

The Kocurious case of Noodlococcus: genomic insights into *Kocuria rhizophila* from characterisation of a laboratory contaminant

Gregory E. McCallum^{1,2,†}, Siu Fung Stanley Ho^{1,3,†}, Elizabeth A. Cummins^{1,4}, Alex J. Wildsmith¹, Ross S. McInnes¹, Christoph Weigel⁵, Lok Yee Sylvia Tong⁶, Joshua Quick¹, Willem van Schaik¹ and Robert A. Moran^{1,7,*}

Abstract

The laboratory contaminant strain Noodlococcus was named for its coccoid cells and unusual colony morphology, which resembled a pile of noodles. Along with laboratory characterisation and electron microscopy, we generated a complete Noodlococcus genome sequence using Illumina and Oxford Nanopore data. The genome consisted of a single, circular, 2,732,108 bp chromosome that shared 97.5% average nucleotide identity (ANI) with the *Kocuria rhizophila* type strain TA68. We identified genomic features involved in replication (*oriC*), carotenoid synthesis (*crt*) and genome defence (CRISPR-Cas) and discovered four novel mobile elements (*ISKrh4-7*). Despite its environmental ubiquity and relevance to food production, bioremediation and human medicine, there have been few genomic studies of the *Kocuria* genus. We conducted a comparative, phylogenetic and pangenomic examination of all 257 publicly available *Kocuria* genomes, with a particular focus on the 56 that were identified as *K. rhizophila*. We found that there are two phylogenetically distinct clades of *K. rhizophila*, with within-clade ANI values of 96.7–100.0% and between-clade values of 89.5–90.4%. The second clade, which we refer to as *Kocuria pseudorhizophila*, exhibited ANI values of <95% relative to TA68 and should constitute a separate species. Delineation of the two clades would be consistent with the rest of the genus, where all other species satisfy the 95% ANI threshold criteria. Differences in the *K. rhizophila* and *K. pseudorhizophila* pangenomes likely reflect phenotypic as well as evolutionary divergence. This distinction is relevant to clinical and industrial settings, as strains and genomes from both clades are currently used interchangeably, which may lead to reproducibility issues and phenotype–genotype discordance. Investigating an innocuous laboratory contaminant has therefore provided useful insights into the understudied species *K. rhizophila*, prompting an unexpected reassessment of its taxonomy.

Impact Statement

Bacterial genome sequence databases are dominated by a relatively small number of medically relevant genera, while most of the global bacterial population's diversity is largely uncharacterised. *Kocuria* is a widespread bacterial genus with industrial

[Continued on next page]

Received 11 June 2025; Accepted 16 September 2025; Published 06 October 2025

Author affiliations: ¹Institute of Microbiology and Infection and Department of Microbes, Infection and Microbiomes, School of Infection, Inflammation and Immunology, College of Medicine and Health, University of Birmingham, Birmingham B15 2TT, UK; ²Department of Evolution, Ecology and Behaviour, Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK; ³Department of Microbiology, School of Clinical Medicine, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR; ⁴Ineos Oxford Institute for Antimicrobial Research, Department of Biology, University of Oxford, Oxford, UK; ⁵Institute of Biotechnology, Technical University of Berlin, Ackerstraße 76, 13355 Berlin, Germany; ⁶Department of Infectious Diseases and Public Health, City University of Hong Kong, Kowloon, Hong Kong SAR; ⁷School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia.

*Correspondence: Robert A. Moran, robert.moran@sydney.edu.au

Keywords: contaminant; genomics; *Kocuria rhizophila*; pangenome; taxonomy.

Abbreviations: ANI, average nucleotide identity; BHI, brain heart infusion; ContamClub, Contamination Club; IS, insertion sequence; NCBI, National Centre for Biotechnology Information; OD₆₀₀, OD at 600 nm; ONT, Oxford Nanopore Technologies; QC, quality control; SEM, scanning electron microscopy; TSB, tryptic soy broth; TSDs, target site duplications.

†These authors contributed equally to this work

Five supplementary figures and three supplementary tables are available with the online version of this article. All supporting data, code and protocols have been provided within the article or through supplementary data files.

001526 © 2025 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

and medical relevance. Despite its ubiquity, only 22 complete and 235 draft *Kocuria* genomes were publicly available at the outset of this study. Our phylogenetic and pangenomic examination of all available *Kocuria* genomes was the first for this genus, providing insights into its diversity and taxonomy. Most notably, we found that *Kocuria rhizophila* is comprised of two clades that are sufficiently divergent to constitute different species but are frequently used interchangeably in experimental and genomic research. The complete, high-quality Noodlococcus genome generated and characterised here can serve as a reference for true *K. rhizophila*, particularly while there is only a draft genome sequence available for type strain TA68.

DATA SUMMARY

Sequencing reads and the assembled Noodlococcus genome are available from NCBI BioProject accession PRJNA835814 and BioSample accession SAMN28111796. The complete sequence of the Noodlococcus chromosome can be found in the GenBank nucleotide database under accession number CP097204.1. Entries for the novel insertion sequences ISKrh4 to ISKrh7 can be found in the ISFinder database (<https://isfinder.biotoul.fr>).

INTRODUCTION

Kocuria is a bacterial genus in the family *Micrococcaceae* that was first distinguished from *Micrococcus* in 1995 [1]. As of September 2024, there were 27 *Kocuria* species listed in the National Centre for Biotechnology Information (NCBI) Taxonomy Browser, 22 of which were represented by genome sequences. All *Kocuria* are Gram-positive cocci that arrange in mixtures of short chains and disordered clusters, forming colonies that are usually pigmented [1, 2], with yellow [3], orange [4] and pink [5] colonies described. *Kocuria* species have been isolated from diverse terrestrial and marine environments, and strains have exhibited a wide array of phenotypes, including tolerance of cold temperatures, radiation, pH extremes or salinity [4, 6–10].

Kocuria has been found as a contaminant in clinical and research laboratories [11], likely because of its environmental ubiquity and presence on human skin. Mistaking *Kocuria* in clinical samples for contaminants can lead to misdiagnoses in rare cases where *Kocuria* causes human infections [12]. Reported *Kocuria* infections have included cases of peritonitis [13, 14], bacteraemia [15–18], meningitis [19], endocarditis [20–22], postsurgical infections [23] and infections of various other body sites [2, 24]. Infections are typically responsive to antimicrobial chemotherapy as *Kocuria* are generally susceptible to antibiotics, apart from their intrinsic resistance to nitrofurantoin and furazolidone [2, 3, 5].

Kocuria rhizophila was named in 1999 following characterisation of the type strain TA68, which was isolated in 1995 from the rhizosphere of *Typha angustifolia* (narrowleaf cattail) growing on a floating mat in the Soroksár tributary of the Danube River, Hungary [3]. *K. rhizophila* has since been isolated from a wide range of terrestrial and aquatic environments such as permafrost [25], peat soil [26], waterfalls [27] and deep-sea sponges [28]. It has been found on food products [29], and in close association with humans [30], plants [3, 7, 31, 32] and animals [33–35]. Strains of *K. rhizophila* have been used routinely as controls for antibiotic sensitivity testing [36].

Despite its ubiquity and relevance to human health, there is a sparsity of *Kocuria* genome data in public databases. The first complete *Kocuria* genome was the soil-derived *K. rhizophila* strain DC2201, which was published in 2008 [37]. As of September 2024, there were 257 *Kocuria* genomes in the NCBI database, 22 of which were complete (Table S1, available in the online Supplementary Material). Complete genome sequences have revealed that *K. rhizophila* has one of the smallest actinobacterial chromosomes, ranging from 2.6 to 2.8 million base pairs [33, 37]. Given their vast potential for diverse metabolite production, the need for expanded knowledge of actinobacterial genomics has been expressed recently [38]. While the number of publicly available *Kocuria* genomes has been increasing over recent years, we are aware of only a single study that has performed comparative genomic analysis of six *Kocuria* genome sequences from four species [39].

In August 2019, we found an unusual colony growing on an agar plate that had been left on a laboratory bench for 11 days. The colony was raised, yellow-pigmented and looked like a pile of noodles. After a Gram stain revealed that it was a Gram-positive coccus, we named the strain ‘Noodlococcus’ and sought to identify it. Generating a complete Noodlococcus genome sequence revealed that it was a strain of *K. rhizophila*. We characterised the Noodlococcus genome and used it as a vehicle for comparative analyses of *Kocuria*, focussing particularly on the phylogeny and pangenome of *K. rhizophila*. Our investigation uncovered previously unrecognised diversity in *K. rhizophila*, with consequences for the classification of widely used type strains and genome sequences.

METHODS

Media, isolation and culture conditions

The original Noodlococcus colony was found on a brain heart infusion (BHI) agar plate that had been used to culture *Enterococcus faecium* in a laboratory at the University of Birmingham, UK. An overnight *E. faecium* culture in BHI broth had been diluted and spread on the BHI plate to obtain single colonies. After overnight incubation at 37 °C, *E. faecium* colonies were picked using

sterile pipette tips and transferred to new agar plates. The plate that colonies had been picked from was left in ambient conditions on a laboratory bench for 11 days before Noodlococcus and re-grown *E. faecium* colonies were observed. Noodlococcus was subsequently re-streaked onto a BHI agar plate and incubated at 37 °C overnight. A single colony was then picked and grown in BHI broth at 37 °C with shaking at 200 r.p.m. Gram staining was performed on a single phase of growth as described by Coico [40].

A variety of conditions were tested to determine the optimal growth conditions for Noodlococcus. Growth in tryptic soy broth (TSB), BHI broth, lysogeny broth and nutrient broth was compared by adding 5 µl of overnight culture to 100 ml of broth in 500 ml conical flasks. The flasks were then incubated for 24 h at 30 °C with shaking at 200 r.p.m. Aggregated Noodlococcus cells were broken up by pipetting before the OD at 600 nm (OD₆₀₀) was measured from 1 ml of culture. Three technical replicates were used for each condition.

To determine the optimal temperature for growth, culture turbidity was measured at 4, 10, 15, 20, 25, 28, 30, 37, 40, 45 and 50 °C. A single Noodlococcus colony was picked and added to 50 µl TSB. After mixing by pipetting, 5 µl of this inoculum was added to 6 ml TSB in a 30 µl universal container for each temperature condition (in biological triplicate, each with 3 technical replicates). OD₆₀₀ measurements were taken for 1 ml of the samples to establish a baseline reading. Samples were incubated with shaking at 200 r.p.m. After 24 h, cultures were mixed by pipetting to break up cell aggregates, and OD₆₀₀ was measured again. Growth in a range of NaCl concentrations (0, 1, 2, 3, 5, 7, 10, 15 and 20% weight/volume) and pH conditions (pH 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12) was determined in the same way in TSB at 30 °C. For salinity and pH tests, OD₆₀₀ was also measured after 96 h of growth. To evaluate growth in anaerobic conditions, Noodlococcus was streaked onto TSB agar plates and incubated for 96 h in a MACS MG-500 Anaerobic Chamber Workstation (Don Whitley) set to 37 °C and connected to a 5% carbon dioxide, 5% hydrogen/nitrogen (anaerobic) cylinder (BOC). For all subsequent experiments, Noodlococcus was cultured in TSB at 30 °C with 200 r.p.m. shaking, and on TSB agar plates at 30 °C, unless stated otherwise.

Electron microscopy

Overnight Noodlococcus culture was fixed in 2.5% glutaraldehyde for 30 min at 4 °C. Scanning electron microscopy (SEM) was conducted with a SEM-Zeiss EVO15 VP ESEM microscope at the Centre for Electron Microscopy at the University of Birmingham.

Biochemical analysis

Noodlococcus was sent to DSMZ (Braunschweig, Germany) for fatty acid composition analysis. An oxidase test was carried out by smearing a colony onto an oxidase test strip (Merck) and observing for a change in colour after 10 s. *Pseudomonas aeruginosa* PAO1 was used as a positive control, and *Escherichia coli* DH5α as a negative control. A catalase test was performed by smearing a colony onto sterile glass, adding one drop of 3% hydrogen peroxide and observing for the production of bubbles. *E. coli* DH5α was used as a positive control, and *E. faecium* 64/3 as a negative control.

Antibiotic susceptibility testing

The susceptibility of Noodlococcus to ampicillin, cefotaxime, ceftazidime, ciprofloxacin, colistin, erythromycin, nitrofurantoin and tetracycline was measured using the broth microdilution method [41] and interpreted against the EUCAST non-species-specific PK/PD breakpoints [42]. Where breakpoints were listed as ‘insufficient evidence’, breakpoints for *Staphylococcus* spp. were used as a surrogate. Assays were performed in biological triplicate, and the mode of the three replicates was recorded.

DNA extraction

A total of 1 ml of overnight culture was homogenised by vigorous vortexing for 1 min, and DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega) using their Gram-positive protocol with the inclusion of lysozyme (10 mg ml⁻¹; Sigma-Aldrich). DNA concentrations were quantified using the Qubit dsDNA BR assay kit (Thermo Fisher). DNA quality was assessed using a NanoDrop 2000 spectrophotometer.

Genome sequencing and assembly

Short-read DNA libraries were prepared with the Nextera XT library prep kit (Illumina). Shotgun sequencing was carried out by MicrobesNG (Birmingham, UK) using the HiSeq 2500 sequencing platform (Illumina) which generated 150 bp paired-end reads. Fastp v0.23.2 [43] was used to trim adapter sequences and to remove both low-quality and duplicate reads (--dedup).

Long-read sequencing libraries were constructed with the Ligation Sequencing Kit SQK-LSK109 [Oxford Nanopore Technologies (ONT)] with the following minor adjustments. An additional Ampure bead clean-up was carried out before DNA repair and end prep to improve ligation efficiency. To further increase ligation efficiency, incubation times for end repair, dA-tailing and ligation were increased to 30 min [44]. Long-read libraries were sequenced on a GridION (ONT) using a FLO-MIN106D R9.4.1 flow cell (ONT), MinKNOW 5.0.5 (ONT) and a 72 h run script with active channel selection enabled. The sequencing signal was

basecalled using Guppy v6.0.1 (ONT) super accuracy mode (--chunk-size 3000). The bottom 5% of reads by quality score and reads less than 1 kb were removed with Filtlong v0.2.1 [45].

Long-read assembly was performed using the consensus assembler Tricycler v0.5.3 [46]. Briefly, reads were subset into 12 different samples which were fed into three assemblers: Flye v2.9 [47], Miniasm+Minipolish v0.3, v0.1.3 [48, 49] and Raven v1.6.0 [50] (4 samples each). Contigs from all three assemblies were clustered, reconciled, aligned and partitioned to generate a consensus assembly. The accuracy of the consensus assembly was increased with Medaka v1.0.6 [51]. Finally, the assembly was polished with the quality-controlled Illumina reads using PolyPolish v0.5 [52] and then POLCA (MaSuRCA suite v4.0.9) [53]. The genome was annotated using Bakta v1.9.4 [54].

Genome feature identification and annotation

CRISPR repeats were identified manually in the region downstream of Cas genes. Replication origin prediction was performed as outlined previously [55]. Insertion sequences (IS) were found by sequentially comparing 100 kb segments of the *Noodlococcus* chromosome to its entire sequence in order to identify regions of >500 bp that occurred at multiple positions and shared >99% nucleotide identity. These regions were checked for the presence of transposase genes, and putative transposase sequences were used to query the ISFinder database [56]. IS were assigned to families based on transposase identities, with inverted repeats and target site duplications (TSDs) identified manually when relevant. The complete genome of *K. rhizophila* 28R2A-20 (GenBank accession CP072262 [28]) was used to identify naïve insertion positions that did not contain IS.

Phylogenetic analysis

All published *Kocuria* genomes ($n=257$) were downloaded using the NCBI Datasets platform on 21 September 2024. Genome completeness and contamination were assessed using CheckM2 [57]. Genomes with <90% completeness and/or >5% contamination were filtered out. Average nucleotide identities (ANIs) were determined by FastANI v1.33 [58]. Each genome was annotated using Bakta v1.9.4 [54]. Panaroo v1.5.0 [59] (--clean-mode moderate) was used to generate a core-genome alignment of the genomes, including the *Noodlococcus* genome, and IQ-Tree v2.3.6 [60] was then used to infer a maximum likelihood phylogenetic tree using the best-fit nucleotide substitution model (GTR+F+I+R5 for core-genome phylogenies) as determined by ModelFinder [61], and 1,000 ultrafast bootstrap replicates [62]. A core-genome alignment was generated (--clean-mode set to strict) with all the genomes that were monophyletic to the *K. rhizophila* strain TA68 or NCTC8340. Two strains of *Kocuria tytonicola* (strains 473 and DSM 104133) were included in this alignment for use as an outgroup. Representative full-length 16S rRNA genes, annotated by Bakta, were extracted from the genomes that passed CheckM2 quality control (QC) and aligned with MAFFT's [63] G-INS-i algorithm before a phylogeny was inferred as described above. Genomes containing only partial 16S rRNA genes were removed from this analysis. Phylogenetic trees were visualised using TreeViewer v2.2.0 [64]. A heatmap of the ANI results was generated using the pheatmap R package v1.0.12 [65].

ANI comparisons

To compare intra- and inter-species ANI values, all genomes that passed CheckM2 QC ($n=230$) were first classified using Genome Taxonomy Database and associated taxonomic classification toolkit (GTDB-Tk) v2.3.2 [66] classify workflow using release 214 of the GTDB-Tk reference package [67]. GTDB-Tk adds alphabetic suffixes to the end of genus or species names if the classification is ambiguous; however, for the sake of grouping the *Kocuria* genomes into species for ANI comparisons, these suffixes were removed. Genomes classified outside of the *Kocuria* genus ($n=13$) and genomes only classified as *Kocuria* to the genus level ($n=16$) were discarded. Any *Kocuria* species made up of less than three genomes ($n=6$ species out of 16: *Kocuria coralli*, *Kocuria dechangensis*, *Kocuria polaris*, *Kocuria soli*, *K. tytonicola* and *Kocuria tytonis*) were also discarded ($n=9$ genomes), leaving a total of $n=192$ genomes classified as a *Kocuria* species with >2 other genomes sharing the same species classification. ANIs for these genomes calculated using FastANI were used to make intra- and inter-species comparisons.

Pangenome analysis

A pangenome was generated from the 51 *K. rhizophila* genomes using Panaroo (--clean-mode strict) with a 98% sequence identity threshold. Panaroo's default definitions of core ($99 \leq x \leq 100\%$), soft core ($95 \leq x \leq 99\%$), shell ($15 \leq x \leq 95\%$) and cloud ($0 \leq x \leq 15\%$) were used. The twilight analysis package [68] was used to perform population structure-aware gene classification between the two clades. Functional annotation was performed by eggNOG-mapper v2.0 [69] with the eggNOG v5 database [70]. Prophages were detected using geNomad v1.11.0 [71], with quality and completeness determined using CheckV v1.0.3 [72].

RESULTS

Morphological characteristics of *Noodlococcus*

The original *Noodlococcus* colony was round, raised and sulphur yellow. It measured ~9 mm across and 2 mm tall, with a complex secondary structure that resembled a pile of noodles (Fig. 1a, b). Subsequent streaking on BHI agar produced small

(1 mm) yellow colonies after incubation at room temperature or 37 °C for 24 h, which developed a raised central ring structure when left at room temperature for seven days (Fig. S1), and took between two and three weeks to form secondary structures that resembled the original colony. In BHI broth incubated with shaking (200 r.p.m.) at 37 °C overnight, *Noodlococcus* produced a single non-diffuse colony-like structure.

Gram staining revealed that *Noodlococcus* was a Gram-positive coccus arranged in short chains or irregular clusters (Fig. 1c). SEM showed that individual cocci were non-uniform, generally ovoid and ~0.7–1.0 µm long and 0.5–0.8 µm wide (Fig. 1d–f).

Growth conditions and biochemistry

Of the liquid media assessed, *Noodlococcus* grew optimally in TSB under aerobic conditions (Fig. S2a). It did not grow anaerobically. *Noodlococcus* grew at temperatures between 20 and 40 °C, with optimal growth at 28 °C (Fig. S2b). It grew between pH 6 and pH 11, with optimum growth at pH 7 (Fig. S2c). Growth was reduced in media containing >3% NaCl (Fig. S2d).

Noodlococcus was oxidase-negative and catalase-positive. Fatty acid composition analysis revealed that the major fatty acids present in the *Noodlococcus* cell wall were anteiso- $C_{15:0}$ (45.7%), anteiso- $C_{17:0}$ (16.9%) and iso- $C_{15:0}$ (14.9%), with closest matches identified in the *Micrococcus-luteus*-GC subgroup C (sim index score 0.521). *Noodlococcus* was sensitive to penicillin, cephalosporin, fluoroquinolone, macrolide and tetracycline antibiotics, but resistant to colistin and nitrofurantoin, with MICs of 32 and >256 µg ml⁻¹, respectively (Table S2).

Complete genome of *K. rhizophila* *Noodlococcus*

The complete *Noodlococcus* genome sequence was assembled from a combination of Illumina and Nanopore reads using Tricycler. The genome consisted of a single, circular 2,732,108 bp chromosome with an overall G+C content of 70.6%. The *Noodlococcus* 16S rRNA gene was 99.9% identical to that of the *K. rhizophila* type strain TA68, with 96% query coverage. Confirming its species assignment, we found that the ANI of the *Noodlococcus* genome relative to that of TA68 was 97.5%. Bakta annotation identified 2,343 ORFs, 46 tRNAs and 9 rRNAs in the *Noodlococcus* chromosome. A putative origin of replication (*oriC*) was identified between the *dnaA* and *dnaN* genes (positions 601,211–601,811 of GenBank accession CP097204). It includes a DNA unwinding

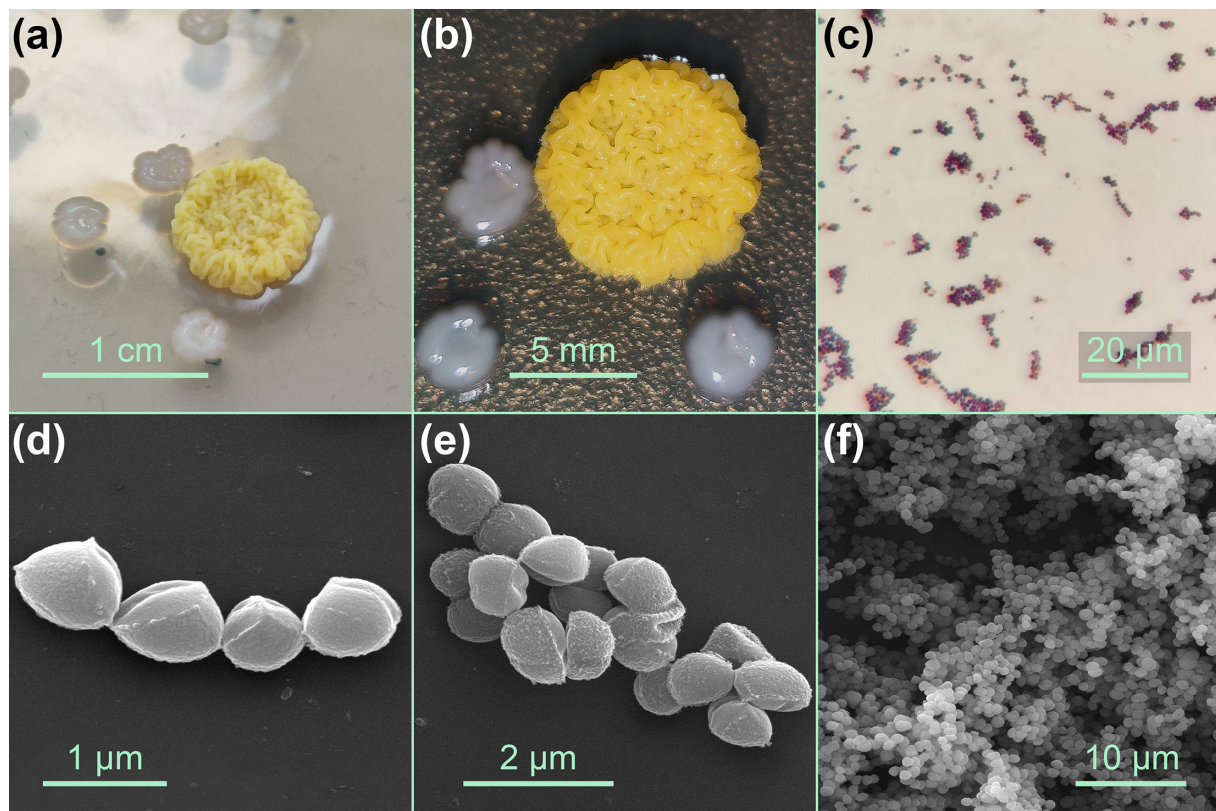


Fig. 1. Morphology of laboratory contaminant *Noodlococcus*. (a, b) Photographs of the original *Noodlococcus* colony that was found growing on BHI agar. The smaller, white colonies were *E. faecium* that had re-grown after being picked for further culturing 11 days prior. (c) Phase-contrast micrograph of Gram-stained *Noodlococcus*. (d–f) Scanning electron micrographs of *Noodlococcus* cells.

element, DnaA-trio motifs and two arrays of DnaA boxes, resembling previously characterised actinobacterial chromosomal replication origins [55, 73] (Fig. 2a).

As its colour was such a distinctive aspect of its colony morphology, we searched the *Noodlococcus* genome for pigment determinants. Carotenoids are naturally occurring pigments produced by a wide range of organisms for a variety of purposes [74]. We found a putative carotenoid synthesis cluster that contained six *crt* genes and resembled the gene cluster of *Micrococcus luteus* NCTC 2655 that directs synthesis of the γ -cyclic C_{50} carotenoid sarcinaxanthin, which has been experimentally characterised [75]. The *Noodlococcus* cluster contains *crtE*, *B*, *I*, *Yg*, *Yh* and *E2* genes (Fig. 2b) that encode proteins with amino acid identities that range from 48.1 to 75.6% to their equivalents from NCTC 2655 but does not contain a gene equivalent to *crtX*. Genes for a MarR-family regulator and fructosamine kinase lie immediately upstream of the *crt* genes, separating them from a set of four genes for a putative ABC transporter (Fig. 2b).

A CRISPR locus was identified downstream of determinants for a type I-E CRISPR-Cas system (Fig. 2c). The locus contained 98 spacer sequences of 33 or 34 bp, interspersed with 99 copies of a 28 bp repeat unit. The spacer sequences were used to query the GenBank non-redundant nucleotide database, and 64 out of 98 returned matches to non-*Kocuria* sequences (Table S3). Fifty-six spacers matched bacteriophage genomes derived from urban environment metagenomic datasets with identities ranging from 90.9 to 100%, and 8 matched chromosomal sequences from various bacterial genera with identities of 84.8–93.9%. All 15 of the bacteriophage genomes matched by *Noodlococcus* spacers were *Caudoviricetes* and have previously been predicted to be lytic [76]. To account for lysogenic bacteriophage, the *Noodlococcus* chromosome was screened for the presence of prophage regions, but none were found.

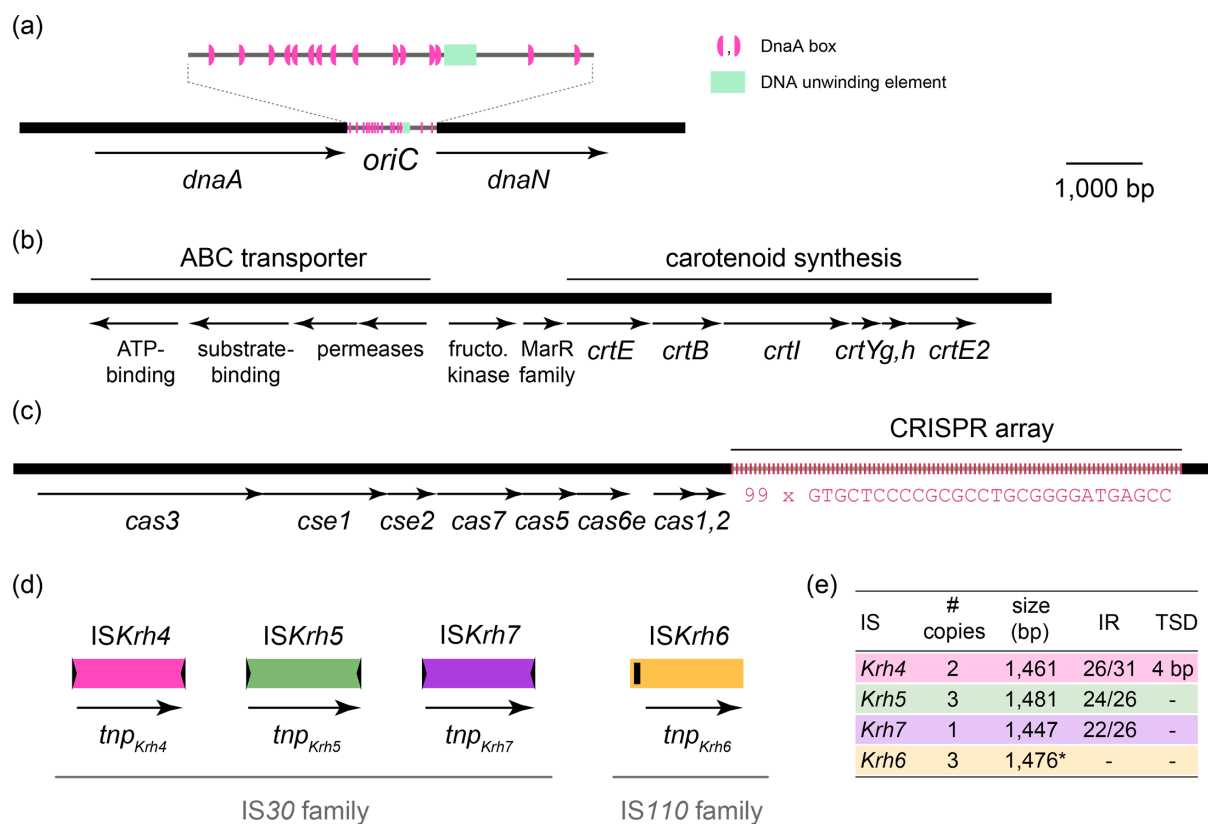


Fig. 2. Genomic features of *Noodlococcus*. All parts drawn to the same scale from GenBank accession CP097204. The extents and orientations of genes are indicated by labelled arrows beneath the horizontal lines that represent segments of the *Noodlococcus* genome. (a) Origin of replication. The *oriC* region is magnified 4.5× above to display fine-scale features as indicated in the key to the right. (b) Carotenoid synthesis region. The extents of regions that determine carotenoid synthesis and a putative ABC transporter are marked by labelled lines above. (c) CRISPR-Cas locus. The extent of the CRISPR locus is marked above. Each short, vertical pink line represents a copy of the 28 bp repeat unit, the sequence of which is shown below. (d) IS. Coloured vertical lines represent novel IS found in the *Noodlococcus* chromosome, grouped according to family membership. Terminal inverted repeats are shown as short black arrows and a seekRNA-determining region as a small black rectangle. (e) IS characteristics. Table outlining the features of novel IS characterised here. *, size estimate that requires experimental validation (see text); IR, terminal inverted repeats (identical bases/total bases).

Three novel IS were identified based on their presence at multiple positions in the *Noodlococcus* chromosome (Fig. 2d, e). They included two IS30 family elements, ISKrh4 (at two chromosomal positions, flanked by distinct 4 bp TSDs) and ISKrh5 (at three positions, no TSDs). Another novel element, ISKrh7, present at a single chromosomal position, was identified by homology to ISKrh5, with which it shared 74.2% nucleotide identity (1,044 of 1,407 bp). The third IS found at multiple positions was ISKrh6, which belongs to the IS110 family. It has recently been demonstrated that the non-coding region upstream of the transposase gene in IS110 family elements determines a seekRNA that directs transposition [77]. The size of ISKrh6 was estimated as 1,476 bp based on the extent of conserved sequence at all three positions in *Noodlococcus*, which might include the element's target site (<10bp), but this would need to be distinguished from the ends of the element experimentally.

Diversity of publicly available *Kocuria* genomes

A total of 257 *Kocuria* genomes were retrieved from the NCBI database (last search: 21 September 2024). These included representatives of 24 *Kocuria* species, with 80 genomes unclassified. A total of 230 (89.5%) passed CheckM2 QC ($\geq 90\%$ completeness, $\leq 5\%$ contamination). A genus-wide core genome phylogeny revealed a structure that was largely concordant with existing species definitions (Fig. S3). The phylogeny was supplemented by using FastANI to compare ANI across all genomes. ANI is used widely for assessing species boundaries, with $>95\%$ ANI a typical threshold for genomes belonging to the same species [58]. Most *Kocuria* species definitions were supported by ANI values (Fig. S3). Apart from exceptions detailed below, inter-species ANI values ranged from 77.7 to 88.8%, while intra-species ANI values ranged from 94.5 to 100%. Two genomes were outliers – *Kocuria palustris* DE0549 and *Kocuria rosea* TA28 – which had ANI values of 85.9 and 86.0% to their respective type strains (Figs S3 and S4). These were likely mislabelled genomes that either represented novel *Kocuria* species or were a result of assembly chimaeras, but comparison to further closely related genome sequences would be needed to confirm this. A major exception to the largely concordant data was *K. rhizophila*, which included *Noodlococcus* and accounted for 22.2% of *Kocuria* genomes examined here.

Fifty-six genomes were labelled as *K. rhizophila*, with a wide range of isolation sources including human, animals or animal products and diverse environments (Table 1). Six of these genomes did not pass QC, including four which were flagged or suppressed by GenBank due to contamination, inappropriate genome size or other genome content issues (Table S4). In the core genome phylogeny, all genomes labelled *K. rhizophila* clustered in two monophyletic clades except for L3_129_000G1_dasL3_129_000G1_metabat.metabat.108 and TNDT1 which were located within the *Kocuria salsicia* and *Kocuria flava* clusters (Fig. S3). These genomes also had whole-genome ANIs of 85.7 and 80.7% relative to *K. rhizophila* type strain TA68, respectively, providing further evidence that they have been mislabelled. Three genomes (HMSC066H03, BT304 and APC 4018) that had been labelled as unclassified *Kocuria* sp. in GenBank also clustered within the two clades (Fig. S3), leaving a total of 51 *K. rhizophila* genomes (Table 1).

Distinction of two monophyletic clades in *K. rhizophila*

Genomes that have been labelled *K. rhizophila* clearly separated into two monophyletic clades in the genus-wide core genome phylogeny (Fig. S3). This topology remained when a core genome phylogeny was inferred for *K. rhizophila* genomes alone (Fig. 3), as well as in the full 16S rRNA phylogeny (Fig. S5). Type strain TA68, *Noodlococcus*, and commonly used reference strain NBC_01227 were located together in one monophyletic clade, while the genomes of *K. rhizophila* available from major culture collections (NCTC 8340 and NBRC 12708) and other commonly used reference strains (DC2201, FDAARGOS_302) were located in the other clade. FastANI revealed inter-clade ANI values of 89.5–90.4% and intra-clade ANI values of 96.7–100% (Figs 3 and 4). These ANI values clearly contrast with those observed for all other *Kocuria* species (Fig. S4), which suggests that the *K. rhizophila* clades constitute two distinct species. We refer to the clade containing *K. rhizophila* type strain TA68 as *K. rhizophila*, and the other clade as *Kocuria pseudorhizophila*. We identified a total of 35 *K. rhizophila* and 16 *K. pseudorhizophila* genomes (Fig. 3, Table 1).

Within clades, we observed two clusters of genomes that exhibited near-identical ANIs. The first of these was in the *K. pseudorhizophila* clade and contained all four genomes derived from the widely used reference strain ATCC 9341 (DC2201, NBRC 12708, FDAARGOS_302 and NCTC8340), which had ANIs around 100% (Fig. 3). The second cluster included five genomes in the *K. rhizophila* clade (strains D2, MGYG-HGUT-02537, p3-SID209, p3-SID208 and NPDC051498), with ANIs ranging between 99.9 and 100% (Fig. 3). Interestingly, these isolates originated from a range of sources and geographical locations (Tables 1 and S5). MGYG-HGUT-02537, collated and uploaded to NCBI as part of the Unified Human Gastrointestinal Genome collection [30], was identical to and is likely a duplicate genome of strain D2 (ANI 100%), isolated from a human stool sample in India. However, NPDC051498, from the Natural Products Discovery Centre collection [78], was isolated from soil in Peru, and strains p3-SID209 and p3-SID208 were isolated from a human skin swab in the USA (Tables 1 and S5). There were nine further instances where strains from different studies shared ANIs $>99\%$ (Fig. 3, Table S5).

Pangenome analysis further distinguishes *K. rhizophila* clades

To further interrogate the differences between the two *K. rhizophila* clades, we examined their combined pangenome. A pangenome, composed from 51 genomes from the two *K. rhizophila* clades, was constructed using Panaroo [59]. The pangenome sample

Table 1. *K. rhizophila* genomes in the NCBI database (last search: 21 September 2024)

Strain	Isolation source	Geographic location	Collection date	Clade	Assembly accession	BioProject accession	Reference
DC2201 (=NBR103217)	ATCC 9341 derivative	Unknown	2008	<i>pseudorhizophila</i>	GCA_000010285.1	PRJDA27833	[37]
P7-4	Intestine of <i>Siganus doliatus</i>	Pacific Ocean	≤2011	<i>rhizophila</i>	GCA_000214115.2	PRJNA66631	[34]
TPW45	Waterfall	Malaysia	2013	<i>rhizophila</i>	GCA_000813865.1	PRJNA269640	[27]
14ASP	Soil	Pakistan	2013	<i>rhizophila</i>	GCA_001038535.1	PRJNA286912	[10]
RF	Soil	Iran	2011	<i>rhizophila</i>	GCA_001190985.2	PRJNA256070	[8]
G2	Meat grinder	Denmark	≤2015	<i>pseudorhizophila</i>	GCA_001499775.1	PRJEB9947	[88]
HMSC066H03	Vaginal/rectal	USA	≤2016	<i>rhizophila</i>	GCA_001812045.1	PRJNA299987	Unpublished
FDAARGOS_302*	Soil†	USA‡	≤1949§	<i>pseudorhizophila</i>	GCA_002208685.2	PRJNA231221	[89]
UMB0131	Catheter	USA	2015	<i>rhizophila</i>	GCA_002861865.1	PRJNA316969	[90]
D2	Stool	India	2013	<i>rhizophila</i>	GCA_002879775.1	PRJNA428934	Unpublished
BT304	Cow intestine	South Korea	2015	<i>pseudorhizophila</i>	GCA_003290245.1	PRJNA475186	[33]
TA68	Rhizosphere of <i>T. angustifolia</i> ; Danube	Hungary	≤1995	<i>rhizophila</i>	GCA_003667225.1	PRJNA495744	[3, 80]
4R-34	Alder plant root nodule	USA	2017	<i>rhizophila</i>	GCA_004563775.1	PRJNA480027	[31]
4R-31				<i>rhizophila</i>	GCA_004563945.1		
DE0228	Environmental (Duke University campus)	USA	2018	<i>pseudorhizophila</i>	GCA_007677595.1	PRJNA543692	[91]
DE0203				<i>rhizophila</i>	GCA_007677895.1		
DE0200				<i>rhizophila</i>	GCA_007677945.1		
H09	Ground beef	USA	2018	<i>pseudorhizophila</i>	GCA_016620465.1	PRJNA688548	Unpublished
28R2A-20	Marine sponge	Atlantic Ocean	2015	<i>rhizophila</i>	GCA_017723815.1	PRJNA716289	[28]
ndh24	Meat (Nuodeng ham)	China	2019	<i>pseudorhizophila</i>	GCA_020829765.1	PRJNA694819	[29]
ndh45				<i>pseudorhizophila</i>	GCA_020829785.1	PRJNA694823	
ACRRQ	Skin	USA	2018	<i>rhizophila</i>	GCA_022347075.1	PRJNA694925	Unpublished
Noodlocooccus	Laboratory contaminant	UK	2019	<i>rhizophila</i>	GCA_023373505.1	PRJNA835814	This study
6.1.26	Skin (zebrafish skin mucus)	USA	2022	<i>rhizophila</i>	GCA_024655785.1	PRJNA849171	Unpublished
p3-SID1135	Skin (nares swab)	USA	2019	<i>rhizophila</i>	GCA_025144265.1	PRJNA803478	Unpublished
p3-SID292	Skin (axillary-groin swab)	USA	2018	<i>rhizophila</i>	GCA_025147725.1		
p3-SID284			2018	<i>rhizophila</i>	GCA_025147935.1		
p3-SID209			2018	<i>rhizophila</i>	GCA_025148985.1		

Continued

Table 1. Continued

Strain	Isolation source	Geographic location	Collection date	Clade	Assembly accession	BioProject accession	Reference
p3-SID208	Skin (axillary-groin swab)	USA	2018	<i>rhizophila</i>	GCA_025149045.1	PRJNA803478	Unpublished
p3-SID1455	Skin (volar forearm swab)		2019	<i>rhizophila</i>	GCA_025149885.1		
p3-SID1414	Skin (toe web space swab)		2019	<i>rhizophila</i>	GCA_025151645.1		
p3-SID1383			2019	<i>rhizophila</i>	GCA_025151745.1		
SLM-01	Blood	South Korea	2022	<i>rhizophila</i>	GCA_027889775.1	PRJNA921965	[18]
ZSM2	Iron mine rock	Algeria	2022	<i>rhizophila</i>	GCA_030403625.1	PRJNA983803	Unpublished
APC 4018	Snub-nosed spiny eel (skin)	Atlantic Ocean	2017	<i>rhizophila</i>	GCA_030407675.1	PRJNA883941	[35]
BE309	Sorghum-associated	USA	2022	<i>rhizophila</i>	GCA_031457995.1	PRJNA709325	Unpublished
CTOTU47604	Urban	USA	2013	<i>pseudorhizophila</i>	GCA_032017965.1	PRJNA850115	[92]
CTOTU48865			2013	<i>rhizophila</i>	GCA_032060705.1		
CTOTU50493			2013	<i>pseudorhizophila</i>	GCA_032101205.1		
NU1901-B013	Gut (Atlantic salmon)	Norway	2019	<i>pseudorhizophila</i>	GCA_033042015.1	PRJNA675623	Unpublished
NBC_01227	Inside anthill	Denmark	2020	<i>rhizophila</i>	GCA_035990065.1	PRJNA747871	[32]
NBC_00781	Strawberry plants	USA	2016	<i>pseudorhizophila</i>	GCA_036347115.1		
UNHI	Laboratory contaminant	USA	2024	<i>rhizophila</i>	GCA_040028385.1	PRJNA1116229	Unpublished
NBRC 12708	ATCC 9341 derivative	Unknown	2022	<i>pseudorhizophila</i>	GCA_040372225.1	PRJDB16551	Unpublished
NPDC051498	Soil	Peru	2024	<i>rhizophila</i>	GCA_040681985.1	PRJNA1106891	[78]
155	Meat processing factory	Russia	2021	<i>pseudorhizophila</i>	GCA_041021375.1	PRJNA1134931	[83]
PT10	Roots of <i>Panicum turgidum</i>	Tunisia	2018	<i>rhizophila</i>	GCA_900576785.1	PRJEB29164	[7]
NCTC8340	ATCC 9341 derivative	USA	1951	<i>pseudorhizophila</i>	GCA_900637835.1	PRJEB6403	Unpublished
MGYG-HGUT-02557	Human gut	Unknown	2019	<i>rhizophila</i>	GCA_902388135.1	PRJEB33885	[30]
Marseille-P2672	Permafrost	Siberia	2019	<i>rhizophila</i>	GCA_904423785.1	PRJEB40293	[25]
ERR9968735_bin.6_MAG	Oral cavity	Australia	2014	<i>pseudorhizophila</i>	GCA_963519965.1	PRJEB65351	[93]

≤ = on or before this date.

*Genome of ATCC 934.1.

†Incorrectly listed as 'Rhizosphere of narrowleaf cattail (*T. angustifolia*) from a floating mat in Danube River' in GenBank.

‡Incorrectly listed as 'Hungary' in GenBank.

§Incorrectly listed as '1999' in GenBank. ATCC 934.1 first mentioned in a publication in 1949 [84].

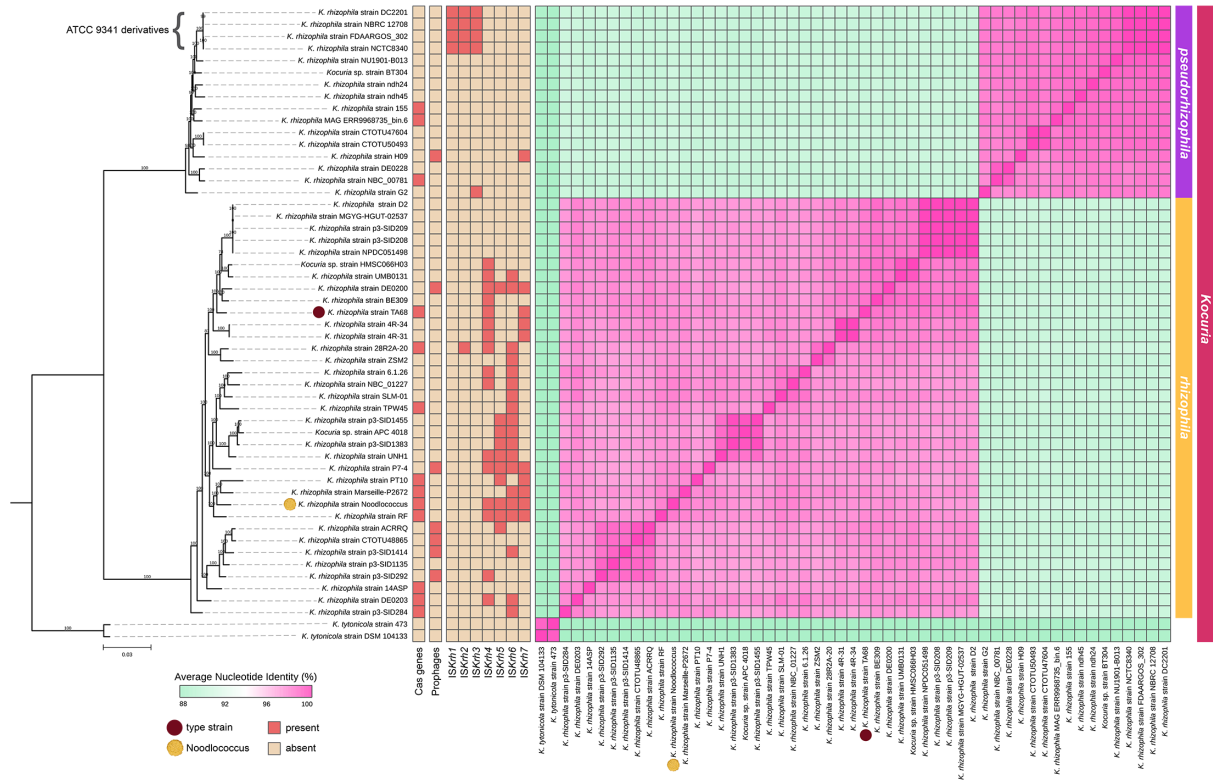


Fig. 3. Core genome phylogeny and pairwise ANIs of *K. rhizophila* genomes ($n=51$). Maximum likelihood phylogeny was inferred with the GTR+F+I+R5 substitution model using IQTREE. Ultrafast bootstrap supports are labelled on each branch, and the scale bar represents substitutions per site. *K. tytonicola* strains 473 and DSM 104133 were used as outgroups. Pairwise ANIs were calculated with FastANI and visualised with heatmap. Proposed species demarcation is indicated by the group labels on the right side of the figure. Presence/absence of CRISPR (Cas genes), prophages and IS are shown to the right of the tree. Type strain TA68 is highlighted with a brown circle, and Noodlococcus is highlighted with a yellow circle.

consisted of 1,013 core, 711 soft core, 951 shell and 2,275 cloud genes. Within the pangenome, we identified clade-exclusive gene sets (Fig. 5). For example, *K. rhizophila* and *K. pseudorhizophila* possessed 114 and 90 clade-specific core genes, respectively. Clade-specific core genes are present in all representative genomes of one clade and never present in the other clade. A large proportion ($n=53/114$ and $n=37/90$) of clade-specific core gene clusters was unable to be assigned to clusters of orthologous genes categories with known function, highlighting the need for further functional studies in these organisms. Core genes specific to the *K. rhizophila* clade included those for putative siderophore transporters and molybdenum cofactor biosynthesis (*moeA*, *mobA*). Core genes specific to the *K. pseudorhizophila* clade included those for a putative xanthine dehydrogenase (*xdhABC*) and catabolism of protocatechuate (*pcaBCGHJ*). Sets of clade-specific shell and clade-specific cloud genes were also identified (*K. rhizophila*: 318 clade-specific shell, 1,192 clade-specific cloud; *K. pseudorhizophila*: 180 clade-specific shell, 521 clade-specific cloud).

To examine the conservation of characterised regions of the Noodlococcus chromosome (Fig. 2c) in the rest of *K. rhizophila*, we screened all 51 genomes for their presence. The putative carotenoid synthesis region found in Noodlococcus was conserved across clades, with variation in nucleotide identities consistent with ANIs. The Cas genes were less well conserved, detected in 13 genomes spread across the phylogeny (Fig. 3). The Cas gene segments in other genomes ranged from 99.4 to 100% identical to the region from Noodlococcus, with the exception of the segment in strain 14ASP, which was 90.6% identical. Screening for IS revealed that the four new elements identified here were largely confined to the *K. rhizophila* clade, while *ISKrh1-3* were almost exclusively present in ATCC 9341 derivatives (Fig. 3).

Having found no prophages in Noodlococcus, we extended our analysis to the remaining 50 genomes. Prophages were detected in six genomes in the *K. rhizophila* clade and one genome in the *K. pseudorhizophila* clade (Fig. 3). All prophages were classified within the *Caudoviricetes* class, and their completeness scores ranged from 33.3 to 84.0%. Notably, three *K. rhizophila* genomes (strains DE0200, ACRRQ and p3-SID1414) harboured the same prophage, sharing over 99% nucleotide identity. All seven prophage sequences exhibited high nucleotide identities (86.2–97.5%) to bacteriophage sequences identified in an urban

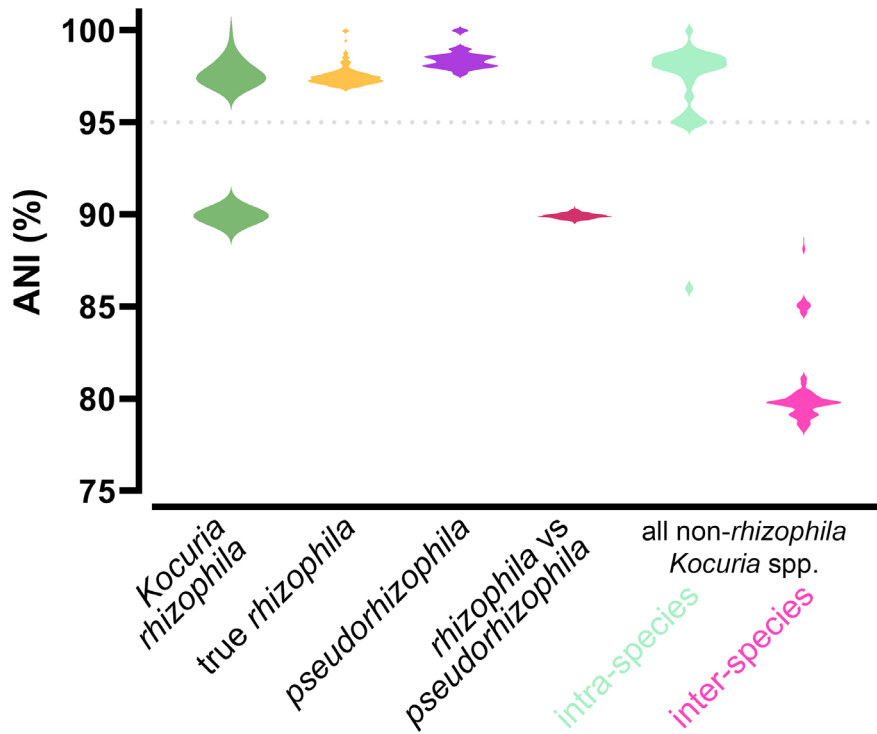


Fig. 4. ANIs between various *Kocuria* taxa. *K. rhizophila* (n=51), *true rhizophila* (n=35) and *pseudorhizophila* (n=16). All non-*rhizophila* *Kocuria* spp. only include genomes of species groups that contained >2 genomes (n=141).

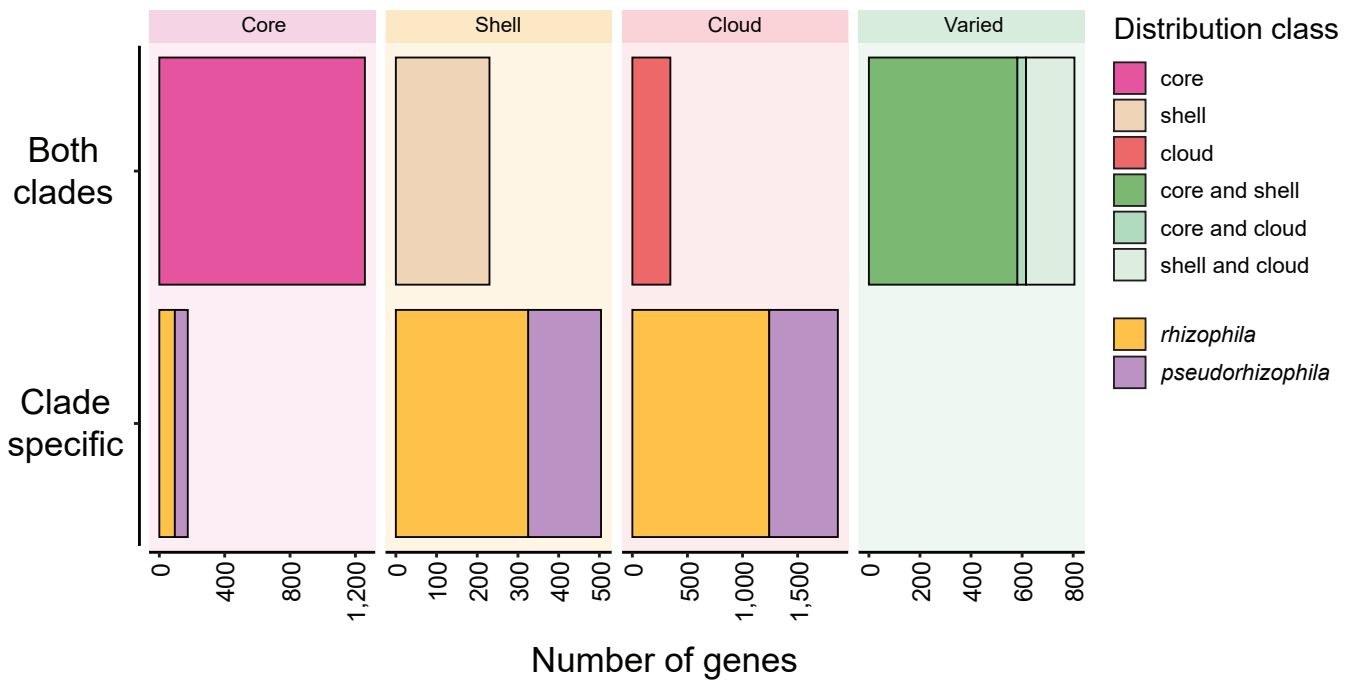


Fig. 5. Population-structure-aware pangenome of *K. rhizophila*. Number of gene clusters of the *K. rhizophila* pangenome (both clades) from each distribution class. Clade-specific gene clusters exclusively present in either *rhizophila* (yellow) or *pseudorhizophila* (purple) are shown for each distribution class. Core: >95% presence; shell: <95% and >15% presence; cloud: <15% presence. Distribution class definitions taken from [68].

environment metagenomic study [76] – the same study from which phage genomes that matched *Noodlococcus* CRISPR spacer sequences were derived. None of the seven prophage-containing genomes contained Cas genes (Fig. 3).

DISCUSSION

Our serendipitous isolation and subsequent characterisation of laboratory contaminant *Noodlococcus* led to this large-scale genomic assessment of the *Kocuria* genus. Despite its ubiquity in nature, *Kocuria* is obscure relative to the medically relevant bacterial genera that account for the vast majority of genomes in public sequence databases [79]. *Noodlococcus* was determined to be a strain of *K. rhizophila*, and this prominent *Kocuria* species became the focus of our analyses.

The *K. rhizophila* type strain, TA68, was isolated in the 1990s [3], and its draft genome sequence is available [80]. At the time of our study, there were 50 draft and just 9 complete *K. rhizophila* genomes in NCBI, including *Noodlococcus*. We have shown here that, of the complete genomes available, only 28R2A-20, NBC_01227, UNH1 and *Noodlococcus* represent the true *K. rhizophila* clade that includes TA68 (Fig. 3). The *Noodlococcus* genome is the first of these to be characterised and described. Our annotations of genomic features associated with replication, carotenoid synthesis and defence are the first for this species (Fig. 2). By examining CRISPR spacers, we have identified putative *Kocuria* phages (Table S4). These phages have previously been predicted to be lytic but were also predicted to be *Arthrobacter* phage using vHULK [76]. Given that vHULK was modelled from a relatively small number of genera that did not include *Kocuria* [81], we expect that the identification of these phage sequences in the *Noodlococcus* spacer region provides stronger evidence that *Kocuria* is their natural host. We also identified several putative lysogenic phages in *K. rhizophila* genomes. We used the *Noodlococcus* genome to explore IS, which are arguably the simplest self-mobile genetic elements. This led to the identification of four novel elements, IS*Krh4-7* (Fig. 2), which are the first examined in *K. rhizophila*, as the first three *Kocuria* IS in ISFinder (IS*Krh1-3*; all IS481 family) were found in DC2201, which is in the *K. pseudorhizophila* clade (Fig. 3). IS*Krh4-7* therefore expands the number of IS families recognised in *Kocuria* to include IS30 and IS110, the latter of which has recently been shown to have unique transposition properties and biotechnological potential [77].

Taking a broad view of the genus, we generated the largest *Kocuria* phylogeny created to date, which supported species assignments and placed some previously unidentified genomes with their closest relatives. We found that *K. rhizophila* is not a monophyletic species and is made up of two distinct clades, which we refer to as *K. rhizophila* and *K. pseudorhizophila*. Our analyses strongly suggest that these should be classified as two different species. This was indicated by both our core genome and full 16S rRNA phylogenies. ANI comparisons also showed a clear distinction between genomes in the *K. rhizophila* and *K. pseudorhizophila* clades, with intra-clade ANI values of >96.7% and >97.5%, respectively. The universal species boundary of 95% ANI has been disputed, with many species exhibiting intra-species ANI values below this [82]. However, we found that for *Kocuria*, non-*rhizophila* intra-species values were above 94.5%, with all inter-species ANI values <90%. Thus, our analyses support a species boundary of around 95% ANI within the *Kocuria* genus. Splitting the two *K. rhizophila* clades into separate species would maintain taxonomic consistency with the rest of the *Kocuria* genus, with intra-clade ANI values >96% and inter-clade values <90.5%. The identification of distinct clade-dependent gene sets may reflect exposure to independent gene pools, possibly due to habitation of differing ecological niches. It appears that these two *K. rhizophila* clades have undergone different evolutionary trajectories, adding weight to the argument that they should be considered as two separate species.

Comparing phenotypic differences between *K. rhizophila* and *K. pseudorhizophila* was not feasible within the scope of this work. However, the fatty acid composition profile of ATCC 9341 (in the *K. pseudorhizophila* clade) has been determined [36] and is the same as that determined for *Noodlococcus* here. Other reported phenotypes for *K. pseudorhizophila* contrast with those of *Noodlococcus*. For example, strains ndh24 and ndh45 (both in the *K. pseudorhizophila* clade) have been shown to tolerate considerably higher salinity (>10% w/v) [29], than *Noodlococcus* (3% w/v). This tolerance reflects the high-salinity environment that ndh24 and ndh25 were isolated from (dry cured ham, see Table 1), and might represent strain-specific rather than clade-wide phenotypic divergence. Interestingly, ndh24 and ndh25 colonies were noted to have turned orange [29]. We found that the putative carotenoid region found in *Noodlococcus* is conserved across *K. rhizophila* and *K. pseudorhizophila* but varies in nucleotide identity consistent with ANIs. We cannot predict the effects of this divergence on pigment production and characteristics. Clearly, further characterisation of more strains from both clades will be required to assess clade-specific phenotypic differences.

K. rhizophila is commonly used as a reference strain in industrial applications, including sterility and antimicrobial susceptibility tests. However, there are multiple reference strains available to purchase, which we have shown are genetically distinct. Type strain TA68, from the true *K. rhizophila* clade (Fig. 3), is available in various culture collections under the names ATCC BAA-50, DSM 11926, IFO 16319, CCM 4950 and NBRC 16319. Genomes for the other widely used reference strain ATCC 9341 (including pseudonyms and derivatives NBRC 103217, NBRC 12708, NCTC 8340, DSM 348, DC2201 and FDAARGOS_302) clustered in the *K. pseudorhizophila* clade (Fig. 3). ATCC 9341 is widely available in commercial products for QC testing, such as *K. rhizophila* Culti-Loops™ from Thermo Fisher Diagnostics (catalogue #R4604075). This strain was originally deposited to ATCC as *Sarcina lutea*, before being reclassified as *M. luteus*. In 2003, it was reclassified again to *K. rhizophila* after DNA hybridisation experiments

indicated that it was more closely related to TA68 than to the *M. luteus* type strain [36]. This reclassification may have provided the basis for subsequent mislabelling of various strains. In 2008, ATCC 9341 derivative DC2201 was the first complete *K. rhizophila* genome published [37], and its use as a species reference has led to strains in the *K. pseudorhizophila* clade being misclassified as *K. rhizophila*. Adding to the confusion, the genome for ATCC 9341 (FDAARGOS_302) was uploaded to NCBI in 2018 with identical metadata to TA68 (Table 1), which is clearly a different strain. Strains from the two clades have been unknowingly used interchangeably in multiple studies. For example, a recent study of antimicrobial resistance in *Kocuria* spp. used ATCC 9341 (*K. pseudorhizophila* clade) during laboratory experiments but used the genome of strain 4R-31 (*K. rhizophila* clade) during bioinformatic analyses [83]. Such misrepresentation could lead to conflicting results and inaccurate conclusions. This will continue to be an issue in both research and industrial applications whilst culture collections and QC products list strains such as ATCC 9341 as *K. rhizophila*, with no indication that they are genetically distinct from true *K. rhizophila*. Standardising the use of *K. rhizophila* reference strains and genomes is therefore essential.

ATCC 9341 was first mentioned (as *S. lutea* PCI 1001) in a 1949 publication by Randall *et al.* that described its use in antibiotic sensitivity testing [84]. A 1954 publication confirms that this strain was originally isolated by W.A. Randall [85]. As our study strongly indicates the need for the separation of *K. rhizophila* into two species, we suggest, as a provisional designation, that an appropriate species name for the *K. pseudorhizophila* clade that includes ATCC 9341 would be *Kocuria randallii* (ran.dal'li.i. N.L. gen. masc. n. *randallii*, of Randall, named in honour of Dr William A. Randall Sr. for contributions to antibiotic assay development and regulatory microbiology).

Given the sporadic and widespread derivations of genomes captured in this relatively small dataset, we were surprised to find that several *K. rhizophila* isolates from disparate sources shared ANI values >99% (Fig. 3, Table S5). This suggests that environmental *K. rhizophila* clones can have extensive geographic distributions. These might be explained by associations with human and animal migration, or by the detection of *Kocuria* in air and clouds [86, 87], where cells would be subjected to global air currents that could contribute to their dissemination. While this is an intriguing possibility, stronger evidence supporting the distribution of individual clones will be required before it can be considered more seriously.

This study, prompted by our characterisation of a laboratory contaminant, exemplifies how chance findings, common yet often undervalued in biological research, can yield novel insights. The discovery of Noodlococcus led to the creation of 'Contamination Club' (ContamClub), a social media initiative that has been a useful vehicle for professional and public science engagement. We hope that ContamClub and the story of Noodlococcus will continue to promote investigation of the unusual and understudied, particularly in the genomics era, where relatively low-cost comparative studies can yield significant findings.

CONCLUSIONS

We have found that *K. rhizophila* is not a monotypic species but is comprised of two clades with distinct ANIs and pangenomic profiles. Distinguishing these clades has important implications for the use of *K. rhizophila* strains as controls in research and industry. The complete genome sequence of laboratory contaminant Noodlococcus has been the basis for our description of previously uncharacterised features of the *K. rhizophila* genome. The Noodlococcus genome and our large-scale genomic evaluation can serve as a baseline for future studies into the distribution, diversity and evolution of this ubiquitous species.

Funding information

The authors received no specific grant from any funding agency.

Acknowledgements

We thank all followers of ContamClub for their encouragement, inspiration, and support. Thanks to Paul Stanley for assistance with electron microscopy.

Author contributions

R.A.M. serendipitously isolated Noodlococcus. The project was conceived by R.A.M., G.E.M., S.F.S.H., R.S.M. and W.v.S. Wet lab experiments were carried out by A.J.W., under the supervision of G.E.M., S.F.S.H., R.S.M. and R.A.M. Long-read sequencing was carried out by J.Q. Sequence annotation was performed by R.A.M. and C.W. Bioinformatic analyses were carried out by G.E.M., S.F.S.H., E.A.C. and R.A.M. Data visualisation was performed by G.E.M., S.F.S.H., R.A.M., E.A.C. and L.Y.S.T. The manuscript was drafted by R.A.M., G.E.M., S.F.S.H. and E.A.C. and was edited and revised by all authors.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Stackebrandt E, Koch C, Gvozdiak O, Schumann P. Taxonomic dissection of the genus *Micrococcus*: *kocuria* gen. nov., *Nesterenkonia* gen. nov., *Kytococcus* gen. nov., *Dermacoccus* gen. nov., and *Micrococcus* cohn 1872 gen. emend. *Int J Syst Evol Microbiol* 1995;45:682–692.
- Kandi V, Palange P, Vaish R, Bhatti AB, Kale V, *et al.* (n.d.) Emerging bacterial infection: identification and clinical significance of *Kocuria* species. *Cureus*;8:e731.
- Kovács G, Burghardt J, Pradella S, Schumann P, Stackebrandt E, *et al.* *Kocuria palustris* sp. nov. and *Kocuria rhizophila* sp. nov., isolated

- from the rhizoplane of the narrow-leaved cattail (*Typha angustifolia*). *Int J Syst Evol Microbiol* 1999;49:167–173.
4. Reddy GSN, Prakash JSS, Prabakar V, Matsumoto GI, Stackebrandt E, et al. *Kocuria polaris* sp. nov., an orange-pigmented psychrophilic bacterium isolated from an Antarctic cyanobacterial mat sample. *Int J Syst Evol Microbiol* 2003;53:183–187.
 5. Purty S, Saranathan R, Prashanth K, Narayanan K, Asir J, et al. The expanding spectrum of human infections caused by *Kocuria* species: a case report and literature review. *Emerg Microbes Infect* 2013;2:1–8.
 6. Youn H-Y, Seo K-H. Isolation and Characterization of Halophilic *Kocuria salsicia* Strains from Cheese Brine. *Food Sci Anim Resour* 2022;42:252–265.
 7. Guesmi S, Pujic P, Nouioui I, Dubost A, Najjari A, et al. Ionizing-radiation-resistant *Kocuria rhizophila* PT10 isolated from the Tunisian Sahara xerophyte panicum turgidum: polyphasic characterization and proteogenomic arsenal. *Genomics* 2021;113:317–330.
 8. Mehrabadi JF, Mirzaie A, Ahangar N, Rahimi A, Rokni-Zadeh H. Draft genome sequence of *Kocuria rhizophila* RF, a radiation-resistant soil isolate. *Genome Announc* 2016;4:00095–16.
 9. Horiuchi A, Kubota N, Hidaka E, Shimabukuro A, Yasukochi S, et al. Notable alkaline tolerance of *Kocuria marina* isolate from blood of a pediatric patient with continuous intravenous epoprostenol therapy. *J Infect Chemother* 2015;21:680–686.
 10. Afridi MS, Van Hamme J d., Bundschuh J, Khan MN, et al. Biotechnological approaches in agriculture and environmental management - bacterium *Kocuria rhizophila* 14ASP as heavy metal and salt-tolerant plant growth-promoting strain. *Biologia* 2021;76:3091–3105.
 11. Sharma G, Khatri I, Subramanian S. Draft genome sequence of *Kocuria palustris* PEL. *Genome Announc* 2014;2:e01261–13.
 12. Becker K, Rutsch F, Uekötter A, Kipp F, König J, et al. *Kocuria rhizophila* adds to the emerging spectrum of micrococcal species involved in human infections. *J Clin Microbiol* 2008;46:3537–3539.
 13. Dotis J, Printza N, Papachristou F. Peritonitis attributable to *Kocuria rosea* in a pediatric peritoneal dialysis patient. *Perit Dial Int* 2012;32:577–578.
 14. Ishihara M, Nagao Y, Nishida Y, Morimoto N, Fujieda M. The first case report of *Kocuria rhizophila* peritonitis in a 3-year-old Japanese girl. *Pediatr Int* 2021;63:1523–1524.
 15. Sohn KM, Baek J-Y, Kim SH, Cheon S, Kim Y-S. Catheter-related bacteremia caused by *Kocuria salsicia*: the first case. *J Infect Chemother* 2015;21:305–307.
 16. Altuntas F, Yildiz O, Eser B, Gündogan K, Sumerkan B, et al. Catheter-related bacteremia due to *Kocuria rosea* in a patient undergoing peripheral blood stem cell transplantation. *BMC Infect Dis* 2004;4:62.
 17. Moissenet D, Becker K, Mérens A, Ferroni A, Dubern B, et al. Persistent bloodstream infection with *Kocuria rhizophila* related to a damaged central catheter. *J Clin Microbiol* 2012;50:1495–1498.
 18. Kim Y, Kim TS, Park H, Yun KW, Park JH. The first case of catheter-related bloodstream infection caused by *Kocuria rhizophila* in Korea. *Ann Lab Med* 2023;43:520–523.
 19. Grama A, Sirbe C, Fufezan O, Pop TL. *Kocuria varians* meningitis in a child with chronic granulomatous disease. *Turk Arch Pediatr* 2021;56:278–279.
 20. Srinivasa KH, Agrawal N, Agarwal A, Manjunath CN. Dancing vegetations: *Kocuria rosea* endocarditis. *BMJ Case Rep* 2013;2013:bcr2013010339.
 21. Citro R, Prota C, Greco L, Mirra M, Masullo A, et al. *Kocuria kristinae* endocarditis related to diabetic foot infection. *J Med Microbiol* 2013;62:932–934.
 22. Aleksic D, Miletic-Drakulic S, Boskovic-Matic T, Simovic S, Toncev G. Unusual case of stroke related to *Kocuria kristinae* endocarditis treated with surgical procedure. *Hippokratia* 2016;20:231–234.
 23. Mathy V, Chousterman B, Munier A-L, Cambau E, Jacquier H, et al. First reported case of postneurosurgical ventriculoperitonitis due to *Kocuria rhizophila* following a ventriculoperitoneal shunt placement. *Infect Dis Clin Pract* 2020;28:169–170.
 24. Napolitani M, Troiano G, Bedogni C, Messina G, Nante N. *Kocuria kristinae*: an emerging pathogen in medical practice. *J Med Microbiol* 2019;68:1596–1603.
 25. Afouda P, Dubourg G, Lévassieur A, Fournier P-E, Delerce J, et al. Culturing ancient bacteria carrying resistance genes from permafrost and comparative genomics with modern isolates. *Microorganisms* 2020;8:1522.
 26. Isaac P, Mutusamy P, Su Yin L, Jing Wei Y, Mohd Salleh F, et al. Complete genome sequences of *Lactococcus lactis* D1_2, a bacterium with antimicrobial properties isolated from peat soil. *Microbiol Resour Announc* 2023;12:e00680–23.
 27. Adrian T-G-S, Tan P-W, Chen J-W, Yin W-F, Chan K-G. Draft genome sequence of *Kocuria rhizophila* strain TPW45, an actinobacterium isolated from freshwater. *J Genomics* 2016;4:16–18.
 28. Williams SE, Stennett HL, Back CR, Tiwari K, Ojeda Gomez J, et al. The Bristol sponge microbiome collection: a unique repository of deep-sea microorganisms and associated natural products. *Antibiotics (Basel)* 2020;9:509.
 29. Shi Q, Wang X, Ju Z, Liu B, Lei C, et al. Technological and safety characterization of *Kocuria rhizophila* isolates from traditional ethnic dry-cured ham of nuodeng, Southwest China. *Front Microbiol* 2021;12:761019.
 30. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* 2021;39:105–114.
 31. Davis I, Sevigny J, Kleiner V, Mercurio K, Pesce C, et al. Draft genome sequences of 10 bacterial strains isolated from root nodules of alnus trees in new hampshire. *Microbiol Resour Announc* 2020;9:01440–19.
 32. Jørgensen TS, Mohite OS, Sterndorff EB, Alvarez-Arevalo M, Blin K, et al. A treasure trove of 1034 actinomycete genomes. *Nucleic Acids Res* 2024;52:7487–7503.
 33. Whon TW, Kim HS, Bae J-W. Complete genome sequence of *Kocuria rhizophila* BT304, isolated from the small intestine of castrated beef cattle. *Gut Pathog* 2018;10:42.
 34. Kim W-J, Kim Y-O, Kim D-S, Choi S-H, Kim D-W, et al. Draft genome sequence of *Kocuria rhizophila* P7-4. *J Bacteriol* 2011;193:4286–4287.
 35. Uniacke-Lowe S, Collins FWJ, Hill C, Ross RP. Bioactivity screening and genomic analysis reveals deep-sea fish microbiome isolates as sources of novel antimicrobials. *Mar Drugs* 2023;21:444.
 36. Tang JS, Gillevet PM. Reclassification of ATCC 9341 from *Micrococcus luteus* to *Kocuria rhizophila*. *Int J Syst Evol Microbiol* 2003;53:995–997.
 37. Takarada H, Sekine M, Kosugi H, Matsuo Y, Fujisawa T, et al. Complete genome sequence of the soil actinomycete *Kocuria rhizophila*. *J Bacteriol* 2008;190:4139–4146.
 38. Seshadri R, Roux S, Huber KJ, Wu D, Yu S, et al. Expanding the genomic encyclopedia of *Actinobacteria* with 824 isolate reference genomes. *Cell Genom* 2022;2:100213.
 39. Sun W, Liu C, Zhang F, Zhao M, Li Z. Comparative Genomics Provides Insights Into the Marine Adaptation in *Sponge-Derived Kocuriaflava* S43. *Front Microbiol* 2018;9:1257.
 40. Coico R. Gram Staining. *Curr Protoc Microbiol* 2006.
 41. Andrews JM. Determination of minimum inhibitory concentrations. *J Antimicrob Chemother* 2001;48 Suppl 1:5–16.
 42. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters, version 12.0; 2022. http://www.eucast.org/clinical_breakpoints/
 43. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* 2023;2:e107.

44. Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience* 2019;8:giz043.
45. Wick R. Filtlong v0.2.1; (n.d.). <https://github.com/rrwick/Filtlong>
46. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, et al. Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* 2021;22:266.
47. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
48. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res* 2021;8:2138.
49. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 2016;32:2103–2110.
50. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci* 2021;1:332–336.
51. Medaka v1.0.6. Python; Oxford Nanopore Technologies; (n.d.). <https://github.com/nanoporetech/medaka>
52. Wick RR, Holt KE. Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* 2022;18:e1009802.
53. Zimin AV, Salzberg SL. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput Biol* 2020;16:e1007981.
54. Schwengers O, Jelonek L, Dieckmann M, Beyvers S, Blom J, et al. Bakta: rapid & standardized annotation of bacterial genomes via alignment-free sequence identification. *Bioinformatics* 2021;7:000685. DOI: 10.1101/2021.09.02.458689.
55. Płachetka M, Żyta-Uklejewicz D, Weigel C, Donczew R, Donczew M, et al. *Streptomyces* origin of chromosomal replication with two putative unwinding elements. *Microbiology (Reading)* 2019;165:1365–1375.
56. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;34:D32–6.
57. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023;20:1203–1212.
58. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.
59. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, et al. Producing polished prokaryotic pangenomes with the panaroo pipeline. *Genome Biol* 2020;21:180.
60. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
61. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
62. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35:518–522.
63. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
64. Bianchini G, Sánchez-Baracaldo P. TreeViewer: flexible, modular software to visualise and manipulate phylogenetic trees. *Ecol Evol* 2024;14:e10873.
65. Kolde R. pheatmap: Pretty Heatmaps. R; (n.d.). <https://github.com/raivokolde/pheatmap>
66. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 2022;38:5315–5316.
67. Parks DH, Chuvpochina M, Rinke C, Mussig AJ, Chaumeil P-A, et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;50:D785–D794.
68. Horesh G, Taylor-Brown A, McGimpsey S, Lassalle F, Corander J, et al. Different evolutionary trends form the twilight zone of the bacterial pan-genome. *Microbial Genomics* 2021;7:000670.
69. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;38:5825–5829.
70. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;47:D309–D314.
71. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, et al. Identification of mobile genetic elements with geNomad. *Nat Biotechnol* 2024;42:1303–1312.
72. Nayfach S, Camargo AP, Schulz F, Eloë-Fadrosch E, Roux S, et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* 2021;39:578–585.
73. Richardson TT, Harran O, Murray H. The bacterial DnaA-trio replication origin element specifies single-stranded DNA initiator binding. *Nature* 2016;534:412–416.
74. Maoka T. Carotenoids as natural functional pigments. *J Nat Med* 2020;74:1–16.
75. Netzer R, Stafsnes MH, Andreassen T, Goksøyr A, Bruheim P, et al. Biosynthetic pathway for γ -cyclic sarcinaxanthin in *Micrococcus luteus*: heterologous expression and evidence for diverse and multiple catalytic functions of C(50) carotenoid cyclases. *J Bacteriol* 2010;192:5688–5699.
76. Flores VS, Amgarten DE, Iha BKV, Ryon KA, Danko D, et al. Discovery and description of novel phage genomes from urban microbiomes sampled by the MetaSUB consortium. *Sci Rep* 2024;14:7913.
77. Siddiquee R, Pong CH, Hall RM, Ataide SF. A programmable seekRNA guides target selection by IS1111 and IS110 type insertion sequences. *Nat Commun* 2024;15:5235.
78. Kalkreuter E, Kautsar SA, Yang D, Bader CD, Teijaro CN, et al. The natural products discovery center: release of the first 8490 sequenced strains for exploring actinobacteria biosynthetic diversity. *Microbiology*. DOI: 10.1101/2023.12.14.571759
79. Blackwell GA, Hunt M, Malone KM, Lima L, Horesh G, et al. Exploring bacterial diversity via a curated and searchable snapshot of archived DNA sequences. *PLoS Biol* 2021;19:e3001421.
80. Braun MS, Wang E, Zimmermann S, Boutin S, Wagner H, et al. *Kocuria tytonicola*, new bacteria from the preen glands of American barn owls (*Tyto furcata*). *Syst Appl Microbiol* 2019;42:198–204.
81. Amgarten D, Iha BKV, Piroupo CM, da Silva AM, Setubal JC. vHULK, a new tool for bacteriophage host prediction based on annotated genomic features and neural networks. *PHAGE* 2022;3:204–212.
82. Murray CS, Gao Y, Wu M. Re-evaluating the evidence for a universal genetic boundary among microbial species. *Nat Commun* 2021;12:4059.
83. Pleshko EM, Zhurina MV. *Kocuria* Species Antibiotic Resistance Genes. *Microbiology* 2024;93:S126–S130.
84. Randall WA, Kirshbaum A, Nielsen JK, Wintermeyer D. Diffusion plate assay for chloramphenicol and aureomycin. *J Clin Invest* 1949;28:940–942.
85. Brown JW, Binkley SB. Growth requirements for a penicillin sensitive *Sarcina lutea*. *J Bacteriol* 1954;68:390–391.
86. Vařilingom M, Attard E, Gaiani N, Sancelme M, Deguillaume L, et al. Long-term features of cloud microbiology at the puy de Dôme (France). *Atmospheric Environment* 2012;56:88–100.
87. González-Martín C, Pérez-González CJ, González-Toril E, Expósito FJ, Aguilera Á, et al. Airborne bacterial community composition according to their origin in Tenerife, Canary Islands. *Front Microbiol* 2021;12:732961.

88. Herschend J, Raghupathi PK, Røder HL, Sørensen SJ, Burmølle M. Draft genome sequences of two *Kocuria* isolates, *K. salsicia* G1 and *K. rhizophila* G2, isolated from a slaughterhouse in Denmark. *Genome Announc* 2016;4:e00075-16.
89. Sichtig H, Minogue T, Yan Y, Stefan C, Hall A, *et al.* FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science. *Nat Commun* 2019;10:3313.
90. Putonti C, Thomas-White K, Crum E, Hilt EE, Price TK, *et al.* Genome investigation of urinary *Gardnerella* strains and their relationship to isolates of the vaginal microbiota. *mSphere* 2021;6:e00154-21.
91. Zhang C, Song W, Ma HR, Peng X, Anderson DJ, *et al.* Temporal encoding of bacterial identity and traits in growth dynamics. *Proc Natl Acad Sci USA* 2020;117:20202–20210.
92. Magnúsdóttir S, Saraiva JP, Bartholomäus A, Soheili M, Toscan RB, *et al.* Metagenome-assembled genomes indicate that antimicrobial resistance genes are highly prevalent among urban bacteria and multidrug and glycopeptide resistances are ubiquitous in most taxa. *Front Microbiol* 2023;14:1037845.
93. Mitchell AL, Almeida A, Beracochea M, Boland M, Burgin J, *et al.* MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res* 2020;48:D570–D578.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org