Zwartia hollandica gen. nov., sp. nov., Jezberella montanilacus gen. nov., sp. nov. and Sheuella amnicola gen. nov., comb. nov., representing the environmental GKS98 (betIII) cluster

Hahn, Martin W.; Pitt, Alexandra; Schmidt, Johanna; Koll, Ulrike; Wolf, Jacqueline; Whitman, William B.; Bodelier, Paul L.E.; Neumann-Schaal, Meina

published in
International Journal of Systematic and Evolutionary Microbiology
2022

DOI (link to publisher)
10.1099/ijsem.0.005513

document version
Publisher's PDF, also known as Version of record

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC DESCRIPTION

Hahn *et al.*, *Int. J. Syst. Evol. Microbiol.* 2022;72:005513

DOI 10.1099/ijsem.0.005513

MICROBIOLOGY SOCIETY

Zwartia hollandica gen. nov., sp. nov., *Jezberella montanilacus* gen. nov., sp. nov. and *Sheuella amnicola* gen. nov., comb. nov., representing the environmental GKS98 (betIII) cluster

Martin W. Hahn[1,*], Alexandra Pitt[1], Johanna Schmidt[1], Ulrike Koll[1], Jacqueline Wolf[2], William B. Whitman[3], Paul L. E. Bodelier[4] and Meina Neumann-Schaal[5]

Abstract

We present two strains affiliated with the GKS98 cluster. This phylogenetically defined cluster is representing abundant, mainly uncultured freshwater bacteria, which were observed by many cultivation-independent studies on the diversity of bacteria in various freshwater lakes and streams. Bacteria affiliated with the GKS98 cluster were detected by cultivation-independent methods in freshwater systems located in Europe, Asia, Africa and the Americas. The two strains, LF4-65^T (=CCUG 56422^T=DSM 107630^T) and MWH-P2sevCIIIb^T (=CCUG 56420^T=DSM 107629^T), are aerobic chemoorganotrophs, both with genome sizes of 3.2 Mbp and G+C values of 52.4 and 51.0 mol%, respectively. Phylogenomic analyses based on concatenated amino acid sequences of 120 proteins suggest an affiliation of the two strains with the family *Alcaligenaceae* and revealed *Orrella amnicola* and *Orrella marina* (= *Algicoccus marinus*) as being the closest related, previously described species. However, the calculated phylogenomic trees clearly suggest that the current genus *Orrella* represents a polyphyletic taxon. Based on the branching order in the phylogenomic trees, as well as the revealed phylogenetic distances and chemotaxonomic traits, we propose to establish the new genus *Zwartia* gen. nov. and the new species *Z. hollandica* sp. nov. to harbour strain LF4-65^T and the new genus *Jezberella* gen. nov. and the new species *J. montanilacus* sp. nov. to harbour strain MWH-P2sevCIIIb^T. Furthermore, we propose the reclassification of the species *Orrella amnicola* in the new genus *Sheuella* gen. nov. The new genera *Zwartia*, *Jezberella* and *Sheuella* together represent taxonomically the GKS98 cluster.

INTRODUCTION

The family *Alcaligenaceae* established by De Ley *et al.* in 1986 [1] currently contains, according to the LPSN database, 28 validly described genera [2]. Half of the type strains of type species contained in the family were obtained from clinical samples or from plants and animals. This includes, for instance, various clinical samples (e.g., *Bordetella* [3], *Advenella* [4]), a sample of marine algae (*Orrella marina* (=*Algicoccus marinus*) [5]), terrestrial plants (e.g. *Orrella dioscoreae* [6]) and ant guts (*Saccharedens* [7]). The other type strains were obtained from samples lacking a clinical or host-associated context. This includes five type strains obtained from soil (e.g. *Azohydromonas* [8]) four type strains originating from sewage (e.g. activated sludge, *Caenimicrobium* [9]), two type strains from natural freshwater (*Pusillimonas* [10] and *Parvibium* [11]), one type strain from bottled mineral water (*Ampullimonas* [12]) and one type strain from deep subseafloor sediment (*Verticiella* [13, 14],).

At present, most newly described bacterial taxa are only rarely detected in cultivation-independent investigations, e.g. by studies based on environmental 16S rRNA gene amplicon sequencing. This indicates that most of these taxa represent non-abundant bacteria. The two above-mentioned *Alcaligenaceae* genera from natural freshwater do not represent abundant freshwater bacteria [15, 16]. However,

Author affiliations: [1]Research Department for Limnology, University of Innsbruck, Mondseestrasse 9, A-5310 Mondsee, Austria; [2]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany; [3]Department of Microbiology, University of Georgia, Athens, GA 30602-2605, USA; [4]Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands; [5]Junior Research Group Bacterial Metabolomics, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.
***Correspondence:** Martin W. Hahn martin.hahn@uibk.ac.at
Keywords: bacterioplankton; freshwater; genome size; GKS98; *Orrella*.
Abbreviations:: AAI, average amino acid identity; AF, alignment fraction; DAM, dilution-acclimatization method; FAM, filtration-acclimatization method; gANI, whole genome average nucleotide identity; IMG, Integrated Microbial Genomes; NSY, nutrient broth–soyotone–yeast extract; RLBH, reverse line blot hybridization; SD, standard deviation.
The accession numbers of the whole genome and 16S rRNA sequences deposited at DDBJ/ENA/GenBank are for strain LF4-65^T are AJ964889 and JAHXRI000000000.1, respectively, and for strain MWH-P2sevCIIIb^T are AJ938029 and PVTV00000000.1, respectively.
Five supplementary tables and three supplementary figures are available with the online version of this article

it is well known that bacteria affiliated with this family are abundant pelagic freshwater bacteria [15, 16]. The taxon representing these abundant freshwater *Alcaligenaceae* bacteria is known as GKS98 [15] or betIII cluster [16]. The GKS98 cluster was described and named by Zwart and colleagues [15] based on phylogenetic analyses of almost full-length and partial 16S rRNA gene sequences originating from several cultivation-independent investigations on the diversity of bacterial communities in freshwater systems (mainly lakes and streams). The name GKS98 was derived from the name of the lake, i.e. Gossenköllesee, located in the Austrian Alps, from which the first 16S rRNA sequence of bacteria affiliated with that cluster was obtained by a cultivation-independent method [17]. Another meta study on diversity of freshwater bacteria by Newton and colleagues [16] improved the phylogenetic analyses and used the name betIII for the taxon largely being represented by the previously established GKS98 cluster.

The GKS98 cluster seems to lack any validly described species currently (but see below). To the best of our knowledge, none of the previous species descriptions mentioned that the proposed taxon is affiliated with the GKS98 cluster. Members of the GKS98 cluster were detected in various freshwater systems by many cultivation-independent investigations, including, for instance, Manzallah Lake, Egypt [18], Yellowstone Lake, Wyoming, USA [19], Adirondack Lake, New York, USA [20], Lake Piburger See, Austria [21], and Marathonas Reservoir, Greece [22]. In addition, cultivation of strains affiliated with the GKS98 cluster was reported previously [23, 24]. Here we describe two previously isolated strains affiliated with the GKS98 cluster and propose that they represent two new genera, *Zwartia* gen. nov. and *Jezberella* gen. nov. During the course of the performed phylogenetic analyses, we realized that the type strain of the previously described species *Orrella amnicola* [25] is, on the one hand, also affiliated with the GKS98 cluster, but, on the other hand, phylogenetically separated from the type species of the genus *Orrella* [6]. Therefore, we propose reclassification of the type strain representing *O. amnicola* in the new genus *Sheuella* gen. nov.

HABITATS AND ISOLATION

Strain LF4-65^T was isolated from shallow Lake Loosdrecht located at 52.204419°N and 5.081272°E (inferred by using Google Earth) at an altitude of about 0 m in the Netherlands. The lake was sampled on 21 January 2004 and the water sample was characterized by a pH of 8.0 and a conductivity of 420 µS cm^{-1}. Samples were cooled, and 1 day after sampling an isolation experiment was started. The dilution-acclimatization method (DAM) was used for the isolation experiment. Basically, DAM is similar to the filtration-acclimatization method (FAM [26]) but differs from FAM by using dilution instead of filtration steps. Tenfold dilution series were established, and 100 µl dilutions were inoculated in wells of 24-well-plates containing 500 µl medium each. The medium used in the wells and used for preparation of the dilution series were either IBM medium (inorganic basal medium) [26] or 0.1 µm filtered water from Lake Loosdrecht. In total, six 24-well plates were inoculated with samples from 10^{-2} to 10^{-7} dilutions. Thus, 12 wells were inoculated for each of the two media and each of the six dilutions. The plates were incubated at 15 °C and stepwise acclimatized to higher substrate concentrations by adding nutrient broth–soyotone–yeast extract (NSY) medium according to the standard acclimatization protocol [26]. After 18 days of acclimatization and incubation, 78.5% of the wells showed a visible turbidity. One and a half millilitres was harvested from each turbid well, DNA was extracted, and the samples were subjected to reverse line blot hybridization (RLBH) with probes specific for groups of abundant freshwater bacteria [27]. In parallel to sampling of the well plates for RLBH, 10 µl samples were transferred to NSY agar plates and incubated at 15 °C. Cultures corresponding to wells tested positive with RLBH probes were purified and stored as glycerol stocks at −70 °C. Strain LF4-65^T was isolated from a well positive for probe GKS98-442 [27] specific for the GKS98 cluster [15].

Strain MWH-P2sevCIIIb^T was isolated from a water sample taken from Pond-2 on 15 October 2003. This freshwater pond is located in the Austrian Alps at an altitude of 1291 m at the geographic coordinates 47.738586°N and 13.301633°E (inferred by using Google Earth) in Austria. Pond-2 is small, with a surface area about 0.1 hectare, shallow with a maximum depth of about 1 m and contains slightly humic water. Pond-2 is located only 100 m away from Pond-1, from which two new species were isolated previously [28, 29]. Both ponds are located in a traditional mountain pasture area used only during summer for pasturing of cows and horses. The water sample from which the strain was isolated had a temperature of 4.2 °C, pH 5.4, a conductivity of 41 µS cm^{-1} and an oxygen saturation of 82.1%.

Strain MWH-P2sevCIIIb^T was isolated by using the standard protocol of the FAM [26]. Briefly, 0.2 µm-filtered water samples from Pond-2 were used for inoculation of four 24-well plates. Each well received either 50, 100, 200 or 400 µl inoculum. IBM medium [26] was used as basal medium, and NSY medium [26] was used for stepwise acclimatization of the cultures to higher substrate concentrations by following the standard protocol. After the acclimatization procedure was completed, cultures were transferred to full-strength NSY medium and subsequently purified by three rounds of alternate growth in liquid NSY medium and plating of dilutions on NSY agar plates. The purified strain was stored as glycerol stock at −70 °C.

PHENOTYPIC CHARACTERIZATION

Besides the substrate utilization experiments (see below), the phenotypic characterization of the two strains was performed as described previously [30]. Both strains form rods but differ in the length of the typically formed rods (Table 1). Most cells of strain LF4-65^T appeared as short rods of less than 1 µm length, but some cell division stages appeared as short chains of up to six

Table 1. Phenotypic and chemotaxonomic characteristics of strains LF4-65^T, MWH-P2sevIIICb^T and their closest relative *Orrella amnicola* strain NBD-18^T (*Sheuella amnicola* gen. nov., comb. nov.)

The three type strains can be distinguished by the four discriminative traits (bold face): temperature range of growth, tolerated salinity range, oxidase and catalase reactions. Major fatty acids are considered with >2%. Q, ubiquinone; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; DPG, diphosphatidylglycerol; GL, glycolipids; L, unidentified lipid; APL, aminophospholipid; PDE, phosphatidyldimethylethanolamine; +, trait present (detected); –, trait absent (not detected); (i), inconclusive; ND, not determined.

Characteristic	**LF4-65^T**	**MWH-P2sevIIICb^T**	**NBD-18^T**
Cell shape	Short rods	Rods	Ovoid to rod-shaped
Cell size (length×width; µm)	0.6×0.4	1.4×0.4	1.4–1.8×0.5–0.9
Colony colour	Beige (unpigmented)	Beige (unpigmented)	White
Temperature range for growth (°C)	5–32	5–32	25–37
Salinity range for growth (% NaCl, w/v)	0–0.3	0–0.6	0–1.0
Oxidase	–	+	+
Catalase	–	+	–
Anaerobic growth (without/with nitrate)	–/–	–/(i)	–/ND
Motility (0.4% soft agar)	–	–	–
Ubiquinones (Q)	Q7 (72.5%), Q8 (27.5%)	Q8 (50.5%), Q7 (49.5%)	Q8 (100%)
Major fatty acids	$C_{16:1}$ ω7*c* (44.3%), $C_{16:0}$ (27.7%), $C_{18:1}$ ω7*c* (16.0%), $C_{14:0}$ 3-OH (4.8%), 11-methyl $C_{18:1}$ ω7*c* (2.4%), $C_{16:1}$ ω7*c* 2-OH (2.2%)	$C_{16:1}$ ω7*c* (36.9%), $C_{18:1}$ ω7*c* (23.7%), $C_{16:0}$ (22.4%), $C_{12:0}$ 3-OH (4.0%), $C_{17:0}$ cyclo ω7*c* (3.9%), $C_{16:1}$ ω7*c* 2-OH (2.9%), $C_{14:0}$ 3-OH (2.8%)	$C_{16:0}$ (31.6%), summed feature 3* (26.5%), $C_{17:0}$ cyclo (12.0%), summed feature 8† (10.5%), summed feature 2‡ (8.0%), $C_{14:0}$ (3.8%), $C_{16:1}$ 2-OH (2.8%)
Polar lipids	PE, DPG, PG, APL, 2×L	PE, DPG, PG, APL, GL, L	PE, PG, DPG, PDE, 3×APL, 5×L
DNA G+C content (mol%)	52.4	51.0	50.9
Genome size (Mbp)	3.2	3.2	3.2
Isolation source	Freshwater lake	Freshwater pond	Freshwater river
pH of water sample	8.0	5.4	7.2

*Summed feature 3 includes $C_{16:1}$ *ω7c* and/or $C_{16:1}$ *ω6c*.
†Summed feature 8 includes $C_{18:1}$ *ω7c* and/or $C_{18:1}$ *ω6c*.
‡Summed feature 2 includes $C_{14:0}$ 3-OH and/or iso-$C_{16:1}$.

cells. By contrast, a small fraction of the rods formed by strain MWH-P2sevCIIIb^T did not appear as elongated rods of more than 1 µm cell length but as curved rods of similar size. Salinity tolerance of the strains was tested with NSY agar plates supplemented with 0 (control), 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 1.0, 1.5 and 2.0% NaCl (w/v). The strains strongly differed in their salinity tolerance (Table 1). The temperature range enabling growth of the strains was tested by growth on NSY agar plates at 5, 10, 15, 20, 25, 30, 32, 33 and 34 °C. No differences were found between the two strains (Table 1). Additional growth tests were performed at 2 °C, which required long incubation periods and was not reliable because the temperature control was poor at this low temperature. Nevertheless, both strains showed weak growth under this condition. Anaerobic growth was tested by using NSY agar plates without and with supplementation with 2 g l^{-1} $NaNO_3$ [31] and incubation in an anaerobic jar. Strain MWH-P2sevCIIIb^T showed no anaerobic growth on standard NSY plates. On plates supplemented with nitrate, only inconclusive results were obtained. In a first experiment, very weak anaerobic growth was observed; however, in a second experiment, this very weak growth could not be reproduced, therefore the strain is considered rather negative for denitrification. Strain LF4-65^T showed in two experiments no anaerobic growth on both media (Table 1). Obviously, the investigated strains are not suitable for testing for some phenotypic traits, and the results of such tests should only be interpreted cautiously.

Motility was tested on NSY soft agar plates solidified with 0.4% agar. Both strains did not show any indications of motility in these tests. However, in contrast to strain LF4-65^T, in the genome of strain MWH-P2sevCIIIb^T no genes were detected putatively encoding synthesis of flagella and chemotaxis (see below). Potential utilization of organic substrates was tested in experiments with Biolog plates (Gen III MicroPlate, Biolog). The inoculation OD (590 nm) was 0.07 and the plates were incubated for 24 h at 25 °C. Both strains were only positive for a few test substrates but showed partially different responses. The obtained results are given in Table S1 (available with the online version of this article).

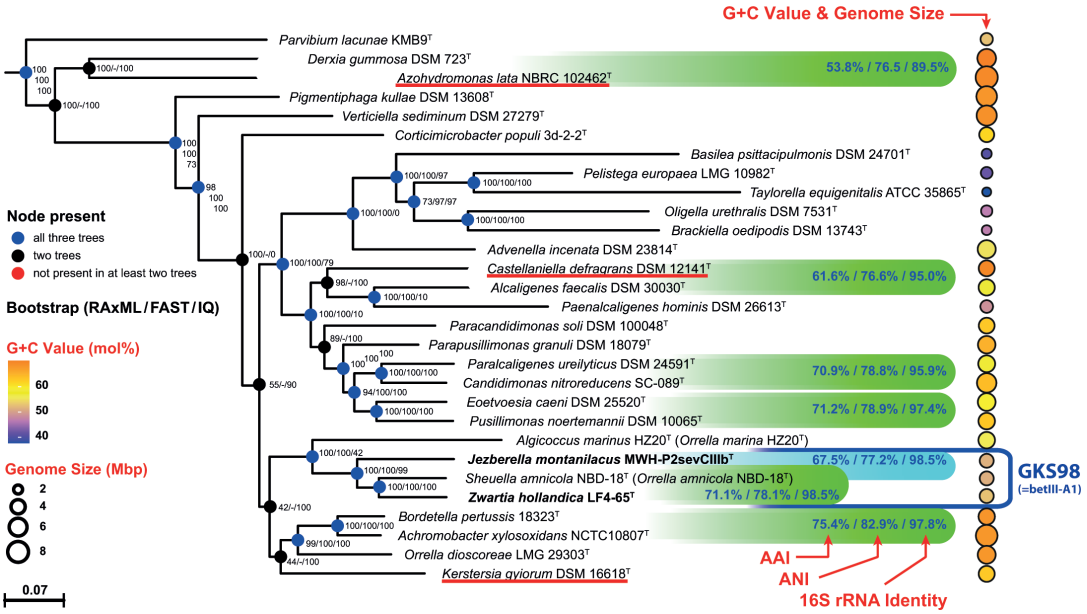

Fig. 1. Phylogenomic tree based on concatenated full-length amino acid sequences of 120 proteins. All type species of the family *Alcaligenaceae* with genome sequences available are shown (three genera are missing; compare Fig. S2 and Table S5). Due to the proposed taxonomic revisions, the type strains of two current *Orrella* species were included, which do not represent the type species of the genus. An RAxML tree supplemented with bootstrap values obtained from a FastTree and an IQ-Tree is shown. The three trees differ in branching order mainly regarding the position of three type strains (underlined in red). Note that the RAxML and the IG-Tree are highly similar in branching order of taxa. Furthermore, the figure displays data on size and G+C values of the genomes of the type strains. For selected neighbouring sibling taxa pairwise AAI, gANI and 16S rRNA sequence similarity values are displayed (green or blue areas indicate the respective pairs). All three trees were rooted by an outgroup consisting of the type strains of the *Burkholderiaceae* species *Cupriavidus necator*, *Pandoraea apista* and *Burkholderia cepacia* (all not shown). The accession numbers of the genomes of all taxa used for the reconstruction of the tree can be found in Table S5.

CHEMOTAXONOMIC CHARACTERIZATION

In general, the chemotaxonomic characterization of the investigated strains was performed as described previously [32]. The biomasses of both strains LF4-65^T and MWH-P2sevCIIIb^T contained the ubiquinones Q7 and Q8 but differed in proportional contributions of the two substances (Table 1). In the biomass of strain LF4-65^T were the major fatty acids $C_{16:1}$ ω7*c*, $C_{16:0}$, $C_{18:1}$ ω7*c*, $C_{14:0}$ 3OH, 11-methyl $C_{18:1}$ ω7*c* and $C_{16:1}$ ω7 c 2OH detected (Table 1). Furthermore traces (<2% of total fatty acids) of $C_{18:0}$, $C_{14:0}$, $C_{14:1}$ ω5*c* and $C_{16:0}$ 3OH were found. In the biomass of strain MWH-P2sevCIIIb^T were the major fatty acids $C_{16:1}$ ω7*c*, $C_{18:1}$ ω7*c*, $C_{16:0}$, $C_{12:0}$ 3OH, $C_{17:0}$ cyclo ω7*c*, $C_{16:1}$ ω7*c* 2OH and $C_{14:0}$ 3OH (Table 1). In addition, the fatty acids 11-methyl $C_{18:1}$ ω7*c*, $C_{18:0}$, $C_{14:1}$ ω5*c*, $C_{14:0}$, $C_{16:0}$ 3OH and $C_{10:0}$ 3OH were detected in traces (<2% of total fatty acids). The two investigated strains differed regarding the composition of polar lipids mainly in the presence of a glycolipid in the biomass of strain MWH-P2sevCIIIb^T and the presence of two instead of one unidentified lipid in the biomass of strain LF4-65^T (Table 1, Fig. S1). The G+C content of the DNA of the two strains differed only by 1.4 mol% (Table 1) but large differences were observed in comparison with several type species of genera affiliated with the family *Alcaligenaceae* (Fig. 1).

GENOMIC CHARACTERIZATION

The genome sequence of strain LF4-65^T was obtained by paired-end sequencing of a shotgun library by an Illumina MiSeq instrument (read length 300 bp). The obtained reads were assembled using the software Velvet version 1.2.10 [33]. The assembled genome is characterized by a coverage of 35× and consists of 25 scaffolds (including five contigs <1 000 bp). The genome size is 3.2 Mbp and the genomic G+C content is 52.4 mol%. The assembled genome sequence was annotated with the Integrated Microbial Genomes (IMG) annotation pipeline [34]. This process excludes all contigs with sequence lengths <1 000 bp. The annotated genome was integrated in the IMG database under the IMG Genome ID 2645728115 and deposited in the DDBJ/ENA/GenBank databases under the accession number JAHXRI000000000.1, respectively.

In the case of strain MWH-P2sevCIIIb^T, a shotgun library was paired-end sequenced by an Illumina HiSeq instrument (read length 150 bp), and the resulting reads were assembled by using the software SPAdes version 3.6.2 [35]. The assembled genome possesses a coverage of 469× and consists of 19 scaffolds (contig size >1 000 bp). The genome size is 3.2 Mbp and the genomic G+C content is 51.0 mol%. The genome was annotated by the IMG annotation pipeline. The annotated genome was integrated

in the IMG database under the IMG Genome ID 2681812960 and deposited in the DDBJ/ENA/GenBank databases under the accession number PVTV00000000.1, respectively.

In comparison to other type species of genera of the family *Alcaligenaceae* the two investigated strains possess medium sized genomes (Fig. 1). The majority of type strains representing type species of *Alcaligenaceae* genera possess larger genome sizes of up to 7.1 Mbp, while a group of host-associated or pathogenic strains including *Taylorella equigenitalis* possess much smaller genome sizes in the range of 1.7–2.3 Mbp (Fig. 1).

It is interesting to compare the gene content of the three strains LF4-65^T, MWH-P2sevCIIIb^T and *Orrella amnicola* NBD-18^T [25], which together represent the GKS98 cluster of abundant freshwater bacteria [15]. The genome assembly of strain LF4-65^T contains two fully assembled copies of the rRNA operons, which contain identical 16S rRNA genes. Each of the two copies is flanked up- and downstream by several individual genes; thus, the genome of this strain contains at least two copies of the ribosomal RNA operon. In the cases of strains MWH-P2sevCIIIb^T and NBD-18^T one complete rRNA operon was assembled, respectively; however in both cases these operons are contained in their own contigs. In contrast to the assembly of the LF4-65^T genome, these contigs lack in both cases any flanking genes not belonging to the rRNA operons. Thus, in the case of strains MWH-P2sevCIIIb^T and NBD-18^T the copy number of the ribosomal genes cannot be concluded from the genome assemblies. In order to get further insights in the copy number of the rRNA operon, mapping experiments were performed by using the software Bowtie2 [36]. Reads obtained by Illumina sequencing were mapped (end-to-end mode, ≥99% sequence identity) on the 16S rRNA genes and six single-copy housekeeping genes. Prior to mapping all genes were trimmed to a length of 1535 bp. In the case of strain LF4-65^T, the mapping coverage of the 16S rRNA gene was by average 201.8% (standard deviation, SD, of the six pairwise comparisons 39.0%) of the six single copy genes, in case of strain MWH-P2sevCIIIb^T the average percentage was 191.9% (SD 4.8%). This suggests that the genomes of both strains contain two copies of the rRNA operon, respectively. Such mapping analyses could not be performed with the genome of *Orrella amnicola* NBD-18^T because of the lack of public access to the read data obtained by the previous genome sequencing.

In all three genomes, genes involved in anoxygenic photosynthesis or putatively encoding proteorhodopsins were not detected. Since in other groups of abundant freshwater bacteria, such as *Limnohabitans* or *Polynucleobacter*, only a portion of the strains contain in their genomes genes putatively encoding photosynthesis systems [29, 37] or light-driven proton pumps like proteorhodopsins [30], these observations cannot ensure that all GKS98 strains lack such systems. Nevertheless, they do indicate that utilization of light energy is not widespread in this taxon. The three strains differ in the presence of several ecologically important genes and functions. Only in the genome of strain MWH-P2sevCIIIb^T were genes detected putatively encoding the synthesis of flagella and chemotaxis systems. The strains also differ in the presence of genes putatively encoding iron uptake systems. Since bioavailable iron is frequently a scarce resource in the water column of freshwater systems, differences in iron uptake systems may be of ecological relevance [38]. In contrast to the other two strains, strain LF4-65^T lacks genes putatively encoding a FeoAB iron-(II) transporter. Such transporters are assumed to be suitable for uptake of Fe^{2+} under acidic or anoxic conditions [38]. The genomes of strains LF4-65^T and NBD-18^T contain genes putatively encoding Afu ABC transporters assumed to be suitable for uptake of Fe^{3+} ions under alkaline conditions. Strain MWH-P2sevCIIIb^T lacks genes encoding such an ABC transporter. By contrast, all three strains possess genes putatively encoding iron complex ABC transporter for uptake of chelated iron ions. These patterns of genes involved in iron uptake may suggest that strain MWH-P2sevCIIIb^T and LF4-65^T are adapted to freshwater systems with either acidic or alkaline pHs, respectively, while strain NBD-18^T is potentially more flexible regarding pH or is adapted to alkaline as well as anoxic conditions. Note that analyses based on data of cultivation-independent investigations seem to indicate differences in pH adaptation among the three genera (see below).

Furthermore, the three strains differ in genes involved in acquisition of inorganic nitrogen. While all three strains contain putative ammonium transporter genes of the Amt family, only strains MWH-P2sevCIIIb^T and NBD-18^T possess genes putatively encoding ABC transporters for uptake of nitrate and nitrite. This pattern fits well to the presence of genes involved in the assimilatory nitrate and nitrite reduction, which were only found in these two strains but not in the genome of strain LF4-65^T. All three strains contain genes putatively encoding one or two ureases for cleaving urea to ammonia and carbon dioxide; however, only the annotations of the genome sequences of strain MWH-P2sevCIIIb^T and NBD-18^T suggested the presence of urea transporter genes. BLASTp searches with the protein sequences of those transporter genes against the proteome of strain LF4-65^T confirmed the absence of such genes (sequence identity >70%, coverage >70%) in the latter strain. It is assumed that strain LF4-65^T probably contains genes encoding an unrecognized urea transporter, otherwise the presence of urease-encoding genes is unexplained.

In contrast to strain LF4-65^T, the genomes of strains MWH-P2sevCIIIb^T and NBD-18^T contain genes suggesting that the strains are able to perform dissimilatory nitrate reduction and their putative ability to perform nitrate respiration under anoxic conditions. This would fit to the very weak anaerobic growth of strain MWH-P2sevCIIIb^T on medium supplemented with nitrate (Table 1). Strain NBD-18^T was reported to be strictly aerobic; however, anaerobic growth on media enriched with nitrate was not tested [25]. No hints on fermentative metabolisms were found in the genomes of the three strains.

All three strains contain genes putatively encoding cytochrome bd complexes, suggesting that all three strains are adapted to growth at low oxygen concentrations. However, as already mentioned, only two of the three strains possess genes potentially enabling anaerobic growth.

PHYLOGENY

Phylogenetic analyses were performed by calculating phylogenomic trees based on amino acid sequences of the set of 120 single-copy marker genes recommended by Parks *et al.* [39]. For this analysis, only the type strains of 25 of the 28 type species of validly described *Alcaligenaceae* genera could be considered because genome sequences of the remaining genera *Ampullimonas*, *Caenimicrobium* and *Saccharedens* were not available. The IMG database [40] was used for searching in genomes for the sequences of the 120 proteins. The genes encoding the proteins were selected based on the pfam or TIGR annotations listed by Parks *et al.* [39]. If a genome lacked a gene or contained more than one gene in a searched pfam or TIGR category, a BLASTp search was performed by using the amino acid sequence of the protein in the respective pfam or TIGR category of the genome sequence of the *Alcaligenes faecalis* type strain and the best BLAST hit was selected. The full-length nucleotide sequences of the genes were downloaded, concatenated and translated to amino acid sequences. The concatenated sequences were aligned by the software MAFFT version 7 [41] with the G-INS-1 option. This resulted in an amino acid data set with 56657 alignment positions. The alignment was filtered with GBlocks version 091b [42] in order to mask highly variable sequence positions. The following block parameters were used for the masking. Minimum number of sequences for a conserved position, 17; minimum number of sequences for a flank position, 27; maximum number of contiguous non-conserved positions, 8; minimum length of a block, 10; allowed gap positions, none. This resulted in a filtered alignment with 39190 positions in 492 blocks representing 69% of the initial alignment. The alignment was then used to calculate phylogenetic trees with the programs RAxML [43], IQ-Tree [44] and FastTree [45]. The obtained RAxML tree supplemented by bootstrap results from the other two trees and additional data on genome size and genomic G+C content of the type strains is shown in Fig. 1. The FastTree tree mainly differed in the position of three type strains from the other two trees; however, strains LF4-65^T and MWH-P2sevCIIIb^T form in all three trees a cluster with the type strains of *Orrella amnicola* [25] and *Orrella marina* [25] (*Algicoccus marinus* [5]). This cluster is well separated from other taxa and is well supported by high bootstrap values. This cluster contains two of the currently three species affiliated with the genus *Orrella*; however, the type strain of *Orrella dioscoreae*, which is the type species of the genus appears in a separate cluster. This clearly suggests that the current genus *Orrella* is polyphyletic and needs a taxonomic revision.

Because the phylogenomic trees lacked three *Alcaligenaceae* genera due to missing genome sequences, their phylogenetic position within the family was estimated by phylogenetic analyses of 16S rRNA gene sequences by using the software MEGA X [46]. The sequences representing all type species of the 28 genera affiliated with the family were aligned and trimmed to a uniform length, which resulted in an alignment length of 1382 bp. The program ModelTest [47] suggested Tamura-Nei TN+G+I to be the best substitution model. A maximum-likelihood tree was calculated (Fig. S2), which strongly differs from the phylogenomic tree in the low number of nodes supported by bootstrap values >70%. Importantly, this tree does not suggest that the three genera, which could not be added to the phylogenomic tree (Fig. 1), are closely related with the GKS98 cluster. Thus, their lack in the phylogenomic tree is not expected to strongly impact the reconstruction of the phylogenetic relationship of the three GKS98 taxa and their closest relatives. Importantly, both gene sets used for phylogenetic reconstructions suggest that the current genus *Orrella* represents a polyphyletic taxon (Figs 1 and S2).

In the phylogenomic and the 16S rRNA trees, the new strains LF4-65^T and MWH-P2sevCIIIb^T and the type strain of *Orrella amnicola* form a group well supported by high bootstrap values. This group represents the previously described GKS98 cluster [15]. The three strains of this cluster differ by lower G+C content and slightly smaller genome sizes from the majority of the *Alcaligenaceae* genera (Fig. 1). Only a cluster of five genera including *Basilea*, *Pelistega*, *Taylorella*, *Oligella* and *Brackiella* and the separately clustering genus *Paenalcaligenes* are characterized by smaller genome sizes and G+C content than the three GKS98 strains. All these taxa most likely represent obligate pathogens or at least obligately host-associated organisms. Thus, their low values potentially reflect lifestyle characteristics.

The three investigated GKS98 strains are characterized by rather high 16S rRNA gene sequence similarities of >98% when compared among each other (Fig. 1 and Table S2). By contrast, in the phylogenomic tree the three taxa appear with branches, which are in length similar or longer as many branches connecting the type species of related *Alcaligenaceae* genera (Fig. 1). This potentially hints on a rather slow molecular evolution of the 16S rRNA gene in the GKS98 strains compared to the protein-encoding part of their genomes. In order to further test if the 16S rRNA gene similarity of the GKS98 strains is unusually high in comparison to their genomic similarity, we plotted for all type species of the family *Alcaligenaceae* (Fig. 1) their respective minimal phylogenetic distance (120 proteins alignment) to the closest related taxon against the 16S rRNA gene similarity values of the same pairing. (Fig. 2). This plot indeed revealed that the molecular evolution of the 16S rRNA genes of GKS98 strains unusually lacks behind the evolution of the 120 protein-encoding genes used for the phylogenetic reconstruction. In order to explore if the genome-wide dissimilarities between the GKS98 strains rather reflects inter-genus or within-genus dissimilarities, we followed the approach by Barco and colleagues and plotted pairwise average nucleotide identity (gANI) and alignment fraction (AF) values of genomes

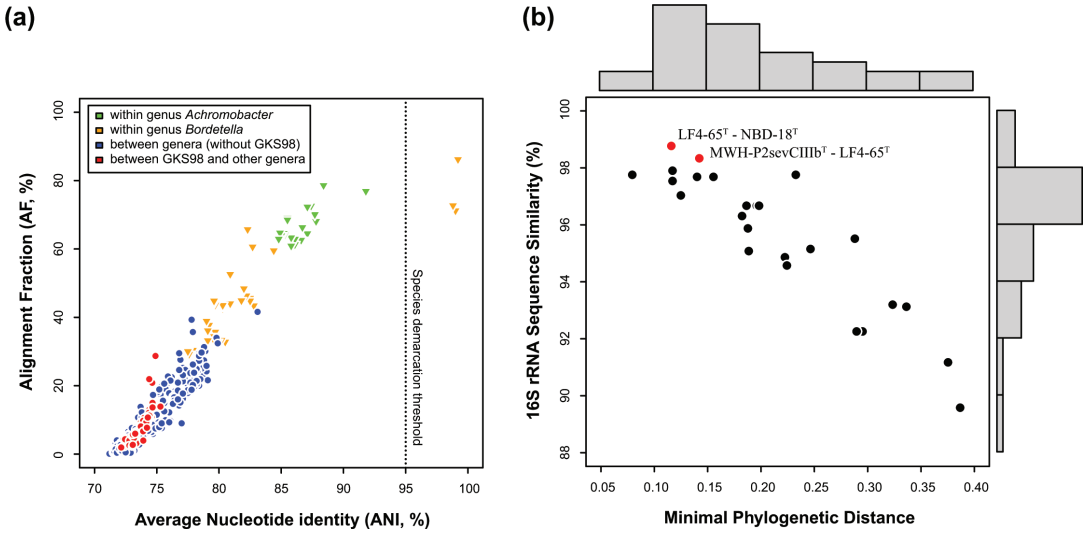

Fig. 2. (a) Pairwise data of average nucleotide identity (gANI) and alignment fractions (AF) for all type species of *Alcaligenaceae* genera represented by genome sequences and the three strains affiliated with the GKS98 cluster. In addition, data for within genus comparisons between type strains were plotted. Pairwise data of type species are represented by filled circles and within genera data are represented by filled triangles. (b) Comparison of the minimal phylogenetic distance of 120 protein sequences of each type species to another *Alcaligenaceae* type species and the corresponding 16S rRNA sequence similarity between the two type strains. The phylogenetic distance data and the *Alcaligenaceae* taxa were the same as those shown in Fig. 1. Along the upper *x*-axis and the right-hand *y*-axis the frequency distributions of the phylogenetic distances and the 16S rRNA similarity data are shown, respectively. Data for the two strains LF4-65^T and MWH-P2sevCIIIb^T are shown as red dots.

representing *Alcaligenaceae* type species [48]. The plotted data were obtained by using the ANIcalculator implemented in the JGI-Integrated Microbial Genomes and Microbiomes system [40]. In contrast to most other software used for gANI calculations, this software generates additionally AF values. This analysis (Fig. 2) suggests that the genomic dissimilarities between the three GKS98 strains are very well within the typical range observed for inter-genera comparisons within the family *Alcaligenaceae*. Obviously, the three GKS98 strains are characterized by rather unusually high 16S rRNA sequence similarity values, which do not reflect the phylogenetic distance between the three taxa appropriately. It is well known that 16S rRNA gene sequences are in some cases insensitive to evolutionary changes in the rest of the genome of a given organism [48].

ECOLOGY

The frequent detection of bacteria affiliated with the GKS98 cluster in the pelagic zone of freshwater systems and the isolation of strains LF4-65^T and MWH-P2sevCIIIb^T from the water column of two freshwater systems clearly suggests that these strains represent planktonic freshwater bacteria. This assumption is further supported by the detection of bacterial cells with GKS98-specific fluorescence *in situ* hybridization probes in the pelagic zone of Lake Gossenköllesee [17] and other freshwater systems [49].

Strains LF4-65^T and MWH-P2sevCIIIb^T were isolated from an alkaline and an acidic system, respectively. These pH conditions fit well to the above-mentioned differences in iron uptake systems, which suggests that strain LF4-65^T is adapted to alkaline and strain MWH-P2sevCIIIb^T to acidic freshwater systems. Interestingly, the majority of uncultivated bacteria forming a clade with strain LF4-65^T in a 16S rRNA tree (Fig. 3) were also obtained from circumneutral or alkaline waters, while the uncultivated bacteria forming a clade with strain MWH-P2sevCIIIb^T originated from acidic and slightly acidic waters. These differences in pH adaptation suggests that it could have been one of the driving forces for the evolution of two lineages. However, in other genera of freshwater bacteria, species adapted to acidic and alkaline conditions are present in a single genus [50]. Genome sequencing and analysis of more strains affiliated with the GKS98 cluster, and more environmental investigations are needed to confirm the suggested ecological tendencies.

BIOGEOGRAPHY

Sequences of cultured and uncultured bacteria affiliated with the GKS98 cluster were searched by performing BLAST analyses with 16S rRNA sequences of the two strains LF4-65^T and MWH-P2sevCIIIb^T. The affiliation with the cluster was confirmed by phylogenetic analyses (Fig. 3). Organisms originating from the same habitat were only included if they clustered differently within the GKS98 cluster. This was the case for two uncultured bacteria (FukuN65 and NE01) represented by three sequences, all obtained from Lake Große Fuchskuhle but in different years and in different studies [51, 52]. Most of the found sequences

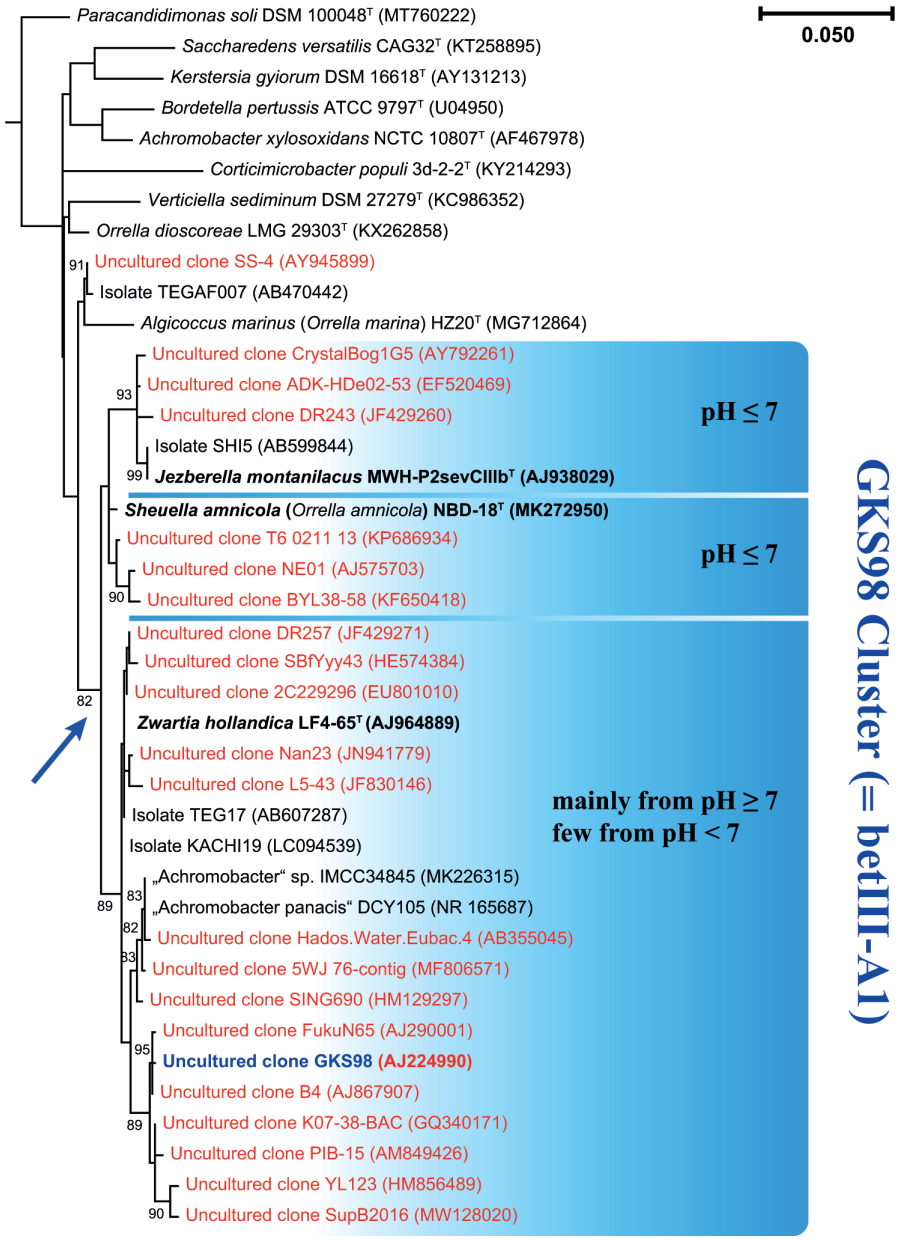

Fig. 3. Environmental sequences and isolated strains affiliated with the GKS98 cluster. The figure shows a part of a maximum-likelihood tree based on the 16S rRNA gene sequences of all type species of *Alcaligenaceae* genera (compare Fig. 2) and sequences of uncultured bacteria and isolated strains affiliated with the GKS98 cluster. The tree was truncated above the genus *Paracandidimonas*. The GKS98 cluster is highlighted in blue and the node defining this taxon is indicated by a blue arrow. Cultivated strains are shown with black font and uncultured bacteria are shown with red font. The uncultured bacterium after which the whole cluster was named is shown with blue font. Potential boarders between the three proposed genera affiliated with the cluster are shown, however, due to the limited phylogenetic power of 16S rRNA sequences, these borders need to be treated cautiously. For each of the three groups within the cluster, tendencies regarding pH of the habitats from which the isolates or sequences were obtained are given. The tree shows (as requested by the Editor) all NCBI accession numbers (Sanger sequences). The accession numbers of the genes or genomes used as sources for the 16S rRNA gene sequences of taxonomic reference taxa can be found in Table S5.

represent uncultured bacteria, however some cultured strains were also found. Most of the cultured strains were obtained by Keiji Watanabe and colleagues [53, 54] and originate from freshwater systems in Japan (e.g. strains KACHI19 and TEG17, Figs 3 and 4). This includes strains closely related with strain MWH-P2sevCIIIb^T, as well as a couple of strains affiliated with strain LF4-65^T. The only other found cultivated strains affiliated with the GKS98 cluster are two strains originating from South Korea, which were both classified as 'Achromobacter' strains. For one of these strains, Singh *et al.* proposed to establish the new species 'Achromobacter panacis' [55]. However, this species description was never validly published. In contrast to the other members

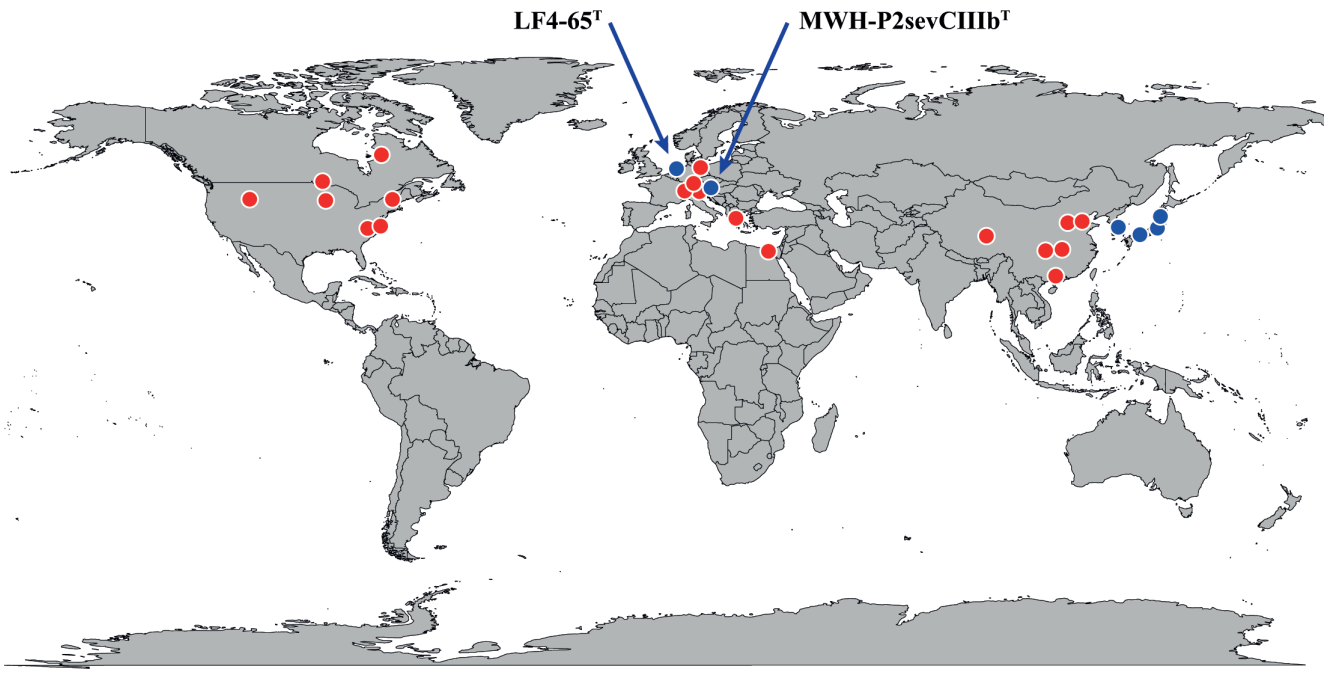

Fig. 4. World map showing the geographic origin of the cultured strains and the environmental sequences affiliated with the GKS98 cluster shown in Fig. 3. Sites from which uncultured bacteria were reported are depicted by red dots, while the sites from which cultured strains were obtained are depicted by blue dots. Note that only long sequences (>1300 bp) could be considered for this analysis and even not all such sequences are included. Short read sequences obtained by next generation sequencing methods were due to their limited phylogenetic resolution not considered. The map was constructed by using R [63] and the package Maps [64].

of the GKS98 cluster, the strain representing this taxon was not obtained from a freshwater system but from the rhizosphere of a ginseng plant. In addition, this strain differs in G+C content of its DNA, which was reported as 64.4 mol% [55]. This value is about 10 mol% higher than the values of the three characterized members of the GKS98 cluster (Fig. 1). Consistently with the previous description of the proposed species '**Achromobacter** panacis', the strain representing this taxon did not cluster in our 16S rRNA trees together with type strains of *Achromobacter* species ([55] and Fig. 3).

All bacteria affiliated with the GKS98 cluster shown in Fig. 3 were reported from locations of the Northern hemisphere (Fig. 4). Note that only long 16S rRNA sequences (>1 300 bp) were considered in this analysis, thus detections based on next generation sequencing techniques yielding rather short sequences were due to their limited phylogenetic resolution not included. These detections based on long sequences include cold habitats like two mountain lakes located in the European Alps at altitudes >2000 m [17] and permafrost thaw ponds located in the Canadian High Arctic [56]. By contrast, detections were also reported from freshwater systems in warm climates, e.g. from Egypt [18] and from Southern China [57]. Furthermore, habitats with detection of GKS98 bacteria include various rivers [58], large oligotrophic lakes (e.g. Lake Superior and Yellowstone Lake [19]), as well as small lakes [52] and an estuary (Delaware Bay [59]). Thus, bacteria of the GKS98 cluster seem to dwell in a geographically and ecologically broad variety of freshwater systems. It seems to be unlikely, that the cluster is not present in freshwater systems located in the tropical climate zone or located on the Southern hemisphere.

PROPOSAL OF NEW GENERA AND SPECIES

The branching order of the phylogenomic tree and comparisons of branch lengths indicate that strains LF4-65^T and MWH-P2sevCIIIb^T represent new genera within the family *Alcaligenaceae* (Fig. 1). The shortest phylogenetic distances to related type strains determined for these two strains are in the same range as the distances of other sibling genera of this family (Fig. 2). The notion that these two strains each represent a new genus is further supported by the observed gANI to AF ratios (Fig. 2), as well as by AAI values, which are at the same level or even below the values of other *Alcaligenaceae* sibling genera (Fig. 1 and Table S3). Complete gANI data sets calculated with the ANI Matrix Calculator of the Kostas Lab [60] and with the software JSpecies [61] are provided as Tables S3 and S4, respectively. Accession numbers of all 16S rRNA genes and genome sequences of the taxonomic reference taxa can be found in Table S5.

As outlined above, the 16S rRNA sequence similarity values observed for the GKS98 strains obviously do not reflect the phylogenetic and genomic distance between these taxa and are therefore not suitable for genus demarcation. The mismatch between distances based on amino acid sequences and ribosomal sequences hints either on a faster evolution of the proteome or a slower evolution of the 16S rRNA gene in the two strains compared to rates in other type strains of *Alcaligenaceae* species. However, decreased substitution rates in the 16S rRNA genes of GKS98 strains seem to be the evolutionary more parsimonious mechanism.

In conclusion, the establishment of two new genera for strains LF4-65^T and MWH-P2sevCIIIb^T is justified. Furthermore, the polyphyletic (Fig. 2) genus *Orrella* [6, 25] needs a taxonomic revision. We propose to transfer the type species of *Orrella amnicola* [25] to the new genus *Sheuella* gen. nov.

TAXONOMY OF THE GKS98 (BETIII) CLUSTER

The GKS98 cluster was phylogenetically defined by Zwart and colleagues [15] based on almost full-length and partial 16S rRNA gene sequences of uncultured bacteria. Almost a decade later, the phylogenetic characterization of the cluster was slightly refined by increasing the set of sequences affiliated with the cluster to twelve 16S rRNA sequences [16]. This new characterization resulted in the dividing of the cluster in the two subclusters betIII-A1 and betIII-A2 (Fig. S3). Demarcation of the GKS98 cluster from related taxa is difficult due to the small number of sequences initially used for defining the cluster. One opportunity for demarcation is provided by the specificity of oligonucleotide probes used for detection of GKS98-affiliated bacteria [17, 27, 49]. Interestingly, those probes pronouncedly differ in phylogenetic breadth (Fig. S3). If the probe with the broadest phylogenetic range, i.e. probe GKS98-646 [17], is considered for demarcation, the type strain of *Algicoccus marinus* (*Orrella marina*) is clearly excluded from the GKS98 cluster. This fits well to differences in ecology between bacteria affiliated with the cluster, which were predominantly described from freshwater systems, and the marine *Algicoccus marinus* (*Orrella marina*) type strain. From the todays point of view, betIII-A2 should not be considered as a part of the GKS98 (betIII) cluster, because betIII-A1 and betIII-A2 form a paraphyletic taxon and the two sequences defining betIII-A2 seem to be affiliated with the genus *Achromobacter* [62].

Based on the three genera proposed above, the GKS98 cluster could be divided in those tree genera, however, the 16S rRNA based phylogeny of the cluster may even suggest the presence of a potential fourth genus (Fig. S3). Interestingly, two cultured strains ('Achromobacter' spp.) affiliated with the potential fourth cluster were presented previously [55]. These strains could be used to test if a fourth genus is present in the GKS98 cluster. Until this open question is solved, we recommend to consider all taxa appearing in Figs 3 and S3 in parts of the GKS98 cluster, which also contains strain LF4-65^T as members of the genus *Zwartia* gen. nov. A detailed suggestion for preliminary genus assignment of strains affiliated with the GKS98 cluster is given in Fig. S3.

DISCRIMINATING FEATURES OF THE THREE PROPOSED GENERA

The three type strains representing the type species of the three proposed genera can be discriminated by the traits cell size, temperature range, NaCl tolerance, and oxidase/catalase reactions. The cells of strain LF4-65^T are smaller (cell length <1 µm) than the cells of the other two strains (>1 µm) (Table 1). Strain NBD-18^T differed from the other two strains by a lack of growth below 25 °C. Strain LF4-65^T did not show growth on NSY agar plates supplemented with more than 0.3% NaCl (w/v), while strain NBD-18^T was still able to grow with 1.0% NaCl (w/v). Strain MWH-P2sevIIICb^T showed both positive oxidase and catalase reactions, while strain LF4-65^T was negative for both reactions and strain NBD-18^T showed an intermediate response (Table 1).

DESCRIPTION OF *ZWARTIA* GEN. NOV.

Zwartia (Zwar'ti.a. N.L. fem. N., *Zwartia*, named after the Dutch scientist Gabriel Zwart, who pioneered the phylogenetic analyses of freshwater bacteria and was involved in the isolation of the strain representing the type species of the genus).

Strain LF4-65^T representing this new genus forms short, unpigmented rods with cell lengths <1.0 µm. Its biomass contains ubiquinones Q7 and Q8 and the polar lipids phosphatidylethanolamine, diphosphatidylglycerol, phosphatidylglycerol, an aminophospholipid and two unidentified lipids. The major fatty acid is $C_{16:1}$ ω7*c*, and the G+C content of the DNA is 52.4 mol%. The genome size is 3.2 Mbp. The genus is affiliated with the family *Alcaligenaceae* and the type species is *Zwartia hollandica* gen. nov., sp. nov.

DESCRIPTION OF *ZWARTIA HOLLANDICA* SP. NOV.

Zwartia hollandica (hol.lan'di.ca. N.L. fem. adj. *hollandica*, from Holland (The Netherlands), pertaining to the country of origin of the type strain).

Cells form short rods, approximately 0.6 µm long and 0.4 µm wide. They grow chemoorganotrophically and aerobically. Cells grown on NSY agar form beige (unpigmented) circular colonies. Growth was observed in the temperature range 5–32 °C, at higher temperatures growth was negative, at lower temperatures growth is likely. Growth occurs at NaCl concentrations up to 0.3% (w/v). Positive in Biolog assimilation and oxidation tests with the substrates dextrin, pectin, glucuronamide, and mucic acid.

Ubiquinones Q7 and Q8 were the only detected quinones. The two quantitatively most important fatty acids are $C_{16:1}$ ω7*c* and $C_{16:0}$. The polar lipid profile contains phosphatidylethanolamine, diphosphatidylglycerol, phosphatidylglycerol, an aminophospholipid and two unidentified lipids. The type strain is LF4-65^T (=CCUG 56422^T=DSM 107630^T), isolated from slightly alkaline water of Lake Loosdrecht, The Netherlands. The genome of the type strain is characterized by a DNA G+C content of 52.4 mol% and a size of 3.2 Mbp. The GenBank accession numbers for the 16S rRNA gene and the draft genome sequences of type strain LF4-65^T are AJ964889 and JAHXRI000000000.1, respectively.

DESCRIPTION OF *JEZBERELLA* GEN. NOV.

Jezberella (Jez.ber.el'la. N.L. fem. n. *Jezberella*, named after the deceased Czech microbial ecologist Jan Jezbera, who intensively worked on ecology and diversity of freshwater bacteria and investigated the freshwater system from which the type strain of the type species of the genus was obtained).

Strain MWH-P2sevIIICb^T representing this new genus forms unpigmented rods with cell lengths >1.0 µm. Its biomass contains ubiquinones Q7 and Q8 and the polar lipids phosphatidylethanolamine, diphosphatidylglycerol, phosphatidylglycerol, an aminophospholipid, a glycolipid and an unidentified lipid. The major fatty acid is $C_{16:1}$ ω7*c*, and the G+C content of the DNA is 51.0 mol%. The genome size is 3.2 Mbp. The genus is affiliated with the family *Alcaligenaceae* and the type species is *Jezberella montanilacus* gen. nov., sp. nov.

DESCRIPTION OF *JEZBERELLA MONTANILACUS* SP. NOV.

Jezberella montanilacus (mon.ta.ni.la'cus. L. masc. adj. *montanus*, of a mountain; L. gen. masc. n. *lacus*, lake; N.L. gen. masc. n. *montanilacus*, of a mountain lake).

Cells form rods, approximately 1.4 µm long and 0.4 µm wide. They grow chemoorganotrophically and aerobically. Cells grown on NSY agar form beige (unpigmented) circular colonies. Growth was observed in the temperature range 5–32 °C, at higher temperatures growth was negative, at lower temperatures growth is likely. Growth occurs at NaCl concentrations up to 0.6% (w/v). Positive in Biolog assimilation and oxidation tests with the substrates dextrin, pectin, glucuronamide, mucic acid, β-hydroxy-D,L-butyric acid and acetoacetic acid. Weak reaction in the Biolog test with L-lactic acid. Ubiquinones Q7 and Q8 are the only detected quinones. The two quantitatively most important fatty acids are $C_{16:1}$ ω7*c* and $C_{18:1}$ ω7*c*. The polar lipid profile contains phosphatidylethanolamine, diphosphatidylglycerol, phosphatidylglycerol, an aminophospholipid, a glycolipid and an unidentified lipid.

The type strain is MWH-P2sevCIIIb^T (=CCUG 56420^T=DSM 107629^T), isolated from slightly acidic water of Pond-2, Austria. The genome of the type strain is characterized by a DNA G+C content of 51.0 mol% and a size of 3.2 Mbp. The GenBank accession numbers for the 16S rRNA gene and the draft genome sequences of type strain MWH-P2sevCIIIb^T are AJ938029 and PVTV00000000.1, respectively.

PROPOSAL TO RECLASSIFY *ORRELLA AMNICOLA* SHEU *ET AL*. 2020

As already mentioned, the current genus *Orrella* [6] consisting of the three species *O. dioscoreae* (type species) [6, 25], *O. amnicola* [25] and *O. marina* (= *Algicoccus marinus*) [5, 25] consistently appeared in the phylogenomic tree (Fig. 1) and in the 16S rRNA tree (Fig. S2) to be a polyphyletic taxon. Furthermore, the current genus appears to be a chemotaxonomically quite heterogenous taxon, for instance, the type strains of the three species differ in G+C content in up to 16.5 mol% [25]. In order to resolve the phylogeny of the genus *Orrella*, we propose to establish for the species *O. amnicola* the new genus *Sheuella* gen. nov. The placement of the three type strains NBD-18^T, LF4-65^T and MWH-P2sevCIIIb^T in three separate genera is well justified by the observed phylogenetic distances separating these taxa, which are in the usual range of inter-genus relationships in the family *Alcaligenaceae* (Figs 1 and 2).

DESCRIPTION OF *SHEUELLA* GEN. NOV.

Sheuella (Sheu.el'la. N.L. fem. n., *Sheuella*, named after the Taiwanese microbiologist Shih-Yi Sheu, who previously described the species representing the type species of the genus).

The description of the new genus is based on the previous description of the species *Orrella amnicola* and the type strain NBD-18^T (=BCRC 81197^T=LMG 31338^T) by Sheu and colleagues [25].

Strain NBD-18^T representing this new genus forms rods with cell lengths >1.4 µm. Its biomass contains ubiquinone Q8 and the polar lipids phosphatidylethanolamine, diphosphatidylglycerol, phosphatidylglycerol, phosphatidyldimethylethanolamine. three aminophospholipids and five unidentified lipids. The major fatty acid is $C_{16:0}$, and the G+C content of the DNA is 50.9 mol%. The genome size is 3.2 Mbp. The genus is affiliated with the family *Alcaligenaceae* and the type species is *Sheuella amnicola* comb. nov.

DESCRIPTION OF *SHEUELLA AMNICOLA* COMB. NOV.

Sheuella amnicola (am.ni'co.la L. gen. masc. n. *amnis*, a stream, a small river; L. masc./fem. suff. *-cola*, a dweller, an inhabitant; from L. masc./fem. n. *incola*, dweller; N.L. masc./fem. n. *amnicola*, an inhabitant of a river).

The characteristics of the species are as given by Sheu and colleagues for the type strain of *Orrella amnicola* [25].

Funding information
Genome sequencing of strain MWH-P2sevCIIIb^T was performed in the GEBA III project by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, and supported by the Office of Science of the U.S. Department of Energy under contract number DE-AC02-05CH11231.

Acknowledgements
We thank Anika Wasner, Anja Frühling, Birgit Grün and Gesa Martens for excellent technical assistance.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. De Ley J, Segers P, Kersters K, Mannheim W, Lievens A. Intra- and intergeneric similarities of the bordetella ribosomal ribonucleic acid cistrons: proposal for a new family, Alcaligenaceae. *Int J Syst Bacteriol* 1986;36:405–414.
2. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic Names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–5612.
3. Vandamme PA, Peeters C, Cnockaert M, Inganäs E, Falsen E. *Bordetella bronchialis* sp. nov. *Bordetella flabilis* sp. nov. and *Bordetella sputigena* sp. nov., isolated from human respiratory specimens, and reclassification of *Achromobacter sediminum* Zhang *et al* 2014 as *Verticia sediminum* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 2015;65:3674–3682.
4. Coenye T, Vanlaere E, Samyn E, Falsen E, Larsson P, *et al*. *Advenella incenata* gen. nov., sp. nov., a novel member of the *Alcaligenaceae*, isolated from various clinical samples. *Int J Syst Evol Microbiol* 2005;55:251–256.
5. Ying J-J, Zhang S-L, Huang C-Y, Xu L, Zhao Z, *et al*. *Algicoccus marinus* gen. nov. sp. nov., a marine bacterium isolated from the surface of brown seaweed *Laminaria japonica*. *Arch Microbiol* 2019;201:943–950.
6. Carlier A, Cnockaert M, Fehr L, Vandamme P, Eberl L. Draft genome and description of *Orrella dioscoreae* gen. nov. sp. nov., a new species of *Alcaligenaceae* isolated from leaf acumens of *Dioscorea sansibarensis*. *Syst Appl Microbiol* 2017;40:11–21.
7. Lin JY, Hobson WJ, Wertz JT. *Saccharedens versatilis* gen. nov., sp. nov., a sugar-degrading member of the Burkholderiales isolated from *Cephalotes rohweri* ant guts. *Int J Syst Evol Microbiol* 2017;67:447–453.
8. Xie C-H, Yokota A. Reclassification of Alcaligenes latus strains IAM 12599T and IAM 12664 and *Pseudomonas saccharophila* as *Azohydromonas lata* gen. nov., comb. nov., *Azohydromonas australica* sp. nov. and *Pelomonas saccharophila* gen. nov., comb. nov., respectively. *Int J Syst Evol Microbiol* 2005;55:2419–2425.
9. Felföldi T, Schumann P, Mentes A, Kéki Z, Máthé I, *et al*. *Caenimicrobium hargitense* gen. nov., sp. nov., a new member of the family *Alcaligenaceae (Betaproteobacteria)* isolated from activated sludge. *Int J Syst Evol Microbiol* 2017;67:627–632.
10. Stolz A, Bürger S, Kuhm A, Kämpfer P, Busse H-J. *Pusillimonas noertemannii* gen. nov., sp. nov., a new member of the family *Alcaligenaceae* that degrades substituted salicylates. *Int J Syst Evol Microbiol* 2005;55:1077–1081.
11. Chen W-M, Xie P-B, Hsu M-Y, Sheu S-Y. *Parvibium lacunae* gen. nov., sp. nov., a new member of the family *Alcaligenaceae* isolated from a freshwater pond. *Int J Syst Evol Microbiol* 2018;68:1291–1299.
12. França L, Albuquerque L, Sánchez C, Fareleira P, da Costa MS. *Ampullimonas aquatilis* gen. nov., sp. nov. isolated from bottled mineral water. *Int J Syst Evol Microbiol* 2016;66:1459–1465.
13. Zhang Z, Fan X, Gao X, Zhang X-H. *Achromobacter sediminum* sp. nov., isolated from deep subseafloor sediment of South Pacific Gyre. *Int J Syst Evol Microbiol* 2014;64:2244–2249.
14. Vandamme PA. Proposal of *Verticiella* gen. nov. as replacement for the illegitimate prokaryotic genus name *Verticia* Vandamme et al. 2015. *Int J Syst Evol Microbiol* 2016;66:5099–5100.
15. Zwart G, Crump B, Kamst-van Agterveld M, Hagen F, Han S. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol* 2002;28:141–155.
16. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 2011;75:14–49.
17. Pernthaler J, Glockner FO, Unterholzner S, Alfreider A, Psenner R, *et al*. Seasonal community and population dynamics of pelagic bacteria and archaea in a high mountain lake. *Appl Environ Microbiol* 1998;64:4299–4306.
18. El Saied HE. Molecular genetic monitoring of bacterial communities in manzala lake, Egypt, based on 16S rRNA gene analysis. *Egyptian Journal of Aquatic Research* 2007;33:179–194.
19. Clingenpeel S, Macur RE, Kan J, Inskeep WP, Lovalvo D, *et al*. Yellowstone Lake: high-energy geochemistry and rich bacterial diversity. *Environ Microbiol* 2011;13:2172–2185.
20. Percent SF, Frischer ME, Vescio PA, Duffy EB, Milano V, *et al*. Bacterial community structure of acid-impacted lakes: what controls diversity? *Appl Environ Microbiol* 2008;74:1856–1868.
21. Salcher MM, Pernthaler J, Zeder M, Psenner R, Posch T. Spatio-temporal niche separation of planktonic betaproteobacteria in an oligo-mesotrophic lake. *Environ Microbiol* 2008;10:2074–2086.
22. Lymperopoulou DS, Kormas KA, Karagouni AD. Variability of prokaryotic community structure in a drinking water reservoir (Marathonas, Greece). *Microbes Environ* 2012;27:1–8.
23. Wu QL, Zwart G, Schauer M, Kamst-van Agterveld MP, Hahn MW. Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. *Appl Environ Microbiol* 2006;72:5478–5485.
24. Watanabe K, Komatsu N, Kitamura T, Ishii Y, Park H-D, *et al*. Ecological niche separation in the *Polynucleobacter* subclusters linked to quality of dissolved organic matter: a demonstration using a high sensitivity cultivation-based approach. *Environ Microbiol* 2012;14:2511–2525.
25. Sheu S-Y, Chen L-C, Yang C-C, Carlier A, Chen W-M. *Orrella amnicola* sp. nov., isolated from a freshwater river, reclassification of *Algicoccus marinus* as *Orrella marina* comb. nov., and emended description of the genus *Orrella*. *Int J Syst Evol Microbiol* 2020;70:6381–6389.
26. Hahn MW, Stadler P, Wu QL, Pöckl M. The filtration-acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. *J Microbiol Methods* 2004;57:379–390.

27. Zwart G, van Hannen EJ, Kamst-van Agterveld MP, Van der Gucht K, Lindström ES, *et al*. Rapid screening for freshwater bacterial groups by using reverse line blot hybridization. *Appl Environ Microbiol* 2003;69:5875–5883.
28. Hahn MW, Huymann LR, Koll U, Schmidt J, Lang E, *et al*. *Polynucleobacter wuianus* sp. nov., a free-living freshwater bacterium affiliated with the cryptic species complex PnecC. *Int J Syst Evol Microbiol* 2017;67:379–385.
29. Hahn MW, Schmidt J, Pitt A, Taipale SJ, Lang E. Reclassification of four *Polynucleobacter necessarius* strains as *Polynucleobacter asymbioticus* comb.nov., *Polynucleobacter duraquae* sp. nov., *Polynucleobacter yangtzensis* sp. nov., and *Polynucleobacter sinensis* sp. nov., and emended description of the species *Polynucleobacter necessarius*. *Int J Syst Evol Microbiol* 2016;66:2883–2892.
30. Pitt A, Schmidt J, Lang E, Whitman WB, Woyke T, *et al*. *Polynucleobacter meluiroseus* sp. nov., a bacterium isolated from a lake located in the mountains of the Mediterranean island of Corsica. *Int J Syst Evol Microbiol* 2018;68:1975–1985.
31. Pitt A, Schmidt J, Koll U, Hahn MW. *Aquirufa antheringensis* gen. nov. sp. nov. and *Aquirufa nivalisilvae* sp. nov., representing a new genus of widespread freshwater bacteria. *Int J Syst Evol Microbiol* 2019;69:2739–2749.
32. Hahn MW, Pitt A, Koll U, Schmidt J, Maresca JA, *et al*. *Aurantimicrobium photophilum* sp. nov., a non-photosynthetic bacterium adjusting its metabolism to the diurnal light cycle and reclassification of *Cryobacterium mesophilum* as *Terrimesophilobacter mesophilus* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 2021;71.
33. Zerbino DR. Using the Velvet *de novo* assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics* 2010;11:11.5.
34. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, *et al*. The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). *Stand Genomic Sci* 2015;10:86.
35. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
36. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
37. Kasalický V, Zeng Y, Piwosz K, Šimek K, Kratochvilová H, *et al*. Aerobic anoxygenic photosynthesis is commonly present within the genus *Limnohabitans*. *Appl Environ Microbiol* 2018;84:02116–02117.
38. Hahn MW, Jezberová J, Koll U, Saueressig-Beck T, Schmidt J. Complete ecological isolation and cryptic diversity in *Polynucleobacter* bacteria not resolved by 16S rRNA gene sequences. *ISME J* 2016;10:1642–1655.
39. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, *et al*. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.
40. Chen I-MA, Markowitz VM, Chu K, Palaniappan K, Szeto E, *et al*. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res* 2017;45:D507–D516.
41. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2019;20:1160–1166.
42. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;56:564–577.
43. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
44. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
45. Price MN, Dehal PS, Arkin AP. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5:e9490.
46. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
47. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 1998;14:817–818.
48. Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, *et al*. A genus definition for bacteria and archaea based on a standard genome relatedness index. *mBio* 2020;11:e02475–02419.
49. Salcher MM, Posch T, Pernthaler J. *In situ* substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J* 2013;7:896–907.
50. Hahn MW, Huemer A, Pitt A, Hoetzinger M. Opening a next-generation black box: ecological trends for hundreds of species-like taxa uncovered within a single bacterial >99% 16S rRNA operational taxonomic unit. *Mol Ecol Resour* 2021;21:2471–2485.
51. Glöckner FO, Zaichikov E, Belkova N, Denissova L, Pernthaler J, *et al*. Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. *Appl Environ Microbiol* 2000;66:5053–5065.
52. Burkert U, Warnecke F, Babenzien D, Zwirnmann E, Pernthaler J. Members of a readily enriched beta-proteobacterial clade are common in surface waters of a humic lake. *Appl Environ Microbiol* 2003;69:6550–6559.
53. Watanabe K, Ishii Y, Komatsu N, Kitamura T, Watanabe M, *et al*. Growth rates and tolerance to low water temperatures of freshwater bacterioplankton strains: ecological insights from shallow hypereutrophic lakes in Japan. *Hydrobiologia* 2016;792:67–81.
54. Watanabe K, Komatsu N, Ishii Y, Negishi M. Effective isolation of bacterioplankton genus *Polynucleobacter* from freshwater environments grown on photochemically degraded dissolved organic matter. *FEMS Microbiol Ecol* 2009;67:57–68.
55. Singh P, Kim YJ, Singh H, Farh M-A, Yang D-C. *Achromobacter panacis* sp. nov., isolated from rhizosphere of *Panax ginseng*. *J Microbiol* 2017;55:428–434.
56. Negandhi K, Laurion I, Lovejoy C. Bacterial communities and greenhouse gas emissions of shallow ponds in the high Arctic. *Polar Biol* 2014;37:1669–1683.
57. Liu Z, Huang S, Sun G, Xu Z, Xu M. Phylogenetic diversity, composition and distribution of bacterioplankton community in the Dongjiang River, China. *FEMS Microbiol Ecol* 2012;80:30–44.
58. Balmonte JP, Arnosti C, Underwood S, McKee BA, Teske A. Riverine bacterial communities reveal environmental disturbance signatures within the *Betaproteobacteria* and *Verrucomicrobia*. *Front Microbiol* 2016;7:1441.
59. Shaw AK, Halpern AL, Beeson K, Tran B, Venter JC, *et al*. It's all relative: ranking the diversity of aquatic bacterial communities. *Environ Microbiol* 2008;10:2200–2210.
60. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* 2016. DOI: 10.7287/peerj.preprints.1900v1.
61. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 2009;106:19126–19131.
62. Yabuuchi E, Yano I. *Achromobacter* gen. nov. and *Achromobacter xylosoxidans* (ex Yabuuchi and Ohyama 1971) nom. rev. *Int J Syst Bacteriol* 1981;31:477–478.
63. Team RC. R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing*. Vienna, Austria URL, 2019. https://www.R-project.org/
64. Minka TP, Deckmyn A. Maps: Draw geographical maps. R package version 3.3.0; 2018. https://CRAN.R-project.org/package=maps