

Whole-Genome Sequences of a Cluster of 14 Unidentified Related *Veillonella* sp. Strains from Human Clinical Samples and Type Strains of 3 *Veillonella* Validated Species

Fabien Aujoulat, Michel Popoff, Laure Diancourt, Alexis Criscuolo, H el ene Jean-Pierre, Estelle Jumas-Bilak, H el ene Marchandin

► **To cite this version:**

Fabien Aujoulat, Michel Popoff, Laure Diancourt, Alexis Criscuolo, H el ene Jean-Pierre, et al.. Whole-Genome Sequences of a Cluster of 14 Unidentified Related *Veillonella* sp. Strains from Human Clinical Samples and Type Strains of 3 *Veillonella* Validated Species. *Microbiology Resource Announcements*, American Society for Microbiology, 2019, 8 (12), 10.1128/MRA.01743-18 . pasteur-02448615

HAL Id: pasteur-02448615

<https://hal-pasteur.archives-ouvertes.fr/pasteur-02448615>

Submitted on 22 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.





Whole-Genome Sequences of a Cluster of 14 Unidentified Related *Veillonella* sp. Strains from Human Clinical Samples and Type Strains of 3 *Veillonella* Validated Species

Fabien Aujoulat,^a Michel R. Popoff,^b Laure Diancourt,^b Alexis Criscuolo,^c H el ene Jean-Pierre,^d Estelle Jumas-Bilak,^e H el ene Marchandin^f

^aHydroSciences Montpellier, CNRS, IRD, Universit e de Montpellier, Montpellier, France

^bCentre National de R eference des bact eries ana erobies et du botulisme, Institut Pasteur, Paris, France

^cInstitut Pasteur–Bioinformatics and Biostatistics Hub–C3BI, USR 3756 IP CNR, Paris, France

^dCHU Montpellier, Laboratoire de Bact riologie, Montpellier, France

^eHydroSciences Montpellier, CNRS, IRD, Universit e de Montpellier, CHU Montpellier, Laboratoire d'Ecologie Microbienne Hospitali ere, Montpellier, France

^fHydroSciences Montpellier, CNRS, IRD, Universit e de Montpellier, CHU N imes, N imes, France

ABSTRACT We report 17 draft genomes for 14 unidentified *Veillonella* sp. strains closely related in 16S rRNA gene-based phylogeny and type strains of 3 *Veillonella* species with the aims of deciphering relationships between related species, evaluating the accuracy of current thresholds for species delineation, and robustly describing new species in the genus.

Veillonella spp. are anaerobic Gram-negative cocci and important representatives of the microbiota of humans and animals. Currently, 14 species in the genus *Veillonella* are validly described. In the absence of discriminative phenotypic characteristics, their identification requires molecular-based methods. However, all species are not discriminated by 16S rRNA gene (*rrs*) analysis because several pairs of species are closely related ($\geq 99\%$ of identical *rrs* nucleotides), such as *V. denticariosi* and *V. rodentium*, *V. ratti* and *V. criceti*, *V. ratti* and *V. seminalis*, and *V. dispar* and *V. parvula* (1–4). In addition, intrachromosomal heterogeneity between *rrs* copies (up to 1.43%) and intraspecific *rrs* variability that may surpass interspecific variability have been demonstrated in this genus (3, 5). This impairs the 16S rRNA gene-based identification of closely related species and suggests that applying the proposed revised threshold for a new species description, which is less than 98.7% of 16S rRNA gene identity, may not be adapted to the genus *Veillonella* (6). Therefore, molecular-based identification methods based on housekeeping genes such as *dnaK*, *rpoB*, and *gltA* were successively developed and used for the description of novel species in addition to or without associated DNA-DNA hybridization (DDH), which is the reference method for novel species description (2, 4).

We present draft genome sequences of 14 human clinical isolates and type strains of 3 *Veillonella* species that were not available at the time of our study. Isolates were recovered from human clinical samples from 14 patients attending the University Hospital in Montpellier, France. The study was approved by the institutional review board of the N imes University Hospital under the approval number 19.01.07. All strains were cultured on Columbia sheep blood agar (bioM erieux) at 37 C in an anaerobic jar with the AnaeroGen system (Oxoid Unipath) for 2 to 5 days (1, 4).

DNA was extracted with the MasterPure DNA purification kit (Epicentre Biotechnologies). Libraries were constructed with the Nextera XT DNA library preparation kit (Illumina, San Diego, CA) and sequenced on a NextSeq 500 instrument with a 2 imes 150-bp paired-end protocol (on average, 1,372,528 read pairs per sample and

Citation Aujoulat F, Popoff MR, Diancourt L, Criscuolo A, Jean-Pierre H, Jumas-Bilak E, Marchandin H. 2019. Whole-genome sequences of a cluster of 14 unidentified related *Veillonella* sp. strains from human clinical samples and type strains of 3 *Veillonella* validated species. *Microbiol Resour Announc* 8:e01743-18. <https://doi.org/10.1128/MRA.01743-18>.

Editor John J. Dennehy, Queens College

Copyright   2019 Aujoulat 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 H el ene Marchandin, helene.marchandin@umontpellier.fr.

Received 8 January 2019

Accepted 20 February 2019

Published 21 March 2019

TABLE 1 Characteristics of the 17 *Veillonella* sp. draft genome sequences^a

Species	Strain	Accession no.	BioSample no.	No. of paired-end reads	Coverage (x)	Genome size (bp)	No. of contigs	N ₅₀ (bp)	G+C content (%)
<i>V. ratti</i>	ATCC 17746 ^T	RQXU000000000	SAMN10465589	1,163,748	156	2,236,286	49	95,000	42.8
<i>V. caviae</i>	DSM 20738 ^T	RQYU000000000	SAMN10465590	1,391,318	211	1,963,997	57	83,095	38.4
<i>V. seminalis</i>	ADV 4313.2 ^T = CIP 107810 ^T	RQZU000000000	SAMN10465591	1,399,442	184	2,252,690	27	152,260	42.2
<i>Veillonella</i> sp.	CHU110 = 14/09/02-B-6149	RQVA000000000	SAMN10465592	973,834	155	1,881,921	43	107,364	39.4
<i>Veillonella</i> sp.	Va140 = CHU594 = 03/11/05-B-4237	RQVB000000000	SAMN10465593	1,206,819	199	1,822,704	23	147,662	39.3
<i>Veillonella</i> sp.	Va138 = CHU732 = 06/09/06-B-3252	RQVC000000000	SAMN10465594	740,997	112	1,972,307	42	104,904	39.5
<i>Veillonella</i> sp.	Va143 = CHU740 = 09/10/06-B-1508	RQVD000000000	SAMN10465595	1,518,558	240	1,893,839	35	130,014	39.3
<i>Veillonella</i> sp.	3891 = 12/03/14-B-3204	RQVE000000000	SAMN10465596	1,295,195	205	1,890,765	38	161,698	39.4
<i>Veillonella</i> sp.	3913 = 09/03/14-B-7013	RQVF000000000	SAMN10465597	698,249	118	1,770,519	15	394,883	39.4
<i>Veillonella</i> sp.	3627 = CHU4040 = 99141203627	RQVG000000000	SAMN10465598	1,308,145	226	1,738,782	34	132,446	39.3
<i>Veillonella</i> sp.	CHU3735 = CNR 79/14 = 14/01/14-B-2378	RQVH000000000	SAMN10465599	1,123,247	190	1,776,932	51	93,431	39.4
<i>Veillonella</i> sp.	3960 = CHU4076 = 99141423960	RQVI000000000	SAMN10465600	746,487	120	1,860,088	69	60,297	39.7
<i>Veillonella</i> sp.	3310 = CHU5981 = 99162773310	RQVJ000000000	SAMN10465601	1,640,876	258	1,870,378	16	311,068	39.6
<i>Veillonella</i> sp.	Va137 = 19/07/06-3404	RQVK000000000	SAMN10465602	1,966,481	310	1,902,403	36	128,606	39.6
<i>Veillonella</i> sp.	Va139 = 10/01/06-3360	RQVL000000000	SAMN10465603	2,226,458	303	2,172,276	76	67,981	39.5
<i>Veillonella</i> sp.	Va141 = 20/06/06-2457	RQVM000000000	SAMN10465604	2,034,756	333	1,806,031	12	229,161	39.2
<i>Veillonella</i> sp.	Va142 = 08/06/05-3381	RQVN000000000	SAMN10465605	1,898,363	281	1,969,382	66	75,004	39.1

^a BioProject number, PRJNA506647.

~212× sequencing depth; Table 1). All sequenced paired-end reads were clipped and trimmed with AlienTrimmer v. 0.4.0 (7), corrected with Musket v. 1.1 (8), and subjected to a digital normalization procedure with khmer v. 1.3 (9). For each sample, the remaining processed reads were assembled and scaffolded with SPAdes v. 3.11.0 (10). All programs were used with their default settings. Strains and characteristics of their whole-genome sequences (WGSs) are presented in Table 1.

Because genomic sequencing is being progressively incorporated into the taxonomy of *Bacteria* (11) and considering that it was not used before in the genus *Veillonella*, taxogenomic metrics such as *in silico* DDH, average nucleotide identity, and phylogenomics will be valuable for (i) deciphering intrageneric relationships between related species and between strains affiliated with a species or of unknown taxonomic status, (ii) evaluating the accuracy of current thresholds for species delineation, (iii) supporting reappraisal of the taxonomy, (iv) robustly describing new species, and (v) proposing minimal standards for description of new species in the genus *Veillonella*. Besides taxonomic purposes, WGS will also allow the characterization of clinically relevant gene content, particularly virulence and antibiotic resistance genes, in *Veillonella* spp. considered opportunistic human pathogens.

Data availability. This whole-genome shotgun project was deposited at DDBJ/ENA/GenBank under the accession numbers [RQUX00000000](https://doi.org/10.1016/j.resmic.2005.01.013) to [RQVN00000000](https://doi.org/10.1016/j.resmic.2005.01.013) as listed in Table 1 (BioProject number [PRJNA506647](https://doi.org/10.1099/ijms.0.65096-0); BioSample numbers [SAMN10465589](https://doi.org/10.1099/ijms.0.65096-0) to [SAMN10465605](https://doi.org/10.1099/ijms.0.65096-0)). The versions described in this paper are versions RQUX01000000 to RQVN01000000. The Sequence Read Archive accession numbers are [SRX5189907](https://doi.org/10.1099/ijms.0.65096-0) to [SRX5189923](https://doi.org/10.1099/ijms.0.65096-0).

REFERENCES

1. Marchandin H, Teyssier C, Jumas-Bilak E, Robert M, Artigues A-C, Jean-Pierre H. 2005. Molecular identification of the first human isolate belonging to the *Veillonella ratti*-*Veillonella criceti* group based on 16S rDNA and *dnaK* gene sequencing. Res Microbiol 156:603–607. <https://doi.org/10.1016/j.resmic.2005.01.013>.
2. Byun R, Carlier J-P, Jacques NA, Marchandin H, Hunter N. 2007. *Veillonella denticariosi* sp. nov., isolated from human carious dentine. Int J Syst Evol Microbiol 57:2844–2848. <https://doi.org/10.1099/ijms.0.65096-0>.
3. Michon A-L, Aujoulat F, Roudière L, Soulier O, Zorogniotti I, Jumas-Bilak E, Marchandin H. 2010. Intrageneric and intraspecific heterogeneity in *rrs* may surpass interspecific variability in a natural population of *Veillonella*. Microbiology 156:2080–2091. <https://doi.org/10.1099/mic.0.038224-0>.
4. Aujoulat F, Bouvet P, Jumas-Bilak E, Jean-Pierre H, Marchandin H. 2014. *Veillonella seminalis* sp. nov., a novel anaerobic Gram-stain-negative coccus from human clinical samples, and emended description of the genus *Veillonella*. Int J Syst Evol Microbiol 64:3526–3531. <https://doi.org/10.1099/ijms.0.064451-0>.
5. Marchandin H, Teyssier C, Siméon de Buochberg M, Jean-Pierre H, Carriere C, Jumas-Bilak E. 2003. Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus *Veillonella*: implications for phylogeny and taxonomy. Microbiology 149:1493–1501. <https://doi.org/10.1099/mic.0.26132-0>.
6. Stackebrandt E, Ebers J. 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 33:152–155.
7. Criscuolo A, Brisse S. 2013. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. Genomics 102:500–506. <https://doi.org/10.1016/j.ygeno.2013.07.011>.
8. Liu Y, Schröder J, Schmidt B. 2013. Musket: a multistage *k*-mer spectrum-based error corrector for Illumina sequence data. Bioinformatics 29:308–315. <https://doi.org/10.1093/bioinformatics/bts690>.
9. Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edvenson G, Fay S, Fenton J, Fenzl T, Fish J, Garcia-Gutierrez L, Garland P, Gluck J, González I, Guermont S, Guo J, Gupta A, Herr JR, Howe A, Hyer A, Härpfer A, Irber L, Kidd R, Lin D, Lippi J, Mansour T, McA'Nulty P, McDonald E, Mizzi J, Murray KD, Nahum JR, Nanlohy K, Nederbragt AJ, Ortiz-Zuazaga H, Ory J, Pell J, Pepe-Ranney C, Russ ZN, Schwarz E, Scott C, Seaman J, Sievert S, Simpson J, Skennerton CT, Spencer J, Srinivasan R, Standage D, Stapleton JA, Steinman SR, Stein J, Taylor B, Trimble W, Wiencko HL, Wright M, Wyss B, Zhang Q, Zyme E, Brown CT. 2015. The khmer software package: enabling efficient nucleotide sequence analysis. F1000Res 4:900. <https://doi.org/10.12688/f1000research.6924.1>.
10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
11. Chun J, Rainey FA. 2014. Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*. Int J Syst Evol Microbiol 64:316–324. <https://doi.org/10.1099/ijms.0.054171-0>.