Legionella felix sp. - A novel Legionella species isolated in Israel from hot tap water

Gal Zizelski Valenci[1], Shosanit Ohad[2], Mor Robinstein[1], Shereen Assaly[1], Ina Kutikov[1], Ludmila Groisman[1], Omer Murik[3], David A. Zeevi[3], Ravid Ben David[1], Alona Farber[1], Zeev Dveyrin[1], Efrat Rorman[1], Israel Nissan[1*]

1. National Public Health Laboratory, Tel Aviv. Public Health Directorate, Ministry of Health, Tel Aviv, Israel
2. Department of Laboratories, Public Health Directorate, Ministry of Health, Tel Aviv, Israel
3. Translational Genomics Lab, Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem

Keywords: *Legionella*, whole genome sequencing, plasmid, biosynthesis of secondary metabolites, metal resistance genes.

*Corresponding author: Israel Nissan, Israel.nissan@gmail.com

Abstract

Bacteria of the genus *Legionella* are natural pathogens of the environment that can cause legionellosis, which can result in severe human pneumonia called Legionnaires' disease. Here we describe a novel *Legionella* species isolated from hot water source in Israel. High quality complete genome was generated using a combination of Nanopore and Illumina sequencing approaches. Our novel *Legionella* isolate possesses a 4,071,709 bp chromosome and a single novel large *incF* conjugative plasmid of 160,437 bp. Interestingly, the chromosome encodes for 14 secondary metabolites biosynthetic gene clusters, more than any known other *Legionella* sp. The novel plasmid encodes for multiple genes that confer resistance to heavy metals. Bioinformatics analysis revealed that the genome of the new isolate differs from any previously described *Legionella* species. The closest related species to our isolate is *Legionella cherrii*. The name proposed for the new specie is *Legionella felix* in honor of Dr. Arthur Felix (1887-1956), a pioneering microbiologist, and member of the Royal Society of UK, who established the National Public Health Laboratory in Tel Aviv.

Introduction:

Legionella spp. are Gram-negative bacteria associated with respiratory infections [1], which at the time of writing this article (Aug 2023), include 72 species, summarized in the "List of Prokaryotic names with Standing in Nomenclature" (www. bacterio. Net).

Bacteria of the *Legionella* genus are opportunistic human pathogens that can cause a serious type of pneumonia called Legionnaires' disease and a mild flu-like illness, called Pontiac fever [2].

The most vulnerable population for Legionnaires' disease are immunocompromised patients, including individuals treated with immunosuppressive therapy, chronic renal disease, age over 50 years, chronic cardiovascular disease, underlying respiratory disease, smoking, and diabetes [3, 4].

In nature, *Legionella spp*. are very common in the environment, populate fresh water and rarely cause illness. The bacteria parasitize and replicate within free-living eukaryotic phagotrophs, and protozoa species primarily *Amoebozoa* [1, 5].

According to the Centers for Disease Control and Prevention (CDC), *Legionella* can grow in natural as well as in man-made settings, especially when not properly maintained. *Legionella* can thrive in various water sources, including heated water, spa pools, air-conditioning cooling towers, premise plumbing, building water systems and geothermal water. *Legionella* persists in natural and man-made environments due to colonization of interspecies biofilms [2, 6]. Nevertheless, not much is known about *Legionella* in saline environments or seawater [7, 8].

Legionella infection occurs almost exclusively via breathing small droplets of water that contain the bacteria. In rare cases, while drinking goes mistakenly into the lungs, it can cause an infection (https://www.cdc.gov/legionella/downloads/fs-legionnaires.pdf). In general, person-to-person transmission is very rare [9].

In spite of the wide variety of *Legionella* species, only about 50% have been associated with human disease [10]. The species *L. pneumophila* is the cause for most Legionnaires' disease (about 90%). It successfully adapted to new and challenging niches, such as Antarctic freshwater lakes, hot water sources (temperature over 60°C), and extremely acidic habitats [11-14]. The next most common etiological agents disease (2-7%) are; *L. longbeachae*, *L.*

bozemanii and *L. micdadei* [12, 14]. The first detection of *L. cherrii* species in the respiratory samples of patients with pneumonia symptoms was published by *Atosa Ghorbani et al*., 2021 [15].

Legionella species encode a highly conserved type IVB secretion system (T4SS) called Dot/Icm. This system is essential for intracellular multiplication, and is encoded for about 29 different dot/icm genes [16]. The Dot/Icm system enables the bacteria to translocate effector proteins into the host cytosol, taking over cell functions and creating a specialized organelle, called the *Legionella* containing vacuole [17, 18].

Results and discussion:

During a routine monitoring test of *Legionella* spp. from hot water spa, performed at the National Public Health Laboratory, Tel Aviv, isolate #227 showed special phenotypic characters. It was isolated from a hot water shower (tap water, temperature above 55° C). The sample was tested in accordance with standard procedures ISO 11731, for the isolation and quantitation of *Legionella* in water. This includes culture on Blood agar and Buffered Charcoal Yeast Extract agar (BCYE) at 36º C, for 48 hours in 2.5% CO_2. No growth was observed in BCYE lacking L-cysteine and iron, as anticipated from *Legionella* spp. We noticed that the colonies that grew on the selective plates were glowing under UV light (Figure 1). PCR analysis using the microproof® Legionella Quantification LyoKit (BIOTECON Diagnostics) identified it as *Legionella* sp.

A matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) profile for isolate #227 was acquired on a Bruker Microflex system and interpreted with Bruker *Legionella* database using the MTB Compass, and Flex control software. Inactivation and preparation of the isolate for MALDI-TOF MS analysis were performed according to the manufacturer's instructions. The MALDI-TOF-MS identified this isolate as *Legionella cherrii* at a maximal score of 2.03.

For Whole Genome Sequencing (WGS), 4 representative colonies were inoculated and suspended in 2 ml of Legionella - Pages Saline Buffer, followed by resuspension of 250 µl of bacterial suspension in 200 µl of Lysis buffer (Hylabs, catalogue number BP007). DNA was isolated using magLEAD® 12gC Automated Nucleic Acid Extraction System (Precision System Science Co., Ltd), following the manufacturer's instructions. WGS of isolate #227 was carried out applying two technologies as described; short-read sequencing using the Illumina MiSeq platform, and Long-read sequencing using the Oxford Nanopore platform.

The sequences obtained from the two technologies were combined into two complete contigs (chromosome and plasmid). For the Illumina MiSeq platform, Paired-end library of isolate #227 was generated using Illumina DNA Prep library preparation kit according to Illumina protocols. The DNA library was sequenced using a 250-bp paired-end read MiSeq Reagent Kit v2 (500-cycles, catalog number MS-102-2003). All bioinformatics analyses were performed utilizing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (https://www.bv-brc.org/) with default parameters, unless otherwise noted [19, 20]. The qualities of the raw reads were checked using Fastq Utilities of BV-BRC [20], confirming that the fastq files were of good quality.

For Oxford Nanopore (ONT) sequencing, DNA from isolate #227 was converted into a sequencing library using Rapid Barcoding Kit 96 (SQK-RBK110.96) chemistry according to the manufacturer's protocol. Sequencing was performed on a MinION sequencer with an R9.4.1 (FLO-MIN106D) flow cell and base calling was carried out by Guppy version 6.0.1. Raw sequencing data were filtered with NanoFilt [21], keeping reads longer than 500bp and with quality > 10. Flye version 2.9 [22] was used for de novo assembly, and the assembly completeness was assessed with Quast version 5.2.0 and Busco version 5.4.2 [22-24]. The ONT sequences assembly results in two big contigs that serve as the template for Illumina reads mapping. A Consensus sequence was obtained with the BV-BRC Variation Analysis Service [20]. The schematic map of isolate #227 chromosome and plasmid sequence is presented in Figure 2. The assembled consensus genome was annotated using The BV-BRC Genome Annotation tool [20]. The genome comprises a 3,913,681 bp chromosome (GC content 38.78%) that encodes for 3505 CDS and a 160,500 bp large plasmid (GC content 37.97%) that encodes for 179 CDS, figure 2. Of the total 3,684 predicted protein-coding sequences, 1482 and 95 are hypothetical on the chromosome and plasmid respectively. A total of 7 rRNAs and 43 tRNAs were identified on the chromosome.
Examination of isolate #227 genome using the VFDB (Virulence Factor Database, as part of the BV-BRC), shows that the *Legionella* isolate #227 harbore many virulence factors, including the dot/icm T4SS (supplementary table S1). Of the total plasmid genes (179), 93 encode for hypothetical proteins, 17 for heavy metal resistance, 18 the *incF* plasmid conjugative transfer genes, and few antibiotic resistance genes (supplementary table S2). Notably, searching for plasmid homologues in the NCBI nr database, using Blastn revealed highly fragmented short homologues sequences in other *Legionella sp.* plasmids, indicating that this is a novel plasmid.

We used the antibiotics and secondary metabolites analysis shell (antiSMASH 6.1.1) to search for the presence of secondary metabolite biosynthetic gene clusters in isolate #227 [25]. The analysis predicts 14 gene clusters encoding the biosynthesis of secondary metabolites and bioactive compounds (Figure 3). For example, the antiSMASH analysis predicts four Non-ribosomal peptide synthetase cluster (NRPS) and NRPS-like region, one of them harbor 50% gene similarity to Legioliulin biosynthetic gene cluster from *Legionella parisiensis* (Figure 3, region 1.2). Legioliulin is a compound responsible for blue-white auto fluorescence in *Legionella dumoffii* under long-wavelength UV light [26]. Another example is the Betalactone biosynthesis coding region (figure 3, region 1.13), containing a gene of 13% similarity to a cluster of genes responsible for the synthesis of the Fengycin compound, a known antifungal lipopeptic antibiotic [27, 28]. Additionally, type III PKS region which contains genes with 12% & 25% similarity to genes involved in the biosynthesis of Nematophin and Ambactin, respectively (Figure 3, region 1.7). In the work of Nicolas J. Tobias et al., 2016 [29] he used antiSMASH for studying biosynthetic gene clusters (BGC) in 15 *Legionella* species. Indeed, *Legionella sp*. possesses variable BGCs, in particular 11 BGCs were identified in *L. anisa*. Therefore, isolate #227 encode for the largest known number of BGCs in a single *Legionella* genome. Notably, all these BGCs are located on the chromosome of isolate #227. The activities and potential utilizations of these BGCs might be studied in the future. The BV-BRC Similar Genome Finder Service, uses the Mash/MinHash algorithm, which presents each genome as a sketch, composed of 1000 of its k-mers [30]. This tool allows searching public genomic databases in the BV-BRC for the closest high-quality representative genome as determined by the NCBI, or, the closest homologue genome regardless of its quality of sequencing. As presented in supplementary table S3, the four most similar genomes strains in the database to isolate 227 are *L. cherrii*, that share between 193-203 common k-mers out of 1000. We used average nucleotide identity (ANI) calculator to estimate average nucleotide identity [31]. The level of similarity between the genome of isolate #227 and the most similar genomes in the NCBI database (top 4 in supplementary table S3) is 93.9667, 93.6396, 93.902, 93.9381 respectively.

We performed genome-based taxonomy at TYPE (strain) GENOME SERVER (TYGS), according to the default parameters (https://tygs.dsmz.de/) [32, 33]. The relevant type-strain genomes were determined automatically by the TYGS server. Genome-based phylogenetic trees of whole-genome, and 16S rDNA gene are shown in Figure 4 respectively. The results of the taxonomic identification performed by the TYPE server are: "Potential new species detected. Isolate #227 does not belong to any species found in TYGS database".

Nevertheless, the most similar representative genome belongs to *Legionella cherrii* DSM 19213. In addition, according to the 16S rDNA based tree (Figure 4B) *L. cherrii* is not the most similar to isolate #227, and the most similar species is *L. anisa*. Pairwise comparisons of isolate #227 genomes vs. type strain genomes and strains in dataset is available in supplementary Tables S4 & S5.

Taxonomic classification was performed using Kraken v2 [34] The taxonomic classification results were inconclusive (Figure 5). Illumina and ONT sequences are available to the public at the NCBI (BioProject number PRJNA1004905).

Based on the above bioindormatics analysis we concluded that MLADI-TOF MS failed to correctly identify isolate #227.

In conclusion, using combination of ONT and Illumina platforms, we achieved a complete and high quality sequence for isolate #227. Based on meticulous bioinformatics analysis we found that *Legionella* isolate #277 represents a possible novel *Legionella* species that harbors large novel plasmid. We proposed the name *Legionella felix* for the new species [35].

Figures legends:

Figure 1. Growth of isolate 227 on Buffered Charcoal Yeast Extract agar (BCYE) at 36° C, for 48 hours in 2.5% CO_2. Photo taken under UV (A) and visible light (B) respectively.

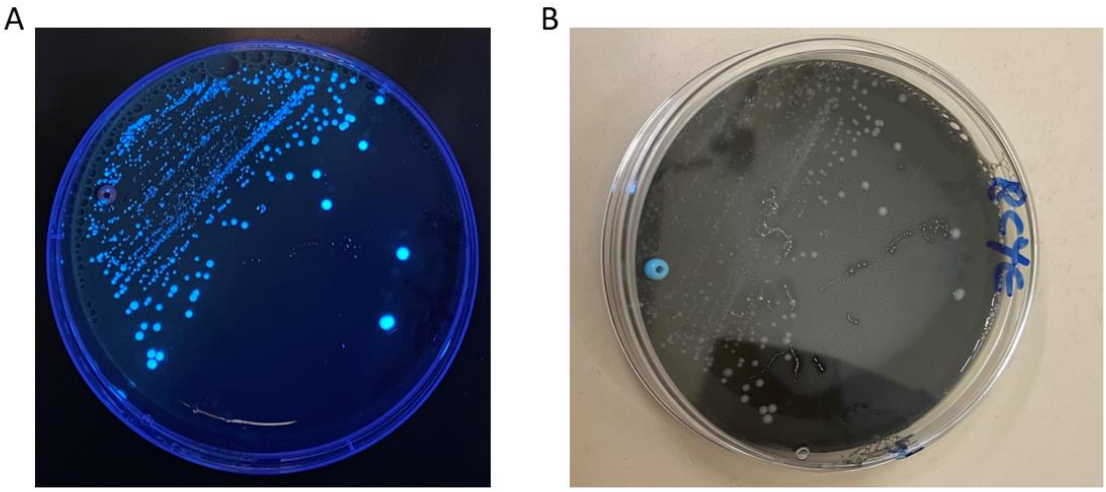


Figure 2. Schematic map of isolate #227 Chromosome and Plasmid. Isolate #227 consensus sequence genome was obtained by the BV-BRC Variation Analysis Service, combining Illumina and Nanopour technologies. ARM- Antimicrobial resistance, VF- virulence factor.

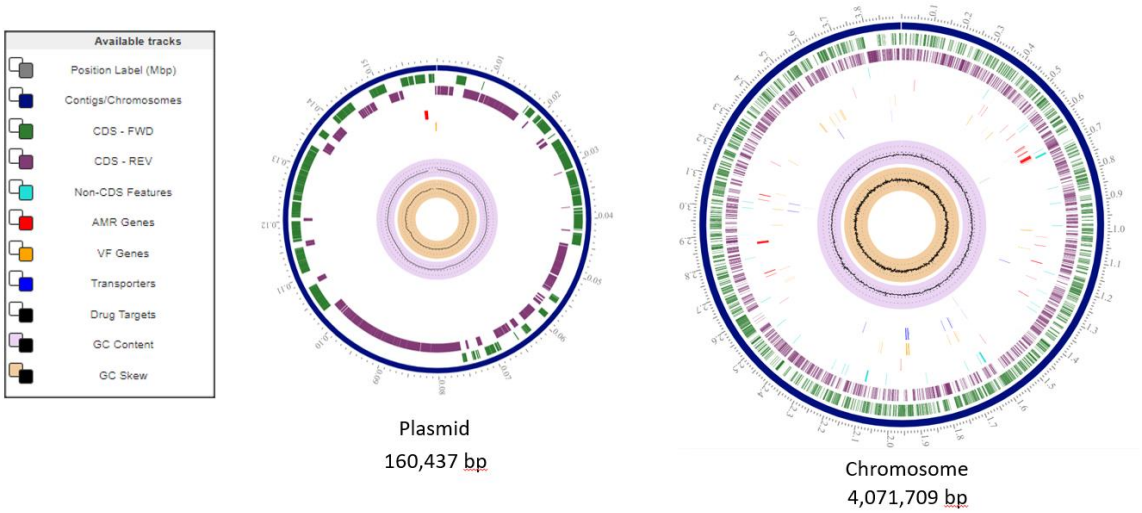


Figure 3. Prediction of gene clusters encoding the BGCs using antiSMASH 6.1.1. Schematic map of the chromosome with location of 14 BGCs (A) and list of regions prediction of gene clusters (B). The different color boxes in A, represents different clusters located on the schematic map of the genome.

A

B

Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	RiPP-like	162,851	173,714			
Region 1.2	NRPS-like	890,165	932,270			
Region 1.3	NAPAA	1,670,692	1,704,696			
Region 1.4	NRPS-like , transAT-PKS-like	1,751,810	1,830,606	legioliulin	Polyketide:Trans-AT type I	50%
Region 1.5	CDPS	1,884,586	1,902,961			
Region 1.6	CDPS	2,265,793	2,286,548			
Region 1.7	T3PKS	2,303,376	2,343,299	nematophin	NRP	12%
Region 1.8	T1PKS , NRPS	2,379,017	2,427,767			
Region 1.9	NRPS	2,816,716	2,860,000			
Region 1.10	NRPS-like	2,943,222	2,986,317			
Region 1.11	T3PKS	3,211,477	3,249,554			
Region 1.12	T3PKS	3,377,230	3,418,348			
Region 1.13	betalactone	3,665,432	3,688,402	fengycin	NRP	13%
Region 1.14	transAT-PKS	3,772,245	3,845,448			

Figure 4. Whole-genome sequence-based phylogenetic tree (A) and 16S rDNA gene sequence-based tree (B). The trees represent the final step of the comprehensive genome-based taxonomic analyses available in the TYGS. Tree inferred with FastME 2.1.6.1. Leaf labels are annotated by affiliation to species and subspecies clusters, genomic G+C content, δ values, overall genome sequence length, number of proteins, and the kind of strain. Percent

G+C 37.8-39.64, Delta statistics 0.299-0.449, Genome size (in bp) 3,573,650-5,227,415, SSU lengths (in bp) 1,092-1533. The colors of scheme boxes were determined automatically by the TYGS server.

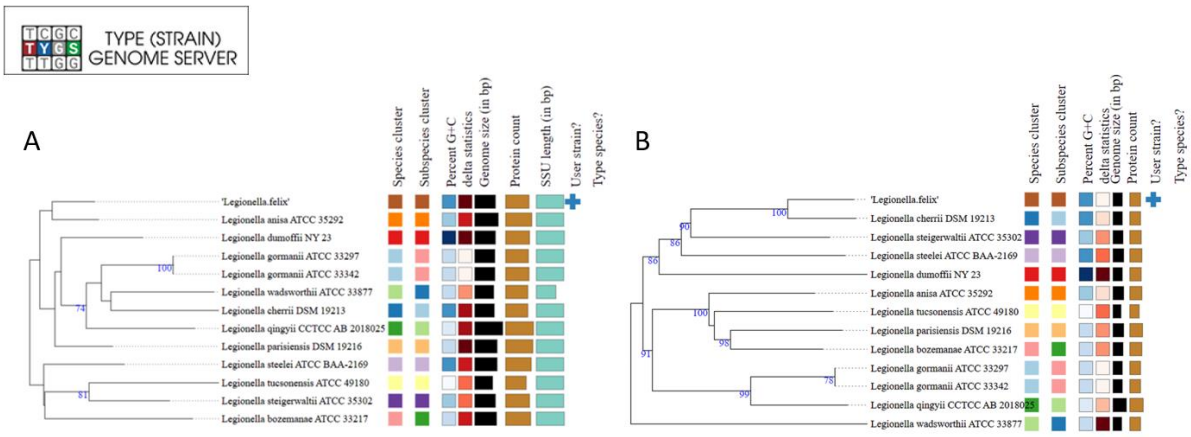
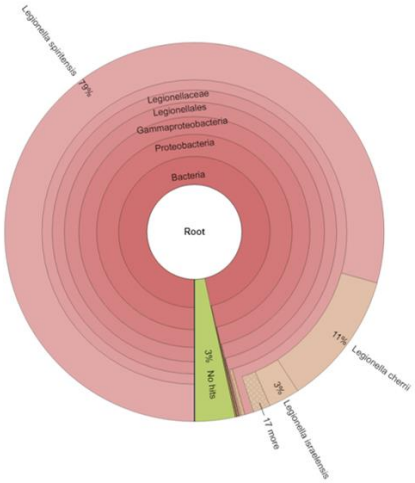


Figure 5. Taxonomic classification of isolate 227# using KRAKEN2.The columns contain data as follows: Percentage of fragments covered by the clade rooted at this taxon, number of fragments covered by the clade rooted at this taxon, number of fragments assigned directly to this taxon, a rank code, indicating (U)nclassified, (R)oot, (D)omain, (K)ingdom, (P)hylum, (C)lass, (O)rder, (F)amily, (G)enus, or (S)pecies. Taxa that are not at any of these 10 ranks have a rank code that is formed by using the rank code of the closest ancestor rank with a number indicating the distance from that rank, NCBI taxonomic ID number, indented scientific name.

Pct Coverage	Frags in Clade	Frags in Taxon	Rank	NCBI Taxon ID	Scientific Name
3.42	25165	25165	U	0	
96.58	710469	3	R	1	
96.58	710457	32	R1	131567	cellular organisms
96.54	710203	172	D	2	Bacteria
96.42	709316	108	P	1224	Proteobacteria
96.36	708865	282	C	1236	Gammaproteobacteria
96.21	707756	4	O	118969	Legionellales
96.21	707743	1255	F	444	Legionellaceae
95.70	704008	5468	G	445	Legionella
79.33	583597	583597	S	452	Legionella spiritensis
11.50	84562	84562	S	28084	Legionella cherrii
2.65	19489	19489	S	454	Legionella israelensis



References:

1. Rowbotham, T.J., *Preliminary report on the pathogenicity of Legionella pneumophila for freshwater and soil amoebae.* J Clin Pathol, 1980. **33**(12): p. 1179-83.
2. Chauhan, D. and S.R. Shames, *Pathogenicity and Virulence of Legionella: Intracellular replication and host response.* Virulence, 2021. **12**(1): p. 1122-1144.
3. Cunha, B.A., A. Burillo, and E. Bouza, *Legionnaires' disease.* Lancet, 2016. **387**(10016): p. 376-385.
4. Barimani, M.J., *Legionella: An uncommon cause of community-acquired pneumonia.* JAAPA, 2022. **35**(10): p. 38-42.
5. Swart, A.L., et al., *Acanthamoeba and Dictyostelium as Cellular Models for Legionella Infection.* Front Cell Infect Microbiol, 2018. **8**: p. 61.
6. Rogers, J., et al., *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, 1994. **60**(5): p. 1585-92.
7. Gast, R.J., et al., *Amoebae and Legionella pneumophila in saline environments.* J Water Health, 2011. **9**(1): p. 37-52.
8. Linsak, D.T., et al., *Sea water whirlpool spa as a source of Legionella infection.* J Water Health, 2021. **19**(2): p. 242-253.
9. Borges, V., et al., *Legionella pneumophila strain associated with the first evidence of person-to-person transmission of Legionnaires' disease: a unique mosaic genetic backbone.* Sci Rep, 2016. **6**: p. 26261.
10. Fields, B.S., R.F. Benson, and R.E. Besser, *Legionella and Legionnaires' disease: 25 years of investigation.* Clin Microbiol Rev, 2002. **15**(3): p. 506-26.
11. McDade, J.E., et al., *Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease.* N Engl J Med, 1977. **297**(22): p. 1197-203.
12. Muder, R.R. and V.L. Yu, *Infection due to Legionella species other than L. pneumophila.* Clin Infect Dis, 2002. **35**(8): p. 990-8.
13. Oliva, G., T. Sahr, and C. Buchrieser, *The Life Cycle of L. pneumophila: Cellular Differentiation Is Linked to Virulence and Metabolism.* Front Cell Infect Microbiol, 2018. **8**: p. 3.

14. Salloum, G., et al., *Identification of Legionella species by ribotyping and other molecular methods.* Res Microbiol, 2002. **153**(10): p. 679-86.
15. Ghorbani, A., et al., *Occurrence of the Legionella species in the respiratory samples of patients with pneumonia symptoms from Ahvaz, Iran; first detection of Legionella cherrii.* Mol Biol Rep, 2021. **48**(11): p. 7141-7146.
16. Burstein, D., et al., *Genomic analysis of 38 Legionella species identifies large and diverse effector repertoires.* Nat Genet, 2016. **48**(2): p. 167-75.
17. Qiu, J. and Z.Q. Luo, *Legionella and Coxiella effectors: strength in diversity and activity.* Nat Rev Microbiol, 2017. **15**(10): p. 591-605.
18. Mace, K., et al., *Proteins DotY and DotZ modulate the dynamics and localization of the type IVB coupling complex of Legionella pneumophila.* Mol Microbiol, 2022. **117**(2): p. 307-319.
19. Wattam, A.R., et al., *Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center.* Nucleic Acids Res, 2017. **45**(D1): p. D535-D542.
20. Olson, R.D., et al., *Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR.* Nucleic Acids Res, 2022.
21. De Coster, W., et al., *NanoPack: visualizing and processing long-read sequencing data.* Bioinformatics, 2018. **34**(15): p. 2666-2669.
22. Mikheenko, A., et al., *Versatile genome assembly evaluation with QUAST-LG.* Bioinformatics, 2018. **34**(13): p. i142-i150.
23. Seppey, M., M. Manni, and E.M. Zdobnov, *BUSCO: Assessing Genome Assembly and Annotation Completeness.* Methods Mol Biol, 2019. **1962**: p. 227-245.
24. Simao, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.* Bioinformatics, 2015. **31**(19): p. 3210-2.
25. Blin, K., et al., *antiSMASH 6.0: improving cluster detection and comparison capabilities.* Nucleic Acids Res, 2021. **49**(W1): p. W29-W35.
26. Amemura-Maekawa, J., et al., *Legioliulin, a new isocoumarin compound responsible for blue-white autofluorescence in Legionella (Fluoribacter) dumoffii under long-wavelength UV light.* Biochem Biophys Res Commun, 2004. **323**(3): p. 954-9.
27. Vanittanakom, N., et al., *Fengycin--a novel antifungal lipopeptide antibiotic produced by Bacillus subtilis F-29-3.* J Antibiot (Tokyo), 1986. **39**(7): p. 888-901.

28. Sur, S., T.D. Romo, and A. Grossfield, *Selectivity and Mechanism of Fengycin, an Antimicrobial Lipopeptide, from Molecular Dynamics.* J Phys Chem B, 2018. **122**(8): p. 2219-2226.
29. Tobias, N.J., et al., *Legionella shows a diverse secondary metabolism dependent on a broad spectrum Sfp-type phosphopantetheinyl transferase.* PeerJ, 2016. **4**: p. e2720.
30. Ondov, B.D., et al., *Mash: fast genome and metagenome distance estimation using MinHash.* Genome Biol, 2016. **17**(1): p. 132.
31. Goris, J., et al., *DNA-DNA hybridization values and their relationship to whole-genome sequence similarities.* Int J Syst Evol Microbiol, 2007. **57**(Pt 1): p. 81-91.
32. Meier-Kolthoff, J.P., et al., *TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes.* Nucleic Acids Res, 2022. **50**(D1): p. D801-D807.
33. Meier-Kolthoff, J.P. and M. Goker, *TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy.* Nat Commun, 2019. **10**(1): p. 2182.
34. Wood, D.E. and S.L. Salzberg, *Kraken: ultrafast metagenomic sequence classification using exact alignments.* Genome Biol, 2014. **15**(3): p. R46.
35. . 1956: Nature Publishing Group. p. 257.