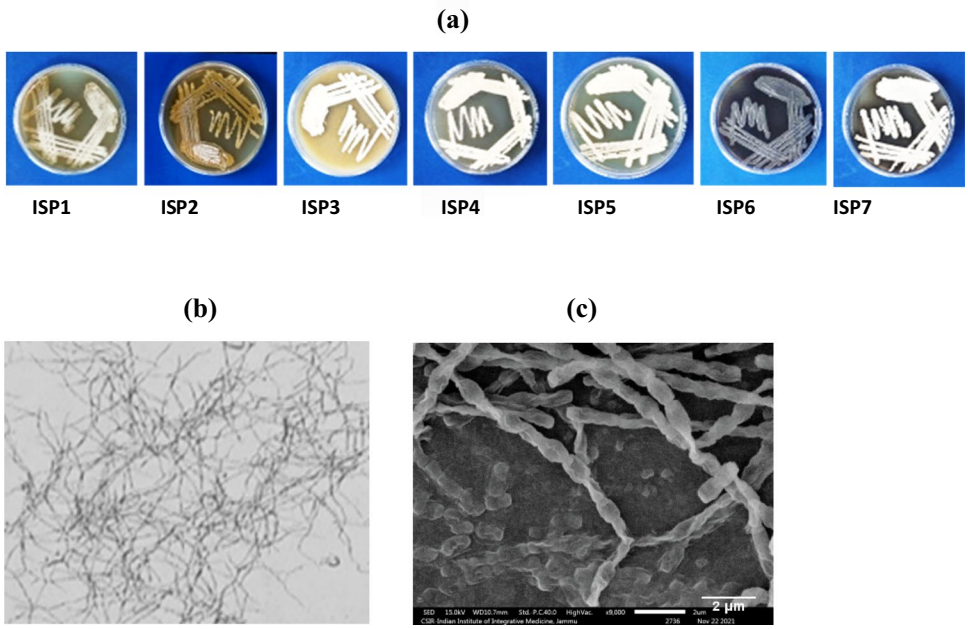

Fig. 2 **a** Growth of isolate RA-WS2 on various ISP media showing smooth and powdery aerial mycelium, mostly white in colour indicating characteristic features of *Streptomyces* genus **b** Microscopic image under light microscope showing branched and flagellated mycelia **c** Scanning Electron Microscopy image showing formation of cylindrical spores arranged in straight chains Scale bar = 2 µm)

property and capable to utilize arginine as a sole nitrogen source. Under the light microscope (Fig. 2b) the isolate RA-WS2 exhibited typical characteristics of the members of *Streptomyces* genus i.e. having well developed filamentous and branched mycelium. Scanning electron microscopy studies revealed that the aerial mycelium produced smooth surfaced cylindrical spores arranged in straight chains as shown in Fig. 2(c). Table 1 demonstrates the phenotypic characteristics of the isolate *Streptomyces* sp. RA-WS2.

Molecular characterization and whole genome analysis

To identify the isolate at molecular level, the forward and reverse sequences obtained by Sanger sequencing were aligned to obtain a consensus sequence of 1419 bp. The 16S partial sequence was deposited in NCBI GenBank with ID OM540363.1. The BLASTn results with reference database indicated 99.4% identity of the isolate *Streptomyces* sp. RA-WS2 with *Streptomyces cyaneochromogenes* having 99% query coverage (Supplementary Figure S1). Further, the phylogenetic tree of 16S sequences of closely related taxa was constructed in MEGA 11 (Fig. 3). However, the morphological characteristics such as aerial and substrate mycelia of the strain were different from *Streptomyces cyaneochromogenes,* which produces blue pigment. The results were reassessed by performing identification on EZbiocloud server to confirm the genetic similarity of *Streptomyces* sp. RA-WS2 and the results indicated that the isolate has closest similarity with *Streptomyces cyaneochromogenes,* followed by *Streptomyces aquilus* (Supplementary Figure S2) which has better morphological similarity with *Streptomyces* sp. RA-WS2. Therefore, biochemical characteristics were compared with *Streptomyces cyaneochromogenes* and *Streptomyces aquilus* (Table 2). The observations of biochemical characteristics were not enough to conclude the identification at species level. Thus, disagreement of the morphological, biochemical and genetic characteristics of the isolate, led us to use Whole genome sequencing approach for species level identification.

The whole genome sequencing performed on Novaseq 6000 provided 20.00100 million initial reads (submitted to SRA Archive with Accession ID SRX16786500). Out of these, 19.223154 million reads were filtered out on the basis of Phred score, (reads with Phred score 36 or above were selected). Table 3 demonstrates the parameters of assembled reads obtained from the Unicycler and Shovill programme. It was observed that Unicycler provided better assembly as compared to Shovill as indicated by N50 score with average depth coverage of 275 times. It was also observed that Shovill based assembly resulted in two times higher misassemblies, while 97.26% of reads could be properly paired to the assembly made by Unicycler. The draft assembly obtained using Unicycler was submitted in NCBI Genome database with Accession number JAMXPJ000000000. The size of the genome of the isolate *Streptomyces* sp. RA-WS2

Table 1 Comparison of morphological characteristics of *Streptomyces* sp. RA-WS2 with closely related taxa

Medium	Macroscopic characteristics				
	Characteristic colour of Mycelium	*Streptomyces* sp. RA-WS2 (Present work)	*Streptomyces cyaneochromogenes* (Tang et al. 2019)	*Streptomyces aquilus* (Li et al. 2020)	*Streptomyces justiciae* (Phongsopitanun et al. 2021)
ISP1	Aerial	White	NA	NA	NA
	Substrate	Brown	NA	NA	NA
ISP2	Aerial	White	Greyish-White	White	NA
	Substrate	Brown	Brown	Brown	NA
ISP3	Aerial	Off White	Glaucous-Grey	White	Greenish yellow/grey
	Substrate	Citrine to greyish yellow	Pink	Citrine	Dark greyish yellow
ISP4	Aerial	White	Greyish-White	White	NA
	Substrate	Light Buff	Creamy	Light Buff	NA
ISP5	Aerial	White	Light purple	White	NA
	Substrate	Buckthorn Brown	Perilla Purple	Buckthorn Brown	NA
ISP6	Aerial	Colourless	Cinnamon-Buff	Colourless	NA
	Substrate	Brown	Cinnamon-Buff	Brown	NA
ISP7	Aerial	White	White	White	NA
	Substrate	Brown	Black	Brown	NA
Microscopic characteristics					
Spore chain morphology		Cylindrical	Spiral and coiled	Long Straight	Cylindrical
Spore shape		Smooth	Smooth	Smooth	Smooth

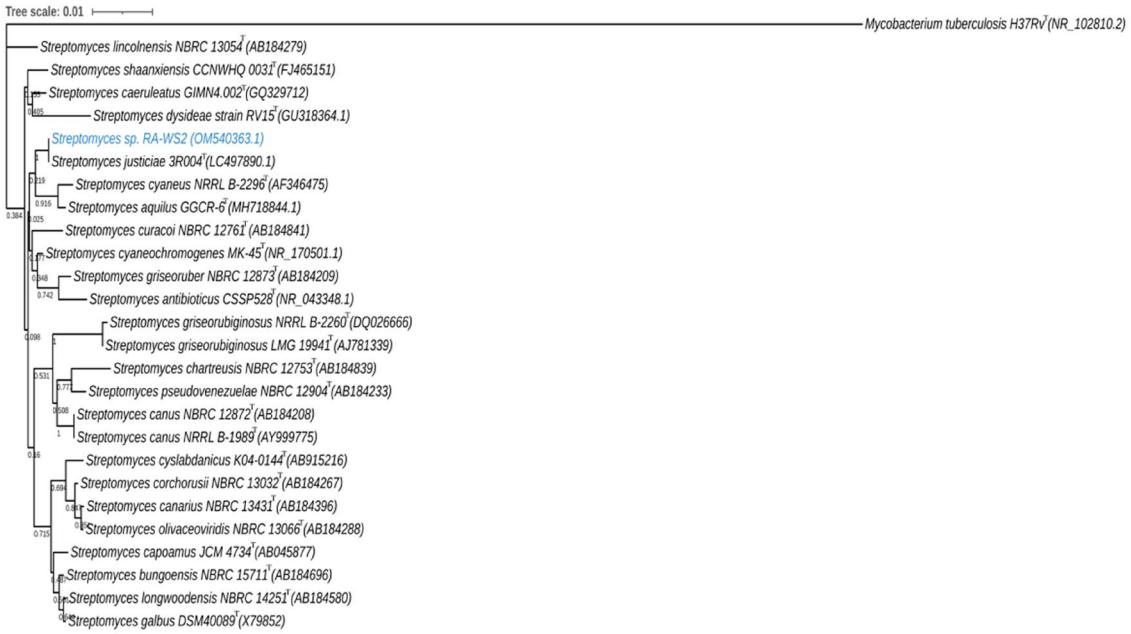

Fig. 3 Neighbour-joining phylogenetic tree based on 16S rRNA partial gene sequence obtained by Sanger sequencing. The relationship between closely related species of the genus *Streptomyces* Type cultures are shown. *Mycobacterium tuberculosis* H37Rv^T was used as an out group. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches in decimal form. The analysis performed using MEGA 11 demonstrates *Streptomyces justiciae* 3R004^T to be its closest neighbour

Table 2 Comparison of biochemical characteristics *Streptomyces* sp. RA-WS2 with closely related taxa

Characteristic feature	*Streptomyces* sp. RA-WS2 (Present work)	*Streptomyces cyaneochromogenes* (Tang et al. 2019)	*Streptomyces aquilus* (Li et al. 2020)	*Streptomyces justiciae* (Phongsopitanun et al. 2021)
Protease test				
Casein hydrolysis	+	NA	NA	+
Gelatin Liquification	+	W	+	+
Amylase test				
Starch hydrolysis	+	+	NA	NA
Assimilation of sole carbon sources (1%, w/v)				
Glucose	+	NA	NA	NA
Adonitol	+	NA	NA	NA
Arabinose	+	NA	NA	W
Lactose	+	NA	NA	NA
Sorbitol	+	NA	NA	NA
Mannitol	+	NA	+	NA
Rhamnose	+	NA	NA	NA
Sucrose	+	+	+	+
Xylose	+	NA	+	NA
Inositol	+	+	NA	NA
Assimilation of sole nitrogen sources (1%, w/v)				
Arginine	+	NA	+	NA
Melanin production				
Melanin	+	-	_	NA
% Salt tolerance (w/v)				
Growth in NaCl	10	6	7	4

Table 3 Quality assessment of assemblies prepared by Denovo assemblers (Unicycler and Shovill)

Reads before filtering	20.00100 million	
Reads after filtering	19.223154 million	
Assembler	Unicycler	Shovill
N50	272,356	175,592
Average coverage depth	275	225
Largest (bp)	704,158	728,680
Total (bp)	10,657,029	10,717,547
Misaassemblies	49	86
Local MA	56	52
Total contigs	76	156

was found to be around 10,657,029 bp. The DNA G+C content was 70.74 mol%. About 9604 protein coding genes were predicted from the assembled genome by Prokka package.

Phylogenetic analysis of the Whole genome assemblies obtained from Shovill, and Unicycler on TYGS server showed a dDDH similarity of 88.3% and 88.4% respectively, with the recommended formula (DDH4) against the closest genome of *Streptomyces justiciae* 3R004^T followed by *Streptomyces aquilus* GGCR-6 ^{T} with a DDH value of 54.3%, the digital DDH value thus indicated the strain *Streptomyces* sp. RA-WS2 to be a strain of *Streptomyces justiciae* as no other species had comparable DDH value (above 70%).

As depicted from the ANI of closely related genome (Supplementary Table S1), the highest similarity was again found with the reference genome of *Streptomyces justiciae* 3R004^T with an ANI of 98.65% having above 95% threshold values, which demarcates the species based on whole genome similarity. The AAI of various close species as tabulated in Supplementary Table S2 also indicated the closest genome of *Streptomyces justiciae* 3R004^T with 96.99% similarity. Therefore, the strain *Streptomyces* sp. RA-WS2 was again found to have closest similarity with *Streptomyces justiciae* based on dDDH, ANI as well as AAI values.

Further, phenotypic and biochemical characteristics of *Streptomyces* sp. RA-WS2 were compared with its genetically close *Streptomyces* species i.e. *Streptomyces justiciae* 3R004^T, *Streptomyces aquilus* GGCR-6 ^{T} and *Streptomyces cyaneochromogenes* MK-45 ^{T} (Table 1 and 2). Based on phenotypic as well as the genetic similarity, the isolate was identified as *Streptomyces justiciae* RA-WS2, however it has higher salt tolerance (upto 10%) as compared to *Streptomyces justiciae* 3R004^T which can grow upto 4% NaCl.

To ascertain the position of *Streptomyces* sp. RA-WS2, on the basis of 16S rRNA gene sequence (NCBI accession no. OP594900), the phylogenetic tree was constructed using complete 16S rRNA gene sequence of 1523 bp extracted from the genome of *Streptomyces* sp. RA-WS2 (Fig. 4). The results obtained were similar to the phylogenetic tree constructed using partial 16S rRNA gene sequence confirming its closest evolutionary relationship with *Streptomyces justiciae*, followed by *Streptomyces aquilus.*

To further assess the evolutionary history of the strain *Streptomyces* sp. RA-WS2 based on the proteins encoded

Fig. 4 Neighbour-joining phylogenetic tree based on 16S rRNA complete gene sequence, extracted from assembled genome. The relationship between closely related species of Type cultures from genus *Streptomyces* are demonstrated. *Mycobacterium tuberculosis* H37Rv^T was used as an out group. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches in decimal form. The analysis performed using MEGA 11 demonstrates *Streptomyces justiciae* 3R004^T to be its closest neighbour

in whole genome, phylogenomics analysis was performed. Various orthologous genes were extracted from genome data set using the Orthofinder package and Phylogenetic tree was constructed by STAG (Fig. 5). The phylogenomic analysis further indicated that *Streptomyces* sp. RA-WS2 was evolutionary closest to *Streptomyces justiciae* 3R004^T.

The secondary metabolite analysis by antiSMASH, revealed the presence of 31 BGCs in the genome of *Streptomyces justiciae* RA-WS2 as elaborated in Table 4. 100% similarity with the already known gene cluster for setomimycin biosynthesis in the genome of *Streptomyces justiciae* RA-WS2, further confirmed it as a potent setomimycin producer as reported earlier by Manhas et al. (2022a, b). The BGC for setomimycin biosynthesis in *Streptomyces justiciae* RA-WS2 has been elaborated in Fig. 6. Presence of phenazine type gene cluster similar to esmeraldin (20464 bp), terpene type gene cluster similar to vazabitide A (20444 bp), Type 1 Polyketide gene cluster similar to maduropeptin (42,419), Type 1 Polyketide type gene cluster similar to tetronasin (58936 bp), butyrolactone type gene cluster similar to scleric acid (19013 bp), Siderophore type gene cluster similar to paulomycin (12758 bp), Type 1 Polyketide type gene cluster similar to rustmicin (44625 bp), butyrolactone type gene cluster (10968 bp) and a NRPS lantipeptide gene cluster, similar to chejuenolide A /chejuenolide B (6454 bp) differentiated the strain *Streptomyces justiciae* RA-WS2 from *Streptomyces justiciae* 3R004^T.

The high salt tolerance (upto 10%) of the strain *Streptomyces justiciae* RA-WS2 can be attributed to the ion transporters and compatible solutes. Presence of BGC of Ectoine, a compatible salt in the genome of our strain *Streptomyces justiciae* RA-WS2 may be attributed to be one of the contributing factors responsible for the halotolerance. However, genetic sequence of the ectoine biosynthesis genes were found to be different from that of *Streptomyces justiciae* 3R004^T. The amino acid sequence of Ectoine ABC transporter protein sequence was also different at 67th position with substitution of arginine instead of histidine in *Streptomyces justiciae* RA-WS2 strain. Serine and aspartic acid at 37th and 39th position in *Streptomyces justiciae* 3R004^T were substituted by proline and glycine respectively, in Ectoine hydroxylase of *Streptomyces justiciae* RA-WS2. These mutations at genetic level might be responsible for enhanced halotolerance of the strain. Comparative BGC also revealed that *Streptomyces justiciae* RA-WS2 possesses Betaine (another osmoprotectant-related cluster), like in *Streptomyces justiciae* 3R004^T.

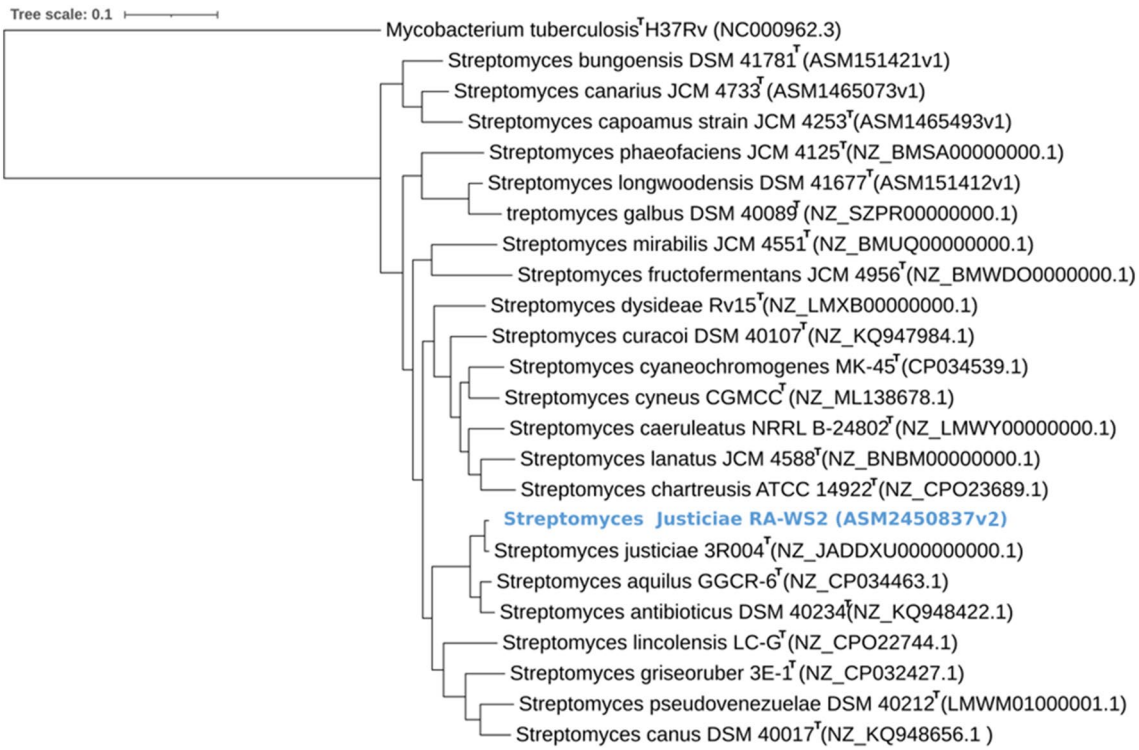

Fig. 5 Phylogenetic tree based on Whole genome sequence of *Streptomyces* sp. RA-WS2 using orthologous genes of closely related genomes. Various orthologs identified were used to construct a consensus phylogenetic tree based on STAG protocol through Orthofinder package Ver 2.5.4. Each multi copy gene was subjected to multiple alignment using MAFFT L-INS-I program and phylogenetic interference was done using IQ-Tree. *Mycobacterium tuberculosis* H37Rv^T was used as an out group

Table 4 Predicted Biosynthetic gene clusters (BGCs) related to biosynthesis of secondary metabolites in *Streptomyces justicae* RA-WS2

Region	Type	Genome Location		Most similar known cluster		Similarity(%)
		From	To			
Region 1	phenazine	206737	227201	esmeraldin	Polyketide + Other:Aminocoumarin	8
Region 2	T2PKS	653951	724444	spore pigment	Polyketide	83
Region 3	terpene	793213	813657	vazabitide A	NRP	4
Region 4	T1PKS	836286	878705	maduropeptin	Polyketide:Iterative type I + Polyketide:Enediyne type I	3
Region 5	melanin	1233379	1241462	melanin	Other	71
Region 6	NRPS,NRPS-like,T1PKS	1564418	1645306	coelichelin	NRP	100
Region 7	T3PKS	1821483	1861690	herboxidiene	Polyketide	9
Region 8	NAPAA	2409226	2443077	Not found	Not found	
Region 9	ectoine	2692342	2700596	ectoine	Other	100
Region 10	other,T1PKS	3268532	3327468	tetronasin	Polyketide	3
Region 11	lanthipeptide-class-i	3909215	3933422	Not found	Not found	
Region 12	melanin	4046061	4056669	melanin	Other	60
Region 13	siderophore	4185891	4196603	desferrioxamin B / desferrioxamine E	Other	83
Region 14	butyrolactone	4804858	4814910	Not found	Not found	
Region 15	T3PKS,NRPS,RiPP-like	4985061	5083319	Feglymycin	NRP	68
Region 16	ectoine	5431021	5441395	showdomycin	Other	35
Region 17	butyrolactone	5603298	5622311	scleric acid	NRP	17
Region 18	T2PKS,RRE-containing	6731865	6812659	setomimycin	Polyketide	100
Region 19	terpene	6930600	6951018	albaflavenone	Terpene	100
Region 20	siderophore	7599236	7609169	Not found	Not found	
Region 21	RiPP-like	7832427	7842843	Not found	Not found	
Region 22	terpene	7924718	7945579	geosmin	Terpene	100
Region 23	siderophore	8158263	8171021	paulomycin	Other	9
Region 24	terpene	8625644	8649824	hopene	Terpene	92
Region 25	RiPP-like,lanthipeptide-class-iii	9069318	9096498	informatipeptin	RiPP:Lanthipeptide	100
Region 26	siderophore	9110129	9124659			0
Region 27	terpene	9394857	9414540	ebelactone	Polyketide	5
Region 28	T1PKS,furan	9461748	9506373	rustmicin	Polyketide:Iterative type I	10
Region 29	butyrolactone	9556546	9567514	marineosin A / marineosin B	Polyketide	9
Region 30	NRPS,lanthipeptide-class-i,lanthipeptide-class-ii	9857615	9922155	chejuenolide A / chejuenolide B	Polyketide	7
Region 31	terpene	10267913	10290543	Not found	Not found	0

Fig. 6 Distribution of secondary metabolite gene clusters (1–31) in *Streptomyces justiciae* RA-WS2 as predicted by antiSMASH server (Bacterial version 6.0). The details of predicted BGC (1–31) are described in Table 4

In conclusion, phenotypic and genotypic characteristics of the strain *Streptomyces justiciae* RA-WS2 evidently indicate that the strain RA-WS2 is a novel strain of *Streptomyces justiciae*. The strain was further characterized as halotolerant actinomycete with tolerance upto 10% of NaCl.

Protologue

16S partial sequence NCBI GenBank OM540363. Genome accession number (GCF_024508375.2).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13205-023-03459-5.

Acknowledgements The authors are thankful to Director, IIIM for his interest. RSM and DC are thankful to ICMR and UGC for their respective fellowships. The manuscript bears institutional communication number CSIR-IIIM/IPR/00499.

Author contributions RSM performed all the wet lab experiments and wrote original manuscript, DC performed all the bioinformatics analysis and contributed in writing the manuscript, AC conceptualized, supervised, finalized the study and corrected the manuscript.

Funding The authors did not receive financial support for submitted work, RSM and DC are thankful for fellowship from ICMR and UGC respectively.

Data availability This whole-genome shotgun project has been deposited at GenBank under the accession number GCF_024508375.2. The raw reads are available in BioProject number PRJNA842579. The Biosample number is SAMN28688843. The version explained in current text is JAMXPJ000000000.2, ranging from JAMXPJ020000001-JAMXPJ020000076.

Declarations

Conflict of interest The authors declare that they have no conflict of interest in publication.

Ethical approval Not applicable.

Consent for publication All the authors have provided their consent for publication.

References

Afgan E, Baker D, Batut B et al (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379

Andrews S (2022) FastQC https://github.com/s-andrews/FastQC

Bentley SD, Chater KF, Cerdeño-Tárraga A-M et al (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–147. https://doi.org/10.1038/417141a

Berdy J (2005) Bioactive microbial metabolites. J Antibiot 58:1–26

Betz N, Strader T (2002) Clean-up with Wizard SV for gel and PCR

Blin K, Shaw S, Kloosterman AM et al (2021) antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49:W29–W35. https://doi.org/10.1093/nar/gkab335

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cole JR, Wang Q, Fish JA et al (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 42:D633–D642. https://doi.org/10.1093/nar/gkt1244

Emms DM, Kelly S (2018) STAG: Species Tree Inference from All Genes. biorxiv. https://doi.org/10.1101/267914

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238. https://doi.org/10.1186/s13059-019-1832-y

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Guo J, Rao Z, Yang T et al (2014) High-level production of melanin by a novel isolate of Streptomyces kathirae. FEMS Microbiol Lett 357:85–91

Ikeda H, Ishikawa J, Hanamoto A et al (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 21:526–531. https://doi.org/10.1038/nbt820

Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301

Li K, Guo Y, Wang J et al (2020) Streptomyces aquilus sp. nov., a novel actinomycete isolated from a Chinese medicinal plant. Int J Syst Evol Microbiol 70:1912–1917

Manhas RS, Tiwari H, Noor M et al (2022a) Setomimycin as a potential molecule for COVID-19 target: in silico approach and in vitro validation. Mol Diversity. https://doi.org/10.1007/s11030-022-10441-5

Manhas RS, Ahmad SM, Mir KB, et al (2022b) Isolation and anticancer activity evaluation of rare Bisaryl anthraquinone antibiotics from novel Streptomyces sp. strain of NW Himalayan region. Chemico-Biological Interactions https://doi.org/10.1016/j.cbi.2022.110093

Meier-Kolthoff JP, Göker M (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3

Parte A, Whitman WB, Goodfellow M, et al (2012) Bergey's manual of systematic bacteriology: volume 5: the Actinobacteria. Springer Science & Business Media

Phongsopitanun W, Kanchanasin P, Sripreechasak P et al (2021) Potential antibiotic production of Streptomyces justiciae sp. Nov., isolated from the root of Justicia subcoriacea. Int J Syst Evol Microbiol 71:005017

Pridham T, Gottlieb D (1948) The utilization of carbon compounds by some Actinomycetales as an aid for species determination. J Bacteriol 56:107–114
Ridgway R (1912) Color standards and color nomenclature. The author
Rodriguez-R LM, Konstantinidis KT (2014) Bypassing cultivation to identify bacterial species. Microbe 9:111–118
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
Seemann T (2022a) Shovill https://github.com/tseemann/shovill
Seemann T (2022b) Barrnap https://github.com/tseemann/barrnap
Shirling ET, Gottlieb D (1966) Methods for characterization of Streptomyces species. Int J Syst Bacteriol 16:313–340
Stanbury PF, Whitaker A, Hall S (2016) Principles of fermentation technology 3rd ed. Burlington MA
Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci 101:11030–11035
Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38:3022–3027
Tang X, Zhao J, Li K et al (2019) Streptomyces cyaneochromogenes sp. nov., a blue pigment-producing actinomycete from manganese-contaminated soil. Int J Syst Evol Microbiol 69:2202–2207
Verma VC, Gond SK, Kumar A et al (2009) Endophytic actinomycetes from Azadirachta indica A. Juss.: isolation, diversity, and antimicrobial activity. Microb Ecol 57:749–756
Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595
Yoon S-H, Ha S-M, Kwon S et al (2017a) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755
Yoon S-H, Ha S, Lim J et al (2017b) A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110:1281–1286. https://doi.org/10.1007/s10482-017-0844-4