



Metagenome to phenome approach enables isolation and genomics characterization of *Kalamiella piersonii* gen. nov., sp. nov. from the International Space Station

Nitin Kumar Singh¹ · Jason M. Wood¹ · Snehit S. Mhatre¹ · Kasthuri Venkateswaran¹

Received: 7 November 2018 / Revised: 26 February 2019 / Accepted: 3 March 2019 / Published online: 23 April 2019

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2019

Abstract

Several evolutionarily distinct, near full-length draft metagenome-resolved genomes (MRG), were assembled from sequences recovered from the International Space Station (ISS) environments. The retrieval of MRGs facilitated the exploration of a large collection of archived strains (~500 isolates) and assisted in isolating seven related strains. The whole genome sequences (WGS) of seven ISS strains exhibited 100% identity to the 4.85×10^6 bp of four MRGs. The “metagenome to phenome” approach led to the description of a novel bacterial genus from the ISS samples. The phylogenomics and traditional taxonomic approaches suggested that these seven ISS strains and four MRGs were not phylogenetically affiliated to any validly described genera of the family *Erwiniaceae*, but belong to a novel genus with the proposed name *Kalamiella*. Comparative genomic analyses of *Kalamiella piersonii* strains and MRGs showed genes associated with carbohydrate (348 genes), amino acid (384), RNA (59), and protein (214) metabolisms; membrane transport systems (108), pathways for biosynthesis of cofactors, vitamins, prosthetic groups, and pigments (179); as well as mechanisms for virulence, disease, and defense (50). Even though *Kalamiella* genome annotation and disc diffusion tests revealed multidrug resistance, the PathogenFinder algorithm predicted that *K. piersonii* strains are not human pathogens. This approach to isolating microbes allows for the characterization of functional pathways and their potential virulence properties that can directly affect human health. The isolation of novel strains from the ISS has broad applications in microbiology, not only because of concern for astronaut health but it might have a great potential for biotechnological relevance. The metagenome to phenome approach will help to improve our understanding of complex metabolic networks that control fundamental life processes under microgravity and in deep space.

Keywords *Kalamiella piersonii* · Metagenome-resolved genomes · Genome-inferred phenotype · Phylogenomics · MLSA

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00253-019-09813-z>) contains supplementary material, which is available to authorized users.

✉ Kasthuri Venkateswaran
kjvenkat@jpl.nasa.gov

Nitin Kumar Singh
nitin.k.singh@jpl.nasa.gov

Jason M. Wood
jason.wood@jpl.nasa.gov

Snehit S. Mhatre
snehit.s.mhatre@jpl.nasa.gov

¹ Jet Propulsion Laboratory, Biotechnology and Planetary Protection Group, M/S 89-2, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109, USA

Introduction

A recent study on the International Space Station (ISS) using an environmental metagenome analysis revealed the presence of 115 validly described microbial genera and 318 species dominated by *Rhodotorula* (35%) and *Pantoea* (10%) that were putatively viable and thriving (Singh et al. 2018b). It has been reported that dead cells can be differentiated from the intact cells before analyzing the samples via state-of-the-art molecular technologies (Singh et al. 2018b; Vaishampayan et al. 2013). In these reports, propidium monoazide (PMA) was used as a viability marker to measure intact microorganisms before DNA was extracted and subjected to downstream molecular analyses (Bonetta et al. 2017; Kibbee and Örmeci 2017; Vesper et al. 2008; Weinmaier et al. 2015). The metagenomic sequences generated after treatment with PMA

were mined and assembled during this study for the retrieval of genomes that could have originated from intact cells.

The draft metagenome-resolved genome (MRG) obtained via binning and assembly validates the read-based finding to determine the uncultivable species (Barnum et al. 2018). A previously published report shows that contemporary workers were able to identify the uncultivable organism *Candidatus Amarolinea aalborgensis* gen. nov., sp. nov., associated with a wastewater treatment plant via fluorescent in situ hybridization microscopy, 16S rRNA microbiome, and shotgun metagenomic sequence analysis but could not cultivate them to validly describe the microbial species (Andersen et al. 2019). Another study, based on 16S rRNA microbiome analysis from a marine saltern pond in southwest Spain revealed the presence of a dominant halophilic bacterium, enabling the researchers to define a suitable culture medium for successfully isolating and describing the novel genus *Spiribacter salinus* (Leon et al. 2014).

The ISS-MRG analyses during this study provided near-complete draft assemblages of seven genomes. Subsequently, based on the average nucleotide identity (ANI; < 95%) and digital DNA-DNA hybridization (dDDH; < 70%) characterizations (Goris et al. 2007), four of these seven MRGs were phylogenetically affiliated to a novel genus. Furthermore, seven strains belonging to this novel genus were retrieved from the ISS culture collection which were originally isolated from various locations of the ISS and archived (Venkateswaran 2017). 16S rRNA gene sequence analyses were not able to resolve taxonomies of these seven ISS strains because they exhibited high similarities with 25 validly described *Pantoea* species (> 97.5%). However, comparative whole genome sequence (WGS) analyses demonstrated that these seven ISS strains were phylogenetically not affiliated with any validly described genus that belong to the family *Erwiniaceae*. This is the first study that utilizes the “metagenome to phenome” approach to describe a novel genus from the ISS samples. The genomic and polyphasic taxonomic characterizations further support the proposal of *Kalamiella* as a novel genus and *K. piersonii* as a type species (IIIF1SW-P2^T).

The objectives of this study are to retrieve novel genomes from metagenome sequences of ISS environmental samples to narrow down the searching for cultured isolates from the archived culture collection of the ISS (~ 500 strains) to generate phenomes. In addition to the taxonomic characterization, comparative genomic analyses of seven *K. piersonii* strains and four MRGs were carried out to elucidate antimicrobial resistance phenotypic properties, multidrug resistance gene profiles, and genes related to potential virulence and pathogenic potential.

Materials and methods

Sample collection and isolation of bacteria

The sampling of ISS surfaces performed for this study took place within the US on-orbit segments. Samples collected during this study were: Node 3 (Location Nos. 1, 2, and 3), Node 1 (Location Nos. 4 and 5), Permanent Multipurpose Module (Location No. 6), US Laboratory (Location No. 7), and Node 2 (Location No. 8 and control). A detailed description of the various locations sampled was published elsewhere (Singh et al. 2018b). Sample collection from ISS environmental surfaces, processing, and cultivation of bacteria were previously reported (Urbaniak et al. 2018; Venkateswaran 2017).

Molecular characterization of pure culture

A loopful of purified microbial culture was subjected to DNA extraction with the UltraClean DNA kit (MO Bio, Carlsbad, CA) or Maxwell Automated System (Promega, Madison, WI) as per manufacturer instructions. The extracted DNA was eluted in 50 µL of molecular-grade water and stored at – 20 °C until further analysis. The 16S rRNA gene was amplified using the forward primer, 27F (5'-AGA GTT TGA TCC TGG CTC AG-3') and the reverse primer, 1492R (5'-GGT TAC CTT GTT ACG ACT T-3') (Checinska et al. 2015; Suzuki et al. 2001). PCR was performed with the following conditions: denaturation at 95 °C for 5 min, followed by 35 cycles consisting of denaturation at 95 °C for 50 s, annealing at 55 °C for 50 s, and extension at 72 °C for 1.5 min, and finalized by extension at 72 °C for 10 min. The amplified products were treated with Antarctic phosphatase and exonuclease (New England Biolabs, Ipswich, MA) to remove 5'- and 3'-phosphates from unused dNTPs before sequencing. Sequencing was performed by Macrogen (Rockville, MD, USA) using 27F and 1492R primers for *Bacteria*. The resulting sequences were assembled using SeqMan Pro from the DNASTar Lasergene package (DNASTAR Inc., Madison, WI). Bacterial sequences were searched against the EzTaxon-e database (Kim et al. 2012) and identified based on the closest percentage similarity (> 97%) to previously identified microbial type strains.

Whole genome sequencing of ISS strains

WGS sequencing was carried out on an Illumina HiSeq 2500 instrument with paired-end sequencing. The NGS QC Toolkit v2.3 (Patel and Jain 2012) was used to filter the data for high-quality vector- and adaptor-free reads for genome assembly (cutoff read length, 80%; cutoff quality score, 20). High-quality vector-filtered reads were used for assembly with SPAdes (Nurk et al. 2013) genome assembler using default parameters. Each genome was subsequently annotated with

the help of Rapid Annotations using Subsystems Technology (RAST) (Aziz et al. 2008). The quality of the final assemblies was checked using the Quast package (Gurevich et al. 2013).

Metagenome sequences and assembly of ISS samples

Metagenomic sequences were generated from PMA-treated samples to enable analyses of the intact portion of microbial cells during this study. The metagenome sequences generated from eight different locations of the ISS during three flights spanning a 15 month period were deposited in NCBI GenBank (Singh et al. 2018b). As already reported, the HiSeq 2500 platform (Illumina) was used for shotgun sequencing, resulting in 100-bp paired-end reads (Singh et al. 2018b). The paired-end 100-bp metagenomic reads were processed with Trimmomatic (Bolger et al. 2014) to trim adapter sequences and low-quality ends, with a minimum Phred score of 20 across the entire length of the read used as a quality cutoff. Reads shorter than 80 bp were removed after trimming. The remaining high-quality reads were subsequently assembled using metaSPAdes (Nurk et al. 2017). Contigs were binned using Metabat2 version 2.11.3 (Kang et al. 2015). Recovered genomes were evaluated with CheckM (Parks et al. 2015), and a recovered genome was considered good with at least 90% completeness and at most 10% contamination. All results were reconfirmed using the MetaWRAP pipeline (Uritskiy et al. 2018).

Phylogenetic analysis

Multilocus sequence analysis (MLSA)-based phylogenetic affiliation was performed as reported elsewhere to interpret the tree positions of the *Erwiniaceae* members considered in this study (Palmer et al. 2018). Representative genomes of *Erwinia*, *Tatumella*, *Mixta*, and *Pantoea* were used to determine the correct phylogenetic position of all ten isolated strains and seven MRGs. Representatives of order *Enterobacteriales* such as *Enterobacteriaceae*, *Pectobacteriaceae*, and *Yersiniaceae* were included in the phylogeny, with *Serratia marcescens* subsp. *marcescens* Db11 selected as the outgroup. Members of genus *Buchnera* and *Wigglesworthia* were not selected due to the absence of full-length genes in the publicly available database (NCBI). Full-length DNA sequences for the genes *atpD*, *gyrB*, *infB*, and *rpoB* (Adeolu et al. 2016) were retrieved from 62 genomes, and an MLSA phylogenetic tree was generated.

NCBI BLASTn (version 2.2.31+) was used to compare the gene similarities (Camacho et al. 2009). Each gene was individually aligned using Clustal-Omega (version 1.0.3) (Sievers et al. 2011), and alignments were visualized for accuracy in SeaView (version 4.5.4) (Galtier et al. 1996). The four gene sequences (*atpD*, *gyrB*, *infB*, and *rpoB*) were concatenated together using a custom Perl script (<https://github.com/>

[sandain/pigeon/blob/master/scripts/concat.pl](https://github.com/sandain/pigeon/blob/master/scripts/concat.pl)) in the order listed. A maximum-likelihood phylogeny was generated using FastTree (version 2.1.10) (Price et al. 2010) from the concatenated nucleotide alignment. Ecotype Simulation 2 (<https://github.com/sandain/ecosim>) was used to root the phylogeny using *Serratia marcescens* subsp. *marcescens* Db11 as the outgroup and to generate the image of the phylogenetic tree. Bootstrapping for the phylogeny was generated by PHYLIP's SEQBOOT (Felsenstein 2004), and the CompareToBootstrap.pl script provided by Price et al. (2010) was used to calculate the bootstrap values.

Phylogenomics and SNP analyses

Pairwise ANI was calculated using the algorithm from Goris et al. (2007) using EzTaxon-e (Kim et al. 2012). The dDDH analysis was performed using the Genome-to-Genome Distance Calculator 2.0 (GGDC 2.0) (Meier-Kolthoff et al. 2013). The single nucleotide polymorphism (SNP)-based phylogenetic tree was generated using CSIPhylogeny (Larsen et al. 2012) version 1.4. Using genome sequences of multiple isolates, CSIPhylogeny calls SNP, filters the SNPs, performs site validation, and infers a phylogeny based on the concatenated alignment of high-quality SNPs.

Differential gene analysis of WGS and MRGs

Genes identified by Palmer et al. (2018) as differentially present in the genomes of *Erwinia*, *Mixta*, *Pantoea*, and *Tatumella* were compared with the WGS and MRGs presented in this study using NCBI BLASTn (version 2.2.31+). A custom reference database created from nucleotide sequences was used to find genes from various *Pantoea* and *Tatumella* genomes used to construct the MLSA phylogeny. Sequences that aligned to the entire reference sequence (or nearly so, > 90%) were considered to be present in these newly sequenced genomes.

Nucleotide sequence deposition

The draft genome sequence of type strain IIF1SW-P2^T (ISS isolate) and ISS-IIF5SW (ISS metagenome) were deposited in NCBI under accession numbers SAMN10096957 and SAMN09635154. The accession number of all WGS and MRGs of ISS strains are given in Table 1. This whole-genome shotgun project has been deposited in GenBank under the BioSample No. SAMN10096957 and SUB ID No. SUB4539219. The version described in this paper is the first version, and the accession number for the isolate *K. piersonii* IIF1SW-P2^T is RARB00000000.

Table 1 Digital DDH (DNA-DNA Hybridization) and ANI (Average Nucleotide Identity) values of ISS strains WG and MRG comparison with members of *Erwaniaceae* family. (WG: Whole Genome of isolated strains; MG: Metagenome Resolved Genome)

Genus	Species	Strain	Accession number	DDDH	ANI
<i>Kalamiella</i> (WG)	<i>Kalamiella piersonii</i>	IIIF1SW-P2 ^T	SAMN10096957	100	100
	<i>Kalamiella piersonii</i>	IIIF2SW-P2		100	100
	<i>Kalamiella piersonii</i>	IIIF5SW-B1		100	100
	<i>Kalamiella piersonii</i>	IIIF5SW-B5		100	100
	<i>Kalamiella piersonii</i>	IIIF5SW-P1		99.9	100
	<i>Kalamiella piersonii</i>	IIIF6SW-B4		99.9	100
	<i>Kalamiella piersonii</i>	IIIF7SW-P2		100	100
<i>Kalamiella</i> (MG)	<i>Kalamiella piersonii</i>	ISS-IIIF1SW		99.2	100
	<i>Kalamiella piersonii</i>	ISS-IIIF5SW	SAMN09635154	99.8	100
	<i>Kalamiella piersonii</i>	ISS-IIIF7SW		99.7	100
	<i>Kalamiella piersonii</i>	ISS-IIIF8SW		99.7	100
<i>Pantoea</i> from ISS	<i>Pantoea brenneri</i>	ISS-IF5SW(MG)		23.4	81
	<i>Pantoea brenneri</i>	ISS-IIF5SW(MG)		23.4	81
	<i>Pantoea dispersa</i>	ISS-IIIF4SW (MG)		23.2	81
	<i>Pantoea brenneri</i>	IF5SW-P1(WG)	MIZY00000000	23.4	81
	<i>Pantoea dispersa</i>	IIIF4SW-B2		23.2	81
	<i>Pantoea dispersa</i>	IIIF4SW-B3		23.2	81
<i>Pantoea</i>	<i>Pantoea agglomerans</i>	NBRC 102470		22.7	81
	<i>Pantoea allii</i>	LMG 24248 ^T	MLFE00000000.1	21.5	80
	<i>Pantoea ananatis</i>	LMG 2665 ^T	JFZU00000000.1	21.4	80
	<i>Pantoea anthophila</i>	11–2	JXXL00000000.1	23.1	81
	<i>Pantoea brenneri</i>	LMG 5343 ^T	MIEI00000000.1	23.4	81
	<i>Pantoea citrea</i>	DSM 13699 RC	NZ_CP015579.1	19.1	78
	<i>Pantoea coffeiphila</i>	342	PDET00000000.1	20.5	79
	<i>Pantoea conspicua</i>	LMG 24534 ^T	MLFN00000000.1	23	81
	<i>Pantoea cypripedii</i>	LMG 2657 ^T	MLJI00000000.1	22.1	81
	<i>Pantoea deleyi</i>	LMG 24200 ^T	MIPO00000000.1	23.5	81
	<i>Pantoea dispersa</i>	EGD-AAK13	AVSS00000000.1	23.3	81
	<i>Pantoea eucalypti</i>	aB	AEDL00000000.1	22.3	80
	<i>Pantoea eucrina</i>	LMG 2781 ^T	MIPP00000000.1	22.3	81
	<i>Pantoea rodasii</i>	LMG 26273 ^T	MLFP00000000.1	21.5	80
	<i>Pantoea rwandensis</i>	LMG 26275 ^T	MLFR00000000.1	21	79
	<i>Pantoea septica</i>	LMG 5345 ^T	MLJJ00000000.1	41.2	90
	<i>Pantoea stewartii</i> subsp. <i>stewartii</i>	DC 283	AHIE00000000.1	21.7	80
	<i>Pantoea vagans</i>	C9-1	CP001894.1	22.8	81
	<i>Pantoea wallisii</i>	LMG 26277 ^T	MLFS00000000.1	22.7	81
	<i>Pantoea</i> sp.	At-9b	CP002436.1,	22	80
	<i>Pantoea</i> sp.	A4	ALXE00000000.1	20.7	79
	<i>Pantoea</i> sp.	GM01	AKUI00000000.1	22.1	80
	<i>Mixta</i>	<i>Mixta alhagi</i>	LTyr-11Z ^T	CP019706.1	21.1
<i>Mixta calida</i>		LMG 25383 ^T	MLFO00000000.1	22.2	81
<i>Mixta gaviniae</i>		LMG 25382 ^T	MLFQ00000000.1	22.9	81
<i>Mixta theicola</i>		QC 88-366 ^T		21.3	80
<i>Erwinia</i>	<i>Erwinia amylovora</i>	LA 636	CBVT00000000.1	21	79
	<i>Erwinia billingiae</i>	NCPPB 661 ^T	FP236826.1	20.4	79
	<i>Erwinia mallotivora</i>	BT-MARDI	JFHN00000000.1	19.7	78
	<i>Erwinia pyriformis</i>	DSM 12163 ^T	FN392236.1	20.9	79
	<i>Erwinia tasmaniensis</i>	Et 1-99 ^T	CU468131.1,	21	79

Table 1 (continued)

Genus	Species	Strain	Accession number	DDDH	ANI
	<i>Erwinia toletana</i>	DAPP-PG 735	AOCZ00000000.1	20.7	79
	<i>Erwinia tracheiphila</i>	PSU-1	APJK00000000.1	19.9	78
	<i>Erwinia</i> sp.	9145	JQNE00000000.1	20.8	79
<i>Tatumella</i>	<i>Tatumella morbirosei</i>	LMG 23360 ^T	CM003276.1	19.1	78
	<i>Tatumella ptyseos</i>	ATCC 33301 ^T	ATMJ00000000.1	19.7	78
	<i>Tatumella saanichensis</i>	NML 06-3099 ^T	ATMI00000000.1	19.1	78
<i>Outgroup</i>	<i>Brenneria goodwinii</i>	OBR 1	CGIG00000000.1	20.8	78
	<i>Cronobacter sakazakii</i>	ATCC 29544 ^T	CP011047.1	21.2	79
	<i>Enterobacter cloacae</i> subsp. <i>cloacae</i>	ATCC 13047 ^T	CP001918.1	21.4	80
	<i>Franconibacter helveticus</i>	LMG 23732 ^T	AWFX00000000.1	20.7	79
	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i>	ATCC 13883 ^T	JOOW00000000.1	20.7	79
	<i>Kluyvera ascorbata</i>	ATCC 33433 ^T	JMPL00000000.1	20.8	80
	<i>Pectobacterium carotovorum</i> subsp. <i>Carotovorum</i>	NCPPB 312 ^T	JQHJ00000000.1	20.8	79
	<i>Serratia marcescens</i> subsp. <i>marcescens</i>	ATCC 13880 ^T	JMPQ00000000.1	20.6	79
	<i>Yokenella regensburgei</i>	ATCC 49455 ^T	JMPS00000000.1	20.2	78

^T denotes that this is the type strain of this species

Phenotypic characterization of ISS strains

Phenotypic characterizations of all ISS strains were performed according to standard protocols (Jones 1981). Morphology, size, and pigmentation were observed on trypticase soy agar (TSA; BD Difco) plates after 24 h of incubation. Salt tolerance was determined using 1% peptone water supplemented with 2.0, 5.0, 7.0, and 10.0% (*w/v*) NaCl. The pH range and pH optimum were tested at pH 5.0, 6.5, 7.0, 8.0, 9.0, and 10.0 by adjusting pH with biological buffers (Xu et al. 2005). Growth at different temperatures was carried out in TSA by incubating the agar plates at 4, 10, 15, 20, 25, 30, 37, and 44 °C for 24 to 48 h. The Gram reaction was determined using the commercial kit (BD Difco), according to the manufacturer's instructions. Motility was checked using the method described by Skerman (Skerman 1967).

All seven isolates were biochemically identified using BioLog (Hayward, CA) carbon substrate utilization profile characterization (Wragg et al. 2014). For cellular fatty acid analysis, all strains were grown on nutrient agar medium at 30 °C for 24 h before collecting biomass. Cellular fatty acids were extracted, methylated, and analyzed by gas chromatography according to the instructions of the Sherlock Microbial Identification System (MIDI version 4.0), as described previously (Müller et al. 1998; Pandey et al. 2002).

Matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry protein analysis was carried out using freshly grown isolates on TSA media (Schumann and Maier 2014). The sample processing for MALDI-TOF, spectra analysis, and comparative analyses steps were followed as described recently (Seuylemezian et al. 2018). The Microflex LT bench-top mass spectrophotometry instrument (Bruker

Daltonics, Billerica, MA, United States) for generating spectra ($n = 3$) and FlexAnalysis software (Bruker Daltonics, Billerica, MA, USA) for processing raw spectral data, were used (Seuylemezian et al. 2018).

Phenotypic antibiotic resistance testing was carried out as described before (Urbaniak et al. 2018). Briefly, the isolates were streaked from glycerol stocks onto TSA plates. A single colony was inoculated into 5-mL Tryptic Soy Broth and grown overnight at 30 °C. Aliquots of 100 µL were plated on TSA. Agar diffusion discs (BD BBL™ Sensi-Disc™, Franklin Lakes, NJ) were placed aseptically on a plate, and the strains were incubated at 30 °C for 24 h. The tested antibiotics included: 30-µg cefazolin; 30-µg cefoxitin, 5-µg ciprofloxacin, 15-µg erythromycin, 10-µg gentamicin, 1-µg oxacillin, 10-µg penicillin, 5-µg rifampin, and 10-µg tobramycin. The diameter of inhibition zones was measured for each antibiotic disk and recorded in millimeters. The resistance results were compared with the zone diameter interpretive charts provided by the manufacturer.

Results

Metagenome-resolved genomes

Analyses of metagenomic sequences from eight different locations of the ISS enabled the recovery of seven nearly full-length MRGs related to members of the family *Erwiniaceae*. The ANI analysis (Table 1) showed high similarities of one MRG with *Pantoea dispersa* (99%), two MRGs with *Pantoea brenneri* (99%), but four MRGs were not phylogenetically affiliated with any of the known species of *Erwiniaceae* due to its lower ANI

values (<81%) and are proposed as novel during this study. Among the 22 validly described *Pantoea* species, only *Pantoea septica* was the closest with the four novel MRGs, and they exhibited a 90% ANI and 41.2% dDDH values which are too low to place them in the same genus (Goris et al. 2007). The four novel MRGs were recovered from ISS location Nos. 1 (Port panel of Cupola), 5 (Overhead-4- Zero-G Stowage Rack), 7 (Overhead-3 panel surface; LAB103), and 8 (Crew quarter bump-out exterior aft wall). Quality filtering of the shotgun metagenome sequences showed that the draft genome size of these novel MRGs were $\sim 4.8 \times 10^6$ bp. Table 2 summarizes assembly statistics and GenBank accession numbers of all MRGs assembled in this study qualifying the minimum information about a metagenome-assembled genome standard (Bowers et al. 2017).

Metagenomes to phenomes

The 16S rRNA gene sequences of ~ 500 strains archived from the ISS were queried against the four novel MRGs. In total, ten strains isolated from Flight No. 3 that exhibited high 16S rRNA gene sequence similarities (>98%) were further sequenced for their whole genomes. The WGS of these ten ISS strains were compared with the four novel MRGs and found that WGS of seven strains showed >99.9% ANI values. The ANI and dDDH analyses of the isolated strains with MRGs, and across the various members of the family *Erwiniaceae* confirm the novel identity of seven out of these ten strains (Table 1). The ANI values of the seven isolated strains that exhibited >99.5% ANI values with MRGs ranged from 79 to 81% for members of *Erwiniaceae* family, except for *Pantoea septica* (90%), which is below the species cutoff (>95%). The identity range represents that seven ISS strains are not related to any of the closely related genera of *Erwiniaceae* family. Additionally, the ANI results were further corroborated by the dDDH value that ranges between 19.1 to 23.4% except for *Pantoea septica* (41.2%), that lies below the species cutoff value of >70%. Lower ANI and dDDH values represent the genomic distance of all isolated strains from other genus members of *Erwiniaceae* family. Table 1 summarizes ANI and dDDH values of the isolated strains against the closely related members of *Erwiniaceae* family. In addition, the WGS of two isolated strains belong to validly described *P. dispersa* (>99% ANI), and one isolate was identified as *P. breneri* (98% ANI). The seven strains belong to novel genus were isolated from ISS location Nos. 1 (Port panel of Cupola), 2 (Forward side panel wall of the Waste and Hygiene Compartment), 5 (Overhead-4-Zero-G Stowage Rack), 6 (Port-2 Rack wall), and 7 (Overhead-3 panel surface; LAB103). A minimum of one strain was isolated in each of these locations with the exception that location No. 5 yielded three strains of the novel genus.

Table 2 Genome characteristics of the assembled metagenome-resolved genome (MRG) from ISS metagenome

Sample	MLST/genome identification	Coarse consistency (%)	Fine consistency (%)	CheckM completeness (%)	CheckM contamination (%)	Contig count	DNA size (bp)	Contigs N50 (bp)	Mean coverage
F3-1P	<i>K. piersonii</i>	92.5	89.2	89.69	0.38	168	4,802,117	61,466	50
F3-5P	<i>K. piersonii</i>	91.5	87.1	88.94	0	107	4,802,460	94,185	50
F3-7P	<i>K. piersonii</i>	90.4	86.3	87.7	0.34	146	4,656,932	59,699	50
F3-8P	<i>K. piersonii</i>	91.1	86.9	89.38	0.16	152	4,749,371	57,018	50
F1-5P	<i>Pantoea breneri</i>	93.8	90.6	100	1.95	139	5,160,744	122,775	165.22
F2-5P	<i>Pantoea breneri</i>	93.4	92.1	100	0.22	86	5,013,168	153,890	231.43
F3-4P	<i>Pantoea dispersa</i>	92.7	90.4	92.92	0.38	60	4,944,646	138,933	50

Genome sequence characteristics of novel *Kalamiella piersonii* strains

The Illumina HiSeq 2500 platform yielded 1.30×10^7 paired-end reads for the sequencing of isolated strains. Subsequently, the draft genome was assembled with a total of 1.27×10^7 reads after performing trimming and quality filtering of the paired-end sequences. The assembled draft genome consisted of 27 contigs with a genome size of $\sim 4.85 \times 10^6$ -bp. The median contig length (N50) size was 5.2×10^5 bp, with a mean coverage of 100×, G + C Mol % was 57.07%, and the number of coding sequences identified after RAST annotation were 4830. Table 3 summarizes assembly statistics of all seven novel ISS strains isolated and sequenced in this study.

MLSA analysis of MRGs and novel ISS strains

The MLSA-based phylogenetic analysis was based on the housekeeping genes identified in the whole genomes (*atpD*, *gyrB*, *infB*, and *rpoB*). Sixty-two genomes of members of family *Erwiniaceae* and related species, including the seven

MRG with all four MLSA genes were used to construct the phylogenetic tree. Concatenated sequences of all genes gave a phylogenetic tree with five clades excluding the members of *Enterobacteriaceae*, *Pectobacteriaceae*, and *Yersiniaceae* (Fig. 1). These clades can be summarized as *Pantoea* clade (22 species +3 MRG), *Kalamiella* clade ($n = 7$ strains +4 MRGs + *Pantoea septica* strain LMG 5345), *Mixta* clade (4 species), *Tatumella* clade (4 species), and *Erwinia* clade (6 species + *Pantoea coffeiphila* strain 342). Differential nesting was seen in the phylogenetic tree for the following organisms: *Pantoea coffeiphila* strain 342 is nested in *Erwinia* clade while *Pantoea septica* strain LMG 5345 is aligned more with the new clade *Kalamiella* proposed in this study.

Differential gene analysis of MRGs and novel ISS strains

A similar set of genes that were reported and identified as differentially present in the genomes of *Erwinia*, *Mixta*, *Pantoea*, and *Tatumella* (Palmer et al. 2018) were found in WGS, and MRGs are presented in this manuscript. Genomes

Table 3 Genome characteristics of the assembled whole genomes from ISS

Statistics	IIIF1SW-P2	IIIF2SW-P2	IIIF5SW-B1	IIIF5SW-B5	IIIF5SW-P1	IIIF6SW-B4	IIIF7SW-P2
# contigs	27	27	27	28	25	27	28
# contigs (≥ 0 bp)	27	27	27	28	25	27	28
# contigs (≥ 1000 bp)	27	27	27	28	25	27	28
# contigs (≥ 5000 bp)	22	22	22	23	21	22	22
# contigs ($\geq 10,000$ bp)	19	19	20	20	19	19	21
# contigs ($\geq 25,000$ bp)	17	17	17	17	17	17	18
# contigs ($\geq 50,000$ bp)	14	14	14	14	14	14	15
Largest contig	1,204,168	1,207,654	1,204,444	1,207,642	1,207,601	1,204,162	1,204,179
Total length	4,851,875	4,851,610	4,852,978	4,852,203	4,854,361	4,853,327	4,842,739
Total length (≥ 0 bp)	4,851,875	4,851,610	4,852,978	4,852,203	4,854,361	4,853,327	4,842,739
Total length (≥ 1000 bp)	4,851,875	4,851,610	4,852,978	4,852,203	4,854,361	4,853,327	4,842,739
Total length (≥ 5000 bp)	4,833,449	4,833,197	4,834,234	4,833,790	4,837,239	4,834,887	4,819,282
Total length ($\geq 10,000$ bp)	4,812,598	4,812,380	4,819,004	4,812,910	4,821,839	4,814,040	4,811,788
Total length ($\geq 25,000$ bp)	4,781,362	4,781,071	4,764,768	4,758,611	4,790,506	4,782,857	4,757,545
Total length ($\geq 50,000$ bp)	4,688,952	4,688,949	4,672,619	4,666,468	4,692,620	4,689,987	4,665,302
N50	522,655	519,786	520,195	519,805	519,762	519,739	479,968
N75	319,040	319,177	325,363	319,077	322,630	319,284	230,470
L50	3	3	3	3	3	3	3
L75	6	6	6	6	6	6	7
GC (%)	57.07	57.07	57.07	57.07	57.07	57.07	57.08
Mismatches							
No. Ns	103	41	173	69	71	87	71
No. Ns per 100 kbp	2.12	0.85	3.56	1.42	1.46	1.79	1.47
Coverage	100×	100×	100×	100×	100×	100×	100X
Coding sequences	4830	4827	4838	4830	4841	4834	4813

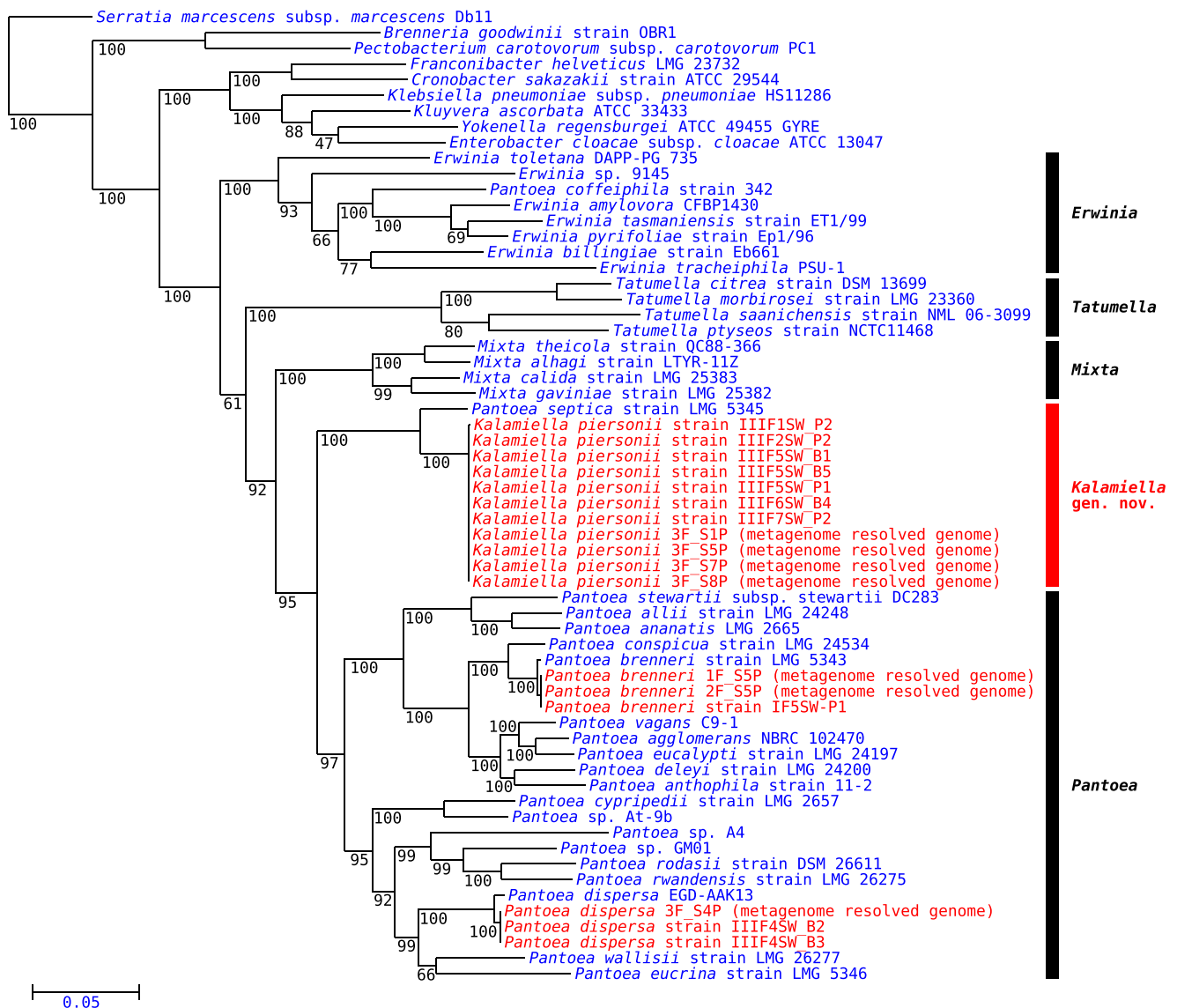


Fig. 1 Maximum-likelihood phylogeny for members of *Kalamiella* gen. nov., *Erwinia*, *Mixta*, *Pantoea*, *Tatumella*, and other genera of the family *Enterobacteriaceae* reconstructed from concatenated, full-length nucleotide sequences of the genes *atpD*, *gyrB*, *infB*, and *rpoB*. The concatenated sequence from *Serratia marcescens* subsp. *marcescens* Db11 was used as the out-group. Newly sequenced isolates and

metagenome-resolved genomes are highlighted in red. The thick black bars on the right indicate the genus of each clade, with *Kalamiella* gen. nov. highlighted in red. The scale bar indicates 5% nucleotide sequence divergence. Bootstrapping values are included on internal nodes in the phylogeny and represent the number of trials (out of 100) that included that particular branching pattern

identified as members of *Kalamiella* gen. nov. shared many genes with *Pantoea* and *Mixta* spp. (Table S4). Most of the genomes identified as *Kalamiella* gen. nov. lack the genes related to purine and pyrimidine metabolism (PRHOXNB or XDH, and *rutD* or *rutC*), whereas most of the genomes identified as *Pantoea* or *Mixta* include these genes, perhaps representing a metabolic difference in these genera. Differences between genomes identified as *Kalamiella* gen. nov. were found mainly in WGS (e.g., *bioF*, *garR*, *hutI*, and *nadA*) which are likely the result of gaps in the assembly and the inability to align an entire sequence to the reference and may not represent biological differences.

Antimicrobial and multidrug-resistant characteristics of ISS strain genome

A detailed genome analysis of type strain IIIF1SW-P2^T was carried out to understand its genetic makeup. Annotated features were classified as carbohydrate metabolism (348 genes), amino acid and derivatives (384 genes), protein metabolism (214 genes), cofactors, vitamins, prosthetic groups, pigments (170 genes), membrane transport (108 genes), and RNA metabolism (59 genes) (Fig. 2). To test antimicrobial resistance at the genomic level, the ISS strain was further compared with nosocomial isolates *Pantoea septica* having 90% ANI identity

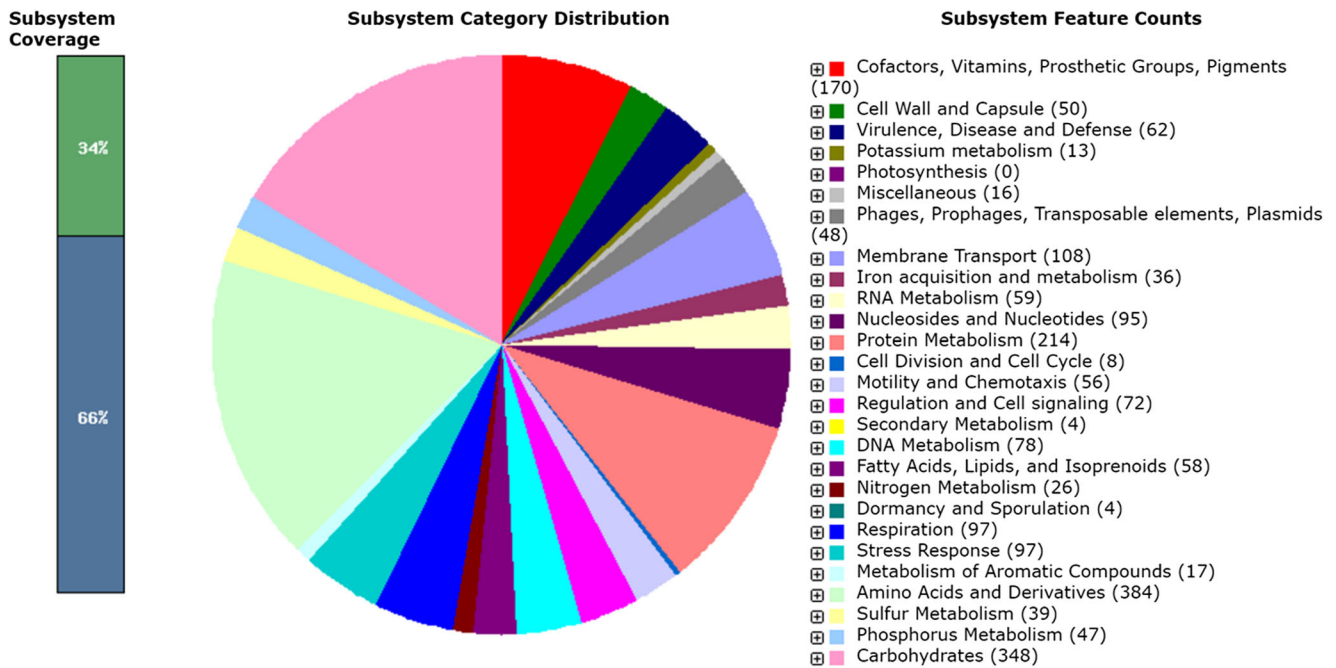


Fig. 2 Metabolic functional profiles and subsystem categories distribution of strain IIF1SW-P2^T (ISS isolate)

with the ISS strain. Features pertaining to virulence, disease, and defense were very similar in both the genomes, accounting for 50 genes in IIF1SW-P2^T and 49 genes in *P. septica*. Both organisms show the presence of cobalt-zinc-cadmium resistance, copper homeostasis, and *Mycobacterium* virulence operon as major virulence factors.

Phenotypic characterization of novel *Kalamiella piersonii* strains

All *Kalamiella* strains are Gram-stain-negative, motile and rod-shaped, and do not form endospores. Colonies are beige in color with circular and smooth edges. All strains grow at temperatures between 12 and 37 °C (with optimum growth at 30 °C), at a pH range between 6.0 and 9.0 (with optimum at 8.0) except for the strain IIF1SW-P2^T and IIF2SW-P2 that showed growth as high as at pH 10. All strains showed growth in the presence of 0–5% NaCl. The phenotypic differentiation of *Kalamiella* strains from other closely related genera belong to the family *Erwiniaceae* is shown in the supplementary Table S1.

The BioLog-based carbon substrate profiles of all members of the family *Erwiniaceae* are depicted in supplementary Table S1. All the species belonging to *Kalamiella*, *Erwinia*, *Mixta*, *Pantoea*, and *Tatumella* did not exhibit differential biochemical characteristics except for their utilization of D-cellobiose and gentiobiose. Both of these sugars were not assimilated by *Kalamiella* and *Erwinia* members, whereas the majority of *Mixta*, *Pantoea*, and *Tatumella* species utilized them as the sole carbon source.

Fatty acid methyl ester (FAME) profiles of *Kalamiella* strains and related genera are shown in supplementary Table S2. All genera of the family *Erwiniaceae* produced palmitic acid (C16:0; ~31 to 39%) in high amounts whereas *Erwinia* species can be differentiated from other genera of *Erwiniaceae* by the high C17:0 cyclo production (~30%) as reported elsewhere (Rezzonico et al. 2016). *Kalamiella* strains produced ~13% of C17:0 cyclo fatty acids, whereas *Mixta* (11%), *Pantoea* (7%), and *Tatumella* (9%) species produced less. *Kalamiella* strains produced C19:0 cyclo ω8c 9-(2-epitylcyclopropyl) nonanoic (2.5%), while only *Pantoea* showed the presence of this fatty acid (0.8%). Similarly, *Kalamiella* and *Mixta* strains were differentiated by not producing vaccenic acid (C18:1 ω7c), but *Pantoea*, *Erwinia*, and *Tatumella* strains showed the presence of this unsaturated fatty acid (Tracz et al. 2015). However, *Kalamiella* and *Mixta* species exhibited similar FAME profiles, but *Mixta* species produced iso-C19:0 which could be used to differentiate from each other (Chen et al. 2017; Palmer et al. 2018).

The MALDI-TOF profiles of the *K. piersonii* strains showed a score value of 1.8 with *P. septica* as the nearest match, while all other known species of *Pantoea* had a MALDI-TOF score < 1.5. This supports that *Kalamiella* strains should be considered a novel genus.

Antibiotic characteristics of novel *Kalamiella piersonii* strains

All seven *K. piersonii* isolates were resistant to ceftioxin, erythromycin, oxacillin, penicillin, and rifampin, while all strains were susceptible to cefazolin, ciprofloxacin, and

tobramycin (Table S3). When compared with the other three *Pantoea* species tested, *K. piersonii* strains did not differ much in its resistance profiles for the antibiotics tested via disc diffusion methods.

Discussion

The integration of the “classical to current” concept in microbiology has given researchers tools which were never available before. Classical microbiology depends on the culture of the organism and studying it as a live entity under a “polyphasic” analysis (Chęcinska Sielaff et al. 2017). On the other side is MRG where the organism is uncultured but the genetic tools give us a predicted genome for downstream analysis (Sangwan et al. 2016). The MRG approach leads to understanding genetic mobility, metabolic interactions (Barnum et al. 2018), expanding the tree of life (Parks et al. 2017), and also decoding the mammalian gut (Stewart et al. 2018). However reliable these tools and approaches are, live organisms are helpful to validate their existence. The fine line of intermediate understanding is where the phenomics lies and plays a crucial role in microbiological studies. Microbial phenomics can be defined as “use of classical knowledge in synchrony with omics data science in deciphering the genetic instructions of the genome, leading to the microbial understanding and cultivation.” Here, we present the very first study utilizing samples collected from ISS where metagenome to phenome approach was undertaken and which led to culturing and description of a novel bacterial genus.

Since the description of novel microbial species requires a cultured organism, an attempt was made to find the strains from the archived bacterial species (~500 strains) isolated from ISS samples that could match with the MRGs. It has been well established that strains that show <97.5% 16S rRNA gene sequence similarities would not exhibit >70% DDH values, a gold standard to describe novel species, thus the 16S rRNA gene was used to screen a large number of strains (Stackebrandt and Ebers 2006). The 16S rRNA gene sequences of 30 out of 500 strains archived from the ISS exhibited higher similarities (>97%) with MRGs, and among them 10 strains (>98%) were selected for WGS. In addition to the 16S rRNA gene (100%), the *gyrB* gene (>99%) that was reported to be the phylogenetic discriminator was amplified (La Duc et al. 2004). Comparative analyses of the 16S rRNA gene (>98%) and the *gyrB* gene (>95%) similarities yielded seven strains (Fig. S1) that were similar to the four MRGs. The WGS of these seven strains showed >99.9% ANI values with the novel MRGs, confirming that the metagenomes to phenomes approach would yield the cultured isolates required for describing novel species. The established threshold values

to delineate bacterial species were >95% for ANI and >70% for dDDH (Goris et al. 2007). Such an approach that enables researchers to isolate microbes also allows them to characterize the functional pathways, and their potential virulence properties that can directly affect human health. The SNPs analyses of cultivated strains and PMA-treated MRGs showed that strains isolated from ISS location Nos. 1 (Port panel of Cupola), 2 (Forward side panel wall of the Waste and Hygiene Compartment), 5 (Overhead-4-Zero-G Stowage Rack), 6 (Port-2 Rack wall), and 7 (Overhead-3 panel surface; LAB103) were identical. The absence of *K. piersonii* strains from location #8, where a novel MRG was recovered, might be due to their low abundance at this site and warrants further study. A quantitative PCR approach that was reported to be successful in measuring specific microbes of interest (Bargoma et al. 2013) should be implemented to selectively monitor for the presence of *K. piersonii*, when necessary.

Redox-active metals such as Co^{2+} , Zn^{2+} , and Ni^{2+} are essential cofactors in bacterial metabolism, but become toxic at higher concentrations (Nies 1992). Hosts to bacterial infection have adapted to this by adjusting their metal homeostasis as a defense mechanism against infection (Braymer and Giedroc 2014; Troxell and Hassan 2013). In an evolutionary arms-race, bacteria have evolved metal resistance mechanisms to overcome these defenses (Barber and Elde 2015). Recent incidences of infection on the ISS (Crucian et al. 2016) in addition to evidence for changes to human gene expression related to immune system function during extended stays in microgravity (unpublished) could explain the high incidence of cobalt-zinc-cadmium resistance genes detected in metagenomes sampled from various locations of the ISS (Singh et al. 2018b). Genes associated with metal resistance have been detected in the MRG and WGS sequenced in this study, including cadmium transporting ATPase (EC 3.6.3.3), cation efflux system proteins CusC and CusF, cobalt-zinc-cadmium resistance proteins CzcA, CzcB, and CzcD, copper sensory histidine kinase CusS, heavy metal-resistant transcriptional regulator HmrR, heavy metal RND efflux CzcC, and zinc transporter ZitB.

Variation lay in the presence of Beta-lactamase (EC 3.5.2.6), DNA-binding heavy metal response regulator and transcriptional regulator MerR family gene in *P. septica*, while the strain IIF1SW-P2^T had unique presence of cobalt-zinc-cadmium resistance protein, mercuric ion reductase (EC 1.16.1.1), FAD-dependent NAD(P)-disulphide oxidoreductase (Table S4). As these have very similar genomic resistance profiles, it should not be concluded that strain IIF1SW-P2^T is a human pathogen similar to *P. septica*. The *K. piersonii* IIF1SW-P2^T strain is a biofilm forming organism which is supported by the presence of genes involved in biofilm formation and quorum sensing. The ISS isolate had more stress-

resistant genes, especially those helping with osmotic stress and detoxification. These genes include betaine aldehyde dehydrogenase (EC 1.2.1.8), choline dehydrogenase (EC 1.1.99.1), high-affinity choline uptake protein BetT, OpgC protein, cell wall endopeptidase, family M23/M37, GST-like protein yncG, RNA-binding protein Hfq, anti-sigma B factor antagonist RsbV, serine phosphatase RsbU, alkanesulfonates ABC transporter ATP-binding protein, and FrmR, a negative transcriptional regulator of the formaldehyde detoxification operon. The above-mentioned genes were not present in the *P. septica*, the nearest Earth analog, which may be the adaptation mechanism of the organism towards the microgravity environment on the ISS or towards the industrial level cleaning regime.

The PathogenFinder (Cosentino et al. 2013) algorithm predicted that the strain IIF1SW-P2^T had ~42% probability of being a human pathogen, and hence *K. piersonii* strains should be rated as non-human pathogens. This was concluded on matches with genes from 15 pathogenic families and 21 nonpathogenic families in the database corresponding with the type strain.

With the inclusion of genomes as prokaryotic taxonomic descriptors (Chun and Rainey 2014), higher level taxons, like class Gammaproteobacteria, saw a prominent reclassification in taxon description. Subsequently, a new order “*Enterobacterales*” was defined in the Gammaproteobacteria and seven families were described including *Erwiniaceae* (Adeolu et al. 2016). The taxonomic description of the family *Erwiniaceae* is being updated with new advancements in genomics and phylogeny (Palmer et al. 2018). The *Erwiniaceae* family, consisting of phytopathogens (Hauben et al. 1998), has undergone multiple taxonomic revisions (Gardan et al. 2004; Mergaert et al. 1999; Rezzonico et al. 2016). The description of seven genera such as *Buchnera*, *Erwinia*, *Mixta*, *Pantoea*, *Phaseolibacter*, *Tatumella*, and *Wigglesworthia* were reported (Adeolu et al. 2016; Palmer et al. 2018). The genus *Pantoea* is ubiquitous and currently consists of 25 validly described species that are pathogenic to plants, insects, and animals (Walterson and Stavrinides 2015). Species associated with *Buchnera* were reported to be a symbiont of aphids (Munson et al. 1991), *Tatumella* consists of human pathogens (Hollis et al. 1981), *Phaseolibacter* members are plant pathogens (Halpern et al. 2013), and *Wigglesworthia* species are an insect endo-symbiont (Aksoy 1995).

In order to discriminate the genus/species combination in the family *Erwiniaceae*, a recent report utilized an MLSA strategy to reassign five *Pantoea* species into a new genus, *Mixta* (Palmer et al. 2018). Gene sequences of *atpD*, *gyrB*, *infB*, and *rpoB* were successfully used to differentiate members of the order *Enterobacterales* (Adeolu et al. 2016). The MLSA tree of 62 genomes that included the ISS genomes clearly shows that 4 MRGs and 7 ISS strains clade separately with *P. septica*, and belong to a new clade with *K. piersonii*.

This opens the possibility for reclassification of *P. septica* into the *Kalamiella* genus, but that is beyond the scope of this study. Phylogenetic distance and cladification of *Kalamiella* is also supported by the ANI and dDDH values where *Kalamiella* closely associates with *P. septica* and is distinctly different from all other members of family *Erwiniaceae*. These phylogenomic characterizations support the proposal of *Kalamiella* as a novel genus and *K. piersonii* as novel species, with the designated type strain of IIF1SW-P2^T that was isolated from the ISS Port panel of the Cupola (location No. 1. which is the observation deck for the crew).

Phenomics also leads to cover the gaps of “great plate anomaly” (Staley and Konopka 1985). A descriptive ISS metagenome analysis was carried out recently that led to the identification of a total of 318 defined microbial species, but only 32 (~10%) of them could be cultured (Singh et al. 2018b). A description of the novel species *Solibacillus kalamii* (Chęcinska Sielaff et al. 2017) found in ISS HEPA filter particulates was possible due to the WGS approach (Seuylemezian et al. 2017), which was found to be very closely related to *Solibacillus silvestris* (Krishnamurthi et al. 2009). Similarly, thorough genomic characterization of several ISS *Enterobacter* strains revealed their close association with clinical strains that caused diseases in neonatal patients (Doijad et al. 2016; Roach et al. 2015) and were also resistant to carbapenem and colistin (Norgan et al. 2016). Isolation of *E. bugandensis* strains helped to characterize its multidrug resistance; otherwise, metagenome and genome sequences could only predict such metabolic pathways (Singh et al. 2018a). Furthermore, the availability of bacterial isolates might allow researchers to analyze them in vivo to discern the influence of microgravity on their pathogenicity, and thorough “omics” characterization of isolates could help to understand the pathogenic potential (Singh et al. 2018a). Microbial phenomics has broad applications in space microbiology not only because of concerns for astronaut health but also for applied and basic microbiology study. Isolation of novel exo-sporium (rich in lipoproteins)-producing spore-forming bacterium, *S. kalamii*, might have great potential for biotechnological relevance (Chęcinska Sielaff et al. 2017). Isolation of enhanced pathogenic variants of known microbes (Singh et al. 2016) will help to improve our understanding of complex metabolic networks that control fundamental life processes under microgravity and in deep space.

***Kalamiella piersonii* gen nov., sp. nov. description**

Etymology: (N.L. fem. dim. n. *Kalamiella*, named after APJ Abdul Kalam (1934–2015), a well-known scientist who advanced space research in India. *piersonii* referring to Duane Pierson, an accomplished American space microbiologist).

Cells are Gram-stain-negative, aerobic, motile short rods (1–1.2 × 2.8 μm), occurring in the single or dual arrangement. Colonies are circular, convex with a diameter of approximately 0.6–1.0 mm and beige in color after 24 h of incubation on TSA medium at 30 °C. Cell growth occurs at 12 and 37 °C but not at 4 or 44 °C. The optimum growth was observed at 30 °C. The pH tolerance is between 6.0 and 10.0, with a pH optimum at 8.0. All of the strains displayed positive growth at 0–5% NaCl. The strains were positive for carbon source utilization of dextrin, D-maltose, *N*-acetyl-D-glucosamine, *N*-acetyl-β-D-mannosamine, α-D-glucose, D-mannose, D-fructose, D-galactose, L-rhamnose, inosine, 1% sodium lactate, D-mannitol, myo-inositol, glycerol, D-glucose-6-PO₄, D-fructose-6-PO₄, troleandomycin, rifamycin SV, glycyl-L-proline, L-alanine, L-arginine, L-aspartic acid, L-glutamic acid, L-histidine, lincomycin, guanidine HCl, niaproof 4, D-galacturonic acid, L-galactonic acid lactone, D-gluconic acid, D-glucuronic acid, glucuronamide, mucic acid, D-saccharic acid, vancomycin, tetrazolium violet, tetrazolium blue, citric acid, D-malic acid, L-malic acid, lithium chloride, and γ-amino-butyric acid. The strains were negative for carbon source utilization of D-trehalose, D-cellobiose, gentiobiose, sucrose, D-turanose, stachyose, D-raffinose, α-D-lactose, D-melibiose, β-methyl-D-glucoside, D-salicin, *N*-acetyl-D-galactosamine, *N*-acetylneuraminic acid, 3-methyl glucose, D-fucose, L-fucose, fusidic acid, D-serine, D-sorbitol, D-arabitol, D-aspartic acid, D-serine, minocycline, gelatin, L-pyroglytamic acid, L-serine, pectin, quinic acid, *p*-hydroxy-phenylacetic acid, methyl pyruvate, D-lactic acid methyl ester, L-lactic acid, α-keto-glutaric acid, bromo-succinic acid, nalidixic acid, potassium tellurite, Tween 40, α-hydroxy-butyric acid, β-hydroxy-D,L-butyric acid, α-keto-butyric acid, acetoacetic acid, propionic acid, acetic acid, formic acid, aztreonam, sodium butyrate, and sodium bromate. Major cellular fatty acids (> 10%) are C16:0, C17:0 cyclo, Summed Feature 3, and Summed Feature 8. Lesser fatty acids are C12:0, C14:0, C14:0 2-OH, C19:0 cyclo ω8c, and Summed Feature 2.

The type strain, IIIIF1SW-P2^T (=DSM 108198=NRRL B-65522^T), was isolated from the ISS Port panel of the Cupola, which is the observation deck for the crew. The DNA G + C content of the type strain is 57.07 mol% (whole genome).

Acknowledgments We thank Aleksandra Checinska Sielaff for isolating the ISS strains, Arman Seuylemezian for MALDI-TOF analysis. Ganesh Babu Malli Mohan, Cynthia Ly, and Tristan Grams for their technical help rendered in reviving and isolating DNA of the ISS strains when needed. © 2019 California Institute of Technology. Government sponsorship is acknowledged.

Authors' contributions KV and NKS conceived and designed the experiments. NKS, JMW, and SSM performed the experiments. NKS analyzed the data. JMW performed analysis of the de novo assemblies, including contig alignment, and annotation checks. SSM carried out the phenotypic assays. Biolog-based biochemical characterization. NKS, JMW, SSM, and KV wrote the paper. All authors read and approved the final manuscript.

Funding Part of the research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with National Aeronautics and Space Administration. This research was funded by a 2012 Space Biology NNN12ZTT001N grant no. 19-12829-26 under Task Order NNN13D111T award to KV, which also funded the post-doctoral fellowships for NKS, JMW and SSM.

Availability of data materials The data described in this manuscript can be freely and openly accessed on NCBI. Please see Table 1 and reference list for details and links to the data.

Compliance with ethical standards

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

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

References

- Adeolu M, Alnajjar S, Naushad S, SG R (2016) Genome-based phylogeny and taxonomy of the ‘*Enterobacteriales*’: proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. Int J Syst Evol Microbiol 66(12):5575–5599. <https://doi.org/10.1099/ijsem.0.001485>
- Aksoy S (1995) *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. Int J Syst Bacteriol 45(4): 848–851. <https://doi.org/10.1099/00207713-45-4-848>
- Andersen MH, McIlroy SJ, Nierychlo M, Nielsen PH, Albertsen M (2019) Genomic insights into *Candidatus Amarolinea aalborgensis* gen. Nov., sp. nov., associated with settleability problems in wastewater treatment plants. Syst Appl Microbiol 42(1):77–84. <https://doi.org/10.1016/j.syapm.2018.08.001>
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formisano K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O (2008) The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75. <https://doi.org/10.1186/1471-2164-9-75>
- Barber MF, Elde NC (2015) Buried treasure: evolutionary perspectives on microbial iron piracy. Trends Genet 31(11):627–636. <https://doi.org/10.1016/j.tig.2015.09.001>
- Bargoma E, La Duc MT, Kwan K, Vaishampayan P, Venkateswaran K (2013) Differential recovery of phylogenetically disparate microbes from spacecraft-qualified metal surfaces. Astrobiology 13(2):189–202. <https://doi.org/10.1089/ast.2012.0917>
- Barnum TP, Figueroa IA, Carlström CI, Lucas LN, Engelbrektson AL, Coates JD (2018) Genome-resolved metagenomics identifies genetic mobility, metabolic interactions, and unexpected diversity in perchlorate-reducing communities. ISME J 12(6):1568–1581. <https://doi.org/10.1038/s41396-018-0081-5>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

- Bonetta S, Pignata C, Bonetta S, Meucci L, Giacosa D, Marino E, Gilli G, Carraro E (2017) Viability of *Legionella pneumophila* in water samples: a comparison of Propidium Monoazide (PMA) treatment on membrane filters and in liquid. *Int J Environ Res Public Health* 14(5):467. <https://doi.org/10.3390/ijerph14050467>
- Bowers RM, Kypides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloie-Fadrosch EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glockner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35(8):725–731. <https://doi.org/10.1038/nbt.3893>
- Braymer JJ, Giedroc DP (2014) Recent developments in copper and zinc homeostasis in bacterial pathogens. *Curr Opin Chem Biol* 19:59–66. <https://doi.org/10.1016/j.cbpa.2013.12.021>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
- Checinska A, Probst AJ, Vaishampayan P, White JR, Kumar D, Stepanov VG, Fox GE, Nilsson HR, Pierson DL, Perry J, Venkateswaran K (2015) Microbiomes of the dust particles collected from the International Space Station and Spacecraft Assembly Facilities. *Microbiome* 3(1). <https://doi.org/10.1186/s40168-015-0116-3>
- Checinska Sielaff A, Kumar RM, Pal D, Mayilraj S, Venkateswaran K (2017) *Solibacillus kalamii* sp. nov., isolated from a high-efficiency particulate arrestance filter system used in the International Space Station. *Int J Syst Evol Microbiol* 67(4):896–901. <https://doi.org/10.1099/ijsem.0.001706>
- Chen C, Xin K, Liu H, Cheng J, Shen X, Wang Y, Zhang L (2017) *Pantoea alhagi*, a novel endophytic bacterium with ability to improve growth and drought tolerance in wheat. *Sci Rep* 7:41564. <https://doi.org/10.1038/srep41564>
- Chun J, Rainey FA (2014) Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int J Syst Evol Microbiol* 64(2):316–324. <https://doi.org/10.1099/ijms.0.054171-0>
- Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O (2013) PathogenFinder—distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One* 8(10):e77302. <https://doi.org/10.1371/journal.pone.0077302>
- Crucian B, Babiak-Vazquez A, Johnston S, Pierson DL, Ott CM, Sams C (2016) Incidence of clinical symptoms during long-duration orbital spaceflight. *Int J Gen Med* 9:383–391. <https://doi.org/10.2147/IJGM.S114188>
- Dojjad S, Imirzalioglu C, Yao Y, Pati NB, Falgenhauer L, Hain T, Foesel BU, Abt B, Overmann J, Mirambo MM, Mshana SE, Chakraborty T (2016) *Enterobacter bugandensis* sp. nov., isolated from neonatal blood. *Int J Syst Evol Microbiol* 66(2):968–974. <https://doi.org/10.1099/ijsem.0.000821>
- Felsenstein J (2004) PHYLIP (phylogeny inference package), version 3.6, vol 5, distributed by the author. Department of Genome Sciences, University of Washington, Seattle
- Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 12(6):543–548
- Gardan L, Christen R, Achouak W, Prior P (2004) *Erwinia papayae* sp. nov., a pathogen of papaya (*Carica papaya*). *Int J Syst Evol Microbiol* 54(Pt 1):107–113. <https://doi.org/10.1099/ijms.0.02718-0>
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57(1):81–91. <https://doi.org/10.1099/ijms.0.64483-0>
- Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8):1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Halpern M, Fridman S, Aizenberg-Gershtein Y, Izhaki I (2013) Transfer of *Pseudomonas flectens* Johnson 1956 to *Phaseolibacter* gen. nov., in the family *Enterobacteriaceae*, as *Phaseolibacter flectens* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 63(1):268–273. <https://doi.org/10.1099/ijms.0.033654-0>
- Hauben L, Moore ER, Vauterin L, Steenackers M, Mergaert J, Verdonck L, Swings J (1998) Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. *Syst Appl Microbiol* 21(3):384–397. [https://doi.org/10.1016/s0723-2020\(98\)80048-9](https://doi.org/10.1016/s0723-2020(98)80048-9)
- Hollis DG, Hickman FW, Fanning GR, Farmer JJ, Weaver RE, Brenner DJ (1981) *Tatumella ptyseos* gen. nov., sp. nov., a member of the family *Enterobacteriaceae* found in clinical specimens. *J Clin Microbiol* 14(1):79–88
- Jones DM (1981) Manual of methods for general bacteriology. *J Clin Pathol* 34(9):1069–1069
- Kang DD, Froula J, Egan R, Wang Z (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. <https://doi.org/10.7717/peerj.1165>
- Kibbee RJ, Örmeci B (2017) Development of a sensitive and false-positive free PMA-qPCR viability assay to quantify VBNC *Escherichia coli* and evaluate disinfection performance in wastewater effluent. *J Microbiol Methods* 132:139–147. <https://doi.org/10.1016/j.mimet.2016.12.004>
- Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, Park SC, Jeon YS, Lee JH, Yi H, Won S, Chun J (2012) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol* 62(Pt 3):716–721. <https://doi.org/10.1099/ijms.0.038075-0>
- Krishnamurthi S, Chakrabarti T, Stackebrandt E (2009) Re-examination of the taxonomic position of *Bacillus silvestris* Rheims et al. 1999 and proposal to transfer it to *Solibacillus* gen. nov. as *Solibacillus silvestris* comb. nov. *Int J Syst Evol Microbiol* 59(Pt 5):1054–1058. <https://doi.org/10.1099/ijms.0.65742-0>
- La Duc MT, Satomi M, Agata N, Venkateswaran K: gyrB as a phylogenetic discriminator for members of the *Bacillus anthracis-cereus-thuringiensis* group. *J Microbiol Methods* 2004, 56(3):383–394
- Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Ponten T, Ussery DW, Aarestrup FM, Lund O (2012) Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50(4):1355–1361. <https://doi.org/10.1128/jcm.06094-11>
- Leon MJ, Fernandez AB, Ghai R, Sanchez-Porro C, Rodriguez-Valera F, Ventosa A (2014) From metagenomics to pure culture: isolation and characterization of the moderately halophilic bacterium *Spiribacter salinus* gen. nov., sp. nov. *Appl Environ Microbiol* 80(13):3850–3857. <https://doi.org/10.1128/AEM.00430-14>
- Meier-Kolthoff JP, Auch AF, Klenk H-P, Goker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. <https://doi.org/10.1186/1471-2105-14-60>
- Mergaert J, Hauben L, Cnockaert MC, Swings J (1999) Reclassification of non-pigmented *Erwinia herbicola* strains from trees as *Erwinia billingiae* sp. nov. *Int J Syst Bacteriol* 49(Pt 2):377–383. <https://doi.org/10.1099/00207713-49-2-377>
- Müller KD, Schmid EN, Kroppenstedt RM (1998) Improved identification of mycobacteria by using the microbial identification system in combination with additional trimethylsulfonium hydroxide pyrolysis. *J Clin Microbiol* 36(9):2477–2480
- Munson MA, Baumann P, Kinsey MG (1991) *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-

- associated, primary endosymbionts of aphids. *Int J Syst Evol Microbiol* 41(4):566–568. <https://doi.org/10.1099/00207713-41-4-566>
- Nies DH (1992) Resistance to cadmium, cobalt, zinc, and nickel in microbes. *Plasmid* 27(1):17–28. [https://doi.org/10.1016/0147-619X\(92\)90003-S](https://doi.org/10.1016/0147-619X(92)90003-S)
- Norgan AP, Freese JM, Tuin PM, Cunningham SA, Jeraldo PR, Patel R (2016) Carbapenem- and colistin-resistant *Enterobacter cloacae* from Delta, Colorado, in 2015. *Antimicrob Agents Chemother* 60(5):3141–3144. <https://doi.org/10.1128/aac.03055-15>
- Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, McLean J, Lasken R, Clingenpeel SR, Woyke T, Tesler G, Alekseyev MA, Pevzner PA (2013) Assembling genomes and mini-metagenomes from highly chimeric reads. *Research in Computational Molecular Biology*. Springer, Berlin, p 158–170
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27(5):824–834. <https://doi.org/10.1101/gr.213959.116>
- Palmer M, Steenkamp ET, Coetzee MPA, Avontuur JR, Chan WY, van Zyl E, Blom J, Venter SN (2018) *Mixta* gen. nov., a new genus in the Erwiniaceae. *Int J Syst Evol Microbiol* 68(4):1396–1407. <https://doi.org/10.1099/ijsem.0.002540>
- Pandey KK, Mayilraj S, Chakrabarti T (2002) *Pseudomonas indica* sp. nov., a novel butane-utilizing species. *Int J Syst Evol Microbiol* 52(Pt 5):1559–1567. <https://doi.org/10.1099/00207713-52-5-1559>
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25(7):1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2(11):1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>
- Patel RK, Jain M (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7(2):e30619. <https://doi.org/10.1371/journal.pone.0030619>
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Rezzonico F, Smits THM, Born Y, Blom J, Frey JE, Goesmann A, Cleenwerck I, de Vos P, Bonaterra A, Duffy B, Montesinos E (2016) *Erwinia gerundensis* sp. nov., a cosmopolitan epiphyte originally isolated from pome fruit trees. *Int J Syst Evol Microbiol* 66(3):1583–1592. <https://doi.org/10.1099/ijsem.0.000920>
- Roach DJ, Burton JN, Lee C, Stackhouse B, Butler-Wu SM, Cookson BT, Shendure J, Salipante SJ (2015) A year of infection in the intensive care unit: prospective whole genome sequencing of bacterial clinical isolates reveals cryptic transmissions and novel microbiota. *PLoS Genet* 11(7):e1005413. <https://doi.org/10.1371/journal.pgen.1005413>
- Sangwan N, Xia F, Gilbert JA (2016) Recovering complete and draft population genomes from metagenome datasets. *Microbiome* 4(1):8. <https://doi.org/10.1186/s40168-016-0154-5>
- Schumann P, Maier T (2014) Chapter 13—MALDI-TOF mass spectrometry applied to classification and identification of Bacteria. In: Michael Goodfellow IS, Jongsik C (eds) *Methods in microbiology*, volume 41. Academic Press, pp 275–306
- Seuylemezian A, Aronson HS, Tan J, Lin M, Schubert W, Vaishampayan P (2018) Development of a custom MALDI-TOF MS database for species-level identification of bacterial isolates collected from spacecraft and associated surfaces. *Front Microbiol* 9(780):780. <https://doi.org/10.3389/fmicb.2018.00780>
- Seuylemezian A, Singh NK, Vaishampayan P, Venkateswaran K (2017) Draft genome sequence of *Solibacillus kalamii*, isolated from an air filter aboard the International Space Station. *Genome Announc* 5(35):e00696–e00617. <https://doi.org/10.1128/genomeA.00696-17>
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>
- Singh NK, Bezdán D, Sielaff AC, Wheeler K, Mason CE, Venkateswaran K (2018a) Multi-drug resistant *Enterobacter bugandensis* species isolated from the International Space Station and comparative genomic analyses with human pathogenic strains. *BMC Microbiol* 18(1):175. <https://doi.org/10.1186/s12866-018-1325-2>
- Singh NK, Blachowicz A, Checinska A, Wang C, Venkateswaran K (2016) Draft genome sequences of two *Aspergillus fumigatus* strains, isolated from the International Space Station. *Genome Announc* 4(4):e00553–e00516. <https://doi.org/10.1128/genomeA.00553-16>
- Singh NK, Wood JM, Karouia F, Venkateswaran K (2018b) Succession and persistence of microbial communities and antimicrobial resistance genes associated with International Space Station environmental surfaces. *Microbiome* 6(1):214. <https://doi.org/10.1186/s40168-018-0609-y>
- Skerman VDB (1967) A guide to the identification of the genera of bacteria, 2nd edn. Williams & Wilkins, Baltimore
- Stackebrandt E, Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 33(4):152–155
- Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* 39:321–346. <https://doi.org/10.1146/annurev.mi.39.100185.001541>
- Stewart RD, Auffret MD, Warr A, Wisner AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M (2018) Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat Commun* 9(1):870. <https://doi.org/10.1038/s41467-018-03317-6>
- Suzuki MT, Beja O, Taylor LT, Delong EF (2001) Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. *Environ Microbiol* 3(5):323–331
- Tracz DM, Gilmour MW, Mabon P, Beniac DR, Hoang L, Kibsey P, Van Domselaar G, Tabor H, Westmacott GR, Corbett CR et al: *Tatumella saanichensis* sp. nov., isolated from a cystic fibrosis patient. *Int J Syst Evol Microbiol* 2015, 65(Pt 6):1959–1966
- Troxell B, Hassan HM (2013) Transcriptional regulation by ferric uptake regulator (fur) in pathogenic bacteria. *Front Cell Infect Microbiol* 3:59. <https://doi.org/10.3389/fcimb.2013.00059>
- Urbaniak C, Sielaff AC, Frey KG, Allen JE, Singh N, Jaing C, Wheeler K, Venkateswaran K (2018) Detection of antimicrobial resistance genes associated with the International Space Station environmental surfaces. *Nat Sci Rep* 8(1):814. <https://doi.org/10.1038/s41598-017-18506-4>
- Uritskiy GV, DiRuggiero J, Taylor J (2018) MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6(1):158. <https://doi.org/10.1186/s40168-018-0541-1>
- Vaishampayan P, Probst AJ, La Duc MT, Bargoma E, Benardini JN, Andersen GL, Venkateswaran K (2013) New perspectives on viable microbial communities in low-biomass cleanroom environments. *ISME J* 7(2):312–324. <https://doi.org/10.1038/ismej.2012.114>
- Venkateswaran K (2017) Team I-MOE Microbial Characteristics of ISS Environmental Surfaces. In: 47th International Conference on Environmental Systems; ICES-2017-177, Charleston, South Carolina
- Vesper S, McKinstry C, Hartmann C, Neace M, Yoder S, Vesper A (2008) Quantifying fungal viability in air and water samples using quantitative PCR after treatment with propidium monoazide (PMA). *J Microbiol Methods* 72(2):180–184. <https://doi.org/10.1016/j.mimet.2007.11.017>

- Walterson AM, Stavriniades J (2015) *Pantoea*: insights into a highly versatile and diverse genus within the *Enterobacteriaceae*. FEMS Microbiol Rev 39(6):968–984. <https://doi.org/10.1093/femsre/fuv027>
- Weinmaier T, Probst AJ, Duc MT, Ciobanu D, Cheng JF, Ivanova N, Rattei T, Vaishampayan P (2015) A viability-linked metagenomic analysis of cleanroom environments: eukarya, prokaryotes, and viruses. Microbiome 3:6. <https://doi.org/10.1186/s40168-015-0129-y>
- Wragg P, Randall L, Whatmore AM (2014) Comparison of Biolog GEN III MicroStation semi-automated bacterial identification system with matrix-assisted laser desorption ionization-time of flight mass spectrometry and 16S ribosomal RNA gene sequencing for the identification of bacteria of veterinary interest. J Microbiol Methods 105: 16–21. <https://doi.org/10.1016/j.mimet.2014.07.003>
- Xu P, Li WJ, Tang SK, Zhang YQ, Chen GZ, Chen HH, Xu LH, Jiang CL: *Naxibacter alkalitolerans* gen. nov., sp. nov., a novel member of the family 'Oxalobacteraceae' isolated from China. Int J Syst Evol Microbiol 55(Pt 3):1149–1153

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