Title

Genome sequencing and multifaceted taxonomic analysis of novel strains of violacein-producing bacteria and non-violacein-producing close relatives

Journal

Microbial Genomics, 9(4)

ISSN

2057-5858

Authors

De León, Marina Estella
Wilson, Harriet S
Jospin, Guillaume
et al.

Publication Date

2023-04-13

DOI

10.1099/mgen.0.000971

MICROBIAL GENOMICS

RESEARCH ARTICLE

De León *et al.*, *Microbial Genomics* 2023;9:000971

DOI 10.1099/mgen.0.000971

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Genome sequencing and multifaceted taxonomic analysis of novel strains of violacein-producing bacteria and non-violacein-producing close relatives

Marina Estella De León[1,*], Harriet S. Wilson[2], Guillaume Jospin[3,4] and Jonathan A. Eisen[3,5,6]

Abstract

Violacein is a water-insoluble violet pigment produced by various Gram-negative bacteria. The compound and the bacteria that produce it have been gaining attention due to the antimicrobial and proposed antitumour properties of violacein and the possibility that strains producing it may have broad industrial uses. Bacteria that produce violacein have been isolated from diverse environments including fresh and ocean waters, glaciers, tropical soils, trees, fish and the skin of amphibians. We report here the isolation and characterization of six violacein-producing bacterial strains and three non-violacein-producing close relatives, each isolated from either an aquatic environment or moist food materials in northern California, USA. For each isolate, we characterized traditional phenotypes, generated and analysed draft genome sequences, and carried out multiple types of taxonomic, phylogenetic and phylogenomic analyses. Based on these analyses we assign putative identifications to the nine isolates, which include representatives of the genera *Chromobacterium*, *Aquitalea*, *Iodobacter*, *Duganella*, *Massilia* and *Janthinobacterium*. In addition, we discuss the utility of various metrics for taxonomic assignment in these groups including average nucleotide identity, whole genome phylogenetic analysis and extent of recent homologous recombination using the software program PopCOGenT.

DATA SUMMARY

Raw reads and genome assemblies for strains HSC-31F16, HSC-77S12, HSC-21Su07, HSC-16F04, HSC-15S17, HSC-2F05, HSC-65S10 and HSC-3S05 are available through NCBI under BioProject accession number PRJNA603836 and those for strain UCD_MED1 are available under BioProject accession number PRJNA546016. Sequence Read Archive (SRA) accession numbers are as follows: HSC-31F16=SRX11406228, HSC-77S12=SRX11406255, HSC-21Su07=SRX12453551, HSC-16F04=SRX11406350, HSC-15S17=SRX11355670, HSC-2F05=SRX11489442, HSC-65S10=SRX11406257, HSC-3S05=SRX11489443 and UCD_MED1=SRR17805880. These Whole Genome Shotgun projects have been deposited at DDBJ/ENA/GenBank under the following accessions (versions described in this paper in parentheses): HSC-31F16, JAAOMA000000000 (JAAOMA010000000); HSC-77S12, JAAOLZ000000000 (JAAOLZ010000000); HSC-21Su07, JACERN000000000 (JACERN010000000); HSC-16F04, JAAOLX000000000 (JAAOLX010000000); HSC-15S17, JAHTGR000000000 (JAHTGR010000000); HSC-2F05, JAHYBX000000000 (JAHYBX0100000000; HSC-65S10, JAAOLY000000000 (JAAOLY010000000); HSC-3S05, JAHYBY000000000 (JAHYBY010000000); UCD_MED1, VDGE00000000 (VDGE01000000). Commands and scripts are available at https://github.com/JonathanEisensLab/9Genomes_paper. The complete ANI output file as well as all supplementary files are also available from the github data repository. Table S1 and Figs S1 and S2 are available with the online version of this article.

Received 10 September 2022; Accepted 31 January 2023; Published 13 April 2023

Author affiliations: [1]Microbiology Graduate Group, University of California, Davis, CA, USA; [2]Department of Biological Sciences, Sierra College, Rocklin, CA, USA; [3]Genome Center, University of California, Davis, CA, USA; [4]AnimalBiome, Oakland, CA, USA; [5]Department of Medical Microbiology and Immunology, University of California, Davis, CA, USA; [6]Department of Evolution and Ecology, University of California, Davis, CA, USA.

***Correspondence:** Marina Estella De León, msdeleon@ucdavis.edu

Keywords: applied microbiology; bacterial species concept; classification and taxonomy; phylogenomics; *Proteobacteria*; microbial diversity; violacein; whole genome sequencing; Janthinobacterium.

Abbreviations: ANI, average nucleotide identity; MHA, Mueller–Hinton agar; NA, nutrient agar; R2A, Reasoner's 2A; SIM, sulfide indole motility; SSD, sum of squared differences; TSA, tryptic soy agar; TSB, tryptic soy broth; YEM, yeast extract malt.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Two supplementary figures and one supplementary tables are available with the online version of this article.

Impact Statement

Bacterial pigments are a subject of investigation that can have a positive impact on economies, the environment and human health. Bacterial pigments are becoming important pharmaceutical and industrial chemicals because the production process of synthetic dyes used for textiles and foods requires hazardous chemicals generating hazardous wastes and dangerous working conditions for employees. Due to their promising applications, the discovery of new bacteria and new bacterial pigments is at the forefront of scientific research and economics. The violet pigment violacein is gaining attention from the scientific community because of its various antibiotic activities, which can be used for human and animal health, and its brilliant colour which can be used for industrial purposes. Pigmented bacteria can be isolated from natural environments, especially freshwater and marine habitats, and should be studied in greater detail because out of all the secondary metabolites having antibiotic activity, pigments are an understudied group. Nine new strains of bacteria were isolated from the environment, including a purple *Janthinobacterium* sp. strain which was observed growing on tofu. Of the nine strains characterized, six showed purple growth and genomic analysis revealed that at least one copy of all of the five genes (*vioABCDEF*) needed to produce violacein were present in each of these genomes. The remaining three strains were closely related to violacein producers, but they do not contain violacein genes nor did they exhibit pigmented colonies. The taxonomy and evolution of bacterial strains that are capable of expressing violacein are still not well understood even though the biosynthesis of the compound has been clearly demonstrated. There is currently no standardized method of taxonomic classification of bacteria to the species level, making correct species identification difficult. Classical methods include metabolic phenotyping, and 16S rRNA gene sequence analysis. Current classification methods include whole genome sequence analysis using average nucleotide identity, and phylogenomic methods based on established gene marker sets. We used classical methods and the currently accepted genomic methods with additional genomic population ecology approaches to maximize confidence in taxonomic placement of the nine novel strains and to accurately assign strains to existing species and describe five novel species of bacteria. This approach contributed to a better understanding of evolutionary relationships between pigmented bacteria and how pigment genes such as for violacein are transferred and retained between and within populations.

INTRODUCTION

The violet pigment violacein is a secondary metabolite produced by bacteria from at least 11 known genera including *Chromobacterium, Janthinobacterium, Iodobacter, Duganella, Collimonas, Pseudoalteromonas, Massilia, Pseudoduganella, Archangium, Microbulbifer* and *Chitinimonas* [1–7]. Such violacein-producing bacteria have been isolated from environments including fresh and ocean waters, soil, ice, and the surfaces of plants, fish and amphibians [3]. Violacein is known to have multiple antimicrobial properties, including antifungal effects [1, 8–10]. Violacein is not essential to cell growth; however, its production probably gives the bacteria some competitive advantages due to its antibacterial, antifungal and antiprotozoal properties [2, 11].

In an effort to better understand the diversity and evolution of violacein-producing bacteria, we sequenced the genomes of nine bacterial isolates. Eight isolates were selected from a culture collection created and maintained at Sierra College; these were chosen because they either produced violet-coloured colonies (five isolates) or were inferred, based on 16S rRNA gene sequence analysis, to be closely related to known violacein-producing strains (three isolates). Additional criteria for selection of these specific isolates included culture viability and access to metadata about the strains. In addition to these eight isolates from the Sierra College collection, one violet-coloured strain isolated from refrigerated tofu (purchased fresh from a farmers market, not packaged) was also selected for whole genome sequencing and analysis. In addition to genome sequencing, observation of cell and colony morphology, plus metabolic testing was used to assist in characterizing the nine isolates.

Proper bacterial species designation can be critical for many purposes including public health, clinical, food safety and biosafety applications because recognizing and mapping out evolutionary relationships can lead to a better understanding of the metabolic potential and survivability of organisms in a given niche [12]. Analysis of the genome data of the nine new isolates revealed that some of the strains could be relatively easily placed into a taxonomic and phylogenetic context with regard to other known bacteria. However, for some of the isolates, taxonomic and phylogenetic placement was not straightforward. In particular, it was challenging to make lower level assignments for these isolates – and especially challenging to make any type of clear species identification.

We discuss various approaches to taxonomic determination for these new isolates including using BLASTn analysis of 16S rRNA genes, average nucleotide identity (ANI), whole genome phylogenies and measures of horizontal versus vertical gene flow (using the PopCOGenT program). Based on a combination of these diverse approaches, we propose novel species designations with high confidence for five of the isolates: *Chromobacterium perflumen* strain HSC-31F16, *Aquitalea aquatica* strain HSC-21Su07, *Iodobacter violacea* strain HSC-16F04, *Massilia hydrophila* strain HSC-2F05 and *Duganella violaceicalia* strain HSC-15S17. We tentatively classify the three new *Janthinobacterium lividum* strains HSC-65S10, HSC-3S05 and UCD_MED1 because this group

Table 1. Strain IDs, isolation details and colony coloration of bacteria in a study of traits, genomics and phylogenetic placement of violacein-producing bacteria and related taxa

Strain isolation location, initial culturing conditions and colony colour. Unique strain IDs are followed by NCBI RefSeq whole genome assembly accession numbers.

Strain ID (RefSeq ID)	Sampling location	Culturing notes	Colony colour
HSC-31F16 (GCF_011602385.1)	American River in Fair Oaks, CA (38.63 N 121.26 W)	Isolated on TSA plates	Pale pinkish-cream
HSC-77S12 (GCF_011601295.1)	Freshwater pond near Cool, CA (38.8 N 121.0 W)	Isolated on TSA plates	Deep violet
HSC-21Su07 (GCF_013911085.1)	Bitney Springs near Nevada City, CA (39.23 N 121.10 W)	Isolated on TSA plates	Milky-tan
HSC-16F04 (GCF_011601265.1)	Hand-dug well in Ophir, CA (38.8 N 121.1 W)	Isolated on NA plates	Violet
HSC-15S17 (GCF_019166075.1)	Water dripping from fern, Point Reyes National Seashore, CA (38.03 N 122.48 W)	Isolated on mannitol-rich, nitrogen-free agar	Milky-beige colonies that became violet with age
HSC-2F05 (GCF_020208145.1)	Freshwater pond in Roseville, CA (38.77 N 121.32 W)	Isolated on TSA plates	Milky-beige
HSC-65S10 (GCF_011601275.1)	Freshwater pond in Rocklin, CA (38.78 N, 121.21 W)	Isolated on NA plates	Deep violet
HSC-3S05 (GCF_020208155.1)	Cooked rice in a home refrigerator in Rocklin, CA (38.78 N 121.20 W)	Isolated on NA plates	Deep violet
UCD_MED1 (GCF_006152155.1)	Violet-coloured tofu found in a home refrigerator in Davis, CA (38.559 N 121.698 W)	Isolated on R2A plates	Deep violet

presents problems in both taxonomy and annotations that require continued investigation, and we discuss the use of a mixed method approach for more accurate taxonomic classification within this group.

METHODS

Isolation of strains used in this study

Sierra College students and staff collected and isolated bacteria from various local environments over many years. A subset of these were then characterized in more detail and chosen for genome sequencing and analysis as a part of this study (the selection process is described in the Results and discussion section). General information about the selected isolates is presented in Table 1. The general growth conditions used for obtaining these isolates is summarized here. Bacterial growth conditions were similar to those previously used to isolate violacein-producing bacteria [13]. Specifically, environmental samples were inoculated onto media containing tryptophan such as tryptic soy agar (TSA), nutrient agar (NA) and/or Reasoner's 2A agar (R2A agar). One exception was strain HSC-15S17 that was initially isolated from mannitol-rich, nitrogen-free agar and subcultured on yeast extract malt (YEM) agar. Some water samples were collected in new Ziploc bags and transferred to TSA or NA plates using sterile pipettes (50–100 µl of water per plate). Other samples were collected with sterile cotton swabs that were used to inoculate plate media directly. Violet-coloured growth was taken from cooked rice using a sterile wire loop and inoculated onto NA plates. Violet-coloured growth was taken from tofu using a sterile loop and inoculated onto R2A plates. Cultures were incubated at room temperature resembling outdoor temperatures (18–20 °C). After incubation at room temperature for 2–3 days, colonies were subcultured for isolation. The subculturing of selected colonies on new agar plates was repeated several times until the strains were purified. All strains were routinely maintained, grown on TSA, NA, R2A or YEM agar media, and preserved as glycerol stocks with a final concentration of 20–25% at −80 °C.

The bacteria maintained during this investigation were frequently sub-cultured either from cryo-vials or isolated colonies and plated on TSA, NA or YEM media. Selective and differential plate media (5% sheep blood agar, MacConkey's agar and Tergitol-7 agar) were used for determining specific characteristics, while Mueller–Hinton agar (MHA) was used to conduct antimicrobial susceptibility testing as described below. All cultures were allowed to grow at room temperature until isolated colonies became visible (typically 48–72 h). Refrigeration or incubation at 37 °C was utilized under some circumstances. Plate cultures were examined visually to determine purity before samples were taken for sub-culturing or enzymatic testing. Colony form, margin, elevation, surface texture, optical character, pigmentation and size (mm) were recorded and variations developing over time were noted. Cell samples were aseptically transferred from isolated colonies on agar plates to new media (broth, agar deeps, agar slants or agar plates) using sterile wire loops. Specific media used for metabolic testing are described below (Bacterial phenotyping and metabolic testing). Tryptic soy broth (TSB) and yeast extract malt broth were used to establish broth cultures for antimicrobial

susceptibility testing. Lawn cultures on the appropriate plate media (MHA or YEM) were prepared from these using sterile cotton swabs.

DNA extraction and 16S rRNA gene PCR and sequencing

Chromosomal DNA for PCR amplification was extracted from fresh plate cultures by mixing visible cell masses with 500 µl Tris buffer (1 mM, pH 8.5) and boiling for 10 min to lyse the cells, denature enzymes present and liberate the template DNA. DNA samples were then used for PCR amplification of 16S rRNA genes as follows: 5 µl aliquots of template DNA were added directly to PCR mixtures containing 25 µl Taq Master Mix (Qiagen), 5 µl primer mix (10 µM Bacteria 8-forward 5′-AGAGTTTGATC-MTGGCTCAG-3′ and Enteric 1530-reverse 5′-AGGAGGTGATCCARCCGCA-3′) and 15 µl sterile water. The PCR specifications were: cycle 1, 94 °C for 4 min; cycles 2–36, 55 °C for 45 s, 72 °C for 2 min and 94 °C for 30 s; cycle 37, 55 °C for 45 s followed by 20 min at 72 °C and then 0 °C until tubes were removed from the device and placed in a freezer. Following PCR amplification and gel electrophoresis, the PCR products were purified using Qiagen DNA purification kits. Sanger sequencing was performed on a 3730 Capillary Electrophoresis Genetic Analyzer and ABI BigDye Terminator v3.1 chemistry at the UC Davis DNA Sequencing Facility. Sequencing primers included the PCR primers listed above as well as Internal 533-forward (5′-CCAGCACGCCGCG-GTAA-3′) and 907-reverse (5′-CCGTCAATTCMTTTRAGTTT-3′). Raw sequence data (.ab1 files) were opened with 4Peaks (RRID:SCR_000015) and traces were observed to determine quality and length (typically 800–900 nt). Reads generated with reverse primers were 'flipped' to obtain reverse complementary sequences and overlapping regions were compared visually. Nucleotide sequences were copied to text files for additional comparison and editing. The four sequences were concatenated and overlapping sections deleted. These 16S rRNA gene sequences were compared with reference RNA sequences (refseq_rna) available in the NCBI database using BLASTn to tentatively assign each isolate to a genus [14].

Genome sequencing, assembly and annotation

DNA for genome sequencing was obtained by transferring single colonies from plate cultures into 5 ml aliquots of TSB and incubating for 24 h on an orbital shaker (220 r.p.m.) at room temperature. DNA was extracted from these cultures using a Qiagen DNeasy Blood and Tissue kit following the manufacturer's instructions. The resulting DNA samples were quantified using a Qubit 4 Fluorometer. Due to the highly viscous biofilm produced by strain HSC-15S17, this strain was submitted to Microbes NG for DNA extraction and sequencing as a subculture on a YEM agar plate. The genomes of all the HSC strains were sequenced by Microbes NG (http://www.microbesng.uk) using Illumina sequencing (HiSeq 2500 or NovaSeq 6000) with 2×250 bp paired-end reads. These genomes were put through a standard analysis pipeline by Microbes NG. Microbes NG identifies the closest available reference genome using Kraken and maps the reads to this using BWA mem to assess the quality of the data. Microbes NG performed a *de novo* assembly of the reads using SPAdes v. 3.7 (December 2019) and mapped the reads back to the resultant contigs using BWA mem again to obtain additional quality metrics.

UCD_MED1 was sequenced at the UC Davis DNA Technologies Core Facility using Illumina Hiseq 4000 technology with a library preparation of PE150. Reads were trimmed for quality using BBDuk v. 37.02, and assembly was performed using the A5-miseq v. May-2019 assembly pipeline [15], which includes a universal Illumina adapter trimming step. CheckM was used to assess assembly quality and contamination. All genomes were eventually annotated by the Prokaryotic Genome Annotation Pipeline (PGAP) when submitted to NCBI [16–18].

Whole genome phylogeny

For phylogenetic analysis we created a data set containing the following: the nine new genomes introduced here, all *Gammaproteobacteria* genomes from the NCBI genome assembly database that were listed as being in the genera *Chromobacterium* (*n*=85), *Janthinobacterium* (*n*=85), *Duganella* (*n*=77), *Massilia* (*n*=71), *Aquitalea* (*n*=15) and *Iodobacter* (*n*=6). This set includes all the genera that the nine new genomes were tentatively assigned to based on 16S rRNA gene sequencing, and one genome to serve as an outgroup. The genome chosen to serve as an outgroup was that of *Archangium violaceum* (formerly *Cystobacter violaceus*) strain Cbvi76, in the order *Myxococcales*, which was selected because this bacterium is a distantly related member of the *Deltaproteobacteria* that is also known to produce violacein (and thus we were examining its genome for features related to violacein production). For this set of genomes, gene marker alignments were generated using GTDB-tk 1.3.0 [19]. Details of the GTDB marker set are available for download at https://data.ace.uq.edu.au/public/gtdb/data/releases/latest/. Then, a maximum likelihood tree was inferred using RAxML 8.2.11 [20], with the following settings: 250 bootstraps, PROTGAMMABLOSUM62 substitution matrix done on amino acids, random seeds for parsimony inference, and rapid bootstrap set to 8 and 47, respectively. Phylogenetic trees were viewed using Dendroscope version 3.5.9 [21].

Average nucleotide identity (ANI)

The whole-genome similarity metric ANI was calculated for all genomes using Fast ANI v1.32 [22]. Fast ANI estimated the ANI between each new genome represented here and all genomes in the NCBI database that share the same genus by using alignment-free approximate sequence mapping. ANI was used with caution as only one part of the entire process of taxonomic

classification because ANI measure does not strictly represent core genome evolutionary relatedness, as orthologous genes can vary widely between pairs of genomes compared [23]. A heat map of the ANI results was generated using the R packages gplots v3.1.1 [24] and Ape v5.4.1 [25] to incorporate the phylogenetic tree, and dplyr v1.0.5 [26] for data preparation and manipulation.

PopCOGenT (Populations as Clusters Of Gene Transfer)

To help delineate populations based on gene flow, an analysis using PopCOGenT was performed on a select subset of clades within the genus *Janthinobacterium* (https://github.com/philarevalo/PopCOGenT) [27]. For the PopCOGenT analysis, 68 *Janthinobacterium* genomes were chosen based on their close phylogenetic similarity to the type strain. After creating a phylogenetic tree, all of the genomes within the two major clades encompassing the three *Janthinobacterium* genomes presented in this study were included in the analysis. Only the phylogenetically most distantly related *Janthinobacterium* clades were excluded.

The subpopulation structure of genomes was investigated based on horizontal gene flow. The program PopCOGenT was used to compare the length of identical DNA sequences in genome pairs. The length bias parameter was calculated for each pair of genomes and a genome network was created to estimate horizontal gene flow between the nodes. The network was then clustered with Infomap to identify subpopulations with higher rates of horizontal gene flow.

Specifically, in PopCOGent (using the default window length of 1000 bp) the magnitude of the sum of squared differences (SSD) is measured to estimate genome homogenization. In addition to SSD, the 'initial divergence' is calculated, which is a measure of the total diversity seen in each genome comparison. The log of the SSD was plotted against the log of the initial divergence in ggplot2 (R version 4.2.1 [28] 3+492 'Prairie Trillium') to view distinct clusters of genomes with high and low rates of recombination [24, 28]. Log_{10} values were used to address skewness towards large values because compared to the large dataset very few points were much larger than the bulk of the data. We then ran a cluster module using Infomap [29] within PopCOGent to observe network formations between clonal pairs or unknown groups. The cluster output was viewed and manipulated using the network visualization software Gephi, with the parameters: layout=ForceAtlas2 [30].

Bacterial phenotyping and metabolic testing

Bacterial morphology includes both cellular (shape, endospore, flagella, inclusion bodies, Gram staining) and colonial (colour, dimensions, form) characteristics. The physiological and biochemical features include data on growth at different temperatures, pH values, salt concentrations or atmospheric conditions, growth in the presence of various substances such as antimicrobial agents, and data on the presence or activity of various enzymes, metabolization of compounds, etc. [31].

All media were initially maintained at room temperature, and data were collected and results determined according to published protocols [32]. Cell morphology was observed by using light microscopy with cells grown in sulfide indole motility medium (SIM) for 24–72 h at 18–24 °C. Observation of violet pigment production was used to infer violacein biosynthesis. Gram staining and KOH testing were performed as previously described [32, 33]. Motility was observed in wet mounts (cells in deionized water) magnified 450× and by stab inoculation in SIM medium (where growth away from the inoculation line indicates motility).

Metabolic testing for all isolates was conducted in triplicate using media and reagents prepared for a general microbiology teaching laboratory and as indicated by previously published species descriptions. Oxidase reactions were performed on filter paper moistened with a solution of *N*,*N*,*N*′,*N*′-tetramethyl-*p*-phenylenediamine and catalase activity was demonstrated using 3% hydrogen peroxide on glass slides. Mode of metabolism was determined using the oxidation/fermentation (O/F) test. Citrate utilization, urea hydrolysis and aesculin hydrolysis were determined on agar slants (Simmon's citrate agar, Christensen's urea agar and aesculin agar respectively). Gelatin hydrolysis was determined in nutrient gelatin deeps, starch hydrolysis was determined on starch agar plates with the addition of Gram's iodine, and methyl-red-Voges–Proskauer (MR-VP) tests were conducted in tubes of MR-VP broth media followed by transfer of 1 ml aliquots of 48–72 h growth to screw-top tubes, application of Barrit's reagents and vortex mixing to test for acetoin. Testing for acid production in MR-VP broth involved addition of six drops of methyl red indicator to cultures minus sample removed for the VP test. The formation of hydrogen sulphide was determined through observation of cultures stabbed into SIM medium, and indole formation was determined in SIM medium with the addition of Kovac's reagent. The reduction of nitrate to nitrite or to nitrogenous gasses was determined in nitrate agar deeps using nitrate reagent A (sulphanilic acid), reagent B (alpha-naphthylamine) and zinc powder (when required). The formation of acid aerobically from carbohydrates (arabinose, glucose, inositol, lactose, maltose, mannitol, raffinose, rhamnose, sorbitol, sucrose and xylose) was determined on the surface of agar deeps containing a bromothymol blue agar base with peptone [32]. Haemolysis reactions and growth on selective and differential media were determined on agar plates (5% sheep blood, MacConkey's, Tergitol-7 and nitrogen-free agar). The presence/absence of genes associated with selected metabolic processes was investigated using the Protein Family Sorter tool of PATRIC-BRC (now BV-BRC) [34]. Antimicrobial susceptibility was determined using the disc diffusion method (Kirby–Bauer test) conducted on MHA plates inoculated with lawn cultures. Because strain HSC-15S17 would not grow on MHA, antimicrobial susceptibility testing for this strain was conducted on YEM. Ampicillin resistance was also determined on TSA plates containing ampicillin. Antimicrobial

discs containing ampicillin (AM-10), bacitracin (B-10), kanamycin (K-30), penicillin G (P10), polymyxin (PB-300), rifampin (RA-5), streptomycin (S-10), tetracycline (TE-30) and vancomycin (VA-30) were applied with a disc dispenser (BD) or with sterile forceps.

Following metabolic testing as described above, single or sole carbon source assimilation was investigated using BiOLOG EcoPlates [35]. Inocula were prepared for each strain by suspending one loopful of cell material (3 mm ball) from a 48 h plate culture in 15 ml of sterile water within a screw-top tube. The cellular material was thoroughly suspended by alternately shaking, vortex-mixing and allowing the tubes to stand over a period of 10–15 min. Each EcoPlate was warmed to room temperature, inoculated under aseptic conditions (100 µl of cell suspension per well), closed, sealed inside a clean Ziploc bag and stored at room temperature. Carbon source assimilation was determined by directly observing each EcoPlate placed on the surface of a portable LED lightbox (plate lid and plastic bag removed) at daily intervals over a period of 4 weeks.

RESULTS

Selection for genome sequencing

Eight strains out of many opportunistically collected environmental isolates were selected. These isolates were considered candidates for potential violacein production. Five were selected because they formed violet- or deep violet-coloured colonies. Three isolates that did not form violet- coloured colonies were selected because 16S rRNA gene sequence analysis indicated they were closely related to known violacein producers (see rRNA analysis details in the next section). In addition, one strain isolated at UC Davis that produced violet-coloured colonies was also included in this study. Table 1 provides additional details about the selected isolates. The isolates chosen came from a diversity of freshwater environments and prepared foods (Table 1).

Key growth features of selected isolates

The cultures grown on plate media all formed colonies that were initially well isolated (Fig. 1); for some strains colony expansion and fusion occurred over time. *HSC-31F16* formed circular, entire, raised, smooth-shiny, semi-opaque, pale pinkish-cream colonies, 1–3 mm in diameter after 48 h on TSA. Colony surfaces became wrinkled with age. This strain also grew well on MAC and T-7 media but colonies were smaller and more translucent. *HSC-77S12* formed circular, entire, low-convexity, smooth-shiny, opaque, deep violet colonies, 1–3 mm in diameter after 48 h on TSA. It also grew well on MAC and T-7 media, but colonies were somewhat smaller. *HSC-21Su07* formed circular, entire, low-convexity, smooth-shiny, semi-opaque, milky-tan colonies 1–3 mm in diameter after 48 h on TSA. It also grew well on MAC and T-7 media but colonies were more translucent and smaller. *HSC-16F04* formed circular, entire, flat, smooth-shiny opaque to semi-translucent, violet colonies 2–4 mm in diameter after 48 h onNA (evidence of swarming appeared over time). It also grew well on TSA, MAC and T-7 media. Colonies formed on TSA were darker violet, more opaque and tended to swarm less than those grown on NA. *HSC-15S17* formed circular, entire, high-convexity, smooth-shiny, semi-opaque, milky-beige colonies that became violet with age and were 1–3 mm in diameter after 48 h on YEM. Colonies became rubbery and difficult to sample over time. This strain grew poorly on NA and not at all on MAC, T-7, MHA or TSA. *HSC-2F05* formed circular, entire, low-convexity, smooth-shiny, semi-opaque, milky-beige colonies 1–3 mm in diameter after 48 h on TSA. It also grew well on MAC and T-7 media but colonies were more translucent and smaller. *HSC-65S10* formed circular to irregular, entire to undulate, low-convexity, wrinkled-shiny, opaque, deep violet colonies 1–3 mm in diameter after 48 h on NA. It also grew well on MAC and T-7 media but new growth was cream-coloured when sub-cultured repeatedly on TSA. *HSC-3S05* formed circular, entire, convex, wrinkled-shiny, opaque, deep violet colonies 1–3 mm in diameter after 48 h on NA. It also grew well on MAC and T-7 media but new growth was cream-coloured when sub-cultured repeatedly on TSA. *UCD_MED1* formed circular to irregular, entire to undulate, low-convexty, wrinkled-shiny, opaque, deep violet colonies 1–3 mm in diameter after 48 h on NA. This strain also grew well on MAC and T-7 media but new growth became cream-coloured when sub-cultured repeatedly on TSA.

Each bacterial strain is cryopreserved at Sierra College and University of California, Davis. The strains are available from the corresponding author upon request. Strain HSC-15S17 is also cryopreserved in the ATCC US National Park Service Special Collection as TSD-229 and with the Culture Collection of Switzerland (CCOS). The other eight strains are available through the Leibniz Institute DSMZ and through the North Regional Research Laboratory (NRRL). Strains HSC-31F16, HSC-21Su07, HSC-16F04, HSC-15S17 and HSC-2F05 represent novel species for which the names *Chromobacterium perflumen* (type strain HSC-31F16=DSM 114492=NRRL B-65644), *Aquitalea aquatica* (type strain HSC-21Su07=DSM 114499=NRRL B-65646), *Iodobacter violacea* (type strain HSC-16F04=DSM 114489=NRRL B-65647), *Duganella violaceicalia* (type strain HSC-15S17=ATCC TSD-229 = CCOS 2056) and *Massilia hydrophila* (type strain HSC-2F05=DSM 114498=NRRL B-65648) are proposed. The strains of *Chromobacterium piscinae* (HSC-77S12=DSM 114490=NRRL B-65645), *Janthinobacterium lividum* (HSC-65S10=DSM 114488=NRRL B-65650), *Janthinobacterium lividum* (HSC-3S05=DSM 114487=NRRL B-65649) and *Janthinobacterium lividum* (UCD_MED1=DSM 114491=NRRL B-65651) represent previously described taxa.

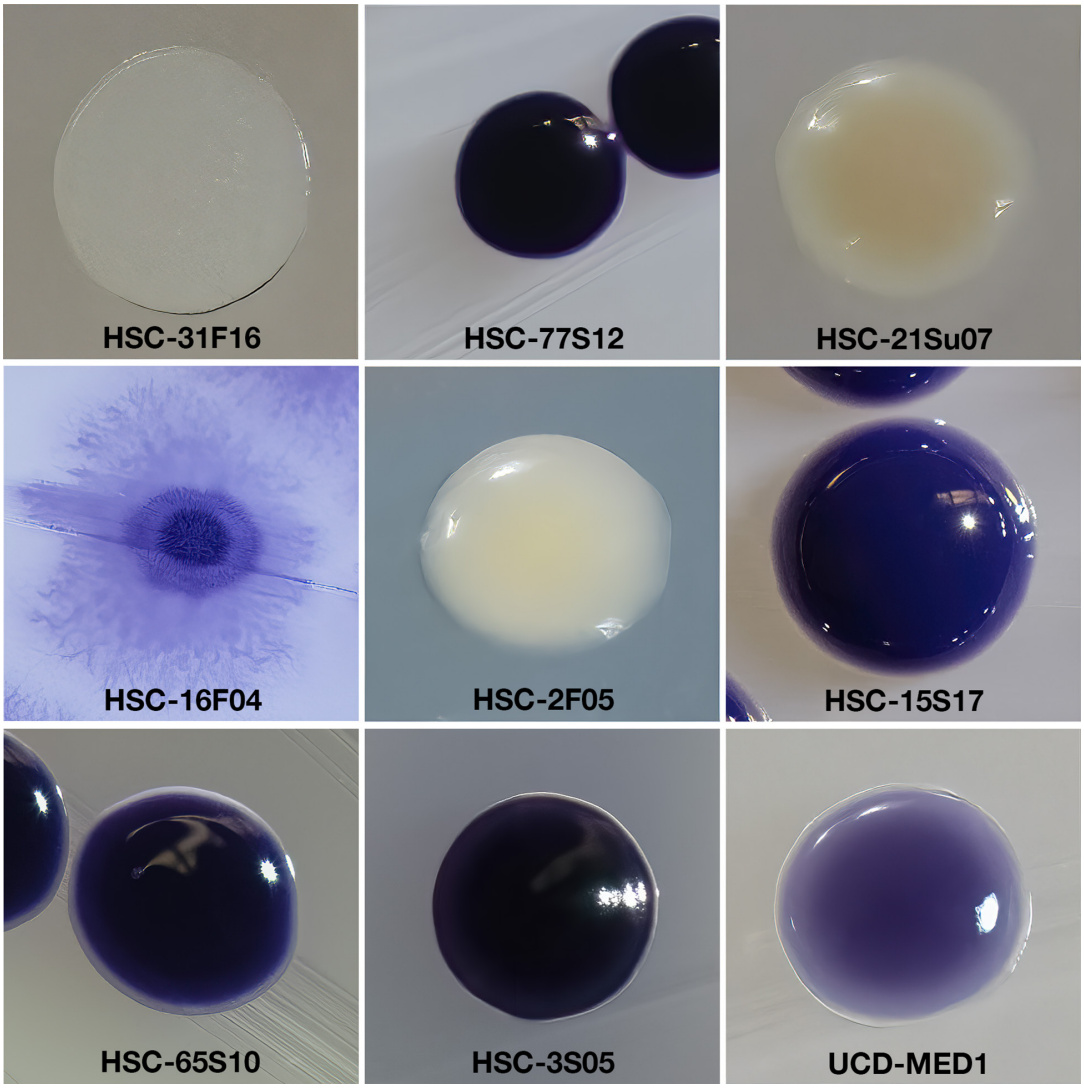

Fig. 1. Colony morphology of the nine new bacterial strains. Strain HSC-16F04 was cultured on NA, strain HSC-15S17 was cultured on YEM, and all other strains were cultured on TSA plates. Photos were taken after 48–72 h of growth at room temperature.

Cell morphology and metabolic testing

A summary of metabolic capabilities including antimicrobial susceptibility and single carbon-source assimilation as determined through the application of BiOLOG EcoPlates is presented in Table 2.

HSC-31F16

Cells were motile, Gram-negative (KOH-positive), non-sporing, fermentative bacilli, 0.7–1×2–5 µm in size. This strain was positive for catalase, oxidase, citrate utilization, aesculin and gelatin hydrolysis, reduction of nitrate to nitrite (but not to nitrogenous gases) plus beta-haemolysis on 5% sheep blood. It was negative for violacein formation, urea and starch hydrolysis, indole and hydrogen sulphide formation (SIM), formation of acid and acetoin in MR-VP medium and growth on nitrogen-free media. Acid was formed aerobically from inositol but none of the other carbohydrates provided; acid was formed through the fermentation of glucose (O/F).

HSC-77S12

Cells were motile, Gram-negative (KOH-positive), non-sporing, fermentative bacilli, 1–1.2×2–3 µm in size. This strain was positive for violacein formation, catalase, oxidase, citrate utilization, gelatin hydrolysis, the ability to reduce nitrate to nitrite (but not to nitrogenous gasses) and beta-haemolysis on 5% sheep blood. It was negative for urea, starch and aesculin hydrolysis, indole and hydrogen sulphide formation (SIM), formation of acid and acetoin in MR-VP medium and growth on nitrogen-free

Table 2. Results of standard metabolic tests including BiOLOG data for each strain

Results indicate consistency over three trials. +, Positive; –, negative; [+], weakly positive; (+), delayed positive and required refrigeration; R, resistant; S, susceptible; I, intermediate or susceptible with increased exposure [68]. (*), Data obtained on yeast extract malt medium (YEM).

Metabolic test performed	HSC-31F16	HSC-77S12	HSC-21Su07	HSC-16F04	HSC-15S17	HSC-2F05	HSC-65S10	HSC-3S05	UCD_MED1
Violacein	–	+	–	+	+	–	+	+	+
(Enzyme:									
Oxidase (cytochrome C oxidase)	+	+	+	–	+	+	+	+	+
Catalase (catalase)	+	+	+	+	+	+	+	+	+
Citrate utilization (citrate permease)	+	+	+	–	–	–	+	+	+
Urea hydrolysis (urease)	–	–	(+)	–	–	–	–	(+)	(+)
Aesculin hydrolysis (aesculinase)	+	–	–	–	–	–	+	+	+
Gelatin hydrolysis (gelatinase)	+	+	–	+	+	+	+	+	+
Starch hydrolysis (amylase)	–	–	–	–	+	+	–	–	–
Nitrate reduction (nitrate reductase)	+	+	+	+	+	+	+	+	+
Nitrite reduction (nitrite reductase)	–	–	–	–	–	–	–	+	+
Indole formation (tryptophanase)	–	–	–	–	–	–	–	–	–
Ampicillin resistance (β-lactamase)	+	+	–	+	+	–	+	+	+
Standard assessments									
Beta-hemolysis (haemolysin)	+	+	+	+	–	+	+	+	+
3% KOH (wall composition)	+	+	+	+	+	+	+	+	+
Growth on Tergitol-7 agar	+	+	+	+	–	+	+	+	+
Growth on MacConkey's agar	+	+	+	+	–	+	+	+	+
Methyl red (acid from glucose)	–	–	–	–	–	–	–	–	–
Voges-Proskauer (acetoin formation)	–	–	–	–	–	–	–	–	–
Glucose acid (O/F test)	+	+	+	+	–	–	+	+	+
Motility (SIM)	+	+	+	+	+	+	+	+	+
Hydrogen sulphide production	–	–	–	–	–	–	–	–	–
Acid production (aerobically) from:									
Arabinose	–	–	–	–	–	–	+	+	+
Inositol	+	–	+	–	–	–	+	+	+
Lactose	–	–	–	–	–	–	–	–	–

Continued

Table 2. Continued

Metabolic test performed	HSC-31F16	HSC-77S12	HSC-21Su07	HSC-16F04	HSC-15S17	HSC-2F05	HSC-65S10	HSC-3S05	UCD_MED1
Maltose	–	–	–	+	–	–	+	+	+
Mannitol	–	–	–	–	–	–	+	+	+
Raffinose	–	–	–	–	–	–	–	–	–
Rhamnose	–	–	–	–	–	–	+	+	+
Sorbitol	–	–	–	–	–	–	+	+	+
Sucrose	–	–	–	–	–	–	+	+	+
Xylose	–	–	–	–	–	–	+	+	+
Antimicrobial susceptibility:									
Ampicillin AM-10	R	R	S	R	R *	S	R	R	R
Bacitracin B-10	R	R	R	R	I *	S	R	R	R
Kanamycin K-30	R	S	S	S	S *	S	S	I	S
Penicillin P-10	R	R	R	R	R *	S	R	R	R
Polymyxin PB-300	S	R	R	R	I *	R	R	R	R
Rifampin RA-5	S	S	S	S	S *	S	S	S	S
Streptomycin S-10	R	R	S	I	S *	R	S	I	S
Tetracycline TE-30	S	S	R	S	S *	S	S	S	S
Vancomycin VA-30	R	R	R	R	S *	R	R	R	I
BiOLOG EcoPlate Data									
Carbon source assimilation:									
Pyruvic acid methyl ester	[+]	+	+	[+]	–	[+]	+	+	+
Tween 40	+	+	+	[+]	[+]	[+]	+	+	+
Tween 80	+	+	+	–	[+]	–	+	+	+
Alpha-cyclodextrin	–	–	–	–	[+]	[+]	–	–	–
Glycogen	+	+	–	–	+	[+]	–	–	–
D-Cellobiose	[+]	[+]	[+]	–	–	–	+	+	+
Alpha-D-lactose	–	[+]	[+]	–	+	–	–	+	–
Beta-methyl-D-glucoside	–	–	[+]	–	–	–	–	–	–
D-Xylose	–	–	[+]	–	–	–	+	+	+
L-Erythritol (meso-erythritol)	–	–	[+]	–	–	–	–	–	–

Continued

Table 2. Continued

Metabolic test performed	HSC-31F16	HSC-77S12	HSC-21Su07	HSC-16F04	HSC-15S17	HSC-2F05	HSC-65S10	HSC-3S05	UCD_MED1
D-Mannitol	–	–	[+]	–	–	–	+	+	+
N-Acetyl-D-glucosamine	+	+	+	[+]	[+]	–	+	+	+
D-Glucosaminic acid	–	–	[+]	–	–	–	–	–	–
Glucose 1-phosphate	+	+	[+]	[+]	–	–	+	+	+
D,L-Alpha-glycerol phosphate	+	+	[+]	–	–	–	+	+	+
D-Galactonic acid gamma-lactone	–	–	[+]	–	–	–	–	–	–
D-Galacturonic acid	–	–	[+]	–	[+]	–	–	–	–
2-Hydroxy benzoic acid	–	–	–	–	–	–	–	–	–
4-Hydroxy benzoic acid	–	–	–	–	–	–	–	–	–
Gamma-hydroxybutyric acid	+	–	+	–	–	–	–	[+]	–
Itaconic acid	–	–	–	–	–	–	–	–	–
Alpha-ketobutyric acid	[+]	+	[+]	–	–	–	–	–	–
D-Malic acid	[+]	+	+	–	–	–	+	–	–
L-Arginine	+	+	+	–	–	–	[+]	+	–
L-Asparagine	+	+	+	–	–	[+]	+	+	+
L-Phenylalanine	+	+	+	–	–	–	[+]	[+]	[+]
L-Serine	+	+	+	–	–	[+]	+	+	+
L-Threonine	+	+	+	[+]	–	[+]	[+]	+	[+]
Glycyl-L-glutamic acid	+	+	+	[+]	–	[+]	+	+	+
Phenylethylamine	–	[+]	–	–	–	–	–	–	–
Putrescine	+	[+]	+	–	–	–	–	–	–

media. Acid was not formed aerobically from any of the carbohydrates utilized, but acid was formed through the fermentation of glucose and maltose (O/F).

HSC-21Su07

Cells were motile, Gram-negative (KOH-positive), non-sporing, fermentative bacilli, 0.5–1.0×1.5–2.5 µm in size. This strain was positive for catalase, oxidase, citrate utilization, urea hydrolysis, the reduction of nitrate to nitrite (but not to nitrogenous gasses) and haemolysis on 5% sheep blood. It was negative for violacein formation, aesculin, starch and gelatin hydrolysis, indole and hydrogen sulphide formation (SIM), the formation of acid and acetoin in MR-VP medium and growth on nitrogen-free media. Acid was formed aerobically from inositol but none of the other carbohydrates utilized; acid was formed through the fermentation of glucose in O/F medium.

HSC-16F04

Cells were motile, Gram-negative (KOH-positive), non-sporing, fermentative bacilli, 0.7–1.0×3–5 µm in size. This strain was positive for violacein formation, catalase, gelatin hydrolysis, reduction of nitrate to nitrite (but not to nitrogenous gasses) and beta-haemolysis on 5% sheep blood. It was negative for oxidase, citrate utilization, urea, starch and aesculin hydrolysis, indole and hydrogen sulphide formation (SIM), the formation of acid and acetoin in MR-VP medium and growth on nitrogen-free media. Acid was formed aerobically from maltose but none of the other carbohydrates utilized; acid was formed through the fermentation of glucose in O/F medium.

HSC-15S17

Cells were motile, Gram-negative (KOH-positive), non-sporing, oxidative coccobacilli, 1.0–1.5×1.5–2 µm in size. This strain was positive for violacein formation, catalase, oxidase, gelatin liquefaction, starch hydrolysis, the reduction of nitrate to nitrite (but not to nitrogenous gases) and growth on nitrogen-free media. It was negative for citrate utilization, urea and aesculin hydrolysis, indole and hydrogen sulphide production (SIM) and the formation of acid and acetoin from glucose fermentation (O/F and MR-VP). It was not haemolytic on sheep blood agar and acid was not formed aerobically from any of the carbohydrates utilized.

HSC-2F05

Cells were motile, Gram-negative (KOH-positive), non-sporing, oxidative bacilli, typically 0.7–1.0×2–5 µm in size (some reached 20 µm in length). This strain was positive for catalase, oxidase, gelatin liquefaction, starch hydrolysis, the reduction of nitrate to nitrite (but not to nitrogenous gasses) and haemolysis on sheep blood agar. It was negative for violacein formation, citrate utilization, urea and aesculin hydrolysis, indole and hydrogen sulphide production (SIM), the formation of acid and acetoin from glucose fermentation (O/F and MR-VP) and growth on nitrogen-free media. Acid was not formed aerobically from any of the carbohydrates utilized.

HSC-65S10

Cells were motile, Gram-negative (KOH-positive), non-sporing, oxidative bacilli, typically 0.5–1.0×3–5 µm in size (some reached 20 µm in length). This strain was positive for violacein formation, catalase, oxidase, citrate utilization, aesculin and gelatin hydrolysis, the ability to reduce nitrate to nitrite (but not to nitrogenous gases), beta-haemolysis on 5% sheep blood and growth on nitrogen-free media. It was negative for urea and starch hydrolysis, indole and hydrogen sulphide formation (SIM), and the formation of acid and acetoin in MR-VP medium. Acid was formed aerobically from arabinose, glucose, inositol, maltose, mannitol, rhamnose, sorbitol, sucrose and xylose in O/F-type media, but not from lactose or raffinose.

HSC-3S05

Cells were motile, Gram-negative (KOH-positive), non-sporing, oxidative bacilli, typically 0.5–1.0×3–5 µm in size (some reached 15–20 µm in length). This strain was positive for violacein formation, catalase, oxidase, citrate utilization, aesculin, gelatin and urea hydrolysis, the ability to reduce nitrate and nitrite, beta-haemolysis on 5% sheep blood and growth on nitrogen-free media. It was negative for starch hydrolysis, indole and hydrogen sulphide formation (SIM) and the formation of acid and acetoin in MR-VP medium. Acid was formed aerobically from arabinose, glucose, inositol, maltose, mannitol, rhamnose, sorbitol, sucrose and xylose in O/F-type media, but not from lactose or raffinose.

UCD_MED1

Cells were motile, Gram-negative (KOH-positive), non-sporing, oxidative bacilli, typically 0.5–1.0×3–5 µm in size (some reached 20 µm in length). This strain was positive for violacein formation, catalase, oxidase, citrate utilization, aesculin, gelatin and urea hydrolysis, the ability to reduce nitrate and nitrite, beta-haemolysis on 5% sheep blood and growth on nitrogen-free media. It was negative for starch hydrolysis, indole and hydrogen sulphide formation (SIM), and the formation of acid and acetoin in MR-VP medium. Acid was formed aerobically from arabinose, glucose, inositol, maltose, mannitol, rhamnose, sorbitol, sucrose and xylose in O/F-type media, but not from lactose or raffinose.

Table 3. 16S rRNA gene sequence BLASTn percentage identities and putative genus assignments

Strain ID	Closest NCBI BLASTn match	Accession no.	Percentage identity	Genus assignment
HSC-31F16	*Chromobacterium aquaticum* CC-SEYA-1	NR_044405	98.92	*Chromobacterium*
HSC-77S12*	*Chromobacterium piscinae* LMG 3947	NR_114953	99.93	*Chromobacterium*
HSC-21Su07	*Aquitalea denitrificans* 5YN1-3	NR_044535	99.22	*Aquitalea*
HSC-16F04*	*Iodobacter arcticus* AsdM4-16	NR_108462	98.56	*Iodobacter*
HSC-15S17*	*Duganella zoogloeoides* IAM 12670	NR_025833	98.55	*Duganella*
HSC-2F05	*Massilia varians* CCUG 35299	NR_042652	98.32	*Massilia*
HSC-65S10*	*Janthinobacterium lividum* DSM 1522	NR_026365	99.93	*Janthinobacterium*
HSC-3S05*	*Janthinobacterium lividum* DSM 1522	NR_026365	99.93	*Janthinobacterium*
UCD_MED1*	*Janthinobacterium lividum* DSM 1522	NR_026365	99.80	*Janthinobacterium*

*Violacein-producing bacteria.

16S rRNA gene sequencing and tentative genus assignment

Nearly complete 16S rRNA gene sequences were obtained for strains HSC-31F16 (1527 bp), HSC-77S12 (1527 bp), HSC-21Su07 (1527 bp), HSC-16F04 (1524 bp), HSC-15S17 (1520 bp), HSC-2F05 (1526 bp), HSC-65S10 (1522 bp), HSC-3S05 (1522 bp) and UCD_MED1 (1522 bp). The 16S gene in bacteria is generally about 1550 bp. These sequences were then compared using BLASTn searches to 16S rRNA gene sequences in the NCBI RefSeq database (Table 3). The results from 16S rRNA gene BLASTn comparisons showed that, though they did not produce violet-coloured colonies, strains HSC-31F16, HSC-21Su07 and HSC-2F05 were closely related to known violacein producers and thus merited further analysis. The results also showed that the six isolates that produced violet colonies (HSC-77S12, HSC-16F04, HSC-15S17, HSC-65S10, HSC-3S05 and UCD_MED1) were in groups recognized as violacein producers. In addition, rRNA gene sequence analysis was used to give each isolate a tentative genus assignment. The 16S rRNA gene sequences were submitted to the NCBI GenBank and assigned the following accession numbers: strain HSC-31F16=MW242637, strain HSC-77S12=ON012846, strain HSC-21Su07=ON013926, strain HSC-16F04=ON013928, strain HSC-15S17=MT579559, strain HSC-2F05=ON013927, strain HSC-65S10=ON012847, strain HSC-3S05=ON013781 and strain UCD_MED1=ON012958.

Genome sequencing and assembly

All isolates were sequenced by Microbes NG except UCD_MED1 which was sequenced by UC Davis (see Methods for details). Results from the whole genome shotgun sequencing and genome assembly for each isolate are shown in Table 4.

Whole genome phylogeny

We created a phylogenetic tree that shows the placement of each of the nine new genomes within the phylogenetic context of a total of 349 genome assemblies. The tree consists of all bacterial strains of the same genera as the new strains presented here which are available within the NCBI database, plus one outgroup. This phylogenetic assemblage included genomes from 87 *Chromobacterium* strains, 16 *Aquitalea* strains, seven *Iodobacter* strains, 78 *Duganella* strains, 72 *Massilia* strains, 88 *Janthinobactertium* strains (incorporating three strains labelled as *Janthinobacterium* but are of inconclusive genus identity) and one *Archangium* strain.

Here we show relevant subsets of each major clade a new strain is grouped within (Fig. 2a–g). Taxonomic annotation (incorrect/misidentified taxonomic annotations are placed in parentheses), strain ID and the GenBank accession number for each genome assembly are included in the branch labels. Bootstrap values are indicated on the branch nodes, and the scale bar (upper left corner) defines the scale of branch lengths. In this case, a line segment with the number '0.01' or '0.001' shows the length of a branch that represents an amount of genetic change of 0.01 (1.0%) or 0.001 (0.1%). Type strains which are descendants of the original isolates used in species and subspecies descriptions, as defined by the Bacteriological Code, and exhibit all of the relevant phenotypic and genotypic properties, and are cited in the original published taxonomic circumscriptions are indicated as [TYPE]

Table 4. Whole genome sequencing and assembly statistics

Sample ID	Mean coverage	No. of reads	No. of reads w/ insert size >300	No. of contigs	Total length (bp)	GC (%)	N50
HSC-31F16	35.07	442350	370682	127	5706489	62.32	115329
HSC-77S12	55.06	639334	426745	49	4929710	63.57	215656
HSC-21Su07	71.02	774903	465627	43	4455205	59.6	209150
HSC-16F04	88.62	1097144	544750	39	4776933	49.41	385521
HSC-15S17	75.85	1212392	1133660	108	7475607	64.09	332609
HSC-2F05	118.62	1206959	969307	40	4686940	66.7	531308
HSC-65S10	46.83	735531	406192	57	6358340	62.58	273432
HSC-3S05	97.11	1277461	1071991	18	6225449	62.47	1041639
UCD_MED1	93.0	6800000	NA	53	6382737	62.87	451255

in each phylogenetic tree. The entire phylogenetic tree containing all publicly available genomes from the six genera in NCBI plus the outgroup used to root the tree, *Archangium violaceum* (349 taxa in total), can be found in Fig. S1.

Implications of the patterns seen in the whole genome phylogeny are discussed in the Discussion section below.

Average nucleotide identity

ANI values were determined for the nine draft genome sequences by comparison with all available genome assemblies annotated as bacteria classified within the same genera. Table 5 shows the strains with the highest ANI values against the strains in this study as well as the type strains with the highest ANI value. The full ANI dataset is available in Table S2.

Comparison of ANI to phylogenetic relationships

A heatmap of the ANI values, with strains ordered by their position in the whole genome phylogeny, was generated (Figs 3a–f, and S2). This ANI heatmap dendrogram shows the association between ANI and each pairwise comparison of genomes in the dendrogram.

Average nucleotide identity percentage (ANI%) for genome assemblies within the six genera representing the strains introduced in this study were determined, where colouring is used to represent increments of ANI percentage. Maroon/brown=95% and above, indicating the standard species-level differentiation. Black=99–99.9%, dark brown=98–98.9%, light brown=97–97.9%, dark orange=96–96.9%, dark pink=95–95.9%, orange=94–94.9%, light orange=93–93.9%, yellow=92–92.9%, light blue=91–91.9%, periwinkle=90–90.9% and white=<90%. Horizontal black lines indicate the position of a new strain introduced here.

Identifying and differentiating distinct populations within the genus *Janthinobacterium*

For the strains that had been assigned to the genera *Aquitalea*, *Chromobacterium*, *Iodobacter*, *Duganella*, and *Massilia*, ANI analysis and whole genome phylogenetics was considered sufficient for making taxonomic assignments (see Discussion). However, for the strains assigned to the genus *Janthinobacterium*, ANI analysis and whole genome phylogenetics was inconclusive due to ambiguous patterns of ANI% versus clusters in the tree. We therefore applied an additional type of analysis, using the program PopCOGenT to augment characterization of the *Janthinobacterium* group. PopCOGenT measures recent gene flow and recombination among genomes and thus serves as a 'reverse ecology' approach that can aid in the identification of genetic boundaries between isolates that are not discernible from standard ANI and phylogenomic methods [27].

PopCOGenT is designed to assess population structure under the assumption that organisms within the same population, compared to organisms that are genetically isolated from each other, should have higher rates of recombination and genomes with higher percentages of regions that have undergone recent recombination. It assesses the rates and extent of recombination by measuring levels of divergence between genomes as well as the lengths of identical stretches of DNA. Recent recombination between lineages should lead to an overabundance in number and length of identical regions between two genomes from those lineages (because there will not have been as much time to accumulate substitutions as compared to non-recombining regions). The method estimates recombination by comparing the lengths of identical regions between genome pairs versus an expectation under a Poisson model of just clonal mutation. This parameter is referred to as 'length bias' and is reported as 'Observed SSD' with higher values indicating more recombination. Generally, length bias is positively (although not linearly) correlated with the amount of gene flow between each pairwise genome comparison [27]. In addition, PopCOGenT measures the 'initial divergence', which is a measure of the total diversity seen in each genome comparison.

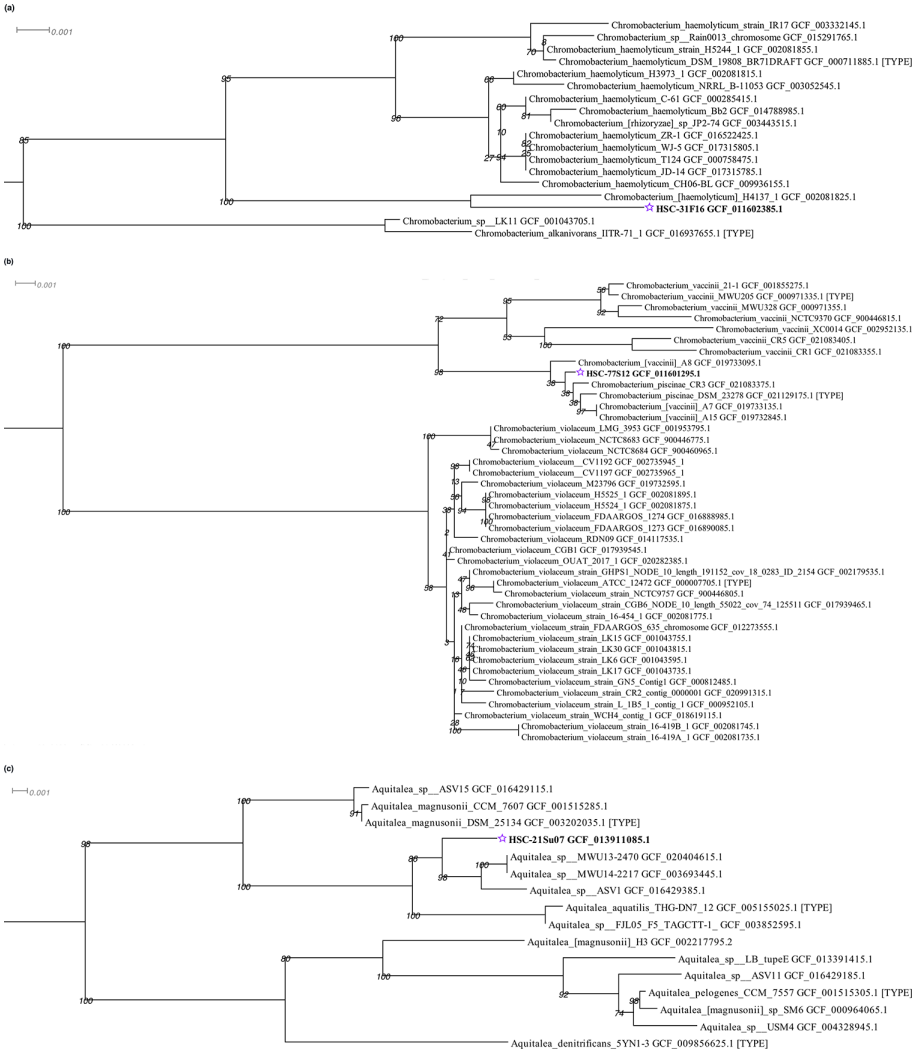

Fig. 2. Continued

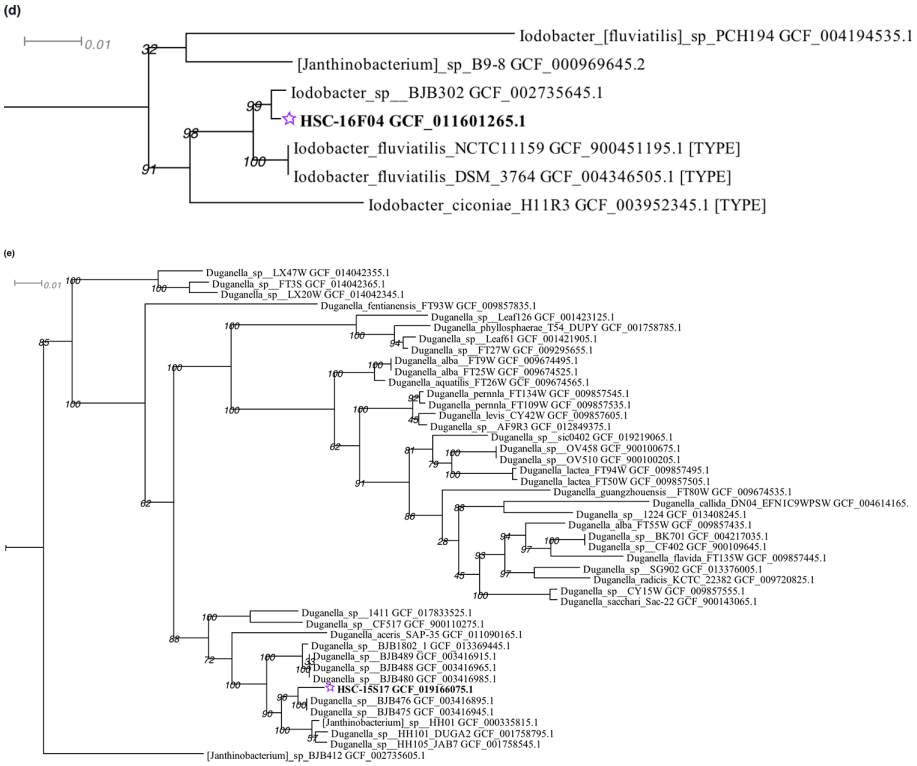

Fig. 2. Continued

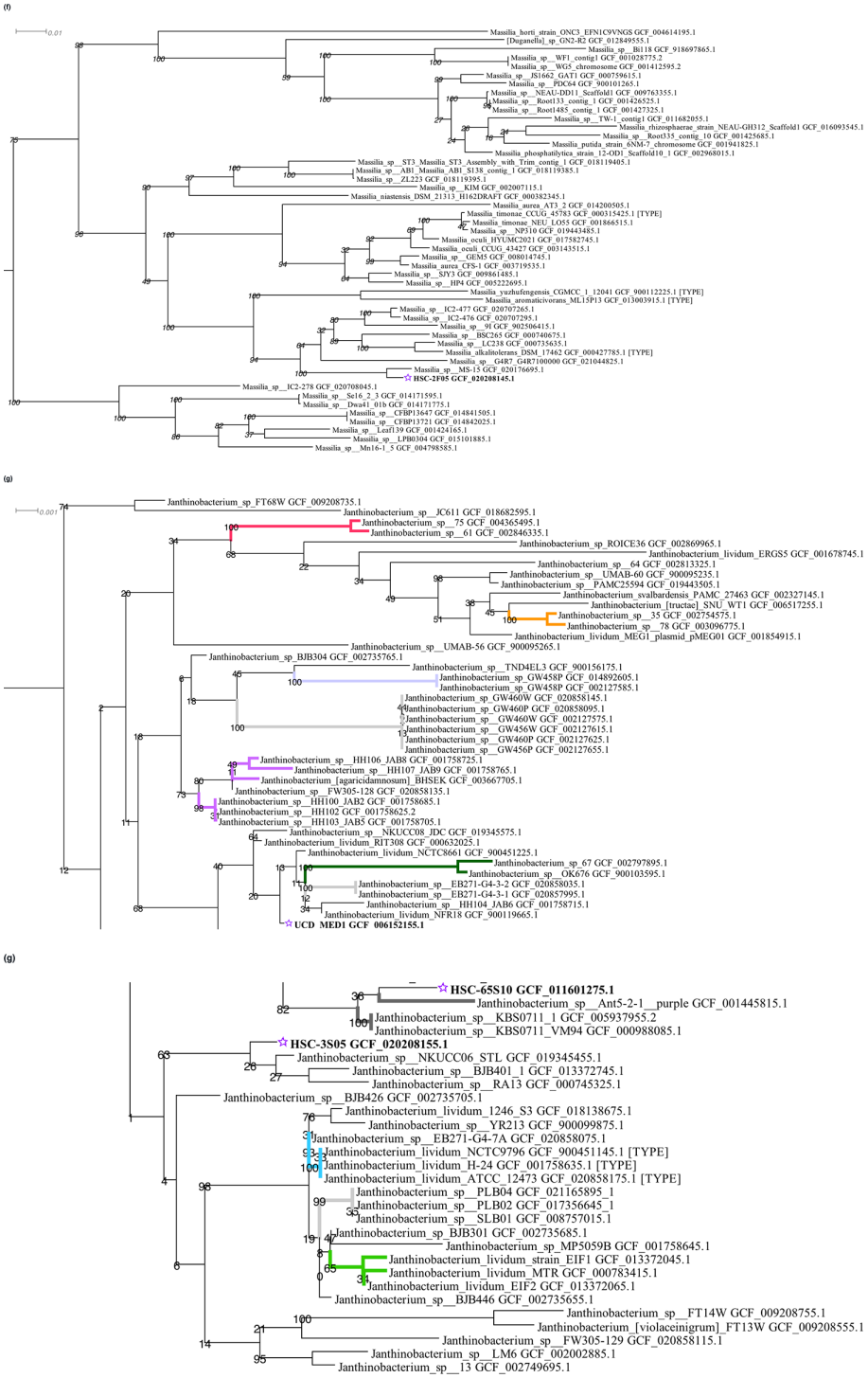

Fig. 2. The subsets shown were extracted from a GTDB-tk alignment-based whole genome tree that includes all publicly available genome assemblies of the associated genera in the NCBI database. (a) Phylogenetic placement of strain HSC-31F16 and surrounding taxa. (b) Phylogenetic placement of strain HSC-77S12 and surrounding taxa.(c) Phylogenetic placement of strain HSC-21Su07 and surrounding taxa. (d) Phylogenetic placement of strain HSC-16F04 and surrounding taxa. (e) Phylogenetic placement of strain HSC-15S17 and surrounding taxa. (f) Phylogenetic placement of strain HSC-2F05 and surrounding taxa. (g) Phylogenetic placement of strains HSC-65S10, HSC-3S05 and UCD_MED1 and surrounding taxa. The genomes shown in this figure were also used for PopCOGenT analysis. The colours of the branches correspond to the colours of the nodes in Figure 5, indicating eight distinct populations. Light grey branches represent clonal complexes that did not form networks.

Table 5. Summary of highest ANI (%) scores

ANI was calculated for each new strain presented here. A summary of the bacterial strains with the highest matching ANI scores to each of the nine new strains is listed. The ANI score with the type strain of the closest matching species is also listed.

Genome ID	Highest ANI for any genome	ANI (%)	Highest ANI for any type strain	ANI (%)
Chromobacterium HSC-31F16 GCF_011602385.1	*Chromobacterium* sp. H4137 GCF_002081825.1	97.46	*Chromobacterium haemolyticum* DSM 19808 GCF_000711885.1	93.92
Chromobacterium HSC-77S12 GCF_011601295.1	*Chromobacterium sp.* A7 GCF_019733135.1	98.85	*Chromobacterium piscinae* DSM 23278 GCF_021129175.1	98.82
Aquitalea HSC-21Su07 GCF_013911085.1	*Aquitalea* sp. MWU14-2470 GCF_020404615.1	94.57	*Aquitalea aquatilis* THG-DN7.12 GCF_005155025.1	92.37
Iodobacter HSC-16F04 GCF_011601265.1	*Iodobacter* sp. BJB302 GCF_002735645.1	94.89	*Iodobacter fluviatilis* DSM 3764 GCF_004346505.1	92.13
Duganella HSC-15S17 GCF_019166075.1	*Duganella* sp. BJB476 GCF_003416895.1	93.78	*Duganella aceris* SAP-35 GCF_011090165.1	87.03
Massilia HSC-2F05 GCF_020208145.1	*Massilia* sp. MS-15 GCF_020176695.1	96.37	*Massilia alkalitolerans* DSM 17462 GCF_000427785.1	88.12
Janthinobacterium HSC-65S10 GCF_011601275.1	*Janthinobacterium* sp. Ant5-2-1 GCF_001445815.1	96.42	*Janthinobacterium lividum* NCTC9796 GCF_900451145.1	93.27
Janthinobacterium HSC-3S05 GCF_020208155.1	*Janthinobacterium* sp. NKUCC06 STL GCF_019345455.1	98.16	*Janthinobacterium lividum* NCTC9796 GCF_900451145.1	93.67
Janthinobacterium UCD MED1 GCF_006152155.1	*Janthinobacterium* sp. NCTC8661 GCF_900451225.1	97.58	*Janthinobacterium lividum* NCTC9796 GCF_900451145.1	93.36

One way of assessing the results of PopCOGenT analysis is to examine the relationship between initial divergence and SSD, which are shown for the *Janthinobacterium* genomes in Fig. 4. Another way of assessing PopCOGenT analysis is to infer networks of recent gene flow, which are shown in Fig. 5. The implications of the PopCOGenT analysis are addressed in the Discussion.

For all pairs of *Janthinobacterium* genomes, the $\log_{10}$ transformation of initial divergence (a measure of the total similarity of the genomes) is plotted versus the log of the SSD score (a measure of the length bias towards longer stretches of identical DNA, which is used as a measure of the extent of recombination). The pairwise comparisons represented by the points in the upper left quadrant of the plot are genomes that are very similar to each other and in some cases may be separate sequences of the same strain. They thus have low levels of initial divergence and very high levels of inferred SSD. The pairwise comparisons represented by the points in the lower right of the plot represent the genome pairs that are significantly more distantly related (note the large jump in log initial divergence) and these generally have lower SSD scores.

DISCUSSION

Taxonomic assignment based on a combination of analyses

A multifaceted approach, including a combination of chemotaxonomic, phenotypic and genotypic data, was used to determine the taxonomic and phylogenetic positions of the nine isolates introduced here. Initial 16S rRNA gene sequence analysis allowed strains to be assigned a putative genus-level taxonomic classification. Species-level classification was then confirmed or designated using a combination of whole genome phylogenetic placement and ANI analysis. ANI values of approximately 95% are proposed by some to represent an accurate threshold for demarcating a species boundary for some bacteria [36, 37]. However, measures of similarity, even when taken across a whole genome, can be misleading regarding relatedness due to factors such as selection, mutation bias, unequal rates of evolution and gene loss; therefore, ANI should be used with caution as a measure of possible species boundaries. It is thus critically important to supplement ANI measures with actual phylogenetic analysis, such as that based on whole genome sequences. We used the GTDB-tk-based system here to provide such a phylogenomic analysis. In addition, morphological features, and the results of metabolic tests and other phenotypic characteristics have been used extensively to characterize bacteria, and although these alone cannot be used for classification or phylogenetic placement they do provide information about new strains and so were included. For most of the taxa found in this study, 16S rRNA gene sequence

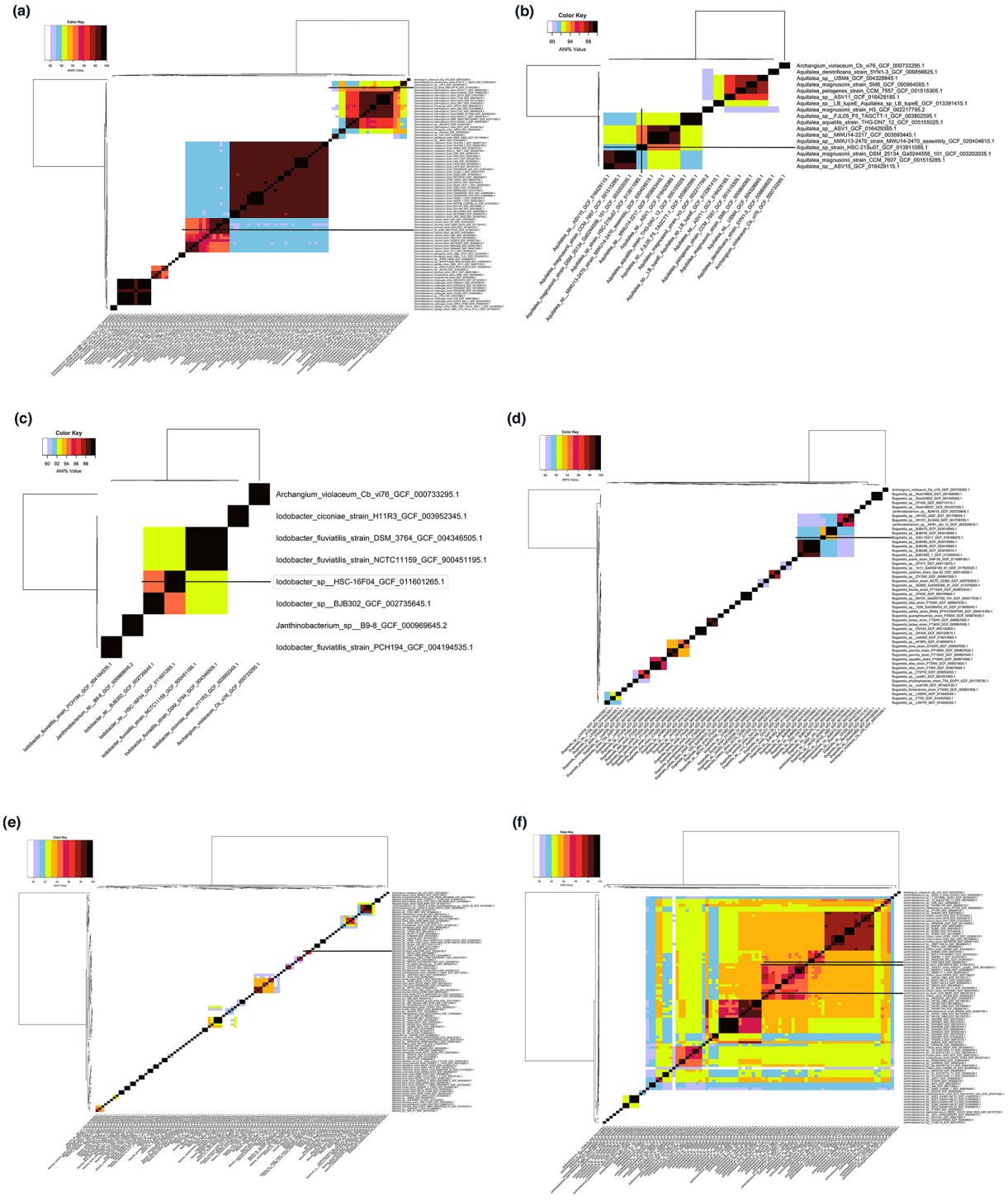

Fig. 3. (a–f) ANI heatmap dendrograms. Bacterial strains are ordered by GTDB-tk based dendrograms and coloured by ANI increments. (a) ANI heatmap dendrogram of the genus *Chromobacterium*. The top horizontal black line represents the position of strain HSC-31F16, and the bottom black line represents the position of HSC-77S12. (b) ANI heatmap dendrogram of the genus *Aquitalea*. The horizontal black line represents the position of strain HSC-21Su07. (c) ANI heatmap dendrogram of the genus *Iodobacter*. The horizontal black line represents the position of strain HSC-16F04. (d) ANI heatmap dendrogram of the genus *Duganella***.** The horizontal black line represents the position of strain HSC-15S17. (e) ANI heatmap dendrogram of the genus *Massilia*. The horizontal black line represents the position of strain HSC-2F05. (f) ANI heatmap dendrogram of the genus *Janthinobacterium***.** The top horizontal black line represents the position of strain HSC-3S05, the middle black line represents the position of strain HSC-65S10 and the bottom black line represents the position of strain UCD_MED1.

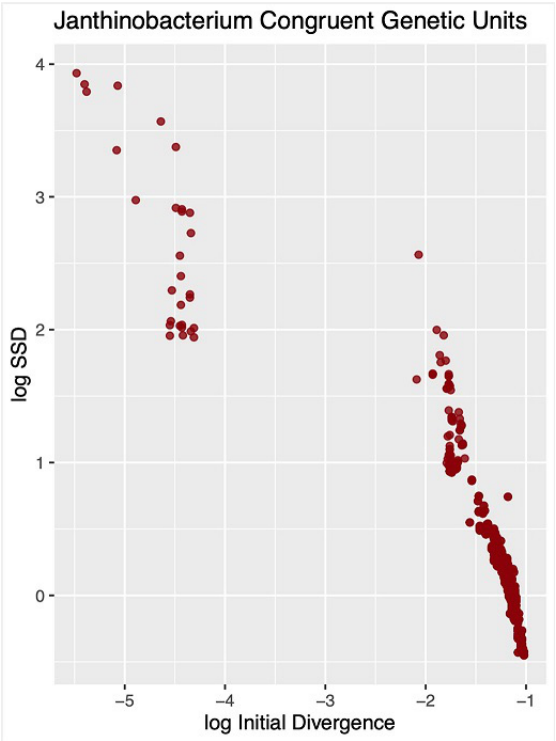

Fig. 4. PopCOGenT-based analysis of 68 *Janthinobacterium* genomes showing two distinct sets of genomes in terms of initial divergence (*x*-axis) versus SSD (*y*-axis).

comparisons, ANI values and whole genome phylogenetic analyses provided a reasonable basis for inference of taxonomy and phylogenetic position. For the taxa determined to be in the genus *Janthinobacterium*, these methods alone were not fully sufficient, and we used PopCOGenT to provide additional information about evolutionary relationships as indicated by rate of gene flow. We discuss the isolates and their species designations in greater detail below. In addition, we also discuss possible needs for taxonomic reassignment of some of the genomes currently in NCBI.

Genus *Chromobacterium*: strains HSC-31F16 and HSC-77S12

Results indicate that *Chromobacterium* strain HSC-31F16 and the reference strain H4137_1 represent a novel species. Although strain HSC-31F16 initially appeared most similar to *C. aquaticum* strain CC-SEYA-1 [38] based on 16S rRNA gene sequences, it presented seven metabolic differences. It was catalase-positive, susceptible to rifampin, unable to assimilate galacturonic acid, glucosaminic acid, itaconic acid and phenylethylamine as single carbon sources, but able to assimilate γ-hydroxybutyric acid; these are all in opposition to the responses shown by *C. aquaticum*. Although genome data for *C. aquaticum* strain CC-SEYA-1 were not available for comparison, it is unlikely that strain HSC-31F16 is a representative of the same species.

The whole genome-based phylogenetic tree Figs 2a and S1) indicated strain HSC-31F16 might represent a *C. haemolyticum* strain; however, this designation was not strongly supported by ANI values. The genome from strain HSC-31F16 shared only 93.92% ANI with the *C. haemolyticum* type strain MDA0585=DSM 19808 [36]. It also shared only 94.03% ANI with a complete genome from *C. haemolyticum* strain Bb2. Because the genome from strain HSC-31F16 is located on a long branch with only one other genome, that from *C. haemolyticum* strain H4137_1, and because it shares 97.46% ANI with that genome, both strains probably represent a species other than *C. haemolyticum*.

Based on results from all methods, we propose that strain HSC-31F16 be designated as representing a novel species. Specifically, we propose that this strain and strain H4137_1 be named *Chromobacterium perflumen* (per.flu'.men *per* L. prep. through or by means of, *flumen* L. n. river, *perflumen* through the river); HSC-31F16 was isolated from water collected along a major river in northern California.

Strain HSC-77S12 should be taxonomically classified as the species *Chromobacterium piscinae*. This assignment is based on the whole genome phylogeny, which shows that it groups into a clade with those from other *C. piscinae* strains, including the type strain *C. piscinae* DSM 23278 (Figs 2b and S1). In addition, this genome shares >95% ANI with those of other members of the

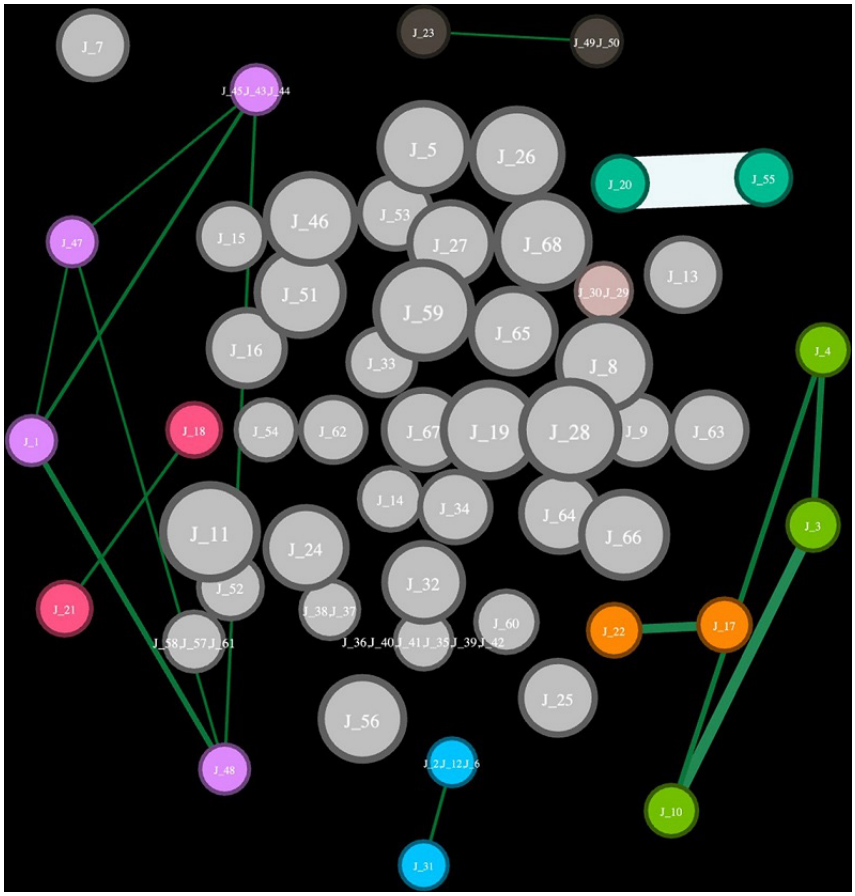

Fig. 5. Gene flow-based networks between *Janthinobacterium* genomes identified using the Infomap cluster module of PopCOGenT. Populations investigated using the cluster module (Infomap) within PopCOGenT. Networks are constructed based on recent gene flow measured by SSD and initial divergence. Infomap applies a standard clustering algorithm to the raw gene flow network. Clonal complexes (identical or nearly identical genomes) are collapsed into single nodes and are indicated by multiple labels on the same node (i.e. genome IDs J45, J43 and J44). Such placements were not used to reveal anything about gene flow. Single genomes and clonal complexes that form gene flow networks with other genomes are connected by strings with weighted edges indicating amount of gene flow. Nodes that do not form gene flow networks with any other genomes in this group are coloured in grey. For nodes with other colours, the colours are used to highlight distinct gene flow networks. The phylogenetic tree branches in Fig. 2g correspond with node colours representing eight subpopulations. Abbreviated genome labels are used here, and corresponding full names are provided in Table S1.

species, including 98.82% ANI with the type strain DSM 23278. We also suggest that the *Chromobacterium* '*vaccinii*' strains A7 (GCF_019733135.1), A8 (GCF_019733095.1) and A15 (GCF_019732845.1) should be renamed as *C. piscinae* due to the 98.85, 98.72 and 98.82% ANI similarity scores, respectively, of their genomes with that of HSC-77S12 and the phylogenetic grouping of these genomes within the *C. piscinae* clade containing the *C. piscinae* type strain. In addition, *C. 'piscinae'* strain ND17 (GCF_000812585.1) should probably be renamed as *C. amazonense* because the genome of this strain is in a clade with that of the *C. amazonense* type strain DSM 26508 (GCF_001855565.1). The metabolic characteristics of HSC-77S12 were generally consistent with the description provided for *Chromobacterium piscinae* LMG 3947 [39], with a few exceptions. Strain HSC-77S12 was positive for catalase and citrate utilization, while *C. piscinae* was described as negative for these traits. In addition, *C. piscinae* was described as being aerobic while strain HSC-77S12 was facultative (able to ferment glucose). *Chromobacterium piscinae* was also described as being negative for putrescine and L-phenylalanine assimilation while HSC-77S12 demonstrated weak assimilation of putrescine and positive assimilation of L-phenylalanine.

Genus *Aquitalea*: strain HSC-21Su07

Results suggest that *Aquitalea* strain HSC-21Su07 and the *Aquitalea* strains MWU13-2470, MWU14-2217 and ASV1 (that lack species designations) also represent a nvel species. Although the 16S rRNA gene sequences of strain HSC-21Su07 showed high similarity (99.22 and 99.00%) with those of the *Aquitalea denitrificans* type strain 5YN1-3 (NR_044535) and *Aquitalea magnusonii* strain TRO-001DR8 (NR_043475) respectively, phylogenetic tree placement and ANI values indicate a lower degree of relatedness.

Within the phylogenetic tree, the genomes of strains HSC-21Su07, MWU13-2470, MWU14-2217 and ASV1 form a small but separate clade, and when compared to one another have ANI values (~94%), just below the current standard species threshold of 95%. The sister clade to this group includes the genome of the type strain *Aquitalea aquatilis* THG-DN7_12 but the ANI value with this strain is moderately low (~92%) and HSC-21Su07 should not be assigned to this species. Strain HSC-21Su07 appears even more distantly related and has lower ANI values when compared to the other *Aquitalea* species for which genomes are available (*A. denitrificans*, *A. magnusonii* and *A. pelogenes*)and (Figs 2c and S1) and is not assigned to these species.

The metabolic features of strain HSC-21Su07 were most consistent with those described for *A. denitrificans* [40], with the exception of glucose fermentation and urease formation (negative for strain 5YN1-3 and positive for HSC-21Su07). Features were less similar to those of *A. magnusonii,* which was described as being positive for urea hydrolysis and glycogen utilization, resistant to ampicillin, rifampin and streptomycin, but susceptible to tetracycline [41]. Strain HSC-21Su07 was also urease-positive but could not utilize glycogen, and was susceptible to ampicillin, rifampin and streptomycin but resistant to tetracycline.

Based on these findings we propose a novel species name for strain HSC-21Su07 and the other strains within the same clade, namely MWU13-2470, MWU14-2217 and ASV1: *Aquitalea aquatica* (a.qua'ti.ca. L. adj. *aquatica* living, growing or found in water, aquatic). HSC-21Su07 was isolated from spring water.

Genus *Iodobacter*: strain HSC-16F04

Results indicate that *Iodobacter* strain HSC-16F04 and *Iodobacter* sp. strain BJB302 also merit a novel species designation. The strain HSC-16F04 was initially assigned to the genus *Iodobacter* based on 16S rRNA gene sequence comparisons including 98.56% identities with the same genes of both *Iodobacter arcticus* strain AsdM4-16 (NR_108462) and *Iodobacter fluviatilis* strain Logan C009 (NR_104727). Although this genus assignment is supported by whole genome phylogenetics and ANI, the genome of strain HSC-16F04 does not group closely with those of any named species within the limited number of genomes in the NCBI genome assembly database (there are only seven). Many more *Iodobacter* strains have been reported in the literature and in other databases such as the Ribosomal Database Project (RDP), but whole genome sequencing and publication for many *Iodobacter* strains are lacking. Phylogenetically, the genome from *Iodobacter* strain HSC-16F04 is sister to that from the violet-coloured strain *Iodobacter* sp. BJB302 and these shared an ANI value of 94.89%. These two genomes form a sister clade to that of the effectively published but not validly published (under the rules of the International Code of Nomenclature of Bacteria), violet coloured *Iodobacter fluviatilis* type strain DSM 3764 (Figs 2d and S1), but the genome of strain HSC-21Su07 shared an ANI value of only 92.1% with that of this type strain. Not all existing *Iodobacter* species are represented in this phylogeny due to lack of genome availability; unrepresented *Iodobacter* species include *I. limnosediminis* and the proposed novel species *I. arcticus*.

The characteristics of strain HSC-16F04 were generally consistent with those described for *I. fluviatilis* [42], except that unlike *I. fluviatilis, Iodobacter* strain HSC-16F04 was oxidase-negative, did not assimilate malic acid and showed weak assimilation of Tween 40. Characteristics were less similar to those of *Iodobacter arcticus*, which was described as being strictly aerobic (respiratory), non-motile, citrate-positive and gelatinase-negative [43]. Strain HSC-16F04 was facultative (fermentative), motile, citrate-negative and gelatinase-positive.

Since the genome from HSC-16F04 does not group with existing *I. fluviatilis* or *I. ciconiae* genomes phylogenetically and shares only low ANI values with these, we suggest that strain HSC-16F04 represents a novel *Iodobacter* species. The genome from this violet pigment producer and that of *Iodobacter* strain BJB302 appear closely related and are significantly unlike other *Iodobacter* whole genomes reported to date. The species name we propose for these two strains is *Iodobacter violacea* (vi.o.la.ce.a. L. adj. *violaceus -a -um* violet; the colonies are deep violet when formed on TSA or R2A agar).

Genus *Duganella*: strain HSC-15S17

Evidence suggests that *Duganella* strain HSC-15S17 represents a novel species. This strain was tentatively assigned to the genus *Duganella* based on 16S rRNA gene sequence similarity [98.55% with *Duganella zoogloeoides* strain IAM 12670 (NR_025833) and 98.28% with *Duganella zoogloeoides* strain NBRC 102465 (NR_114106)]. This assignment was supported by whole genome phylogeny (Figs 2e and S1) and ANI scores of >90% with genomes of other *Duganella* strains.

In terms of species-level assignment, the genome of HSC-15S17 is in a clade with genomes from two violet-coloured strains classified only to genus: *Duganella* sp. BJB475 and BJB476. A genome from the closest named relative to these is from *D. aceris* SAP-35; however, this taxonomic name has been effectively published but not validly published under the rules of the International Code of Nomenclature of Bacteria (Bacteriological Code). In terms of metabolism, the metabolic features of strain HSC-15S17 were only partially consistent with those of described species within this genus and characteristics described for the related species *Duganella zoogloeoides* [44]. Strain HSC-15S17 produces large, gelatinous, violet-coloured colonies when grown on YEM while the colonies of *D. zoogloeoides* were described as pale yellow to straw-coloured. In addition, *D. zoogloeoides* was described as forming acid aerobically from glucose and urease enzymes which strain HSC-15S17 did not.

Strain HSC-15S17 should be considered as representing a novel species and propose that this and the *Duganella* strains BJB475 and BJB476 be given the name *Duganella violaceicalia* (vi.o.la.ce.i.cal.i.a. L. adj. *violaceus -a -um* violet; the purple pigment violacein is produced by this strain; Gr. adj. *kal* or *kall* beauty or beautiful; N.L. fem. adj. *violaceicalia* violet beauty). The striking violet colour of *Duganella* strain HSC-15S17 colonies develops over time as they grow.

Genus *Massilia*: strain HSC-2F05

Results indicate that the *Massilia* strains HSC-2F05 and MS-15 represent a novel species. Strain HSC-2F05 was assigned to the genus *Massilia* based on 16S rRNA gene sequence comparison and showed 98.32% identity with the same genes from *Massilia varians* strain CCUG 35299 (NR_042652). This assignment was supported by its placement within the GTDB-tk-aligned whole genome phylogeny and the ANI values (Figs 2f and S1; Table 5).

At the time of writing, the NCBI genome assembly database contained 98 public genomes with the genus annotation *Massilia*, and these were incorporated into the phylogenetic tree and heatmap dendrogram. Analysis of the GTDB-tk-aligned heatmap dendrogram verified that almost every one of the *Massilia* genome assemblies is not closely related to its nearest phylogenetic neighbours (<90% ANI). Many individual genomes can be considered to represent unique species. The genome of *Massilia* strain HSC-2F05 has a 96.37% ANI value with that of its nearest phylogenetic neighbour *Massilia* sp. strain MS-15, a strain without species-level classification, but has less than 89% ANI values with sister taxa within the whole genome phylogenetic tree. The closest genome with species-level classification shares 84.52% ANI with *Massilia oculi* strain CCUG_43427 (GCF_003143515.1). The metabolic features of strain HSC-2F05 were only partially consistent with characteristics described for the closest 16S rRNA gene match *Massilia varians,* previously known as *Naxibacter varians* [45, 46]. Morphological features were similar, but *Massilia varians* was described as being H_2S-positive and negative for nitrate reduction while strain HSC-2F05 was negative for H_2S production and positive for nitrate reduction. In addition, *M. varians* was described as being positive for aesculin hydrolysis while HSC-2F05 was negative, and as susceptible to polymyxin while HSC-2F05 was resistant to that antibiotic.

Massilia strains HSC-2F05 and MS-15 should be designated as *Massilia hydrophila* (hy.dro'.phi.la. Gr. n. *hydr* water; Gr. adj. *philos* loving; M.L. adj. *hydrophila* water loving). *Massilia* strain HSC-2F05 was isolated from pond water.

Genus *Janthinobacterium*: strains HSC-65S10, HSC-3S05 and UCD_MED1

Strains HSC-65S10, HSC-3S05 and UCD_MED1 were tentatively assigned to the genus *Janthinobacterium* based on analysis of 16S rRNA gene sequences. This was confirmed by the whole genome phylogenetic analysis Figs 2g and S1); however, making species-level assignments for these strains was complicated for multiple reasons discussed here. We describe some of the challenges presented by this genus and possible solutions below.

Challenge 1. Taxonomic misannotation in the genus *Janthinobacterium*

One major challenge in making specieslevel assignments for our strains related to problems in the naming of organisms with genomes listed as being from the genus *Janthinobacterium* in NCBI. There are currently five formally recognized species in the genus *Janthinobacterium*, including the original type strain: *J. lividum* DSM 1522^T=NCTC 9796=H-24=ATCC 12473 [47]; the only pathogenic strain: *J. agaricidamnosum* DSM 9628^T=NBRC 102515 [48]; and three species introduced by Lu *et al.* [49]: *J. violaceinigrum* strain FT13W^T=GDMCC 1.1638=KACC 21319, *J. aquaticum* strain FT58W^T=GDMCC 1.1676=KACC 21468 and *J. rivuli* strain FT68W^T=GDMCC 1.1677=KACC 21469 [49]. One issue with this group is that there is some confusion in the NCBI database regarding the last two genomes. The 16S rRNA gene sequences for strains FT58W and FT68W are listed in the NCBI taxonomy database as *J. aquaticum* and *J. rivuli* but their genomes are not included under their taxonomic entries. Instead, the genomes are listed as being *J.* sp. with strain designations but not species identification. Another complicating factor for this group is that, according to the NCBI bacterial taxonomy database, three additional species names have been effectively published but not validly published under the rules of the International Code of Nomenclature of Bacteria (Bacteriological Code). These are *J.* [*tructae*] strain SNU WT3^T [50], *J. svalbardensis* strain JA-1=DSM 25734 [51] and *J. psychrotolerans* strain S3–2=DSM 102223 [52].

In addition to these general naming issues, there are also many genome assemblies in the NCBI database that are included in the genus *Janthinobacterium* but for which the taxonomic assignment is almost certainly incorrect. We note, for taxa which we have inferred the names are probably incorrect, we write them with square brackets (e.g. [misannotated name]) around what we have found to be the incorrect parts of the names. One type of incorrect taxonomic naming involves multiple genomes that in the GTDB-tk phylogeny clearly do not belong in the genus *Janthinobacterium*. For example, the genome from the isolate listed as [*Janthinobacterium*] sp. strain CG23_25 (RefSeq ID: GCF_001485665.1) falls within the *Massilia* branch of the whole genome phylogenetic tree (Fig. S1). Similarly, the genome labelled as [*Janthinobacterium*] sp. strain B9-8 (RefSeq ID GCF_000969645.2) falls within the *Iodobacter* clade, and the genome labelled as [*Janthinobacterium*] strain HH01 (RefSeq ID: GCF_000335815.1) falls within the genus *Duganella* (Fig. S1). In addition, there are three genome assemblies annotated as '*Janthinobacterium*' that are placed in the GTDB-tk tree outside of the genus but in a group for which no clear genus can be assigned. These genomes are:

[*Janthinobacterium*] sp. 17J80-10 (GCF_004114795.1), [*Janthinobacterium marseille*] P9896 (GCF_903469655.1) and [*Janthinobacterium marseille*] (GCF_000013625.1).

In addition to mis-assignment to the genus *Janthinobacterium*, there are many genomes correctly assigned to the genus *Janthinobacterium* but which have species names assigned that are inconsistent with the placement of type strains in the GTDB-tk tree. For example, the genome assembly labelled *J.* [*agaricidamnosum*] strain BHSEK (RefSeq ID: GCF_003667705.1) shares a more recent branch node with *J. lividum* strains than with the type strain of *J. agaricidamnosum*. The genome assemblies labelled *J. svalbardensis* PAMC 27463 and *J.* [*tructae*] strain SNU WT3^T are sister taxa that group together within a clade consisting of strains named only to genus or labelled as *J. lividum*.

These incorrect taxonomic assignments and the ambiguity regarding formal species names, as well as some taxa for which the 16S rRNA gene data are listed under one name and the genomes are listed under another, all serve to make assigning names to new isolates within this group challenging.

Challenge 2. Inconsistent results from 16S rRNA gene sequence analysis, ANI and whole genome phylogeny

Another key challenge in making species-level assignments for the genus *Janthinobacterium* is that results from analysis of 16S rRNA gene sequences, ANI and whole genome phylogeny were somewhat ambiguous and contradictory; we summarize some of these results here. First, the 16S rRNA genes of all three of our strains showed >99.80% identity to 16S rRNA genes of the *Janthinobacterium lividum* type strain DSM 1522 (NR_026365), and equally high percentage identity scores with many other *Janthinobacterium lividum* strains in the NCBI database. Such high percentage identity of 16S rRNA genes is frequently used to indicate that strains are from the same species. However, the results of ANI analysis and whole genome phylogenetics did not support a simple assignment of these strains to *J. lividum*. In particular, while ANI overall generally tracked well with the whole genome phylogeny for the genus *Janthinobacterium*, the specific ANI levels for the new strains and for strains that could potentially be assigned to *J. lividum* were lower than the standard 95% cutoff used by many to delineate species. For example, if we use a 95% ANI cutoff to delineate species, only the 12 closest strains to the *J. lividum* type strain NCTC 9796 can be classified as *J. lividum*. These include: EB271-G4-7A, YR213, BJB446, 1246, BJB301, MP5059B, EIF1, SLB01, PLB04, PLB02, MTR and EIF2. If we use 94% as the species cutoff, we can include strain SP_13. And if we use 93% as an ANI cutoff species threshold for this species, 34 additional strains could be included in the species *J. lividum*. If we focus solely on the three new strains of *Janthinobacterium* described in this paper, we also see some complexities in species boundary identification. The genome of each of the three new strains reported here has an ANI value of about 93% with each other, and each of them has a 93.2–93.6% ANI to the *J. lividum* type strain. An additional 60 genomes have ANI values between 90 and 95% (averaging 93%) with the *J. lividum* type strain. Included in these 60 are many that have been published as '*J. lividum*'. Individual genomes forming distinct clades such as groups with short branches compared to each other, and separated by a long branch connecting the group to the rest of the tree have around 95% or better ANI similarity with their nearest neighbours.

Using strict ANI value cutoffs can be insufficient to identify species boundaries because of issues including unequal rates of evolution, different gene loss and gain, and the fact that species boundaries are determined by ecological and genetic factors and not percentage identity. In addition, it is important to note that the 95% species threshold is based on analysis of a limited number of genomes generally with high levels of similarity [23]. Regardless of these caveats, the ANI values within this group suggest that drawing species boundaries using ANI in the same way as has been sometimes used for other taxa is potentially invalid.

Improving resolution in the genus *Janthinobacterium* using PopCOGenT

Because the genus *Janthinobacterium* does not clearly fit into the standard ANI model, and the strains investigated here have nearly identical phenotypes (Tables 1 and 2), we used PopCOGenT to assist in investigation of the evolutionary history and species-level classification of this group in greater detail. As mentioned above, PopCOGenT assesses population structure by inferring the extent of recombination relative to the amount of divergence between genomes and this in turn can help determine where natural genetic boundaries exist between strains [27]. PopCOGenT has been used to distinguish recent gene flow events from historical ones and to identify ancestral nodes in phylogenomic trees that lead to speciation events [53]. PopCOGenT can be used to predict the recombination pattern and population structure within bacterial species and reveal distinct populations. Directionality of gene flow from population to population can also be determined [54]. We used PopCOGenT to analyse all genomes for a clade within the genus *Janthinobacterium* that included the type strains, the three new *Janthinobacterium* genomes presented in this study, and other related strains (68 genomes in total, Table S1). The results were assessed in two ways: by examining the relationships between initial divergence and length bias (Fig. 4) and by inferring networks of recent gene flow (Fig. 5). The results shown in Fig. 4 highlight that the *Janthinobacterium* genome comparisons come in two main categories. First, there are genome comparisons where the two genomes are very closely related (on the left along the 'initial divergence' *x*-axis). Some of these pairs appear to represent resequencing of the same strains, either because they were isolated from the same sample or because groups separately sequenced strains obtained from a common source. Other pairs appear to represent just quite closely related separate isolates.

One application of PopCOGenT is to detect gene flow in coexisting microbes of the same species. The program can also be applied to examine gene flow for ancestral strains. PopCoGenT can be used to measure the amount of relatively recent horizontal gene flow compared to the level of divergence among many sets of microbes, whether or not they coexist, and data can provide information about historical genetic boundaries between lineages. In this study, PopCOGenT was used to provide information on patterns of gene flow in lineages of *Janthinobacterium* strains in order to supplement analysis of phylogeny, genomes and ANI.

In addition, there are genome comparisons where the genome pairs are significantly more distantly related (along the right side of the *x*-axis) such that the initial divergence scores 'jump' multiple log values. The discontinuity along the *x*-axis is caused by the genomes that form distinct subpopulations (upper left side) having lower initial divergence than the genomes that do not form populations. This confirms that subpopulation clusters have lower rates of genetic divergence and genomes with higher rates of genetic divergence do not form subpopulations with other genomes in the dataset. The genome pairs with high levels of initial divergence almost always have lower SSD scores (and thus less inferred gene flow) than the genome pairs with low levels of initial divergence. Within this dataset genomes that form subpopulations must have an initial divergence score of <−4, and those that do not form subpopulations have an initial divergence score of >−3.

To infer specific subpopulation structure from the initial divergence and SSD scores shown in Fig. 4, we generated a gene-flow network diagram (Fig. 5). Each genome is represented by a node in this diagram and the edges represent inferred gene flow with the thickness of the edges representing the amount of gene flow. Genomes that do not have any inferred gene flow with other genomes here are coloured in grey. Genomes that are part of gene flow networks are coloured in other colours with each specific colour used representing a distinct gene flow network. Overall, the only genome groups that formed strong gene flow networks were ones that were very closely related to each other as indicated by low initial divergence scores and being very close in the whole genome phylogenetic tree. Of note, no networks were seen that included pairs of strains with ANIs <95%. Overall, the PopCOGenT analysis shows that for the strains in the genus *Janthinobacterium* analysed here, significant amounts of gene flow only occurs among some (but not all) of the strains for which the genome tree shows they are very closely related. In some cases, there is no detectable gene flow even with strains that are quite close. This possibly could be due to physical or ecological separation of strains that are closely related, which could prevent gene flow from occurring. We note that the strains included in this analysis were isolated from a wide diversity of locations and ecosystems and it is not surprising that they may be genetically isolated from each other.

Integrating information from 16S rRNA genes, ANI, whole genome phylogeny, phenotype and gene flow networks for the genus *Janthinobacterium*

In this section we discuss the implications for the taxonomic assignments of the new *Janthinobacterium* isolates by integrating the PopCOGenT results with the analysis of 16S rRNA sequences, ANI, whole genome phylogeny and phenotype.

Phylogenetically, HSC-3S05 groups in a clade in the whole genome tree with three other strains classified only to the genus level. HSC-3S05 has an ANI >95% with those strains. This clade is a sister group to a clade that includes both the *J. lividum* type strain and the type species for *J. violaceinigrum*. HSC-3S05 has 93.67% ANI with the *J. lividum* type strain and 92.39% ANI with the *J. violaceinigrum* type species. The metabolic characteristics of HSC-3S05 are generally consistent with the description provided for 'typical' *Janthinobacterium lividum* [55], except that the assimilation of *N*-acetyl-D-glucosamine, cellobiose, α-lactose and L-phenylalanine were found to be negative in published strains while HSC-3S05 was positive or weakly positive for these.

HSC-65S10 groups in a clade in the whole genome tree with four other strains classified only to the genus level. HSC-65S10 has an ANI >95% with those strains. HSC-65S10 is phylogenetically more distant from the *J. lividum* and *J. violaceinigrum* type strains than HSC-3S05 is. However, the ANI of HSC-65S10 with these two type strains is 93.27 and 92.29 respectively, similar to that of HSC-3S05 with these strains. The PopCOGenT results show that HSC-65S10 is not part of a gene flow network with any other strains in the dataset, providing evidence that strain HSC-65S10 may have a history of ecological isolation from other strains in the dataset. The metabolic characteristics of HSC-65S10 were generally consistent with the description provided for 'typical' *Janthinobacterium lividum* [55], except that the assimilation of *N*-acetyl-D-glucosamine and cellobiose were described as being negative for all published strains while HSC-65S10 was positive.

Strain UCD_MED1 has 97.58, 96.36, and 96.36% ANI with *Janthiobacterium* strains only classified to genus: NCTC8861, EB271-G4-3-1 and EB271-G4-3-2 respectively. The PopCOGenT results show that UCD_MED1 is not part of a gene flow network with any other strains in the dataset, providing evidence that strain UCD_MED1 may have a history of ecological isolation from other strains in the dataset. Whether groups such as this one should be considered separate species or subspecies is debatable. The metabolic characteristics of UCD_MED1 were generally consistent with the description provided for 'typical' *Janthinobacterium lividum* [55], except that the assimilation of *N*-acetyl-D-glucosamine and cellobiose were found to be negative in published strains while UCD_MED1 was positive.

The three new representatives of the *Janthinobacterium* strains presented here could be representatives of novel species as they (1) have little gene flow with other strains represented here, (2) are phylogenetically somewhat distinct from type strains of formally

described species and (3) have ANI values of <95% with any strains that can be readily assigned to a species. However, it may be premature for proposing novel species designations here given that (1) gene flow in the genus *Janthinobacterium* appears to drop off precipitously even between quite closely related strains (see above), (2) ANI is an indirect and imprecise measure of species boundaries and there are other examples of ANIs of <95% occurring within proposed species groups [23] and (3) the metabolic features of the three new strains are generally consistent with those described for typical *J. lividum* strains. We thus propose that without any other evidence all the *Janthinobacterium* strains shown in Fig. 2g be considered in the same species, namely *J. lividum,* and furthermore that a 90% ANI cutoff is more appropriate for species boundaries in the genus *Janthinobacerium*. This thus would mean that several recently published 'species' be designated subspecies within this group.

CONCLUSIONS

Escaping the mire of subjectivity in taxonomy

During this study, we determined the taxonomic status of nine bacterial strains using phylogenomic and population ecology approaches. We found that taxonomically classifying six of these strains using relatively standard methods (a combination of analysis of 16S rRNA gene sequences, ANI and whole genome phylogeny) worked well and we suggested five of the strains be recognized as representing novel species. However, for the three strains in the genus *Janthinobacterium*, these standard approaches gave somewhat inconsistent and ambiguous results. Using these standard approaches we were unable to determine where species boundaries were within the group. This is not particularly surprising given that the methods being used are somewhat indirect ways to assess species boundaries and do not actually address factors such as genetic isolation between groups, which is needed to formally determine where species boundaries exist. We thus added an additional analysis using the program PopCOGenT to examine population structure within the *Janthinobacterium* group. PopCOGenT estimates the extent of recent genetic exchange between genomes by assessing the balance between shared identical genome regions versus total divergence of genomes. This additional analysis showed that between the genomes immediately surrounding the *J. lividum* type strain, low recombination rates dominate throughout the clades, suggesting that existing strains may comprise novel, unnamed species. Using a combination of standard tools as well as population genetic analysis as with PopCOGenT was useful here and will presumably be useful in future studies and will better integrate naming with species concepts for bacteria in general [56]. Experimenting with new bacterial taxonomic classification methods leads to open discussions on how taxonomists can escape the mire of subjectivity and agree on a definitive methodology for classifying bacteria to species level.

The perils of misnaming

Our analysis here was made complicated by mis-classification of strains for which genome sequences are available. Overall, we suggest revision of the taxonomic status of many genome assemblies within all six of the genera evaluated in this study, especially within the genus *Janthinobacterium*. This highlights a major issue for microbiology in that published draft genome assemblies can sometimes be inaccurate [57, 58]. Published mis-classified assemblies are problematic because they can lead to false conclusions and mis-identified strains [59]. Misidentification can also lead to inappropriate ongoing research. We encourage more attention to be paid to the taxonomic annotations of new genomes being uploaded to public databases.

Relevance of metabolic and physiological assays

For this study we recorded classical phenotypic characteristics including cell and colony morphology, metabolic testing and antimicrobial susceptibility testing. Although not all phenotypic characteristics were determined, those that were are consistent with characteristics included previously in the descriptions of genera and species investigated. Phenotypes are observable characteristics of cells [60], and although the term 'phenotype' is sometimes applied more broadly, the classical phenotypic characteristics of bacteria comprise morphological, physiological and biochemical features [61]. Growth phenotypes are of particular interest because they help better understand the selective forces that may have shaped the history of a particular organism and also because they provide insights into an organism's niche and how to work with it in the lab [60]. Many bacteria carry genes that are only expressed under certain circumstances, such as virulence factors regulated through quorum sensing and genes involved in pigment production being influenced by temperature variation [62, 63]. Although it is doubtful that individual phenotypes or small collections of phenotypes can consistently and correctly represent evolutionary relationships [64], organisms are described in phenotypic terms, and the descriptions help define a taxonomic group that may also be recognized at the phylogenetic level [65].

Even with the advent of rapid DNA sequencing, and its growing use in identifying and describing species, verification of bacterial species still requires phenotypic descriptions, whenever possible using experimental cell culture and phenotyping. Importantly, it is still not possible to accurately predict phenotypes of bacteria from complete genome sequences and thus characterizing and describing phenotypes is still a critical component of any species description. A minimal phenotypic description is not only the identity card of a taxon but also a key to its biology. Although they are accepted as necessary, differential phenotypic characters are often hard to find with a reasonable amount of effort and time [31]. This of course does not mean phenotypes are the only thing needed to describe an organism. Phenotypes are prone to convergent evolution and can differ greatly under varying conditions

and even for strains that are genetically identical (e.g. due to epigenetic modifications). Thus, it is important to use both phenotypic and genetic/genomic information in characterizing taxa.

Future directions for investigating violacein-producing bacteria

The violacein biosynthesis pathway relies on the amino acid tryptophan. The first two genes in the pathway, *vioAB*, are homologous with genes used for indole compounds staurosporine and rebeccamycin, which are found in some actinomycetes. *VioC* is homologous with a monooxygenase housekeeping gene used in various processes across a wide range of prokaryotic and eukaryotic organisms, while *vioED* are uniquely used for production of violacein and deoxyvioalcein. The striking violet pigment violacein is the final product of a biosynthetic pathway that also produces prodeoxyviolacein (green) from the first three genes in the pathway (*vioABE*), and deoxyviolacein (violet) from the first four genes (*vioABEC*). The end product, violacein (navy-violet), is produced when all five genes are used (*vioABEDC*). Deoxyviolacein is not cytotoxic, whereas violacein is, and therefore synthesizing deoxyvioalcein can be good for developing products such as biosensors and food colorants [66].

Further research should focus on predicting violacein production in cultured or uncultured bacteria by developing models and using phylogenomic methods. Investigations can elucidate how horizontal gene transfer of violacein genes may have contributed to the pigment's distribution throughout the bacterial tree of life. Models could be useful for future research where production of this pigment can be inferred from any bacterial whole genome without the need for culturing and extraction. The construction of models that allow the user to discover new strains of violacein-producing bacteria and the methods for creating and using these models could easily be co-opted for searching for other pigments in any bacterial whole genome assemblies. In this way, the *vioABCDE* genes can be searched for and compared to determine if they are present in non-pigmented strains (e.g. in *Janthinobacterium*) and also to determine if they were probably acquired independently through multiple horizontal gene transfers by various strains/species.

The taxonomy and evolution of bacterial strains that are capable of expressing violacein, such as *Janthinobacterium*, are not well understood even though the biosynthesis of the compound has been clearly demonstrated in multiple genera. Future studies should focus on understanding concordance between biochemical phenotyping and metabolomic predictions. Phenotyping and genome annotation are important for understanding if the bacteria will be useful, safe and economic for future applications. This evolutionary knowledge can help provide context for future synthetic biology and genetic modification studies using violacein and non-model bacterial species such as these.

The evolutionary or adaptive reasons leading to the *Janthinobacterium* subpopulations inferred by PopCOGenT should be further investigated. Some metabolic tests can be used to define species and biotypes within a species, and the delimitation of biotypes within a specific group has been found to generally agree with observed phylogenetic structure [67]. Future studies should use metabolic tests to observe if splitting a specific species such as *Janthinobacterium lividum* into biotypes based on differences in niche occupancy reflects phylogenetic subdivisions within the species. Codon usage redundancy may cause GC content to shift within members of a genus. Differences in GC content among coding sequences may be caused by mutational bias due to adaptation to different lifestyles or niches among species isolated from very different environments. The variations in metabolic capabilities and GC content between different species and lineages that have adapted to distinct niches suggest that these traits have evolved over a long period of time in response to the demands of their specific niches [67]. Using publicly available whole genome sequences, additional studies on genera with violacein-producing bacteria should also involve comparing GC contents of coding sequences as these are thought to reflect subtle differences in mutational bias as a consequence of long-term niche adaptation.

Funding information
The author(s) received no specific grant from any funding agency.

Acknowledgements
This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1 650 042. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation. We wish to thank the following Sierra College microbiology students for their work in isolating and partially characterizing bacterial strains: Bonnie Anderson (HSC-16F04), Kayleigh MacDonald (HSC-2F05), Brian Harbour (HSC-21Su07), Derrick Hicks (HSC-65S10), Noor Mashal (HSC-31F16) and David Galle (HSC-15S17). We also thank Xiqian (Annie) Yu for assistance with running PopCOGenT. Photo Credit: Jim Wilson, for whom we give special thanks for taking breathtaking photos of the bacterial colonies.

Conflicts of interest
The authors have declared that no competing interest exists.

References

1. Durán N, Justo GZ, Ferreira CV, Melo PS, Cordi L, *et al*. Violacein: properties and biological activities. *Biotechnol Appl Biochem* 2007;48:127–133.
2. Wang H, Wang F, Zhu X, Yan Y, Yu X, *et al*. Biosynthesis and characterization of violacein, deoxyviolacein and oxyviolacein in heterologous host, and their antimicrobial activities. *Biochem Eng J* 2012;67:148–155.

3. Choi SY, Kim S, Lyuck S, Kim SB, Mitchell RJ. High-level production of violacein by the newly isolated *Duganella violaceinigra* str. NI28 and its impact on *Staphylococcus aureus*. *Sci Rep* 2015;5:15598.
4. Lang E, Schumann P, Tindall BJ, Mohr KI, Spröer C. Reclassification of *Angiococcus disciformis*, *Cystobacter minus* and *Cystobacter violaceus* as *Archangium disciforme* comb. nov., *Archangium minus* comb. nov. and *Archangium violaceum* comb. nov., unification of the families *Archangiaceae* and *Cystobacteraceae*, and emended descriptions of the families *Myxococcaceae* and *Archangiaceae*. *Int J Syst Evol Microbiol* 2015;65:4032–4042.
5. Myeong NR, Seong HJ, Kim HJ, Sul WJ. Complete genome sequence of antibiotic and anticancer agent violacein producing *Massilia* sp. strain NR 4-1. *J Biotechnol* 2016;223:36–37.
6. Doing G, Perron GG, Jude BA. Draft genome sequence of a violacein-producing *Iodobacter* sp. from the Hudson Valley Watershed. *Genome Announc* 2018;6:e01428-17.
7. Jude BA. Draft genome sequence of a *Chitinimonas* species from Hudson Valley Waterways that expresses violacein pigment. *Microbiol Resour Announc* 2019;8:e00683-19.
8. Brucker RM, Harris RN, Schwantes CR, Gallaher TN, Flaherty DC, *et al*. Amphibian chemical defense: antifungal metabolites of the microsymbiont *Janthinobacterium lividum* on the salamander *Plethodon cinereus*. *J Chem Ecol* 2008;34:1422–1429.
9. Sasidharan A, Sasidharan NK, Amma D, Vasu RK, Nataraja AV, *et al*. Antifungal activity of violacein purified from a novel strain of *Chromobacterium* sp. NIIST (MTCC 5522). *J Microbiol* 2015;53:694–701.
10. Fang L, Zhang G, Pfeifer BA. Engineering of *E. coli* for heterologous expression of secondary metabolite biosynthesis pathways recovered from metagenomics libraries. In: Charles TC, Liles MR and Sessitsch A (eds). *Functional Metagenomics: Tools and Applications*. Cham: Springer International Publishing; 2017. pp. 45–63. https://doi.org/10.1007/978-3-319-61510-3_3 [accessed 21 April 2022].
11. Harris RN, James TY, Lauer A, Simon MA, Patel A. Amphibian pathogen *Batrachochytrium dendrobatidis* is inhibited by the cutaneous bacteria of amphibian species. *EcoHealth* 2006;3:53–56.
12. Mateo-Estrada V, Graña-Miraglia L, López-Leal G, Castillo-Ramírez S. Phylogenomics reveals clear cases of misclassification and genus-wide phylogenetic markers for *Acinetobacter*. *Genome Biol Evol* 2019;11:2531–2541.
13. Yoon S-H, Baek H-J, Kwon S-W, Lee C-M, Sim J-S, *et al*. Production of violacein by a novel bacterium, *Massilia* sp. EP15224 strain. *Microbiol Biotechnol Lett* 2014;42:317–323.
14. Clarridge III JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 2004;17:840–862.
15. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015;31:587–589.
16. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
17. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, *et al*. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 2018;46:D851–D860.
18. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, *et al*. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 2021;49:D1020–D1028.
19. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2019;36:1925–1927.
20. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
21. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 2012;61:1061–1067.
22. 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;9:1–8.
23. Murray CS, Gao Y, Wu M. Re-evaluating the evidence for a universal genetic boundary among microbial species. *Nat Commun* 2021;12:4059.
24. Warnes MGR, Bolker B, Bonebakker L, Gentleman R, Huber W. Package 'gplots. In: *Var R Program Tools Plotting Data*. 2016.
25. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;35:526–528.
26. Wickham H, François R, Henry L, Müller K. dplyr: a grammar of data manipulation. 2020. R Package Version 08; 2021
27. Arevalo P, VanInsberghe D, Elsherbini J, Gore J, Polz MF. A reverse ecology approach based on a biological definition of microbial populations. *Cell* 2019;178:820–834.
28. Rstudio T. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, Public Benefit Corporation; 2022. http://www.rstudio.com/
29. Rosvall M, Axelsson D, Bergstrom CT. The map equation. *Eur Phys J Spec Top* 2009;178:13–23.
30. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 2014;9:e98679.
31. Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, *et al*. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 1996;60:407–438.
32. Mac Faddin JF. *Biochemical Tests for Identification of Medical Bacteria*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 1980.
33. Carlone GM, Valadez MJ, Pickett MJ. Methods for distinguishing gram-positive from gram-negative bacteria. *J Clin Microbiol* 1982;16:1157–1159.
34. Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, *et al*. PATtyFams: protein families for the microbial genomes in the PATRIC database. *Front Microbiol* 2016;7:118.
35. Rutgers M, Wouterse M, Drost SM, Breure AM, Mulder C. Monitoring soil bacteria with community-level physiological profiles using BiologTM ECO-plates in the Netherlands and Europe. *Appl Soil Ecol* 2016;97:23–35.
36. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, *et al*. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018;68:461–466.
37. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, *et al*. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. *mSystems* 2020;5:e00731-19.
38. Young C-C, Arun AB, Lai W-A, Chen W-M, Chou J-H, *et al*. *Chromobacterium aquaticum* sp. nov., isolated from spring water samples. *Int J Syst Evol Microbiol* 2008;58:877–880.
39. Kämpfer P, Busse HJ, Scholz HC. *Chromobacterium piscinae* sp. nov. and *Chromobacterium pseudoviolaceum* sp. nov., from environmental samples. *Int J Syst Evol Microbiol* 2009;59:2486–2490.
40. Lee C-M, Weon H-Y, Kim Y-J, Son J-A, Yoon S-H, *et al*. *Aquitalea denitrificans* sp. nov., isolated from a Korean wetland. *Int J Syst Evol Microbiol* 2009;59:1045–1048.
41. Lau HT, Faryna J, Triplett EW. *Aquitalea magnusonii* gen. nov., sp. nov., a novel Gram-negative bacterium isolated from a humic lake. *Int J Syst Evol Microbiol* 2006;56:867–871.
42. Logan N, Brenner DJ, Krieg NR, Staley JT, Garrity GM. Genus VII Iodobacter Logan 1989, 455 VP. In: *Bergey's Manual of Systematic Bacteriology*, 2nd ed. Springer, 2005. pp. 833–836.
43. Srinivas TNR, Manasa P, Begum Z, Sunil B, Sailaja B, *et al*. *Iodobacter arcticus* sp. nov., a psychrotolerant bacterium isolated from meltwater stream sediment of an Arctic glacier. *Int J Syst Evol Microbiol* 2013;63:2800–2805.

44. Hiraishi A, Shin YK, Sugiyama J. Proposal to reclassify *Zoogloea ramigera* IAM 12670 (P. R. Dugan 115) as *Duganella zoogloeoides* gen. nov., sp. nov. *Int J Syst Bacteriol* 1997;47:1249–1252.
45. La Scola B, Birtles RJ, Mallet MN, Raoult D. *Massilia* gen. nov. in validation of publication of new names and new combinations previously effectively published outside the IJSEM. *Int J Syst Evol Microbiol* 2000;50 Pt 2:423–424.
46. Kämpfer P, Falsen E, Busse H-J. *Naxibacter varians* sp. nov. and *Naxibacter haematophilus* sp. nov., and emended description of the genus *Naxibacter*. *Int J Syst Evol Microbiol* 2008;58:1680–1684.
47. De Ley J, Segers P, Gillis M. Intra- and intergeneric similarities of *Chromobacterium* and *Janthinobacterium* ribosomal ribonucleic acid cistrons. *Int J Syst Bacteriol* 1978;28:154–168.
48. Lincoln SP, Fermor TR, Tindall BJ. *Janthinobacterium agaricidamnosum* sp. nov., a soft rot pathogen of *Agaricus bisporus*. *Int J Syst Bacteriol* 1999;49 Pt 4:1577–1589.
49. Lu H, Deng T, Cai Z, Liu F, Yang X, *et al*. *Janthinobacterium violaceinigrum* sp. nov., *Janthinobacterium aquaticum* sp. nov. and *Janthinobacterium rivuli* sp. nov., isolated from a subtropical stream in China. *Int J Syst Evol Microbiol* 2020;70:2719–2725.
50. Jung WJ, Kim SW, Giri SS, Kim HJ, Kim SG, *et al*. *Janthinobacterium tructae* sp. nov., isolated from kidney of rainbow trout (Oncorhynchus mykiss). *Pathogens* 2021;10:229.
51. Ambrožič Avguštin J, Žgur Bertok D, Kostanjšek R, Avguštin G. Isolation and characterization of a novel violacein-like pigment producing psychrotrophic bacterial species *Janthinobacterium svalbardensis* sp. nov. *Antonie van Leeuwenhoek* 2013;103:763–769.
52. Gong X, Skrivergaard S, Korsgaard BS, Schreiber L, Marshall IPG, *et al*. High quality draft genome sequence of *Janthinobacterium psychrotolerans* sp. nov., isolated from a frozen freshwater pond. *Stand Genomic Sci* 2017;12:8.
53. Chu X, Wang X, Cheung LS, Feng X, Ang P, *et al*. Coastal transient niches shape the microdiversity pattern of a bacterioplankton population with reduced genomes. *mBio* 2022;13:e0057122.
54. Wang J, Li Y, Pinto-Tomás AA, Cheng K, Huang Y. Habitat adaptation drives speciation of a *Streptomyces* species with distinct habitats and disparate geographic origins. *mBio* 2022;13:e0278121.
55. Gillis M, De Ley J. The genera *Chromobacterium* and *Janthinobacterium*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH and Stackebrandt E (eds). *The Prokaryotes*. New York, NY: Springer New York; 2006. pp. 737–746. http://link.springer.com/10.1007/0-387-30745-1_32 [accessed 5 April 2022].
56. VanInsberghe D, Arevalo P, Chien D, Polz MF. How can microbial population genomics inform community ecology? *Philos Trans R Soc Lond B Biol Sci* 2020;375:20190253.
57. Smith H, Akiyama T, Foreman C, Franklin M, Woyke T, *et al*. Draft genome sequence and description of *Janthinobacterium* sp. strain CG3, a psychrotolerant antarctic supraglacial stream bacterium. *Genome Announc* 2013;1:e00960-13.
58. Valdes N, Soto P, Cottet L, Alarcon P, Gonzalez A, *et al*. Draft genome sequence of *Janthinobacterium lividum* strain MTR reveals its mechanism of capnophilic behavior. *Stand Genomic Sci* 2015;10:110.
59. Dieser M, Smith HJ, Ramaraj T, Foreman CM. Janthinobacterium CG23_2: comparative genome analysis reveals enhanced environmental sensing and transcriptional regulation for adaptation to life in an antarctic supraglacial stream. *Microorganisms* 2019;7:10.
60. Bochner BR. Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* 2009;33:191–205.
61. Balows A. Manual of clinical microbiology 8th edition. *Diagn Microbiol Infect Dis* 2003;47:625–626.
62. Meng X, Ahator SD, Zhang LH. Molecular mechanisms of phosphate stress activation of *Pseudomonas aeruginosa* quorum sensing systems. *mSphere* 2020;5:e00119-20.
63. De León ME, Wilson HS, Jospin G, Eisen JA. Draft genome sequences and genomic analysis for pigment production in bacteria isolated from blue discolored Soymilk and Tofu. *J Genomics* 2021;9:55–67.
64. Trost B, Haakensen M, Pittet V, Ziola B, Kusalik A. Analysis and comparison of the pan-genomic properties of sixteen well-characterized bacterial genera. *BMC Microbiol* 2010;10:1–18.
65. Moore WEC, Stackebrandt E, Kandler O, Colwell RR, Krichevsky MI, *et al*. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Evol Microbiol* 1987;37:463–464.
66. Hui C-Y, Guo Y, Li H, Gao C-X, Yi J. Detection of environmental pollutant cadmium in water using a visual bacterial biosensor. *Sci Rep* 2022;12:6898.
67. Williams DJ, Grimont PAD, Cazares A, Grimont F, Ageron E, *et al*. The genus *Serratia* revisited by genomics. *Nat Commun* 2022;13:5195.
68. Gajic I, Kabic J, Kekic D, Jovicevic M, Milenkovic M, *et al*. Antimicrobial susceptibility testing: a comprehensive review of currently used methods. *Antibiotics* 2022;11:427.