Genome sequence of *Corynebacterium amycolatum* ICIS 99 isolated from human vagina reveals safety and beneficial properties

Irina V. Gladysheva1 · Yuriy A. Khlopko1 · Sergey V. Cherkasov1 · Vladimir Ya. Kataev1

Received: 19 December 2021 / Revised: 14 March 2022 / Accepted: 14 March 2022 / Published online: 29 March 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

**Abstract**

*Corynebacterium amycolatum* ICIS 99 was isolated from vaginal smears of healthy women and showed promising results in antimicrobial screenings. Here, we report the draft genome sequence of this strain and analyze its main features to assess its safety and useful properties. The genome is 2,532,503 bp long and contains 2186 CDSs with an average G + C content of 59.0%. Analyses of the ICIS 99 genome revealed the absence of true virulence factors. The genome contains genes involved in the synthesis of secondary metabolites and bacteriocins of the class sactipeptide. In the genome of ICIS 99, we identified a large number of genes responsible for adaptation and survival in the vaginal environment, including acid and oxidative stress resistance genes. The genomic information of ICIS 99 provides a basis for understanding the safety and useful properties of ICIS 99 and for considering it as a potential probiotic strain. The whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number JAIUSU000000000.

**Keywords** *Corynebacterium amycolatum* · Draft genome sequence · Secondary metabolites · Vaginal microbiome

**Introduction**

The genus *Corynebacterium* was described in 1896 by Lehmann & Neumann and includes straight or slightly curved rod-shaped, Gram-positive, catalase-positive, non-spore-forming, non-motile bacteria that belong to the family *Corynebacteriaceae* of the phylum *Actinobacteria* (Lehmann 2016). Currently, this genus contains 120 valid species (www.bacterio.net/corynebacterium.html). *Corynebacterium* are widespread in the environment and can be isolated from soil, water, and plant material. Certain species of corynebacteria are of great importance in industry and agriculture (Ikeda and Takeno 2013). Some *Corynebacterium* spp. are well-known pathogens of humans or animals; other species are part of the normal microbiota, although they may occasionally cause infections in humans or can be transmitted to humans by zoonotic contact (Tauch and Sandbote 2014).

*Corynebacterium amycolatum* is a typical bacterium of the microbiota on human skin and mucous membranes. It was first isolated by Collins and Burton from clinical specimens in 1988 (Collins et al. 1988). A distinctive feature of this species compared to other corynebacteria is the absence of typical mycolic acids in the cell membrane. According to the 16S rRNA gene phylogenetic analyses of different *Corynebacterium* species, *C. amycolatum* is only distantly related to other corynebacteria (Pascual et al. 1995).

In the literature, studies on *C. amycolatum* are limited, and this species is described as a causative agent of sepsis, endocarditis, meningitis, septic arthritis, and urinary tract infections (Reddy et al. 2012). However, the similarity of non-diphtheria corynebacteria often calls into question the identification of isolates from clinical specimens (Esteban et al. 1999).

In this study, we report the draft genome sequence of *C. amycolatum* ICIS 99, which was isolated from vaginal smears of healthy women during our pilot study aimed...
at investigating the biological properties of isolates of Corynebacterium spp. In a previous study, the strain was identified as Corynebacterium xerosis by biochemical tests of the Api Coryne system. Cell-free supernatants (CFSs) from this strain inhibited growth and biofilm formation and destroyed one-day-developed (grown for 24 h) opportunist pathogens, such as Escherichia coli, Staphylococcus aureus, Klebsiella pneumoniae and Pseudomonas aeruginosa (Gladysheva et al. 2017). A preliminary safety screening revealed that ICIS 99 did not show hemolytic, DNase, or gelatinase activities (data unpublished).

Therefore, the present study was undertaken to study the genome of this strain to assess its safety and useful properties.

Materials and methods

Bacterial strain and genomic DNA extraction

C. amycolatum ICIS 99 was isolated from the vaginal contents of healthy women and is part of the Collection of Microorganisms of the Institute for Cellular and Intracellular Symbiosis UrB RAS (Orenburg, Russia). The strains were grown in Luria–Bertani (LB) Miller broth (Becton Dickinson, USA) at 37 °C for 24 h and sub-cultured once in the same medium at 37 °C for 24 h before use. The strain morphologies were observed by scanning electron microscopy (Supplementary Fig. 1).

Genomic DNA was extracted using a Quick DNA Fun gal/Bacterial Kit (Zymo Research, USA). The quality of the extracted DNA was assessed according to the A260/280 ratio using a Nanodrop 8000 (Thermo Fisher Scientific, USA), and electrophoresis was performed in a 1% agarose gel. DNA concentration was quantified using a Qubit 4 Fluorometer and a dsDNA High Sensitivity Assay Kit (Life Technologies, USA).

Genome sequencing, assembly and annotation

The DNA library for whole-genome sequencing was prepared using a NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina® (New England BioLabs, USA). The library was validated using capillary electrophoresis on a QIAxcel Advanced System (Qiagen, Germany) using the QIAxcel DNA High Resolution Kit and normalized using qPCR on the CFX Connect Real-Time PCR System (Bio Rad, USA). Paired-end sequencing (2 × 250 bp) was carried out on a MiSeq platform (Illumina, USA) using Reagent Kit v.2 (Illumina, USA) in the Center of Shared Scientific Equipment “Persistence of microorganisms” at the Institute for Cellular and Intracellular Symbiosis UrB RAS.

The paired-end reads were filtered and trimmed with the Trimmomatic program (Bolger et al. 2014). De novo genome assembly was performed using the Unicycler v. 0.4.9b genome assembler (Wick et al. 2017). The contigs were analyzed for gene prediction using the NCBI Prokaryotic Genome Annotation Pipeline (PGPA) and rapid annotation using subsystem technology (RAST) (Aziz et al. 2008). RAST data visualization was performed using GraphPad Prism version 9.2.0 and GraphPad Software (La Jolla, CA, USA) (www.graphpad.com). The interactive visualization of the genome of Corynebacterium amycolatum ICIS 99 was performed with GView (Petkau et al. 2010).

Tentative taxonomic identification was carried out using the BLASTn 2.12.0+ program, according to the rRNA/ITS databases Targeted Loci Project. Summary results are presented in Supplementary Table 1.

Average nucleotide identity (ANI) values were calculated with the Orthologous Average Nucleotide Identity tool (Lee et al. 2016), a DDH estimation was performed as described previously (Meier-Kolthoff et al. 2013). The following were used for comparison: the complete reference genome of C. amycolatum FDAARGOS 1108 (CP068168); five draft genomes of C. amycolatum strains currently available from the NCBI database, including SK46 (WGS project: ABZU01), UMB0042 (PKHS01), ICIS 5 (SSOR01), ICIS 9 (MTPT01) and ICIS 53 (MIFV01); and complete reference genomes of the species Corynebacterium xerosis (CP032788), Corynebacterium lactis (CP006841), and Corynebacterium jeikeium (NZ_LS483459).

The bioinformatic tools antiSMASH 5.0 (Blin et al. 2021), NaPDoS (Ziemert et al. 2012) and BAGEL4 (Heel et al. 2018) were used to determine potential clusters of secondary metabolites with antimicrobial activity. Antibiotic-resistance genes in the genomes were predicted using the RGI (Resistance Gene Identifier) tool (Alcock et al. 2020). The presence of putative virulence genes in the genomes was investigated using the Virulence Factor of Bacterial Pathogens Database (VFDB) (Liu et al. 2019). The CRISPR regions were identified with the CRISPR online detection tool CRISPR finder (https://crispr.cas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) (Couvin et al. 2018).

Nucleotide sequence accession number

This whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAIUSU000000000. The version described in this paper is version JAIUSU010000000.
Results and discussion

General genomic features and genome-based phylogeny of ICIS 99

The genome characteristics of strain ICIS 99 are presented in Table 1 and Fig. 1. The draft genome of ICIS 99 was 2,532,503 bp long with an N50 length of 167,044 bp, an L50 of 6, and a G + C content of 59.0%. The final assembled genome consisted of 41 contigs. Genome annotation was performed using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok), and 2186 coding sequences, including 2156 proteins (CDSs), 30 pseudogenes, 3 complete rRNAs (5S, 16S, 23S) and 54 tRNAs, were identified.

The RAST subsystem annotated the coding proteins into 22 functional categories based on their COG classifications (Fig. 2). Most genes belonged to the categories protein metabolism (17.75%), amino acids and derivatives (16.43%), carbohydrates (12.78%), cofactors, vitamins, prosthetic groups, and pigments (9.63%), nucleosides and nucleotides (5.78%), respiration (4.87%) and fatty acids, lipids, and isoprenoids (4.77%).

ANI, OrthoANI and DDH analyses made it possible to definitively determine that ICIS 99 was a C. amycolatum strain. The ANI value between ICIS 99 and seven strains of C. amycolatum ranged from 94.83 to 97.94% (Fig. 3), and the OrthoANI values ranged from 71.73 to 77.05% (Supplementary Fig. 2). Digital DDH similarities showed from 79.60 to 81.90% identity between the ICIS 99 genome and ICIS 5, ICIS 9, ICIS 53 and UMB004 genomes, 59.30% identity between ICIS 99 and FDAARGOS 1108, and 61.30% identity between ICIS 99 and SK46. Between ICIS 99 and the remaining strains used for comparison: C. xerosis, C. lactis, and C. jeikeium, the digital DDH similarities were less 28.8% (Supplementary Table 2).

Unique genomic characteristics of ICIS 99

To assess the safety of this strain, we examined the genome for the presence of genes related to antibiotic resistance and virulence. The RGI tool was used to predict antibiotic-resistance genes in ICIS 99. In the genome of ICIS 99, only one gene was found to code for resistance to the antibiotic chloramphenicol.

VFDB software was used to predict virulence factors in ICIS 99. VFDB predicted 18 virulence factors in ICIS 99 (Supplementary Table 3). These virulence factors can be classified into ten categories: adherence, iron uptake, regulation, amino acid and purine metabolism, anti-phagocytosis, cell surface components, immune evasion, lipid and fatty acid metabolism, protease and secretion system. However, none of the genes identified were true virulence factors. These genes encode only proteins responsible for the structural and functional characteristics of microorganisms of the genus Corynebacterium or ‘niche factors’ (Swierczynski and Ton-That 2006; Wennerhold and Bott 2006; Tauch and Burkovski 2015; Baumgart et al. 2016; Bush 2018). True virulence factors, such as toxin-related genes (diphtheria toxin and phospholipase D) and hemolysin-related genes characteristic of well-known pathogenic corynebacteria, were not identified.

The previously described antibacterial activity of CFSs from ICIS 99 suggests that this strain contains genes responsible for the synthesis of secondary metabolites and bacteriocins. The presence of secondary metabolic gene clusters was examined by the antiSMASH 5.0 platform (Blin et al. 2021) and NaPDoS (Ziemert et al. 2012). AntiSMASH 5.0 revealed 4 putative biosynthetic gene clusters, which were predicted to be related to the biosynthesis of different types of secondary metabolites, including terpene, T3pks (type III polyketide synthases), NAPAA and NRPS (nonribosomal peptide) (Supplementary Table 4). Two gene clusters showed low similarity (≤ 10%) with the most similar clusters, of which T3pks gene clusters were associated with the biosynthesis of meroclorin A–D-like compounds and the

---

Table 1 General genome features of Corynebacterium amycolatum ICIS 99

| Feature           | Value          |
|-------------------|----------------|
| Coverage          | 375 ×          |
| Number of contigs | 41             |
| N50               | 167,044        |
| L50               | 6              |
| Depth of coverage | 375            |
| Genome size (b.p.)| 2,532,503      |
| GC contents (%)   | 59.0           |
| Genes (total)     | 2246           |
| CDSs (total)      | 2186           |
| Genes (coding)    | 2156           |
| CDSs (with protein)| 2156          |
| rRNAs (5S, 16S, 23S)| 3 (1, 1, 1)  |
| Complete rRNAs   | 1, 1, 1 (5S, 16S, 23S) |
| tRNAs             | 54             |
| Pseudo Genes (total)| 30            |
| BioProject        | PRJNA339674    |
| BioSample         | SAMN21545782   |
| GenBank accession no. | JAIUSU000000000 |

OrthoANI value ranged from 94.85 to 97.94% (Supplementary Fig. 2). Relative to other Corynebacterium strains (C. xerosis (CP032788), C. lactis (CP006841), and C. jeikeium (NZ_LS483459), the ANI values ranged from 73.87 to 76.42% (Fig. 3), and the OrthoANI values ranged from 71.73 to 77.05% (Supplementary Fig. 2). Digital DDH similarities showed from 79.60 to 81.90% identity between the ICIS 99 genome and ICIS 5, ICIS 9, ICIS 53 and UMB004 genomes, 59.30% identity between ICIS 99 and FDAARGOS 1108, and 61.30% identity between ICIS 99 and SK46. Between ICIS 99 and the remaining strains used for comparison: C. xerosis, C. lactis, and C. jeikeium, the digital DDH similarities were less 28.8% (Supplementary Table 2).
Nrps gene cluster was associated with the biosynthesis of phthoxazolin-like compounds. The other two gene clusters did not resemble any registered gene clusters. NaPDoS revealed one putative gene cluster, T2pks (type II polyketide synthase), that was involved in the biosynthesis of actinorhodin (similarity 27%). The existence of clusters of secondary metabolic genes indicates that the ICIS 99 strain can be a producer of new natural bioactive products.

To search for genes involved in the synthesis of potential bacteriocins, we used BAGEL4. The genome of strain ICIS 99 had one area of interest (AOI) that included the class sactipeptide (Fig. 4 and Supplementary Table 5). Sactipeptides are a new class of ribosomally synthesized and post-translationally modified peptides (RiPPs) and are a growing class of natural products that have garnered substantial attention because of their structural diversity and biological activities (Balty et al. 2019). The presence of an AOI encoding sactipeptide in the genome of strain ICIS 99 indicates that strain may produce the sactipeptide.

In the genome of ICIS 99, we identified a large number of genes coding for proteins involved in the stress response. These stresses include pH, osmotic pressure, nitrogen stress and oxidative stress. The presence of these genes determines the ability of ICIS 99 to adapt to and colonize the vaginal biotope of healthy women, which is a fairly aggressive environment for most pathogenic and opportunistic microorganisms. The detailed analysis of
Fig. 2 Subsystem category distribution of major protein-coding genes of *Corynebacterium amycolatum* ICIS 99 as annotated by the RAST annotation server. The bar chart shows the subsystem coverage in percentage. The pie chart shows percentage distribution of the 22 most abundant subsystem categories.

Fig. 3 Heatmap generated with Original ANI values calculated from the OAT software.
genes coding for proteins involved in stress response in the genome of ICIS 99 is shown in Supplementary Table 6. The genome of ICIS 99 was analyzed for the occurrence and diversity of CRISPR-Cas systems. The identification of CRISPR loci suggests the presence of a prokaryotic immune system, which confers resistance to foreign genetic elements and provides a form of acquired immunity (Amitai and Sorek 2016). The ICIS 99 genome contained three CRISPR loci (CRISPR1–CRISPR3) (Supplementary Table 7). The detected CRISPR1/CRISPR-associated (Cas) system was type IE (Cas5, Cas7, Cse2, Cas6, Cas3, Cas1, and Cas2).

In conclusion, the genomic information for ICIS 99 provides the basis for understanding the safety and useful properties of ICIS 99 and for considering its use as a potential probiotic strain. However, further research is needed to test its probiotic efficacy in vivo.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1093/NAR/GKZ935.

Acknowledgements The authors are grateful to Marina Ignatenko, employee of the Center for Revealing and Support of Gifted Children “Gagarin”, Orenburg region, Russia, for conducting scanning electron microscopy of the samples.

Author contributions IVG designed a study. VYK carried out the experimental design. YAK carried out bioinformatic analysis. IVG and SVC analyzed data and prepared the manuscript. All authors have read and approved the final manuscript.

Funding This study was conducted as part of a collaborative research effort between the Center for Revealing and Support of Gifted Children “Gagarin” and the Institute of Cellular and Intracellular Symbiosis, Ural Branch of the Russian Academy of Sciences, Orenburg, Russia. No funding was received for conducting this study.

Availability of data and materials All data generated or analyzed during this study are included in this published article and its Supplementary Information files. Additional data are available from the corresponding author upon request.

Fig. 4 Sactipeptide encoded by genes represented by orf00016 predicted in ICIS 99 genome with BAGEL4

Declarations

Conflict of interest The authors have no financial conflicts of interest to declare.

References

Alcock BP, Raphenya AR, Lau TTY et al (2020) CARD 2020: anti-biotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/NAR/GKZ935

Amitai G (2016) Sorek R (2016) CRISPR–Cas adaptation: insights into the mechanism of action. Nat Rev Microbiol 14(2):67–76. https://doi.org/10.1038/nrmicro.2015.14

Aziz RK, Bartels D, Best A et al (2008) The RAST server: rapid annotations using subsystems technology. BMC Genom 9:1–15. https://doi.org/10.1186/1471-2164-9-75/TABLES/3

Balty C, Guillot A, Fradale L et al (2019) Ruminococcin C, an anticrostidial sactipeptide produced by a prominent member of the human microbiota Ruminococcus gnavus. J Biol Chem 294:14512–14525. https://doi.org/10.1074/JBC.RA119.009416

Baumgart M, Schubert K, Bramkamp M, Frunzke J (2016) Impact of LytRCpsA-Psr proteins on cell wall biosynthesis in Corynebacterium glutamicum. J Bacteriol 198:3045–3059. https://doi.org/10.1128/JB.00406-16/SUPPL_FILE/ZJB999904215S01.PDF

Blin K, Shaw S, Klooosterman 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

Bolger AM, Lobse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Bush MJ (2018) The actinobacterial WhiB-like (Wbl) family of transcription factors. Mol Microbiol 110:663–676. https://doi.org/10.1111/MML.14117

Collins MD, Burton RA, Jones D (1988) Corynebacterium amylolatum sp. nov. a new mycolic acid-less Corynebacterium species from human skin. FEMS Microbiol Lett 49:349–352. https://doi.org/10.1111/j.1574-6968.1988.tb02755.x

Couvin D, Bernheim A, Toffano-Nioche C et al (2018) CRISPRCas-Finder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 46:W246–W251. https://doi.org/10.1093/NAR/GKY425
Esteban J, Nieto E, Calvo R et al (1999) Microbiological characterization and clinical significance of Corynebacterium amycolatum strains. Eur J Clin Microbiol Infect Dis 18:518–521. https://doi.org/10.1007/S100960050336

Gladysheva IV, Cherkasov SV, Khlopko Y (2017) Antibacterial activities of metabolites from Corynebacterium spp. strains isolated from the reproductive tract of a healthy woman against human pathogenic bacteria. Int J Pharma Bio Sci. https://doi.org/10.22376/IJPBS.2017.8.3.B549-556

Ikeda M, Takeno S (2013) Amino acid production by Corynebacterium glutamicum. In: Yukawa H, Inui M (eds) Corynebacterium glutamicum Microbiology monographs, vol 23. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29857-8_4

Lee I, Kim YO, Park SC, Chun J (2016) OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/IJSEM.0.000760/CITE/REFWORKS

Lehmann KB (2016) Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik. https://doi.org/10.5962/BHL.TITLE.117384

Liu B, Zheng D, Jin Q et al (2019) VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res 47:D687–D692. https://doi.org/10.1093/NAR/GKY1080

Meier-Kolthoff JP, Auch AF, Klenk HP et al (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform 14:60. https://doi.org/10.1186/1471-2105-14-60

Pascual C, Lawson PA, Farrow JAE et al (1995) Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences. Int J Syst Bacteriol 45:724–728. https://doi.org/10.1099/0027713-45-4-724

Petkau A, Stuart-Edwards M, Stothard P, van Domselaar G (2010) Interactive microbial genome visualization with GView. Bioinformatics 26:3125–3126. https://doi.org/10.1093/bioinformatics/btq588

Reddy BS, Chaudhury A, Kalawat U et al (2012) Isolation, speciation, and antibiogram of clinically relevant non-diphtherial Corynebacteria (Diphtheroids). Indian J Med Microbiol 30:52–57. https://doi.org/10.4103/0255-0857.93033

Swierczynski A, Ton-That H (2006) Type III pilus of corynebacteria: pilus length is determined by the level of its major pilin subunit. J Bacteriol 188:6318. https://doi.org/10.1128/JB.00606-06

Tauch A, Burkovski A (2015) Molecular armory or niche factors: virulence determinants of Corynebacterium species. FEMS Microbiol Lett. https://doi.org/10.1093/FEMSLE/FNV185

Tauch A, Sandbote J (2014) The family Corynebacteriaceae. Prokaryotes Actinobacteria. https://doi.org/10.1007/978-3-642-30138-4_187

Van Heel AJ, De Jong A, Song C et al (2018) BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res 46:W278–W281. https://doi.org/10.1093/NAR/GKY383

Wennerhold J, Bott M (2006) The DtxR regulon of Corynebacterium glutamicum. J Bacteriol 188:2907–2918. https://doi.org/10.1128/JB.188.8.2907-2918.2006/SUPPL_FILE/SUPPLEMENTAL_MATERIAL_WENNERHOLD_AND_BOTT.PDF

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

Ziemert N, Podell S, Penn K et al (2012) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS ONE 7:e34064. https://doi.org/10.1371/JOURNAL.PONE.0034064

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.