



Characterization of the First Cultured Psychrotolerant Representative of *Legionella* from Antarctica Reveals Its Unique Genome Structure

Sho Shimada,^{a,c} Ryosuke Nakai,^b Kotaro Aoki,^a Sakae Kudoh,^{d,e} Satoshi Imura,^{d,e} Norifumi Shimoeda,^f Giichiro Ohno,^g Kentaro Watanabe,^d Yasunari Miyazaki,^c Yoshikazu Ishii,^a Kazuhiro Tateda^a

^aDepartment of Microbiology and Infectious Diseases, Toho University School of Medicine, Tokyo, Japan

^bBioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Hokkaido, Japan

^cDepartment of Respiratory Medicine, Tokyo Medical and Dental University (TMDU), Tokyo, Japan

^dNational Institute of Polar Research, Research Organization of Information and Systems, Tokyo, Japan

^eDepartment of Polar Science, The Graduate University for Advanced Studies, SOKENDAI, Tokyo, Japan

^fTochigi Medical Center, Tochinoki Hospital, Tochigi, Japan

^gTokatsu Hospital, Chiba, Japan

Sho Shimada and Ryosuke Nakai contributed equally to this work. Author order was determined by the contribution to the experiments.

ABSTRACT Culture-independent analysis shows that *Legionella* spp. inhabit a wide range of low-temperature environments, but to date, no psychrotolerant or psychrophilic strains have been reported. Here, we characterized the first cultivated psychrotolerant representative, designated strain TUM19329^T, isolated from an Antarctic lake using a polyphasic approach and comparative genomic analysis. A genome-wide phylogenetic tree indicated that this strain was phylogenetically separate at the species level. Strain TUM19329^T shared common physiological traits (e.g., Gram-negative, limited growth on buffered charcoal-yeast extract α -ketoglutarate [BCYE α] agar with L-cysteine requirements) with its relatives, but it also showed psychrotolerant growth properties (e.g., growth at 4°C to 25°C). Moreover, this strain altered its own cellular fatty acid composition to accumulate unsaturated fatty acid at a lower temperature, which may help maintain the cell membrane fluidity. Through comparative genomic analysis, we found that this strain possessed massive mobile genetic elements compared with other species, amounting to up to 17% of the total genes. The majority of the elements were the result of the spread of only a few insertion sequences (ISs), which were spread throughout the genome by a “copy-and-paste” mechanism. Furthermore, we found metabolic genes, such as fatty acid synthesis-related genes, acquired by horizontal gene transfer (HGT). The expansion of ISs and HGT events may play a major role in shaping the phenotype and physiology of this strain. On the basis of the features presented here, we propose a new species—*Legionella antarctica* sp. nov.—represented by strain TUM19329^T (= GTC 22699^T = NCTC 14581^T).

IMPORTANCE This study characterized a unique cultivated representative of the genus *Legionella* isolated from an Antarctic lake. This psychrotolerant strain had some common properties of known *Legionella* species but also displayed other characteristics, such as plasticity in fatty acid composition and an enrichment of mobile genes in the genome. These remarkable properties, as well as other factors, may contribute to cold hardiness, and this first cultivated cold-tolerant strain of the genus *Legionella* may serve as a model bacterium for further studies. It is worth noting that environmentally derived 16S rRNA gene phylotypes closely related to the strain characterized here have been detected from diverse environments outside Antarctica, suggesting a wide distribution of psychrotolerant *Legionella* bacteria. Our culture- and genome-based findings may accelerate the ongoing studies of the behavior and pathogenicity of

Citation Shimada S, Nakai R, Aoki K, Kudoh S, Imura S, Shimoeda N, Ohno G, Watanabe K, Miyazaki Y, Ishii Y, Tateda K. 2021.

Characterization of the first cultured psychrotolerant representative of *Legionella* from Antarctica reveals its unique genome structure. *Microbiol Spectr* 9:e00424-21. <https://doi.org/10.1128/Spectrum.00424-21>.

Editor Adriana Lopes dos Santos, Nanyang Technological University

Copyright © 2021 Shimada et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yoshikazu Ishii, yishii@med.toho-u.ac.jp.

Received 28 May 2021

Accepted 14 September 2021

Published 20 October 2021

Legionella spp., which have been monitored for many years in the context of public health.

KEYWORDS *Legionella*, psychrotolerant, insertion sequence, horizontal gene transfer, comparative genomic analysis, cellular fatty acids, mobile genetic elements

Legionella spp. are aerobic Gram-negative intracellular bacteria that mainly inhabit aquatic environments. Their natural hosts are protozoa such as free-living amoebae, while the accidental introduction of *Legionella* into humans via aerosols can cause pneumonia or fever-like symptoms known as Pontiac fever. More than 60 species of *Legionella* have been identified to date (1), around half of which are pathogenic to humans. Outbreaks of disease in human-made environments have been reported worldwide (2), and *Legionella* spp. are environmental bacteria that requires public health attention.

The *Legionella* species described to date are mesophilic, with an optimal growth temperature ranging from 25°C to 45°C (3, 4), none of which are psychrotolerant or psychrophilic. However, recent culture-independent analysis has elucidated the unexpected diversity of *Legionella* in low-temperature-treated drinking water, as well as polar lakes (5, 6). Our study also identified multiple uncharacterized *Legionella* lineages, sharing low 16S rRNA gene sequence similarities with known species in Antarctic terrestrial environments (7). From a public health perspective, the cultivation of uncultured *Legionella* species inhabiting cold environments is a crucial step in assessing the cold-tolerant mechanism(s) and potential pathogenicity of this organism.

In this context, we succeeded in isolating and cultivating the first psychrotolerant strain, designated TUM19329^T, from Antarctic freshwater lake sediment (7). Here, our study physiologically and genomically characterized this newly discovered strain and compared its characteristics to other *Legionella* species. We consequently propose the underlying mechanism responsible for cold hardiness and describe a novel species—*Legionella antarctica*—for which this strain is representative.

RESULTS AND DISCUSSION

Phylogenetic position of strain TUM19329^T and related strains. The three full-length 16S rRNA gene sequences (1,544 bp) of strain TUM19329^T, which were retrieved from the genome (8), showed 97.4% to 97.5% sequence identities with the closest strain, *Legionella fallonii* LLAP-10^T (DDBJ/ENA/GenBank accession number [LN614827](#); 9), in a BLASTN search against the NCBI nucleotide collection (nonredundant/nucleotide [nr/nt]) database as of July 2021. According to the identity threshold (<98.65%) for differentiating two species (10), this suggests that strain TUM19329^T is a candidate for a novel species. Note that the genome-wide average nucleotide identity (ANI) between strains TUM19329^T and LLAP-10^T was also relatively low at 77.5% (detailed comparative genomic comparisons are discussed later). Further phylogenetic analysis based on the genome sequences of strain TUM19329^T and other *Legionella* species also revealed that TUM19329^T was relatively close to *L. fallonii* on the phylogenetic tree (Fig. 1). The phylogenetic position in this tree is similar to the trend shown by the 16S rRNA gene sequence-based tree, which places strain TUM19329^T relatively close to *L. fallonii* (see Fig. S1 in the supplemental material), although this tree has potentially low resolution because of the short sequence information compared with the genome-scale tree (Fig. 1).

Strain TUM19329^T was distantly related to all known *Legionella* species, while a BLASTN database search against uncultured/environmental sequences showed that the sequences of this strain were almost identical (99.9% to 100%) to those recovered from a benthic moss colony known as “moss pillars” of another freshwater lake in Antarctica (GenBank accession number [AB630760](#); 11). Moreover, on the basis of the IMNGS platform (12), which is the database search against metagenome-derived 16S rRNA gene amplicon data sets, we further found that the TUM19329^T 16S rRNA gene sequence matched 841 data sets with a sequence similarity threshold of 99%; the data sets contained 259 soil, 100 aquatic, 86 freshwater, 76 rhizosphere, and 76 *Arabidopsis*

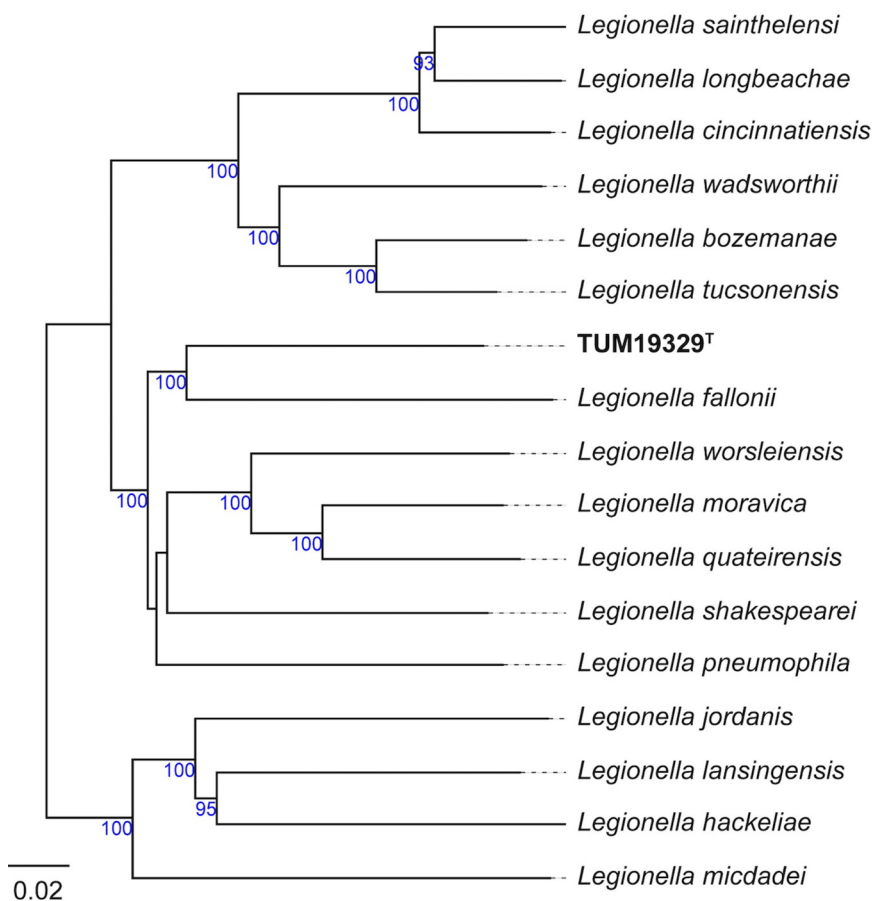


FIG 1 Tree inferred with FastME v2.1.6.1 from whole-proteome-based Genome BLAST Distance Phylogeny (GBDP) distances. The branch lengths are scaled via the GBDP distance formula d_g . Branch values are GBDP pseudobootstrap support values of >60% from 100 replications, with an average branch support of 92.1%. The trees were rooted at the midpoint. The scale bar indicates the branch length value. The strain names of each species are listed in Table 2.

thaliana and other plant metagenomes. This data set includes sequence data derived from samples such as soil crust in the Svalbard (Arctic Norway) (accession number [DRR148112](#)), high-latitude Qinghai Lake sediments ([SRR1303671](#)), and lake water of Little Long Lake ([SRR2962948](#); note that the temperature at the time of sample collection was around 20°C; 13). The habitats of this strain inferred using the ProkAtlas tool (containing multiple 16S rRNA gene sequences labeled by one environmental category [14] with a threshold of 99%) were permafrost (habitat preference score, 25.0%), rice paddy (21.5%), freshwater (19.2%), soil (18.8%), and wetland (15.4%). The results suggested that the relatives of this strain may be distributed not only in Antarctic aquatic habitats but also in a variety of other environments.

Morphological, physiological, infectious, and biochemical characteristics. TUM19329^T is a Gram-stain-negative, non-spore-forming bacterium with a size of 2.2 by 0.3 μm . Growth was observed at 4°C to 25°C but not above 30°C on buffered charcoal-yeast extract α -ketoglutarate (BCYE α) agar. Colonies appeared after 10 days at 25°C. Given such cold-tolerant growth characteristics, initial enrichment cultivation at 4°C rather than 33°C, which is the general growth condition set for previously described *Legionella* spp., may have led to successful isolation of the first psychrotolerant strain. Note that the temperature of the freshwater lake, known as Lake Naga-ike, which strain TUM19329^T inhabits, ranges from 0°C to 10°C (15). The strain only grew at pH 6.5 to 7.0, with a NaCl concentration lower than 1%, in buffered yeast extract (BYE) broth (Fig. S2). The strain required L-cysteine for growth. Both cefinase and oxidase tests were negative, as were gelatinase, hippurate reactions, and all the sugar tests.

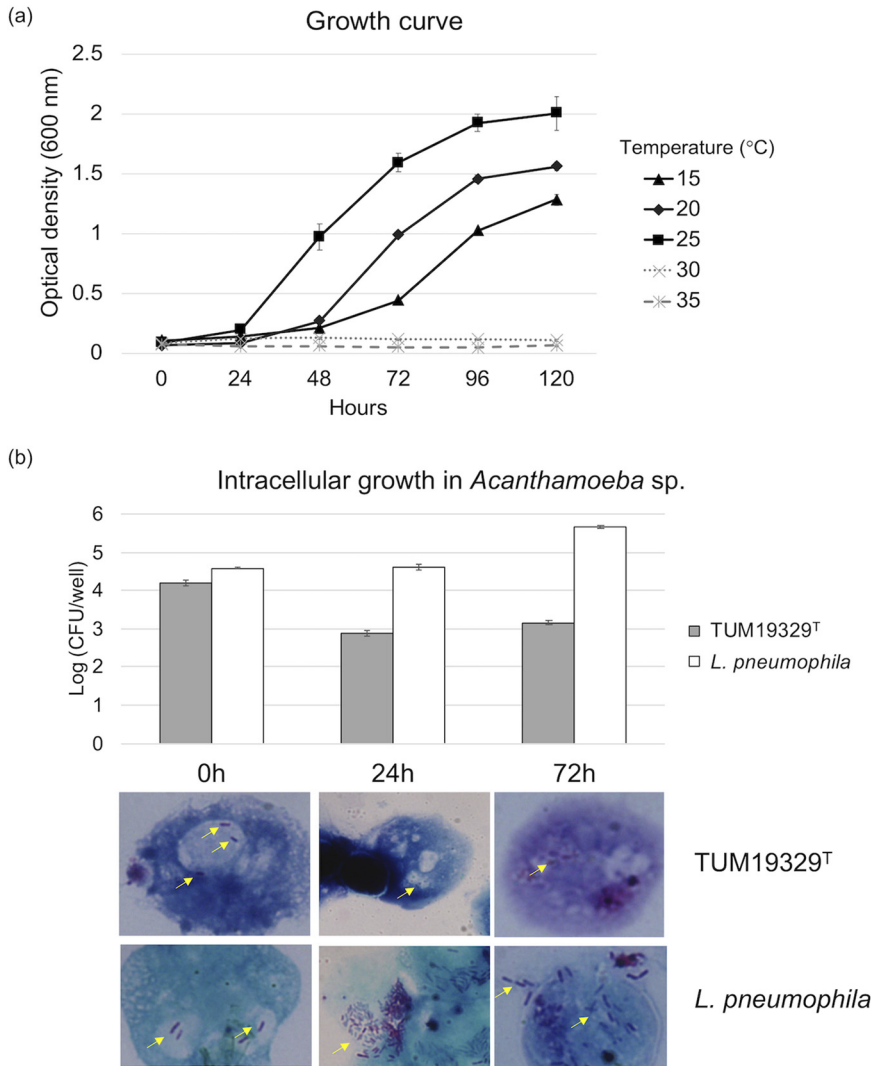


FIG 2 Proliferative potential of strain TUM19329^T in culture medium and eukaryotic cells. (a) Growth curves of strain TUM19329^T in BYE broth at different temperatures. (b) Intracellular growth of strain TUM19329^T and *L. pneumophila* Philadelphia-1 in an *Acanthamoeba* sp. The light micrograph images were obtained by Gimenez staining at each time point. All images were taken at $\times 1,000$ magnification. The yellow arrows indicate bacterial cells. The experiment was conducted twice in triplicate.

To assess the replication capacity of TUM19329^T in host eukaryotes, we compared its intracellular growth with that of *Legionella pneumophila* strain Philadelphia-1 in *Acanthamoeba* spp., a known natural host of *Legionella* spp. Whereas *L. pneumophila* tended to proliferate after intracellular uptake, the bacterial number of strain TUM19329^T decreased after 24 h of uptake, and thereafter, no change was observed until 72 h (Fig. 2b). Although our strain did not show such clear proliferation, there may be a different and more suitable host for this strain in cold environments. We propose that our psychrotolerant strain may be used as bait to screen for and identify its potential host(s). Pursuing its host will accelerate our understanding of the ecology of *Legionella* spp. under low-temperature conditions.

Chemotaxonomic characteristics. The G+C content of the genome of strain TUM19329^T was 39.1%, and the major respiratory quinones were Q-12 (69.1%) and Q-13 (30.9%); these features were shared with the closest strain, *L. fallonii* LLAP-10^T. The dominant cellular fatty acids (CFAs; >10% of the total fatty acids) of the strain grown at 25°C were anteiso-C_{15:0} (36.2%), C_{16:1}ω7c/C_{16:1}ω6c (27.2%), and iso-C_{16:0} (15.2%) (Table 1). The pattern of large amounts of branched CFAs was also similar to that

TABLE 1 CFA compositions of strain TUM19329^T and the most closely related strain, LLAP-10^{Ta}

Fatty acid or characteristic	Data for strain:		
	TUM19329 ^T		LLAP-10 ^T
Cultured temp (°C)	10	25	30
iso-C _{13:0}	–	T	–
anteiso-C _{13:0}	–	T	–
iso-C _{14:0}	1.3	3.6	4-5
C _{14:1} ω5c	1.1	1.2	–
C _{14:0}	1.1	T	2-3
C _{13:0} 3-OH / C _{15:1} iso H (summed feature 1)	–	T	ND
iso-C _{15:0}	1.5	3.9	T
anteiso-C _{15:0}	19.9	36.2	12-15
C _{15:1} ω6c	1.6	1.6	3-4
C _{15:0}	–	–	2-3
C _{14:0} 3-OH	T	T	–
iso-C _{16:0}	9.0	15.2	11-13
C _{16:1} ω7c/C _{16:1} ω6c (summed feature 3)	53.7	27.2	21-27 ^b
C _{16:0}	4.2	1.7	15-19
C _{15:0} 2-OH	–	T	–
iso-C _{17:1} ω9c/C _{16:0} 10-methyl (summed feature 9)	T	T	ND
iso-C _{17:0}	T	T	–
anteiso-C _{17:0}	3.0	4.1	2-3
C _{17:0} cyc	–	–	2-6
C _{17:0}	–	T	2
iso-C _{18:0}	–	T	T
C _{18:0}	1.8	1.0	6-7
C _{18:0} 3-OH	–	T	ND
C _{19:0} cyc ω8c	–	T	ND
C _{19:0}	–	–	1-2
anteiso-C _{19:0}	–	–	T
C _{20:0}	T	T	1-2

^aValues are the percentage of total fatty acids. T, trace (~0.9%); –, not detected; ND, no data. Data of strain LLAP-10^T from Adeleke et al. (9). Note that the psychrotolerant Antarctic strain TUM19329^T from this study does not grow above 30°C.

^bThis is given as the value of C_{16:1} ω7c in the previous study.

observed in *L. fallonii* LLAP-10^T and its close relatives (9); however, the CFA profile drastically changed when cultured at lower temperature. After cultivation at 10°C, the proportion of monounsaturated fatty acids (C_{16:1} ω7c/C_{16:1} ω6c) among the total fatty acids changed from 27.2% to 53.7% (Table 1).

One of the prominent biological strategies to survive at low temperature is believed to be the maintenance of cell membrane fluidity (16). Generally, a decrease in temperature causes membrane viscosity or phase transition in biological membranes due to the tight packing of the fatty acyl chains (17). To overcome this, cold-adapted organisms alter fatty acid production and contain unsaturated fatty acyl chains containing one or more double bonds, which adopt a more expanded conformation and possess a lower melting temperature than their corresponding saturated chains (17); for example, a deep-sea psychrotolerant bacterium *Shewanella electrodiphila* MAR441^T accumulates monounsaturated fatty acids (e.g., C_{16:1} ω7c and C_{18:1} ω7c) to more than half (56%) of the total fatty acids at low temperature (18). Taken together, we believe that the ability of strain TUM19329^T to alter its own fatty acid content is one of the crucial mechanisms of cold tolerance.

Comparative genome analysis. General overview. We further performed a comparative genomic analysis with 16 known mesophilic *Legionella* species to characterize the genomic properties of the cold-tolerant strain TUM19329^T. The general features of strain TUM19329^T compared to other species are summarized in Table 2. Previous comparative genomic studies have shown that some psychrophilic and psychrotolerant bacteria tend to have distinctive genomic features (e.g., altered G+C contents) and amino acid usage frequencies (e.g., reduced proline contents) (19–22). However, in our comparative genomic analysis, although the *Legionella* species compared showed

TABLE 2 Summary of the basic genomic data of the 17 *Legionella* spp. compared in this study

Taxon name	Strain name	GenBank accession ID	No. of contigs	Size (Mbp)	GC (%)	No. of CDS	No. of rRNA	No. of tRNA
<i>Legionella antarctica</i>	TUM19329 ^T	GCA_011764505.1	1	3.75	39.15	3,608	9	42
<i>Legionella bozemanae</i>	WIGAT	GCA_900450575.1	7	4.23	37.97	3,761	9	46
<i>Legionella cincinnatiensis</i>	72-OH-H ^T	GCA_900452415.1	2	4.1	36.86	3,498	9	46
<i>Legionella fallonii</i>	LLAP-10 ^T	GCA_000953135.1	3	4.43	38.32	3,748	12	46
<i>Legionella hackeliae</i>	Lansing 2 ^T	GCA_000953655.1	2	3.57	38.96	3,183	12	43
<i>Legionella jordanis</i>	BL 540 ^T	GCA_900637635.1	1	3.13	41.74	2,865	12	44
<i>Legionella lansingensis</i>	1677-MI-H ^T	GCA_900187355.1	1	2.99	40.7	2,703	12	43
<i>Legionella longbeachae</i>	Longbeach 4 ^T	GCA_004283175.1	38	4.13	37	3,554	5	42
<i>Legionella micdadei</i>	TATLOCK ^T	GCA_000953635.1	1	3.31	40.46	2,953	9	43
<i>Legionella moravica</i>	316-36 ^T	GCA_900452715.1	2	3.83	40.12	3,212	9	41
<i>Legionella pneumophila</i> subsp. <i>pneumophila</i>	Philadelphia-1	GCA_001752765.1	1	3.41	38.33	3,021	9	43
<i>Legionella quateirensis</i>	200/83-1335 ^T	GCA_900452695.1	5	4.28	39.06	3,611	9	43
<i>Legionella sainthelensi</i>	MSH-4 ^T	GCA_900637685.1	1	4.16	37.06	3,746	12	49
<i>Legionella shakespearei</i>	214 ^T	GCA_001468025.1	76	3.51	41.56	2,933	4	40
<i>Legionella tucsonensis</i>	1087-AZ-H ^T	GCA_001468035.1	12	3.36	37.41	2,967	3	41
<i>Legionella wadsworthii</i>	Wadsworth 81-716A ^T	GCA_900452925.1	2	3.6	38.08	3,189	11	43
<i>Legionella worsleiensis</i>	95/83-1347 ^T	GCA_900453045.1	2	3.16	40.4	2,705	9	41

some differences in genome size (ranging from 2.99 to 4.43 Mb), G+C content (36.9 to 41.7%), and protein-coding sequence (CDS) number (2,703 to 3,761), strain TUM19329^T did not have the distinct characteristics reported in previous studies (see details below).

Specifically, the codon usage and amino acid composition of the 17 species were relatively similar (Fig. 3a), and the strains that were related to each other tended to form close clusters on the hierarchically clustered heatmap based on the percentage of codons and amino acids of the entire genome (Fig. 3b). TUM19329^T was placed close to its phylogenetic relative, *L. fallonii*, in the heatmap based on amino acid usage (Fig. 3b, left panel). Meanwhile, in terms of codon usage, TUM19329^T was close to *L. pneumophila* (Fig. 3b, right panel), thus differing from the trend observed in the comparison of amino acid composition. This suggested that there is a difference in the frequency of codon usage between TUM19329^T and *L. fallonii*. Comparing the two species for relative synonymous codon usage, TUM19329^T tended to use CUG more often, whereas *L. fallonii* tended to use UUA to encode leucine (Fig. 3c). To understand whether this genus changes its codon/amino acid usage as part of a cold adaptation strategy, a further collection of cold-tolerant *Legionella* strains is needed.

As one of the other possible factors for cold hardiness, cold shock proteins are generally essential single-stranded nucleic acid binding proteins that respond to cold shock and regulate a variety of cellular processes (23). The TUM19329^T genome possessed four cold shock proteins encoded by the genes TUM19329_00491, TUM19329_01544, TUM19329_02137, and TUM19329_03456. However, the other 16 mesophilic species also possessed four to six cold shock protein-coding genes. Furthermore, three out of the four cold shock proteins of TUM19329^T formed part of the core genome of the 17 species, and the remaining gene was also shared between several species, suggesting that these proteins are well preserved among *Legionella* spp. In conclusion, TUM19329^T showed no specificity in the number or sequence of cold shock proteins compared with the mesophilic species. These results suggest that factors different from the well-known genomic characteristics and cold shock proteins of psychrophilic/psychrotolerant microorganisms contribute to its cold-tolerant nature. We further compared the individual genes among species to search for features that may explain the psychrotolerant characteristics of this strain.

Core genome. The pan-genome of 17 *Legionella* species consisted of 11,135 pan-genome orthologous groups and 1,340 core orthologous groups. The core orthologous

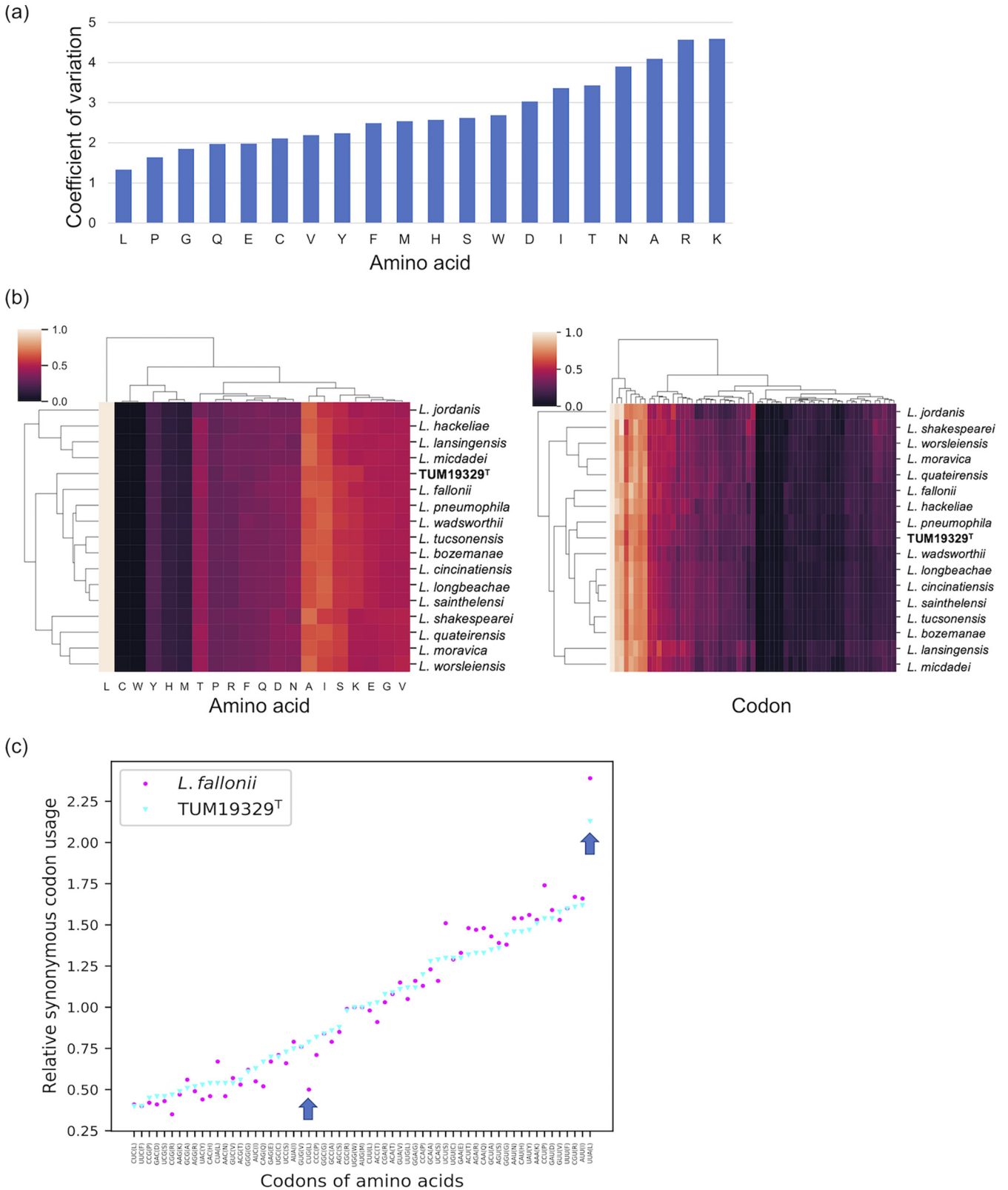


FIG 3 Comparison of the amino acid and codon composition of 17 *Legionella* species. (a) Coefficient of variation of the amino acid compositions between the compared *Legionella* species. (b) An amino acid and codon usage heatmap based on the percentages in each of the genomes of the compared *Legionella* species. A heatmap describing the details of each codon on the horizontal axis is available in Figshare (see “Data Availability”). The strain names of each species are listed in Table 2. (c) Relative synonymous codon usage (RSCU) in TUM19329^T and *L. fallonii*. Stop codons were excluded from the analysis. Codons of amino acids in the x axis were sorted based on the RSCU values of TUM19329^T. The amino acids translated by codon are shown in parentheses following the codon. The arrows point to the synonymous codon for leucine, which showed the most usage bias between the two species.

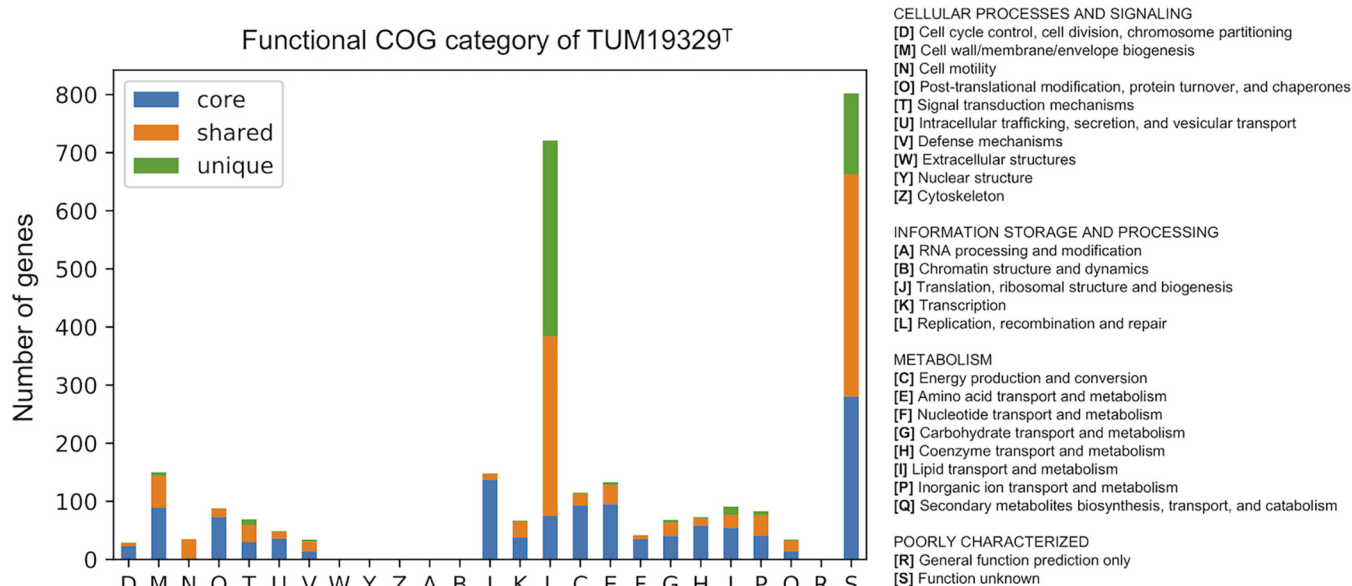


FIG 4 Number of protein coding genes classified into each functional COG category. Genes belonging to each category are colored as core, shared (genes shared between at least one species), or unique. Protein coding genes not annotated with any functional category (729 CDS, 20%) or annotated with multiple COG functional categories (46 CDS, 1.2%) were excluded. COG, clusters of orthologous groups.

groups consisted of the genes of fundamental metabolic pathways, as well as the Dot/Icm-type IVB secretion system and the type II secretion system, which are essential for intracellular replication. Like other *Legionella* species, these genes were conserved in TUM19329^T, although duplicate genes were observed for *icmD*, similar to its most closely related species, *L. fallonii* (24). Interestingly, the region in which the *icmR* gene is located in *L. pneumophila* was replaced with genes that were nonhomologous to the *icmR* gene (TUM19329_00496) in TUM19329^T. IcmR, the gene for which is located between *icmS* and *icmQ*, interacts with IcmQ as a chaperone to block the dimeric form of IcmQ (25). IcmQ is well conserved among *Legionella*, whereas IcmR is hypervariable depending on the species and is thought to be encoded by a fast-evolving gene (26). However, these genes are referred to as functional homologs of IcmRs (FIRs) because they perform the same functions as IcmR. A previous study showed that phylogenetically closely related species mostly share homologous FIRs (27). Indeed, phylogenetic analysis of the FIR gene sequences showed that several phylogenetically related strains (see Fig. 1) formed groups supported with high bootstrap values (~70%; e.g., the group of *L. tucsonensis*, *L. wadsworthii*, and *L. bozemanii*; Fig. S3). In contrast, the putative FIR gene of strain TUM19329^T had the closest phylogenetic relationship with that of *L. fallonii* among the known species, but its bootstrap value was low. Furthermore, direct comparison of the sequences of strain TUM19329^T and *L. fallonii* by BLASTP showed low similarity and coverage of 37% and 41%, respectively. Consistent with the findings observed for other *Legionella* spp., the FIR of TUM19329^T may also have changed in the process of species diversification.

Massive insertion sequences (ISs) and putative horizontal gene transfer (HGT).

Regarding the number of assigned functional genes, 2,879 CDSs of TUM19329^T (79.7%) were annotated to the cluster of the orthologous group (COG) category. While most of the genes in each functional category were core or shared genes among the 17 species, TUM19329^T had a significantly higher proportion of genes classified to the replication, recombination, and repair [L] category, which accounted for 20% (721 genes) of the entire genome (Fig. 4, Table S1) compared with other *Legionella* species ($n = 120$ to 220, 3.8% to 5.9%). The composition of the COGs classified as [L] was mostly mobile genetic elements ($n = 614$) such as transposases, integrases, and plasmids. Among these, the most commonly observed COG IDs were COG2801 ($n = 163$), COG2963 ($n = 156$), COG4584 ($n = 45$), and COG1484 ($n = 44$) (Table S2). To clarify the sequence relationships of the genes classified into these four COG IDs, they were further

analyzed using the phylogenetic tree method. As a result, the gene sequences of each COG ID were clustered into a small number of groups (Fig. 5a). Looking carefully at the position of these sequences in the genome, we found that the genes in certain clusters always appeared in pairs (e.g., cluster 1 and cluster 4 in Fig. 5). As a result of the alignment of these genes, including the up- and downstream sequences, we found six copies (*ISLa1* to *ISLa6*) of the IS structure that encoded several genes and incomplete terminal inverted repeat structures at the end (Fig. 5b, Table 3). These ISs appeared repeatedly, and a total of 169 ISs containing a full-length copy were identified, distributed in various positions throughout the chromosome (Fig. 5c).

ISs are the major driver of genome evolution by interrupting gene functions and reshaping the genome structure by recombination (28). We found multiple IS structures that were fragmented by other ISs (Table S3, Fig. S4). We also identified some of the functional genes other than mobile genes flanking the ISs that were annotated with the same COG ID. These genes are also likely to be remnants of genes fragmented by ISs (Table 4). Considering that the ISs may also regulate the expression of other genes by modifying their promoter regions, the massive ISs found in TUM19329^T may play a major role in shaping the phenotype and physiology of this strain by interference with many of its genes.

Expansion of ISs, as seen in this strain, has been reported in other bacteria and is recognized to play an important role in niche-restricted prokaryotes, such as extremophiles and symbionts, aiding their adaptation to the environment (28). For extremophiles, the enrichment of transposase-encoding genes has been reported in species residing in cold environments, such as *Methanococcoides burtonii* isolated in Antarctica (29). Furthermore, transposases were one of the overrepresented genes detected in metagenomic analysis of cold, 4,000-m-deep ocean samples (30). For symbionts, the expansion of ISs has been observed, especially in species that recently adopted a host-restricted lifestyle compared with long-standing symbionts (31, 32). From an evolutionary viewpoint, this phenomenon has been proposed as a hypothesis that the expansion of ISs reorganizes the genome in the early stages of the evolutionary adaptation process and that unnecessary genes are gradually eliminated, thereby streamlining the genome. In accordance with this hypothesis, strain TUM19329^T might be undergoing genome reduction and streamlining while adapting to its host in a cold environment.

For the detected genes in ISs, a similarity search using BLASTP with default settings (accessed July 2021) confirmed that 9 of the 13 genes were not carried by other *Legionella* spp. (Table 3). Furthermore, the other four genes that had best hits with *Legionella* spp. were found to be acquired by HGT based on a search using HGTector 2 (33), an automated pipeline determining putative horizontally transferred genes. HGTector 2 detects HGT by analyzing BLAST hit distribution patterns based on defined hierarchical evolutionary categories (see Materials and Methods). In accordance with the BLASTP similarity and the results of HGTector 2, 12 out of the 13 genes were likely to have been acquired by HGT relatively recently, rather than being derived from a common ancestor of the genus *Legionella*.

We further searched for genes that this strain may have acquired by HGT events across the whole genome. The number of detected putative genes acquired by HGT was 605 in total, which was higher than for other *Legionella* species (153 to 394 genes), but many of these were mobile genes (373 genes), such as transposases that form massive ISs. This further confirms that many of the mobile genes in this strain have been horizontally acquired and spread throughout the genome by the “copy-and-paste” mechanism.

Unique fatty acid-related genes in a psychrotolerant Antarctic strain. The majority of the high number of genes potentially acquired by HGT were mobile genes, but several putative HGT-derived genes related to metabolic pathways and functions were also identified. These genes were mainly classified into the following COG categories: cell wall/membrane/envelope biogenesis [M] ($n = 25$), amino acid transport and metabolism [E] ($n = 16$), energy production and conversion [C] ($n = 14$), inorganic ion

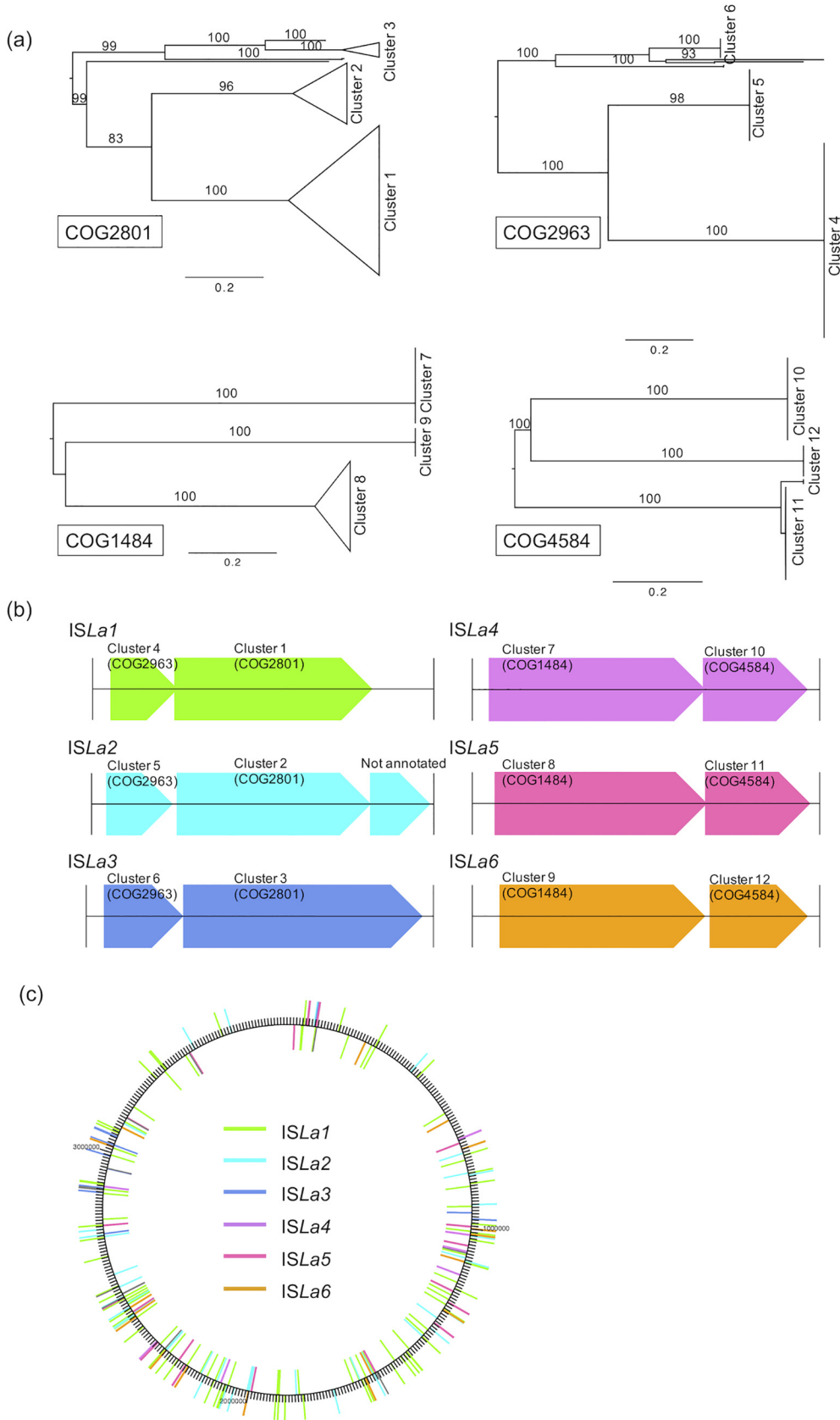


FIG 5 Phylogeny, structure, and distribution of the insertion sequences in the genome of TUM19329^T. (a) Maximum-likelihood tree showing the relationship between the genes classified into the four COG IDs with the highest number (Continued on next page)

transport and metabolism [P] ($n = 12$), lipid transport and metabolism [I] ($n = 11$), and transcription [K] ($n = 10$). It is possible that at least some of these putative exogenous genes were necessary for this strain to adapt to the low temperature conditions of the Antarctic lake.

In particular, several fatty acid synthesis-related genes in COG category [I] were overrepresented in this strain. In addition to the genes shared with other species, TUM19329^T possessed extra genes acquired by HGT annotated as *fabH* (TUM19329_01025), *fabG* (TUM19329_01799 and TUM19329_02687), *fabF* (TUM19329_03119), and *fabZ* (TUM19329_03121). These genes were presumed to be derived from members of the *Actinobacteria*, *Proteobacteria*, and other phyla. Moreover, an extensive BLASTP search against the NCBI nr database (accessed July 2021; option exclude “uncultured/environmental sample sequences”) revealed that, at least for *fabF* and *fabZ*, no *Legionella*-derived sequences were found in the top 100 sequences with high similarities. For example, the top-hit alignments of *fabF* and *fabZ* sequences were with those derived from “*Candidatus Rubidus massiliensis*” of the phylum *Chlamydiae* (54.0% sequence similarity) and *Kibdelosporangium banguiense* of the phylum *Actinobacteria* (41.1%), respectively. Judging from the results based on the HGTector and BLASTP searches, it is more likely that these two genes were exogenously acquired.

Interestingly, it has been proposed that the fatty acid metabolism of *Coxiella burnetii*, which is the etiologic agent of Q fever and belongs to the family *Coxiellaceae* of the order *Legionellales*, is enhanced by HGT-derived genes (34). Strain TUM19329^T also retains Fab-encoding gene sequences common to other *Legionella* species, and the retention of both conventional and potentially exogenous genes might be related to the enhancement of cold tolerance through fatty acid synthesis and alterations, although further verification is needed. Although the mechanisms remain to be resolved, this psychrotolerant strain is capable of the flexibility afforded by altering its CFA composition (Table 1), and future studies should examine the relationship between the expression of lipid synthesis-associated genes and the physiology of this strain under different culture conditions.

Conclusion. Unexplored *Legionella* members have been reported to exhibit high diversity in low-temperature habitats, but it remained unclear whether such strains possess true cold tolerance or whether they reside within a milder environment within their host protist(s). We characterized the first psychrotolerant strain, TUM19329^T, from the sediment of an Antarctic lake. Our polyphasic approach shed light on the mechanism and potential factors responsible for the cold tolerance of this strain. On the basis of its physiological, phylogenetic, and chemotaxonomic properties, we propose naming this strain *Legionella antarctica* sp. nov.

Description of *Legionella antarctica* sp. nov. *Legionella antarctica* (ant.arc'ti.ca. L. fem. adj. *antarctica* southern, pertaining to the Antarctic, where the type strain was isolated).

Cells are 2.2 by 0.3 μm in size, Gram-negative, catalase-negative, oxidase-negative rods that require L-cysteine for growth on BCYE α . Smooth, gray colonies form after 7 to 10 days on BCYE α agar at 25°C, which show slight yellow fluorescence under UV light. Growth is observed at 4°C to 25°C (optimum, 25°C), but not at 30°C, and pH 6.5 to 7.0. No growth occurs in the presence of >1% NaCl. After 14 days of culture, cells test positive for acetoin production with the API 20E test and test positive for esterase, L-arginine arylamidase, and L-aspartic acid arylamidase with the API Campy test. The major fatty acids are anteiso-C_{15:0}, C_{16:1} ω 7C/C_{16:1} ω 6c, and iso-C_{16:0}. However, after cultivation at 10°C, the proportion of monounsaturated C_{16:1} ω 7C/C_{16:1} ω 6c fatty acid among the total fatty acids rose to more than 50%. The major respiratory quinones are Q-12 and Q-13.

FIG 5 Legend (Continued)

of genes, COG2801, COG2963, COG1484, and COG4584. The tree was rooted using midpoint rooting. The scale bar represents the branch length value. Numbers along branches are bootstrap values based on 100 replicates. The genes in each COG ID were grouped in three major clusters. Detailed versions of each tree are available in Figshare (see “Data Availability”). (b) Each of the 12 clusters appeared adjacent to each other in a specific pair in the genome, and analysis, which included the up- and downstream sequences, revealed six IS structures (*ISLa1* to *ISLa6*) that appeared repeatedly. (c) The distribution of *ISLa1* to *ISLa6* in the genome of TUM19329^T.

TABLE 3 Descriptions of the six IS structures with particularly high copy numbers

BLAST similarity search ^b							
IS (length [bp])	No. of full-length copies	Gene ^a (length [bp])	Description of the gene (COG ID)	Host of the gene	Query coverage (%)	Identity (%)	Putative HGT ^c
ISLa1 (1,463)	93	Cluster 4 (285)	Hypothetical protein (COG2963)	Legionella sp.	96	78.02	No
				Legionella geestiana	96	72.53	
				Legionellaceae sp.	97	70.65	
ISLa2 (1,480)	33	Cluster 5 (285)	Hypothetical protein (COG2963)	Legionella sp.	100	98.58	Yes
				Legionellaceae sp.	98	82.31	
				Legionella pneumophila	98	82.31	
ISLa3 (1,260)	6	Cluster 6 (291)	Hypothetical protein (not annotated)	Gammaproteobacteria sp.	100	77.66	Yes
				Desulfobacteraceae sp. strain 4484_190.2	98	59.14	
				Desulfacinum hydrothermale	94	55.06	
				Desulfobacula sp. strain RIFOXYA12_FULL_46_16	99	65.34	Yes
				Desulfacinum infernum DSM 9756	97	66.05	
				Nitrospirae sp.	97	65.56	
				"Candidatus Kentron" sp. TC	58	49.06	No
				"Candidatus Kentron" sp. TC	31	70.37	
				"Candidatus Kentron" sp. LFY	28	70.83	
				Legionella sp.	100	98.58	Yes
				Legionellaceae sp.	98	82.31	
				Legionella pneumophila	98	82.31	
ISLa4 (2,431)	17	Cluster 7 (1,512)	Transposase (COG1484)	Legionella sp.	97	95.09	Yes
				Legionella sp.	100	73.29	
				Legionella pneumophila strain Leg01/53	100	73.04	
				Legionella pneumophila	100	73.04	
				Solemya velum gill symbiont	97	61.05	Yes
				Legionella taurinensis	100	56.26	
				Legionella taurinensis	100	56.26	
				Solemya velum gill symbiont	100	65.29	Yes
				Legionella taurinensis	100	61.57	
				Alphaproteobacteria sp.	99	58.51	
				Methylothera sp.	100	98.24	No
				Legionella sp.	99	73.87	
ISLa5 (2,516)	13	Cluster 8 (1,533)	Putative transposase for insertion sequence element IS5376 (COG1484)	Sphingosinella sp.	99	73.67	
				Methylothera sp.	100	98.8	Yes
				Legionella sp.	100	82.47	
				"Candidatus Saccharibacteria" sp.	97	77.96	
				Gammaproteobacteria sp.	100	74.08	Yes
				Coxiellaceae sp.	98	53.24	
				Deltaproteobacteria sp.	98	51.95	
				Gammaproteobacteria sp.	100	82.86	Yes
				Aquisalmonas sp. strain 2447	97	56.25	
				"Candidatus Kentron" sp. DK	94	57.94	

^aCluster number defined based on the results of phylogenetic analysis of the same COG ID (see Fig. 5a).

^bThe results of the top three hits of the BLASTP similarity search for each gene.

^cGenes presumed to have been acquired by HGT using HGTector.

TABLE 4 Genes interrupted by ISLa1 to ISLa6

Genes ^a	COG ID	Product	IS interrupting the gene
TUM19329_00019, TUM19329_00022	L:COG0507	Hypothetical protein	ISLa5
TUM19329_00269, TUM19329_00272	V:COG0286	Site-specific DNA-methyltransferase (adenine-specific)	ISLa4
TUM19329_00485, TUM19329_00488	G:ENOG410XNQK	Arabinose-proton symporter	ISLa1
TUM19329_00643, TUM19329_00646	S:COG0121	Putative glutamine amidotransferase	ISLa4
TUM19329_00913, TUM19329_00917	I:COG3243	Poly(3-hydroxyalkanoate) polymerase	ISLa2
TUM19329_00959, TUM19329_00962	I:COG1960	Acyl-CoA dehydrogenase family member 11	ISLa3
TUM19329_01188, TUM19329_01191	V:COG1680	Serine-type D-Ala-D-Ala carboxypeptidase	ISLa1
TUM19329_01427, TUM19329_01430	T:ENOG410XNMH	Histidine kinase	ISLa1
TUM19329_01576, TUM19329_01580, TUM19329_01584	T:ENOG410XNMH	Histidine kinase	ISLa2
TUM19329_01604, TUM19329_01607	S:ENOG4112B5D	Hypothetical protein	ISLa1
TUM19329_01615, TUM19329_01618	S:ENOG4112A90	Protein TraM	ISLa1
TUM19329_01623, TUM19329_01626	L:COG2189	Site-specific DNA-methyltransferase (adenine-specific)	ISLa1
TUM19329_01804, TUM19329_01807	I:COG1835	O-acetyltransferase OatA	ISLa1
TUM19329_02422, TUM19329_02428	M:COG2230	Cyclopropane-fatty-acyl-phospholipid synthase	ISLa2
TUM19329_02454, TUM19329_02458	S:COG1988	Hypothetical protein	ISLa1
TUM19329_02485, TUM19329_02491	M:COG0438	1,4- α -glucan branching enzyme	ISLa1
TUM19329_02509, TUM19329_02512	S:COG3173	Hypothetical protein	ISLa1
TUM19329_02618, TUM19329_02621	E:COG0028	Acetolactate synthase	ISLa1
TUM19329_03112, TUM19329_03115	S:ENOG410XUV4	Protein BcsG-like protein	ISLa1
TUM19329_03264, TUM19329_03267	I:COG1597	Undecaprenyl-diphosphate phosphatase	ISLa1

^aMobile genes interrupted by ISs are excluded from the list.

The type strain is TUM19329^T (= GTC 22699^T = NCTC 14581^T), isolated from the freshwater lake sediment of Lake Naga-ike, Skarvsnes, East Antarctica.

MATERIALS AND METHODS

Phylogenetic evolutionary analysis and habitat prediction. The full-length 16S rRNA gene sequences of strain TUM19329^T were retrieved from our previous study (8) and then used in a BLASTN search against the NCBI nucleotide collection (nr/nt) database as of July 2021. To further confirm the phylogenetic placement, we also constructed a maximum-likelihood (ML) phylogenetic tree based on multiple alignments of the 16S rRNA gene sequences obtained using MAFFT v7.450 with the auto option, which selected the L-INS-i algorithm (35) and estimation under the GTRGAMMA model using RAxML v8.2.12 software with 100 bootstrap replicates (36). For the genome-based phylogenetic analysis, the genome sequence data of strain TUM19329^T and its related species were analyzed using the Type (Strain) Genome Server (TYGS) (accessed July 2021) (37) with the following default settings: all pairwise comparisons among the set of genomes were conducted using the Genome BLAST Distance Phylogeny approach (GBDP), and accurate intergenomic distances were inferred under the algorithm “trimming” and distance formula d_s (38); 100 distance replicates were calculated for each genome; digital DNA-DNA hybridization (DDH) values and confidence intervals were calculated using the recommended settings of GGDC 2 (38); the resulting intergenomic distances were used to infer a balanced minimum-evolution tree with branch support via FASTME v2.1.6.1, including subtree pruning and regrafting (SPR) postprocessing (39); branch support was inferred from 100 pseudobootstrap replicates; the trees were rooted at the midpoint (40) and visualized with PhyD3 (41).

The habitability and distribution of strain TUM19329^T and its close relatives were investigated and predicted using the IMNGS platform (12), which conducts a database search against metagenome-derived 16S rRNA gene amplicon data sets, and the ProkAtlas tool, which contains multiple 16S rRNA gene sequences labeled by one environmental category (14). Both tools were performed with a sequence similarity threshold of 99%, using the sequence from strain TUM19329^T as the query.

Morphological, physiological, and biochemical characterization. For morphological and physiological characterization, TUM19329^T was grown on BCYE α agar for 10 days at 25°C. Cell morphology was examined by light microscopy after Gram staining using the Favor G kit (Nissui Pharmaceutical Co., Ltd., Tokyo, Japan). Growth at a range of temperatures (10°C, 15°C, 20°C, 25°C, and 30°C) was assessed in 4 ml of BYE broth supplemented with α -ketoglutarate, L-cysteine, and iron (III) nitrate under constant shaking by transferring a single colony from BCYE α . Growth at various pH values (6.0 to 7.5) and NaCl concentrations (0% to 1.5% [wt/vol]) was also determined with BYE broth. The pH was adjusted by adding either 6 M HCl or 5 M KOH prior to sterilization. The physiological and biochemical characteristics and enzyme activities were tested at 25°C using API 20E and API Campy (bioMérieux, Marcy l’Etoile, France) according to the manufacturer’s instructions, with the exception that the incubation period was extended up to 14 days. The presence of β -lactamase was assessed using Cefinase discs (Becton, Dickinson Microbiology Systems, Sparks, MD, USA).

Infectivity and intracellular growth assay. An *Acanthamoeba* sp. previously isolated from a cooling tower in Japan was cultured in 25-cm² tissue culture flasks with 10 ml of PYG broth [2% proteose

peptone no. 3, 0.1% yeast extract, 0.1 M glucose, 4 mM MgSO₄, 0.4 M CaCl₂, 0.1% sodium citrate, 0.05 M Fe(NH₄)₂·6H₂O, 2.5 mM NaH₂PO₄, and 2.5 mM K₂HPO₄; pH 6.5] at 25°C for 10 days. Cultures of the *Acanthamoeba* strain were transferred to a 15-ml polypropylene tube, centrifuged at 600 × *g* for 10 min, washed twice with fresh PYG broth, and then adjusted to a titer of 2 × 10⁵ cells ml⁻¹. Then, 500 μl of cell suspension was pipetted into each well of a 24-well tissue culture plate (Becton, Dickinson Labware, Franklin Lakes, NJ, USA). After 3 h of incubation at 25°C, the medium was removed and washed three times with *Acanthamoeba* buffer (AC buffer; PYG broth without proteose peptone, glucose, and yeast extract).

The bacterial suspension was prepared as follows: a single colony of strain TUM19329^T that had been incubated on BCYE α agar for 10 days at 25°C was transferred and incubated with continuous shaking in 4 ml of BYE broth supplemented with α -ketoglutarate, L-cysteine, and iron for 5 days until it reached the postexponential phase. *L. pneumophila* Philadelphia-1 was prepared using the same agar and broth as for TUM19329^T but was incubated on the agar for 3 days, followed by shaking in 4 ml of broth for 1 day at 35°C. The bacterial suspension was centrifuged at 3,000 rpm for 10 min and replaced with AC buffer.

Cultures of the *Acanthamoeba* sp. were infected by strain TUM19329^T or Philadelphia-1 at a multiplicity of infection (MOI) of 10 and incubated for 1 h after attachment by centrifugation at 500 × *g* for 5 min. At the end of the infection period, nonphagocytosed and nonadherent bacteria were removed by washing three times with fresh AC buffer. The infected *Acanthamoeba* cells were incubated at 25°C and harvested from the bottom of the wells after 0, 24, or 72 h. Then, 100 μl of each sample was fixed onto a microscope slide by cytospin centrifugation at 550 rpm for 1 min and methanol treatment. The slides were stained using Giménez stain (42) and observed by light microscopy. The remaining cell suspensions were sonicated, and then the number of bacteria in each well was calculated as described previously (43).

Chemotaxonomic characterization. The chemotaxonomic properties of strain TUM19329^T were characterized according to data on the major respiratory quinones, the cellular fatty acid composition, and the genomic GC content. Quinone extraction and determination were performed following a previous method (44). Briefly, cells were grown on BCYE α agar at 25°C for 14 days and were then harvested and freeze-dried. Total lipids were then extracted from the cells using a modified method (45), and the quinones in the crude extract were purified using Sep-Pak plus silica (Waters). The molecular type and concentration of each quinone extracted were analyzed using an ultraperformance liquid chromatography (UPLC) system (Acquity UPLC H-class system, Waters), a photodiode array (PDA) e λ detector (Waters) equipped with an Eclipse plus C₁₈ column (2.1 by 150 mm, 1.8 μm; Agilent Technologies), and Masslynx v4.2 software (Waters). The quinone species was finally determined based on the linear relationship between the logarithm of the UPLC retention time and the number of isoprene units according to the equivalent number of isoprene units of quinone components, as reported by Tamaoka et al. (46). The fatty acid composition was measured when cells were cultured at two different temperatures. Cells grown on BCYE α agar at 25°C for 14 days and 10°C for 48 days were harvested and used for analysis. Fatty acid methyl esters were prepared and analyzed using the protocol of the Sherlock Microbial Identification system v6.0 (Microbial ID; MIDI, Inc.). The fatty acid profile was compared and determined using the clinical bacterial library (CLIN6 v6.20). The genomic GC content was calculated from the complete genome sequence of the strain (8).

Comparative genomic analysis of *Legionella* strains. Comparative genomic analysis in this study followed the method of a previous study (47). Briefly, the genome sequences of the other 16 *Legionella* spp. were obtained from the EzBioCloud database (48). To eliminate potential differences due to previous analyses using different annotation tools, protein CDSs in all of the genomes were again predicted using Prodigal v2.6.2 (49) of the EzBioCloud whole-genome analysis pipeline with the default settings. The CDSs predicted were classified into clusters of orthologous groups (COGs) based on their functional roles following the reference database eggNOG v4.5 (50). For further functional annotation, they were compared with the Swiss-Prot/UniProt (51) and KEGG (52) databases using UBLAST (53). Using the EzBioCloud comparative genomics pipeline with the default settings, pan-genome orthologous groups (POGs) were determined by a combined reciprocal best-hit method using UBLAST (54) with an E value threshold of 1 × 10⁻⁶ and an open reading frame-independent method using nucleotide sequences (55) with at least 70% sequence coverage cutoff. After initial grouping, partial short gene sequences were targeted and used for clustering analysis against the determined POGs using UCLUST (53) with a cutoff of \geq 95% sequence identity. On the basis of comparison of the CDSs identified in each species and the POGs, the core (common to all), shared (common to two or more species), and unique genes were calculated.

Codon and amino acid usage. The codon usage and amino acid usage of the 17 strains were calculated using the Codon Usage Generator v2.4 (<http://bioinfo.ie.niigata-u.ac.jp/?Codon%20Usage%20Generator>) (56, 57). The created data sets were analyzed and visualized using Python v3.7.1 (<https://www.python.org>) and its libraries, such as pandas v1.1.1, matplotlib v3.3.1, and seaborn v0.10.1.

Detection of ISs. Genes in COG category [L] classified into four of the COGIDs (COG2801, COG2963, COG1484, and COG4584), which were especially enriched in TUM19329^T, were phylogenetically analyzed to clarify the sequence relationship. The nucleotide sequences of genes classified in each COG ID were aligned using MAFFT v7.450 using the auto option, which selected the L-INS-i algorithm (35). Subsequently, maximum-likelihood (ML) phylogenetic trees with 100-bootstrap replicates were estimated according to the GTRGAMMA model using RAxML software v8.2.12 (36) and were visualized using FigTree v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). Phylogenetic analysis showed that many of the gene sequences classified into each COGID were duplicated sequences. Duplicated gene sequences that formed clusters in the

phylogenetic tree always had another gene that was detected in pairs in the genome. Using the locus information of these paired gene sequences, the paired regions and their upstream and downstream regions (500 bases each) were extracted using a Python script. The extracted sequences were aligned using MAFFT with the auto option, which selected the L-INS-i algorithm (35), and then manually trimmed to identify the full length of the ISs. The structure of each IS and its location in the genome were visualized using the Python module, GenomeDiagram (58), and Biopython v1.74 (59). The full-length ISs are available on Figshare (see "Data Availability").

Detection of horizontal gene transfer. HGT events were identified using HGTector, a sequence similarity-based HGT prediction pipeline (33). A protein sequence similarity search was performed using DIAMOND v2.0.4 (60) against a database (generated by HGTector) that contains one representative per species from all available nonredundant RefSeq prokaryotic proteomes (October 2019). *Legionella* was set as the self group, and *Legionellales* was set as the close group. Quality cutoffs for valid hits were an E value of $\leq 1e-10$, percentage identity of $\geq 30\%$, and query coverage of $\geq 70\%$. For each protein-coding gene, the top 100 highest-scoring hits from different species were retained.

Data availability. The genome sequence of *Legionella antarctica* TUM19329^T was previously deposited in DDBJ/ENA/GenBank under the accession number AP022839, and the raw sequence data for this sequence were deposited under SRA accession numbers DRR213975 (MiSeq) and DRR213976 (MinION) (8). The data and script for the analysis presented in this article are available on Figshare (doi: 10.6084/m9.figshare.15912081).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.3 MB.

ACKNOWLEDGMENTS

We acknowledge the members of the 60th Japanese Antarctic Research Expedition (JARE 60), especially the terrestrial biology team members for their field support.

This study was supported by the National Institute of Polar Research (NIPR) through General Collaboration Project no. 31-34. This study was also partially funded by a Grant-in-Aid for Scientific Research on Innovative Areas "Post-Koch Ecology" (MEXT KAKENHI no. JP19H05683 and JP19H05679 to R.N.) and a Grant-in-Aid for Early-Career Scientists (JSPS KAKENHI no. JP19K20462 to R.N.).

We thank Aharon Oren of Hebrew University of Jerusalem (Jerusalem, Israel) for helpful comments and suggestions on nomenclature. We also thank Edanz for editing a draft of the manuscript.

REFERENCES

- Mondino S, Schmidt S, Rolando M, Escoll P, Gomez-Valero L, Buchrieser C. 2020. Legionnaires' disease: state of the art knowledge of pathogenesis mechanisms of *Legionella*. *Annu Rev Pathol* 15:439–466. <https://doi.org/10.1146/annurev-pathmechdis-012419-032742>.
- Phin N, Parry-Ford F, Harrison T, Stagg HR, Zhang N, Kumar K, Lortholary O, Zumla A, Abubakar I. 2014. Epidemiology and clinical management of Legionnaires' disease. *Lancet Infect Dis* 14:1011–1021. [https://doi.org/10.1016/S1473-3099\(14\)70713-3](https://doi.org/10.1016/S1473-3099(14)70713-3).
- Wadowsky RM, Wolford R, McNamara AM, Yee RB. 1985. Effect of temperature, pH, and oxygen level on the multiplication of naturally occurring *Legionella pneumophila* in potable water. *Appl Environ Microbiol* 49:1197–1205. <https://doi.org/10.1128/aem.49.5.1197-1205.1985>.
- Rogers J, Dowsett AB, Dennis PJ, Lee JV, Keevil CW. 1994. Influence of temperature and plumbing material selection on biofilm formation and growth of *Legionella pneumophila* in a model potable water system containing complex microbial flora. *Appl Environ Microbiol* 60:1585–1592. <https://doi.org/10.1128/aem.60.5.1585-1592.1994>.
- Wullings BA, van der Kooy D. 2006. Occurrence and genetic diversity of uncultured *Legionella* spp. in drinking water treated at temperatures below 15°C. *Appl Environ Microbiol* 72:157–166. <https://doi.org/10.1128/AEM.72.1.157-166.2006>.
- Carvalho FRS, Nastasi FR, Gamba RC, Foronda AS, Pellizari VH. 2008. Occurrence and diversity of *Legionellaceae* in polar lakes of the Antarctic peninsula. *Curr Microbiol* 57:294–300. <https://doi.org/10.1007/s00284-008-9192-y>.
- Shimada S, Nakai R, Aoki K, Shimoeda N, Ohno G, Kudoh S, Imura S, Watanabe K, Miyazaki Y, Ishii Y, Tateda K. 2021. Chasing waterborne pathogens in Antarctic human-made and natural environments, with special reference to *Legionella* spp. *Appl Environ Microbiol* 87:e02247-20. <https://doi.org/10.1128/AEM.02247-20>.
- Shimada S, Nakai R, Aoki K, Shimoeda N, Ohno G, Miyazaki Y, Kudoh S, Imura S, Watanabe K, Ishii Y, Tateda K. 2020. Complete genome sequence of novel psychrotolerant *Legionella* strain TUM19329, isolated from Antarctic lake sediment. *Microbiol Resour Announc* 9:e00253-20. <https://doi.org/10.1128/MRA.00253-20>.
- Adeleke AA, Fields BS, Benson RF, Daneshvar MI, Pruckler JM, Ratcliff RM, Harrison TG, Weyant RS, Birtles RJ, Raoult D, Halablab MA. 2001. *Legionella drozanskii* sp. nov., *Legionella rowbothamii* sp. nov. and *Legionella fallonii* sp. nov.: three unusual new *Legionella* species. *Int J Syst Evol Microbiol* 51:1151–1160. <https://doi.org/10.1099/00207713-51-3-1151>.
- Kim M, Oh H-S, Park S-C, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 64:346–351. <https://doi.org/10.1099/ijs.0.059774-0>.
- Nakai R, Abe T, Baba T, Imura S, Kagoshima H, Kanda H, Kanekiyo A, Kohara Y, Koi A, Nakamura K, Narita T, Niki H, Yanagihara K, Naganuma T. 2012. Microflora of aquatic moss pillars in a freshwater lake, East Antarctica, based on fatty acid and 16S rRNA gene analyses. *Polar Biol* 35:425–433. <https://doi.org/10.1007/s00300-011-1090-2>.
- Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, Clavel T. 2016. IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep* 6:33721. <https://doi.org/10.1038/srep33721>.
- Schmidt ML, White JD, Deneff VJ. 2016. Phylogenetic conservation of freshwater lake habitat preference varies between abundant bacterioplankton phyla. *Environ Microbiol* 18:1212–1226. <https://doi.org/10.1111/1462-2920.13143>.
- Mise K, Iwasaki W. 2020. Environmental atlas of prokaryotes enables powerful and intuitive habitat-based analysis of community structures. *iScience* 23:101624. <https://doi.org/10.1016/j.isci.2020.101624>.

15. Kimura S, Ban S, Imura S, Kudoh S, Matsuzaki M. 2010. Limnological characteristics of vertical structure in the lakes of Syowa Oasis, East Antarctica. *Polar Sci* 3:262–271. <https://doi.org/10.1016/j.polar.2009.08.002>.
16. Shivaji S, Prakash JSS. 2010. How do bacteria sense and respond to low temperature? *Arch Microbiol* 192:85–95. <https://doi.org/10.1007/s00203-009-0539-y>.
17. Hazel JR, Williams EE. 1990. The role of alterations in membrane lipid composition in enabling physiological adaptation of organisms to their physical environment. *Prog Lipid Res* 29:167–227. [https://doi.org/10.1016/0163-7827\(90\)90002-3](https://doi.org/10.1016/0163-7827(90)90002-3).
18. Zhang J, Burgess JG. 2015. *Shewanella electrodiffila* sp. nov., a psychrotolerant bacterium isolated from Mid-Atlantic Ridge deep-sea sediments. *Int J Syst Evol Microbiol* 65:2882–2889. <https://doi.org/10.1099/ijso.0.000345>.
19. Rabus R, Ruepp A, Frickey T, Rattei T, Fartmann B, Stark M, Bauer M, Zibat A, Lombardot T, Becker I, Amann J, Gellner K, Teeling H, Leuschner WD, Glöckner F-O, Lupas AN, Amann R, Klenk H-P. 2004. The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ Microbiol* 6:887–902. <https://doi.org/10.1111/j.1462-2920.2004.00665.x>.
20. Ayala-del-Río HL, Chain PS, Grzymalski JJ, Ponder MA, Ivanova N, Bergholz PW, Di Bartolo G, Hauser L, Land M, Bakermans C, Rodrigues D, Klappenbach J, Zarka D, Larimer F, Richardson P, Murray A, Thomashow M, Tiedje JM. 2010. The genome sequence of *Psychrobacter arcticus* 273-4, a psychrotolerant Siberian permafrost bacterium, reveals mechanisms for adaptation to low-temperature growth. *Appl Environ Microbiol* 76:2304–2312. <https://doi.org/10.1128/AEM.02101-09>.
21. Metpally RPR, Reddy BVB. 2009. Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: insights into the molecular basis of cold adaptation of proteins. *BMC Genomics* 10:11. <https://doi.org/10.1186/1471-2164-10-11>.
22. Grzymalski JJ, Carter BJ, DeLong EF, Feldman RA, Ghadiri A, Murray AE. 2006. Comparative genomics of DNA fragments from six Antarctic marine planktonic bacteria. *Appl Environ Microbiol* 72:1532–1541. <https://doi.org/10.1128/AEM.72.2.1532-1541.2006>.
23. De Maayer P, Anderson D, Cary C, Cowan DA. 2014. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep* 15:508–517. <https://doi.org/10.1002/embr.201338170>.
24. Gomez-Valero L, Rusniok C, Rolando M, Neou M, Dervins-Ravault D, Demirtas J, Rouy Z, Moore RJ, Chen H, Petty NK, Jarraud S, Etienne J, Steinert M, Heuner K, Gribaldo S, Médigue C, Glöckner G, Hartland EL, Buchrieser C. 2014. Comparative analyses of *Legionella* species identifies genetic features of strains causing Legionnaires' disease. *Genome Biol* 15:505. <https://doi.org/10.1186/s13059-014-0505-0>.
25. Duménil G, Isberg RR. 2001. The *Legionella pneumophila* lcmR protein exhibits chaperone activity for lcmQ by preventing its participation in high-molecular-weight complexes. *Mol Microbiol* 40:1113–1127. <https://doi.org/10.1046/j.1365-2958.2001.02454.x>.
26. Feldman M, Zusman T, Hagag S, Segal G. 2005. Coevolution between nonhomologous but functionally similar proteins and their conserved partners in the *Legionella* pathogenesis system. *Proc Natl Acad Sci U S A* 102:12206–12211. <https://doi.org/10.1073/pnas.0501850102>.
27. Gomez-Valero L, Rusniok C, Carson D, Mondino S, Pérez-Cobas AE, Rolando M, Pasricha S, Reuter S, Demirtas J, Crumbach J, Descorps-Declere S, Hartland EL, Jarraud S, Dougan G, Schroeder GN, Frankel G, Buchrieser C. 2019. More than 18,000 effectors in the *Legionella* genus genome provide multiple, independent combinations for replication in human cells. *Proc Natl Acad Sci U S A* 116:2265–2273. <https://doi.org/10.1073/pnas.1808016116>.
28. Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* 38:865–891. <https://doi.org/10.1111/1574-6976.12067>.
29. Allen MA, Lauro FM, Williams TJ, Burg D, Siddiqui KS, De Francisci D, Chong KKY, Pilak O, Chew HH, De Maere MZ, Ting L, Katrib M, Ng C, Sowers KR, Galperin MY, Anderson IJ, Ivanova N, Dalin E, Martinez M, Lapidus A, Hauser L, Land M, Thomas T, Cavicchioli R. 2009. The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. *ISME J* 3:1012–1035. <https://doi.org/10.1038/ismej.2009.45>.
30. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U, Martinez A, Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311:496–503. <https://doi.org/10.1126/science.1120250>.
31. Bordenstein SR, Reznikoff WS. 2005. Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol* 3:688–699. <https://doi.org/10.1038/nrmicro1233>.
32. Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* 9:218–229. <https://doi.org/10.1038/nrg2319>.
33. Zhu Q, Kosoy M, Dittmar K. 2014. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. *BMC Genomics* 15:717. <https://doi.org/10.1186/1471-2164-15-717>.
34. Moses AS, Millar JA, Bonazzi M, Beare PA, Raghavan R. 2017. Horizontally acquired biosynthesis genes boost *Coxiella burnetii*'s physiology. *Front Cell Infect Microbiol* 7:174. <https://doi.org/10.3389/fcimb.2017.00174>.
35. Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>.
36. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
37. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>.
38. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker 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>.
39. Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol* 32:2798–2800. <https://doi.org/10.1093/molbev/msv150>.
40. Farris JS. 1972. Estimating phylogenetic trees from distance matrices. *Am Nat* 106:645–668. <https://doi.org/10.1086/282802>.
41. Kreft Ł, Botzki A, Coppens F, Vandepoele K, Van Bel M. 2017. PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics* 33:2946–2947. <https://doi.org/10.1093/bioinformatics/btx324>.
42. Gimenez DF. 1964. Staining Rickettsiae in yolk-sac cultures. *Stain Technol* 39:135–140. <https://doi.org/10.3109/10520296409061219>.
43. Ohno A, Kato N, Yamada K, Yamaguchi K. 2003. Factors influencing survival of *Legionella pneumophila* serotype 1 in hot spring water and tap water. *Appl Environ Microbiol* 69:2540–2547. <https://doi.org/10.1128/AEM.69.5.2540-2547.2003>.
44. Watanabe K, Morohoshi S, Kunihiro T, Ishii Y, Takayasu I, Ogata Y, Shindo C, Suda W. 2020. *Fluviibacter phosphoraccumulans* gen. nov., sp. nov., a polyphosphate-accumulating bacterium of *Fluviibacteraceae* fam. nov., isolated from surface river water. *Int J Syst Evol Microbiol* 70:5551–5560. <https://doi.org/10.1099/ijsem.0.004446>.
45. Blich EG, Dyer WJ. 1959. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37:911–917. <https://doi.org/10.1139/o59-099>.
46. Tamaoka J. 1986. Analysis of bacterial menaquinone mixtures by reverse-phase high-performance liquid chromatography. *Methods Enzymol* 123:251–256. [https://doi.org/10.1016/s0076-6879\(86\)23028-1](https://doi.org/10.1016/s0076-6879(86)23028-1).
47. Nakai R, Naganuma T, Tazato N, Morohoshi S, Koide T. 2020. Cell plasticity and genomic structure of a novel filterable *Rhizobiales* bacterium that belongs to a widely distributed lineage. *Microorganisms* 8:1373. <https://doi.org/10.3390/microorganisms8091373>.
48. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 67:1613–1617. <https://doi.org/10.1099/ijsem.0.001755>.
49. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
50. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M, Jensen LJ, von Mering C, Bork P. 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res* 42:D231–D239. <https://doi.org/10.1093/nar/gkt1253>.
51. UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res* 43:D204–D212. <https://doi.org/10.1093/nar/gku989>.
52. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42:D199–D205. <https://doi.org/10.1093/nar/gkt1076>.
53. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.

54. Ward N, Moreno-Hagelsieb G. 2014. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PLoS One* 9:e101850. <https://doi.org/10.1371/journal.pone.0101850>.
55. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, Taviani E, Jeon Y-S, Kim DW, Lee J-H, Brettin TS, Bruce DC, Challacombe JF, Detter JC, Han CS, Munk AC, Chertkov O, Meincke L, Saunders E, Walters RA, Huq A, Nair GB, Colwell RR. 2009. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 106:15442–15447. <https://doi.org/10.1073/pnas.0907787106>.
56. Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 238:143–155. [https://doi.org/10.1016/s0378-1119\(99\)00225-5](https://doi.org/10.1016/s0378-1119(99)00225-5).
57. Suzuki H, Brown CJ, Forney LJ, Top EM. 2008. Comparison of correspondence analysis methods for synonymous codon usage in bacteria. *DNA Res* 15:357–365. <https://doi.org/10.1093/dnares/dsn028>.
58. Pritchard L, White JA, Birch PRJ, Toth IK. 2006. GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics* 22:616–617. <https://doi.org/10.1093/bioinformatics/btk021>.
59. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
60. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>.