frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 December 2019
doi: 10.3389/fmicb.2019.02838

Composition and Diversity of CRISPR-Cas13a Systems in the Genus *Leptotrichia*

*Shinya Watanabe[1], Bintao Cui[1], Kotaro Kiga[1], Yoshifumi Aiba[1], Xin-Ee Tan[1], Yusuke Sato'o[1], Moriyuki Kawauchi[1], Tanit Boonsiri[1], Kanate Thitiananpakorn[1], Yusuke Taki[1], Fen-Yu Li[1], Aa Haeruman Azam[1], Yumi Nakada[2], Teppei Sasahara[1] and Longzhu Cui[1]**

[1] Division of Bacteriology, Department of Infection and Immunity, Faculty of Medicine, Jichi Medical University, Tochigi, Japan, [2] Division of Clinical Laboratory, Tottori University Hospital, Tottori, Japan

OPEN ACCESS

Edited by:
Rakesh Sharma,
Institute of Genomics and Integrative Biology (CSIR), India

Reviewed by:
Yinhua Lu,
Shanghai Normal University, China
Jose Antonio Ibarra,
National Polytechnic Institute, Mexico

***Correspondence:**
Longzhu Cui
longzhu@jichi.ac.jp

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 19 July 2019
Accepted: 22 November 2019
Published: 10 December 2019

Citation:
Watanabe S, Cui B, Kiga K, Aiba Y, Tan X-E, Sato'o Y, Kawauchi M, Boonsiri T, Thitiananpakorn K, Taki Y, Li F-Y, Azam AH, Nakada Y, Sasahara T and Cui L (2019) Composition and Diversity of CRISPR-Cas13a Systems in the Genus Leptotrichia. Front. Microbiol. 10:2838. doi: 10.3389/fmicb.2019.02838

Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas13a, previously known as CRISPR-C2c2, is the most recently identified RNA-guided RNA-targeting CRISPR-Cas system that has the unique characteristics of both targeted and collateral single-stranded RNA (ssRNA) cleavage activities. This system was first identified in *Leptotrichia shahii*. Here, the complete whole genome sequences of 11 *Leptotrichia* strains were determined and compared with 18 publicly available *Leptotrichia* genomes in regard to the composition, occurrence and diversity of the CRISPR-Cas13a, and other CRISPR-Cas systems. Various types of CRISPR-Cas systems were found to be unevenly distributed among the *Leptotrichia* genomes, including types I-B (10/29, 34.4%), II-C (1/29, 2.6%), III-A (6/29, 15.4%), III-D (6/29, 15.4%), III-like (3/29, 7.7%), and VI-A (11/29, 37.9%), while 8 strains (20.5%) had no CRISPR-Cas system at all. The Cas13a effectors were found to be highly divergent with amino acid sequence similarities ranging from 61% to 90% to that of *L. shahii*, but their collateral ssRNA cleavage activities leading to impediment of bacterial growth were conserved. CRISPR-Cas spacers represent a sequential achievement of former intruder encounters, and the retained spacers reflect the evolutionary phylogeny or relatedness of strains. Analysis of spacer contents and numbers among *Leptotrichia* species showed considerable diversity with only 4.4% of spacers (40/889) were shared by two strains. The organization and distribution of CRISPR-Cas systems (type I-VI) encoded by all registered *Leptotrichia* species revealed that effector or spacer sequences of the CRISPR-Cas systems were very divergent, and the prevalence of types I, III, and VI was almost equal. There was only one strain carrying type II, while none carried type IV or V. These results provide new insights into the characteristics and divergences of CRISPR-Cas systems among *Leptotrichia* species.

Keywords: *Leptotrichia*, CRISPR-Cas13a, clustered regularly interspaced short palindromic repeats, CRISPR-Cas, C2c2, crRNA, protospacer, self-targeting spacer

INTRODUCTION

Leptotrichia species, which are strict anaerobic or facultative anaerobic Gram-negative rods found in the oral cavity, intestines, female genital tract and urogenital system of both humans and animals, can ferment carbohydrates to generate multiple products, including lactic acid, and traces of acetic, formic or succinic acid, depending on the substrates and species (Eribe and Olsen, 2017). Some *Leptotrichia* species cause opportunistic infectious diseases, including gingivitis, periodontitis, infectious endocarditis and bacteraemia (Eribe and Olsen, 2008, 2017). With the recent progress and spread of modern techniques, the involvement of *Leptotrichia* species in various infectious diseases has gained considerable attention (Eribe and Olsen, 2017). However, the exact clinical impact of *Leptotrichia* species relevant to infectious diseases remains unclear due to difficulties in bacterial culture, isolation and identification. To date, eight *Leptotrichia* species have been characterized, which include *L. buccalis, L. goodfellowii, L. hofstadii, L. shahii, L. trevisanii, L. massiliensis, L. hongkongensis*, *and L. wadei.* However, many clinical *Leptotrichia* isolates have remained unclassified due to the lack of the whole-genome sequences and biochemical properties of these isolates (Eribe and Olsen, 2017). Currently, *Leptotrichia* genome data are very limited as there are only one complete whole genome sequences of species-identified strain and three unidentified strains available in the GenBank database (Ivanova et al., 2009; Gupta and Sethi, 2014).

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated genes (*cas*) encode proteins involved in the adaptive immune response against archaea and bacteria by preventing an invasion of the host cell by foreign genetic elements, such as bacteriophages and plasmids (Marraffini, 2015). The CRISPR-Cas systems identified thus far show extreme diversity in *cas* gene compositions as well as genomic loci architecture. Despite this diversity, the latest classification system categorizes the CRISPR-Cas systems into two classes based on basic compositions and structures: class 1 systems, consisting of types I and III as well as the putative type IV, which possess multi-subunit effector complexes comprised of multiple Cas proteins, and class 2 systems, which are comprised of types II, V, and VI that are characterized by effector complexes consisting of a single, large Cas protein (Makarova et al., 2018; Shmakov et al., 2018). CRISPR-Cas13a (known previously as CRISPR-C2c2) is the most recently identified CRISPR-Cas system belonging to the type VI class 2 system and is characterized by cleavage activity of a single RNA-guided single-stranded RNA (ssRNA) molecule (Shmakov et al., 2015). An interesting finding recently made by Abudayyeh et al. (2016) is that CRISPR-Cas13a of *L. shahii* can perform promiscuous collateral ssRNA cleavage. The same research group and others recently demonstrated that the CRISPR-Cas13a system is an efficient molecular tool for the cleavage and editing of target RNA in both prokaryotic and mammalian cells (Abudayyeh et al., 2016, 2017; Cox et al., 2017), with highly sensitive detection of nucleic acids (East-Seletsky et al., 2017; Gootenberg et al., 2017).

The genus *Leptotrichia* is the main source of the CRISPR-Cas13a system (Shmakov et al., 2015). While the *L. shahii* CRISPR-Cas13a system has been intensively studied, the variations in the genetic features of CRISPR-Cas systems demonstrated so far have prompted us to carry out this comprehensive study of the occurrence, composition and diversity of CRISPR-Cas13a, and other CRISPR-Cas systems in the genus *Leptotrichia*.

MATERIALS AND METHODS

Leptotrichia Strains and Growth Conditions

Three *Leptotrichia trevisanii* strains, JMUB3870, JMUB3935 and JMUB4039, two *L. wadei* strains, JMUB3933 and JMUB3934 and one *L. hongkongensis* strain, JMUB5056, were isolated between 2012 and 2019 from blood cultures of patients in Japan. *Leptotrichia* strain JMUB3936 was isolated in 2017 from the bronchial lavage fluid of a patient in Japan. The following type strains of *Leptotrichia* species were obtained from the Japan Collection of Microorganisms (JCM): *L. goodfellowii* JCM16774, *L. hofstadii* JCM16775, *L. shahii* JCM16776 and *L. wadei* JCM16777. *Leptotrichia* strains were stored in Gifu Anaerobic Medium (GAM broth; Nissui Pharmaceutical Co., Ltd., Tokyo, Japan) supplemented with 40% glycerol at −80°C, and were recovered for the present study on various agar plates, of which isolates JCM12969, JCM16774, JMUB3933, JMUB3934, JMUB3935, JMUB4039, and JMUB3870 were inoculated on GAM agar, isolates JCM16775, JCM16776, JCM16777, and JMUB3936 were recovered on sheep blood agar (Kohjin Bio Co., Ltd., Saitama, Japan), and isolates JMUB5056 was grown on Anaero Columbia Agar with Rabbit Blood (Nippon Becton Dickinson Company, Ltd., Tokyo, Japan). All plates were incubated anaerobically under conditions of 5% CO_2 with an oxygen absorbent (Anaeropack Kenki, Mitsubishi Gas Chemical Company, Tokyo, Japan) at 37°C for 48–72 h. *Escherichia coli* M1601 cells were grown in Luria-Bertani (LB) medium (Nippon Becton Dickinson Company, Ltd., Tokyo, Japan) with shaking or on LB agar with antibiotic supplementation (chloramphenicol at 10 μg/mL; kanamycin at 50 μg/mL) as needed for plasmid maintenance.

Whole-Genome Sequencing and Sequence Read Assembly

Bacterial genomic DNA was extracted and purified using the NucleoBond AXG kit (Takara Bio, Inc., Otsu, Shiga, Japan). DNA library preparation was performed using a Rapid Barcoding sequencing kit (SQK-RBK004; Oxford Nanopore Technologies, Oxford, United Kingdom). Sequencing was then performed using a MinION Mk-1B device integrated with a FLO-MIN-106 flow cell (Oxford Nanopore Technologies). Base calling was carried out using MinKNOW software (Oxford Nanopore Technologies) and the Porechop tool

(version 0.2.4[1]) was used for demultiplexing of barcoded libraries. The resulting sequence reads were assembled with the Canu tool (version 1.8[2]) (Koren et al., 2017). The assembled genomes were then circularized using the Circlator tool (version 1.5.5[3]) (Hunt et al., 2015), and all of the resulting assemblies that produced contiguous sequences were polished with the nanopolish algorithm (version 0.11.0[4]) (Loman et al., 2015). Genome sequencing of all of the above strains was performed using the MiSeq platform as previously described (Watanabe et al., 2016, 2018). The draft-genomes generated by MinION were then polished again with the reads generated by MiSeq with the Pilon automated genome assembly improvement and variant detection tool (version 1.22; Broad Institute, Cambridge, MA, United States) (Walker et al., 2014). The resulting genome sequences were further trimmed using the CLC genomics workbench (Qiagen, Hilden, Germany).

Gene Annotation and Genome Analysis

Gene extraction and annotation were performed with the Microbial Genome Annotation Pipeline[5] or Prokka (version 1.13.3[6]) (Seemann, 2014). CRISPR-Cas genes were extracted and annotated using the CRISPRCasFinder program[7] (Couvin et al., 2018). Prophages were identified by PHASTER[8] (Arndt et al., 2016). The kSNP trees were constructed with the kSNP alignment-free sequence analysis tool (version 3.021) (Gardner et al., 2015) and then visualized with the FigTree graphical viewer of phylogenetic trees (version 1.4.3[9]). Whole-genome average nucleotide identity (ANI) was calculated by FastANI and the heat map was depicted by ggplot2 of R (Jain et al., 2018). Pairwise blastn analysis for CRISPR-*cas* loci was carried out using the Easyfig application (version 2.2.2[10]) (Sullivan et al., 2011).

Plasmid Construction

To examine the collateral growth inhibition activities of various Cas13a loci of *Leptotrichia* species, two sets of vectors were constructed. One set was for the expression of CRISPR-Cas13a and the other for the generation of their target RNA. The original plasmid pC003 containing a *L. shahii* Cas13a locus. First, the native Cas13a loci including native direct repeats and spacers were isolated from the genome of each *Leptotrichia* strain by polymerase chain reaction (PCR) and then cloned into the pC003 vector by replacing the *shahii*Cas13a locus using the In-Fusion HD Cloning Kit (Takara Bio, Inc., Shiga, Japan), and then kanamycin resistant gene of the vector was replaced by kanamycin-resistant gene (*aphA-3*) to generate pKLC-Cas13a. The pC003-LshC2C2 locus included in the plasmid pACYC184 for spacer cloning was a gift from Dr. Feng Zhang (Addgene plasmid #79152[11]; RRID:Addgene_79152) (Abudayyeh et al., 2016).

To generate arabinose-inducible target RNA expression vector, firstly we constructed arabinose-inducible RFP expression vector (pKLC-RFP) from pC008-pBR322 by replacing tetracycline-inducible element (pBAD33) and ampicillin resistant gene (*bla*) to arabinose-inducible element and chloramphenicol acetyltransferase (CAT). Then the target sequence of the native Cas13a was inserted into just behind the RFP coding sequence. A native spacer sequence from each *Leptotrichia* strain and the protospacer adjacent motif, nucleotide C in this study, were adapted to the 5′ end of the reverse vector primer. The final plasmid backbone was generated by PCR followed by self-ligation using the In-Fusion HD Cloning Kit. All of the generated plasmids were confirmed by Sanger sequencing. The pC008-pBR322 with *tetR*-inducible RFP was a gift from Dr. Feng Zhang (Addgene plasmid # 79157[12]; RRID:Addgene_79157). The primers used in this study are listed in **Supplementary Table S1**.

Cas13a Collateral Growth Inhibition Assay

The assay for *E. coli* growth inhibition as a result of collateral ssRNA cleavage by Cas13a was carried out based on the previous paper (Abudayyeh et al., 2016). The constructed plasmids were chemically transformed into competent *E. coli* MC1601 cells. Briefly, 500 ng of CRISPR-Cas13a expression plasmid and 100 ng of arabinose-inducible target plasmid were mixed and added to the competent cells obtained above. Then, the transformation reaction was incubated on ice for 30 min, followed by heat shock at 42°C in a heat block for exactly 45 s, and then immediately transferred to tubes containing 1 mL of LB broth at room temperature, which were cooled on ice. Afterward, the samples were incubated at 37°C for 1 h to induce antibiotic resistance and then plated on LB agar plates containing kanamycin and chloramphenicol at 37°C until colonies appeared. Then, the colonies were transferred to the LB medium containing kanamycin and chloramphenicol and incubated with gentle shaking until early stationary phase. The optical density at 600 nm (OD 600) of the bacterial culture was adjusted to 0.1 with LB medium containing kanamycin and chloramphenicol, and then 100 μl of the medium were transferred to L-shaped shaking tubes containing 10 ml of LB medium with kanamycin, chloramphenicol and 0.2% L-arabinose to induce target gene expression. During the incubation at 25 rpm at 37°C, the optical density at 600 nm of the bacterial culture were measured every 5 min by TVS126MB (ADVANTEC Toyo Kaisha, Ltd.) for at least 12 h. The Cas13a collateral growth inhibition assay was performed in biological triplicate.

Accession of the Genome Sequence

The genome sequences were deposited to the DNA Data Bank of Japan (DDBJ, Bioproject accession number PRJDB7856) and the

[1]https://github.com/rrwick/Porechop
[2]http://canu.readthedocs.org/
[3]http://sanger-pathogens.github.io/circlator/
[4]https://github.com/jts/nanopolish
[5]http://www.migap.org
[6]https://github.com/tseemann/prokka
[7]https://crisprcas.i2bc.paris-saclay.fr/
[8]http://phaster.ca/
[9]http://tree.bio.ed.ac.uk/software/figtree/
[10]http://mjsull.github.io/Easyfig/
[11]http://n2t.net/addgene:79152
[12]http://n2t.net/addgene:79157

sequence raw data were deposited in the DDBJ Sequence Read Archive (DRA007809).

RESULTS AND DISCUSSION

Comparative Analysis of *Leptotrichia* Genome

CRISPR-Cas13a systems are broadly distributed among *Leptotrichia* species. The estimation of the occurrence, composition, and diversity of CRISPR-Cas13a systems in the genus *Leptotrichia* is particularly challenging because the complete whole-genome information is limited. At present, the genus *Leptotrichia* is composed of eight species. However, the complete whole genome sequences are not available except for *L. buccalis* (Ivanova et al., 2009) and three unidentified *Leptotrichia* strains. In this study, 11 strains of six *Leptotrichia* species were collected, which included *L. goodfellowii*, *L. hofstadii*, *L. hongkongensis, L. shahii*, *L. trevisanii* and *L. wadei*, and their complete whole-gnome sequences were determined (**Supplementary Table S2**). As it was not possible to collect isolates of the remaining species, a draft genome of *L. massiliensis* available from a public database was used. Taken together with the additional draft genome sequences retrieved from the public database, a total of 29 whole-genome sequences were used for comparative genomic analysis. All genomes were surveyed and investigated for the composition, occurrence and diversity of the CRISPR-Cas13a and other CRISPR-Cas systems. The overall features of the genomes, the distribution of the CRISPR-Cas systems, and the numbers of CRISPR-Cas spacers, plasmids and prophages for each strain are summarized in **Table 1**. As shown in **Table 1**, the chromosome size of the genus *Leptotrichia* varies from 2,142,946 to 2,829,322 bp with GC contents of 42% to 48%. Virulence factors were searched against all genomes using Prokka (version 1.13.3), but no class of genes was found to contribute to bacterial pathogenesis.

To depict the relationship among *Leptotrichia* species and strains, a whole-genome single-nucleotide polymorphism (SNP)-based phylogenetic tree was constructed of a total of 29 *Leptotrichia* genomes, which included the complete whole genome-sequencing data determined in the present study. As shown in **Figure 1**, four sequenced *Leptotrichia* type strains (i.e., *L. goodfellowii* JCM16774, *L. hofstadii* JCM16775, *L. shahii* JCM16776, and *L. wadei* JCM16777) belonged to the same clade of the corresponding strains registered as The Department of Energy (DOE) draft-genomes (i.e., *L. goodfellowii* DSM 19756, *L. hofstadii* DSM 21651, *L. shahii* DSM 19757, and *L. wadei* DSM 19758). The analysis also revealed that except for strain JMUB3936 which was originally classified into *L. wadei* by the biochemical property in a clinical microbiology laboratory, all strains were classified into corresponding clades. To further elucidate the species classification of the *Leptotrichia* strains, we performed the whole-genome ANI analysis showing the average nucleotide identities between all orthologous genes shared between the two genomes (**Figure 2** and **Supplementary Table S3**). ANI revealed that the 29 *Leptotrichia* genomes were classified into 14 species including six unclassified species because bacterial strains belonging to the same species typically showing ≥95% ANI among themselves (Konstantinidis and Tiedje, 2005; Goris et al., 2007). Based on the ANI analysis, strain JMUB3936 was classified into new species because the highest ANI was 92.2% (*L. wadei* JMUB3933). Strain JMUB3936 was located between *L. wadei* and *L. shahii*. In addition, 16S rRNA sequence analysis of strain JMUB3936 showed 97.6% and 96.1% identities to those of *L. wadei* type strain DSM19758 and *L. shahii* type strain DSM19757, supporting that JMUB3936 must be classified into a new species of genus Leptotrichia. 16S rRNA sequence analysis is one of the gold standards classifying species and genus, and two isolates having 16S rRNA sequence similarities of ≤99% can generally be classified into different species in the same genus (Clarridge, 2004). Furthermore, JMUB3936 can be considered to be a member of human oral or skin microbiota, since all but one out of 62 16S rRNA sequences came from samples of human oral or skin microbiota deposited in NCBI database (Munson et al., 2004; Täubel et al., 2009; Bik et al., 2010; Bolivar et al., 2012; Kong et al., 2012) had high similarities to that of JMUB3936 (≥99% identity with ≥80% coverage).

To date, eight species of Leptotrichia have been precisely described (Eribe et al., 2004; Woo et al., 2010). Apart from the eight Leptotrichia species, previous analysis based on the 16S rRNA gene identified 10 unclassified Leptotrichia species, indicating that some Leptotrichia strains have not yet being assigned a species name (Eribe and Olsen, 2008, 2017). Species classification based on both phenotype and genotype is an essential starting point to understand the biology, ecology and clinical impacts of Leptotrichia species.

CRISPR-Cas Systems in *Leptotrichia*

To investigate the composition, occurrence and diversity of CRISPR-Cas systems among *Leptotrichia* species, all 29 *Leptotrichia* genomes were analyzed for the presence of CRISPRs and *cas* genes with the use of the CRISPRCasFinder program, which enables the identification of direct repeat consensus boundaries and the extraction of related spacers and *cas* genes (Couvin et al., 2018). This analysis identified a total of 38 CRISPR-Cas systems: 10 (26.3%) type I-B, 1 (2.6%) type II-C, 6 (15.8%) type III-A, 6 (15.8%) type III-D, 3 (7.9%) type III-like and 12 (31.6%) type VI-A, while no type IV and V systems were detected. Of the 29 strains, 8 (27.6%) had no CRISPR-Cas system (**Table 1**). Hence, we concluded that types I, III, and VI were the most dominant CRISPR-Cas systems among *Leptotrichia* species.

Illustrations of all CRISPR-Cas systems identified in this study are summarized in **Figure 3**. Type I-B CRISPR-Cas systems were distributed in five known and one unclassified *Leptotrichia* species (**Figure 3A**). It is well known that CRISPR mediates adaptive immunity according to the following three steps: (1) acquisition of short segments of foreign nucleic acids into CRISPR loci (adaptation); (2) crRNA expression and maturation (crRNA-processing); (3) target interference by recognition of the crRNA and destruction of the target nucleic acid (interference) (Bhaya et al., 2011). The *Leptotrichia* type I-B CRISPR-*cas* loci were found to include full gene sets responsible for the above

TABLE 1 | Genome and CRISPR-Cas system information of genus *Leptotrichia*.

Genome infromation								CRISPR/Cas class						
Species	**Strain**	**Genome sequencing type[b]**	**Chromosome length (bp)**	**GC%**	**GenBank accession no.**	**Number of plasmid**	**Number of prophage/ genomic island**	**I**	**II**	**III**	**IV**	**V**	**VI**	**Number of spacer[c]**
L. wadei	KA00185	Draft genome	n/a	n/a	GCA_001553045.1	n/a	n/a	–	–	–	–	–	–	5
L. wadei	JMUB3933	Complete genome	2361227	42.03%	AP019834	0	2	I-B	–	–	–	–	VI-AI, VI-A2	47
L. wadei	JMUB3934	Complete genome	2414633	42.03%	AP019835	4	2	–	–	–	–	–	VI-A	1
L. wadei	F0279	Draft genome	n/a	n/a	GCA_000469405.1	n/a	n/a	–	–	–	–	–	VI-A	7
L. wadei	JCM16777	Complete genome	2305932	41.74%	AP019829	1	0	–	–	–	–	–	–	3
L. wadei	DSM 19758	Draft genome	n/a	n/a	GCA_000373345.1	n/a	n/a	–	–	–	–	–	–	3
Leptotrichia sp.-I	JMUB3936	Complete genome	2335974	42.99%	AP019841	3	1	–	–	III-AI, III-A2	–	–	–	22
L. shahii	DSM 19757	Draft genome	n/a	n/a	GCA_000373045.1	n/a	n/a	I-B	–	III-A	–	–	VI-A	51
L. shahii	JCM16776	Complete genome	2142946	42.34%	AP019827	1	1	I-B	–	III-A	–	–	VI-A	63
Leptotrichia sp.-2	F0557	Draft genome	n/a	n/a	GCA_000469385.1	n/a	n/a	–	–	–	–	–	VI-A	6
L. hongkongensis	JMUB5056	Complete genome	2261073	29.86%	AP019846	1	1	–	–	–	–	–	–	11
L. massiliensis	P3007	Draft genome	n/a	n/a	GCA_900104625.1	n/a	n/a	–	–	–	–	–	VI-AI, VI-A2	21
L. massiliensis	F0581	Draft genome	n/a	n/a	GCA_000469525.1	n/a	n/a	–	–	–	–	–	VI-A	10
L. hofstadii	JCM16775	Complete genome	2548198	44.14%	AP019823	3	1	–	–	III-A	–	–	–	10
L. hofstadii	DSM 21651	Draft genome	n/a	n/a	GCA_000428965.1	n/a	n/a	–	–	III-A	–	–	–	6
L. hofstadii	F0254	Draft genome	n/a	n/a	GCA_000162955.1	n/a	n/a	–	–	–	–	–	–	0
L. trevisanii	DSM 22070	Draft genome	n/a	n/a	GCA_000482505.1	n/a	n/a	I-B	–	III-D	–	–	–	17
L. trevisanii	JMUB3870	Complete genome	2829322	44.09%	AP019831	2	1	I-B	–	III-D	–	–	–	78
L. trevisanii	JMUB4039	Complete genome	2685755	44.45%	AP019845	0	1	I-B	–	III-D	–	–	VI-A	56
L. trevisanii	JMUB3935	Complete genome	2729392	44.18%	AP019840	0	2	–	–	III-D	–	–	–	14
L. buccalis	C-1013-b	Complete genome	2465610	42.14%	GCA_000023905.1	0	0	I-B	–	III-D	–	–	VI-A	102
Leptotrichia sp.-3	F0260	Complete genome	2194935	42.43%	GCA_001553645.1	0	2	–	–	–	–	–	–	1
Leptotrichia sp.-4	bin_23	Draft genome	n/a	n/a	GCA_003638725.1	n/a	n/a	–	–	–	–	–	–	0
Leptotrichia sp.-4	F0590	Complete genome	2152181	42.09%	GCA_002240055.1	0	1	–	–	III-D	–	–	–	37
L. goodfellowii	F0264	Draft genome	n/a	n/a	GCA_000176335.1	n/a	n/a	–	–	–	–	–	–	11
L. goodfellowii	JCM16774	Complete genome	2290729	46.38%	AP019822	0	1	I-B	–	III-like	–	–	–	38
L. goodfellowii	DSM 19756	Draft genome	n/a	n/a	GCA_000516535.1	n/a	n/a	I-B	–	III-like	–	–	–	39
Leptotrichia sp.-S	W9775	Draft genome	n/a	n/a	GCA_000469505.1	n/a	n/a	–	II-C	–	–	–	–	19
Leptotrichia sp.-6	W10393	Complete genome	2444904	45.80%	GCA_001274535.1	0	0	I-B	–	III-like	–	–	–	201

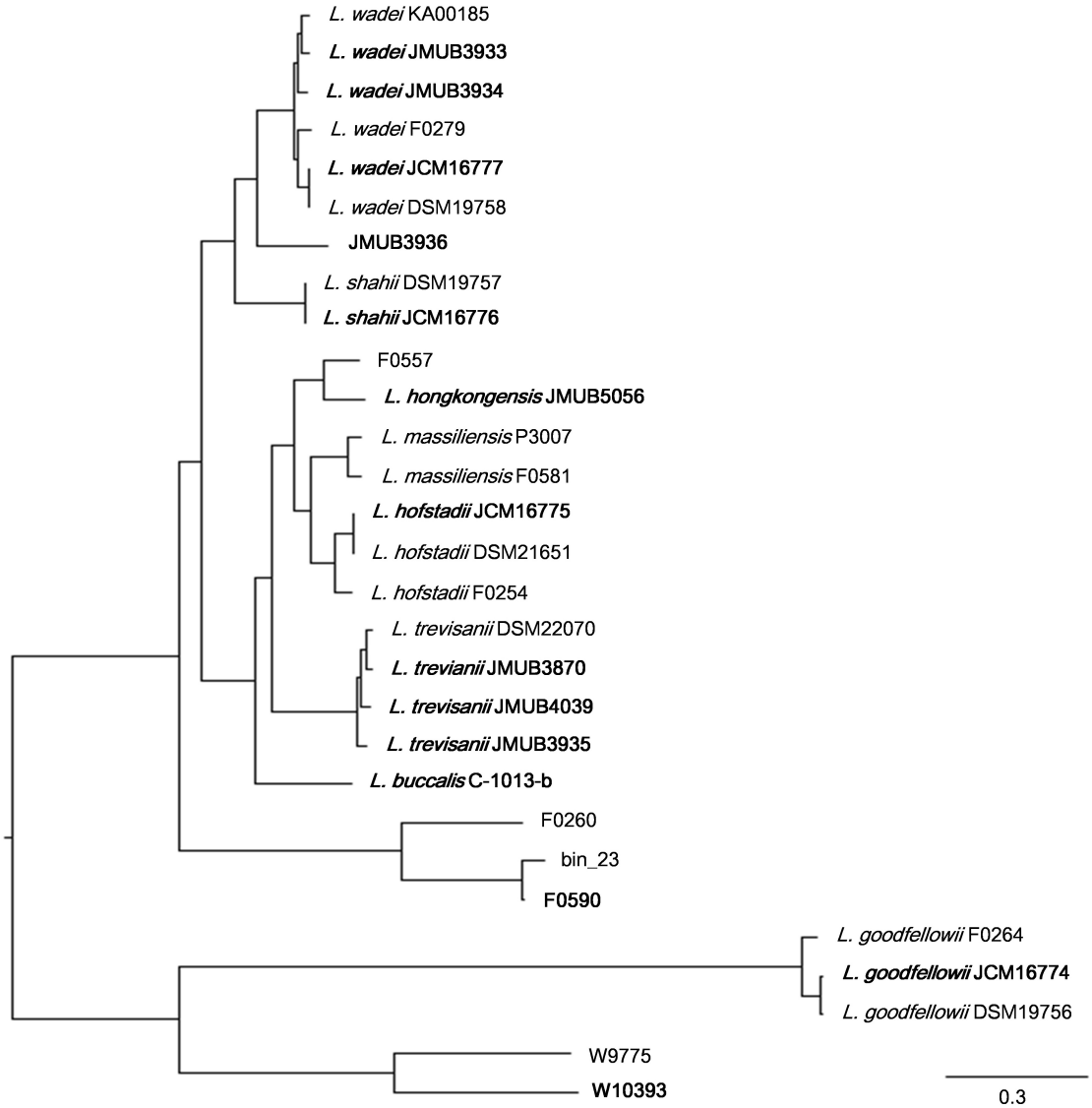


FIGURE 1 | Whole genome SNP-based majority phylogenetic tree of 29 *Leptotrichia* strains generated by kSNA3.0. Horizontal branch lengths indicate changes per number of SNPs. Strains in boldface are the strains their whole genomes were determined in this study. It should be noted that this is a SNPs-based (present in at least 75% strains), not an alignment based tree and no evolutionary direction can be inferred from this tree.

three steps: *cas1*, *cas2* and *cas4* for adaptation, *cas6* for crRNA-processing and *cas3*, *cas5*, *cas7* and *cas8* for interference. All CRISPRs of type I-B CRISPR-Cas systems found in this study were located downstream of the *cas2* gene and/or upstream of the *cas6* gene, which was similar with a recent report of the anaerobic rod *Clostridium thermocellum* by Zöphel et al.[13] The type II-C CRISPR-*cas* system was found in only one *Leptotrichia* strain, W9775 (**Figure 3B**). Type II-C is a class 2 CRISPR-Cas system, and the locus contains a single effector gene (*cas9*) and three adaptation genes (*cas1*, *cas2,* and *csn2*). CRISPR was located downstream of *csn2*. Type III systems have been classified into four main subtypes (III-A to D) (Vestergaard et al., 2014; Makarova et al., 2015). Among the *Leptotrichia* genomes analyzed in this study, six type III-A, six type III-D and three type III-like loci were identified (**Figure 3C**). Type III systems are reported to frequently carry accessory genes either within or immediately bordering the core *cas* gene clusters (Haft et al., 2005; Vestergaard et al., 2014; Shah et al., 2019). The type III-A systems of *Leptotrichia* can be classified into two subtypes: III-A-Lep1 and III-A-Lep2. The III-A-Lep1 loci were found in the genomes of *L. hofstadii* strain JCM16775 and strain JMUB3936, which carried the full set of *cas* genes involving the adaptation, crRNA-processing and interference steps. However, the type III-A-Lep2 system found in the genomes of *L. shahii* strain JCM16776 and strain JMUB3936 did not carry the *csm6* gene, which encodes an HEPN family ribonuclease that

[13]https://archiv.ub.uni-marburg.de/diss/z2015/0381

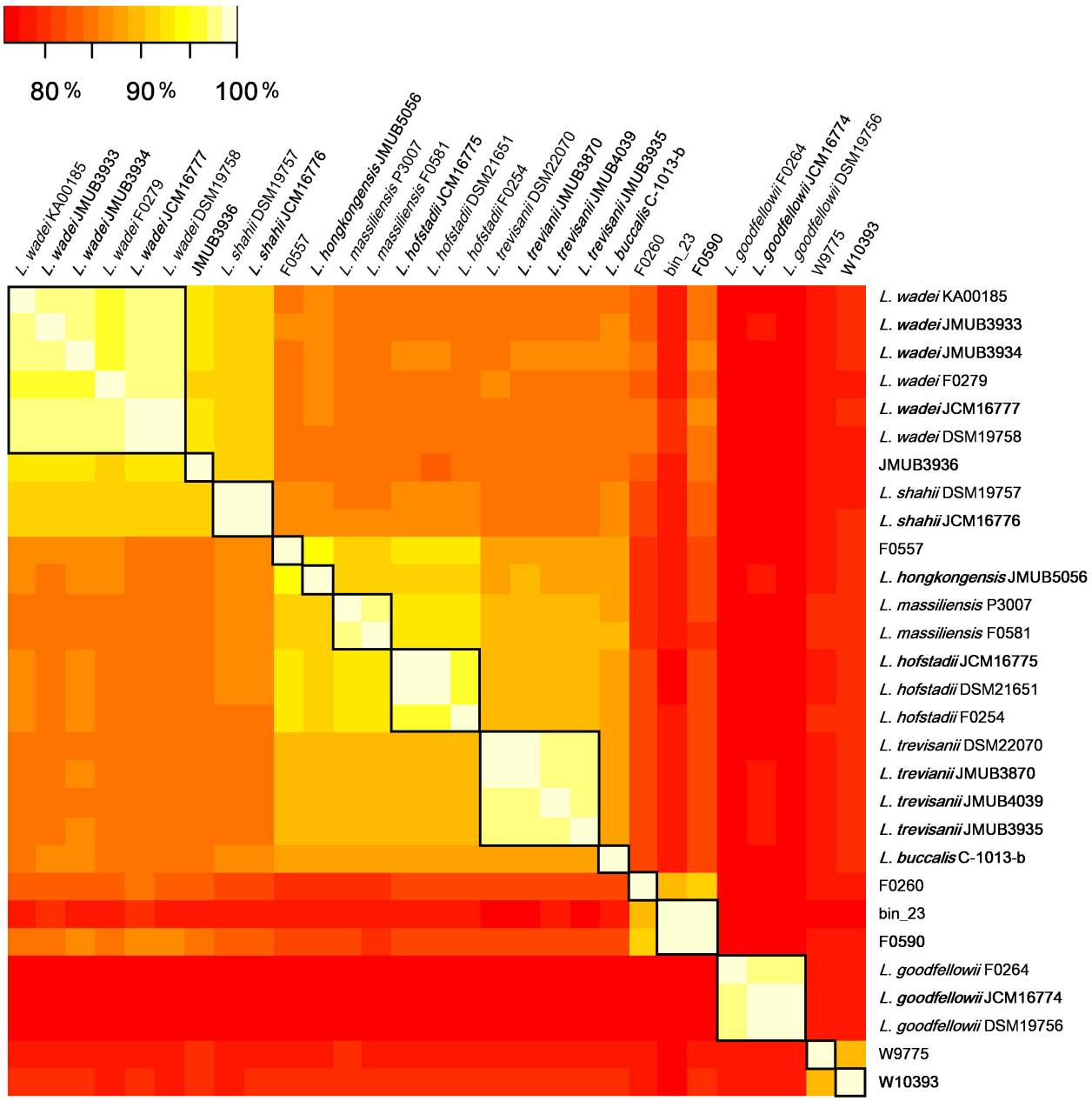


FIGURE 2 | The whole-genome average nucleotide identity (ANI) analysis of 29 *Leptotrichia* genomes generated by FastANI. The genetic relatedness between two genomes was estimated by ANI analysis. ≥95% ANI, which generally indicates genomes belonging to the same species, was highlighted by square in black.

mediates sequence-unspecific ssRNA cleavage and contributes to anti-plasmid immunity (Niewoehner and Jinek, 2016; Foster et al., 2019). Instead of *csm6*, a *csx1* gene, which encodes a Csm6 homologous protein and is known to be involved in ssRNA disruption (Sheppard et al., 2016; Han et al., 2017), was identified upstream of the core *cas* genes in the type III-A-Lep2 loci.

Leptotrichia type III-D CRISPR-Cas systems were identified in four *L. trevisanii* strains, one *L. buccalis* strain and one unclassified *Leptotrichia* strain, F0590 (**Figure 3C**). The genomes of *L. buccalis* strains C-1031-b and F0590, had large regions of the type III-D loci, which included core *cas* genes, *csx1*, CRISPRs and putative accessory genes. Although *L. trevisanii* strains also have type III-D gene clusters similar to those of *L. buccalis* C-1031-b and *L. sp.* F0590, the type III-D loci were split into two clusters. In *L. trevisanii* genomes, the core *cas* genes involved in crRNA-processing and interference steps were located more than 200 kb from the gene cluster of the accessory gene *csx1*, adaptation genes *cas4*, *cas2*, and *cas1* and CRISPRs. We also found type III-like CRISPR-*cas* loci, which include *csm3* genes and CRISPRs in the

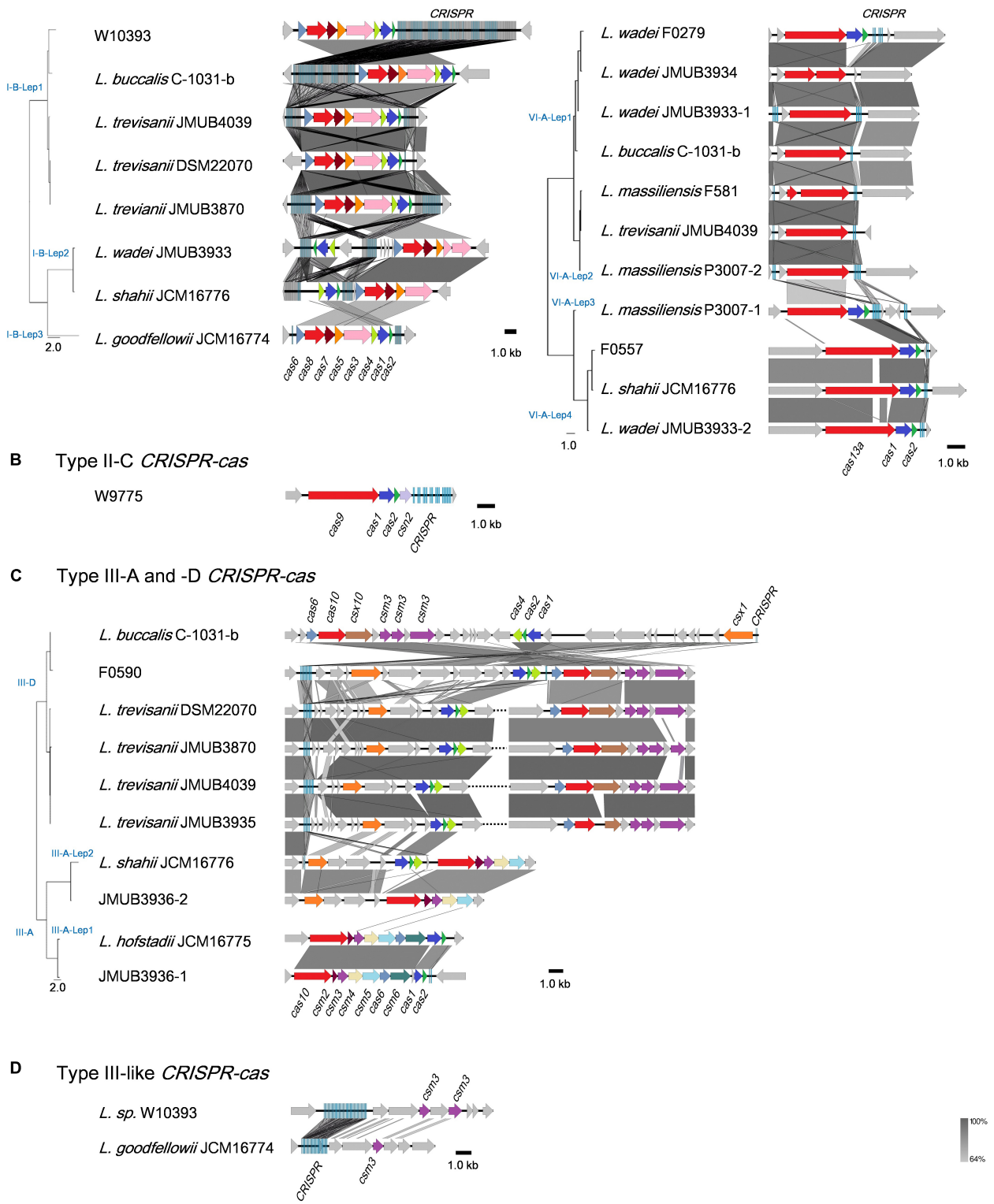


FIGURE 3 | Schematic representations of **(A)** Type I-B, **(B)** Type II-C, **(C)** Type III-A and -D, **(D)** Type III-like, and **(E)** Type VI-A of the CRISPR-Cas systems in *Leptotrichia* genomes. Genetic elements of CRISPR-Cas systems identified in this study are arranged according to their relative position in the chromosome. Different color arrows indicate different categories of *cas* genes: red, genes of *cas8*, *cas9*, *cas10*, and *cas13a*; blue, *cas1*; pink, *cas3*; yellow, *cas4*; orange, *cas5* and *csx1*; cobalt blue, *cas6*; brown, *cas7* and *csx10*; dark purple, *csm2*; purple, *csm3*; light yellow, *csm4*; light blue, *csm5*; deep green, *csm6*; gray, other genes. Left panels at the strain names represent phylogenetic parsimony trees calculated and drawn by kSNP3 and FigTree, respectively, based on DNA sequences covering all CRISPR/cas loci. Note that the regions with homologies of 64% or more show in gray bar.

genomes of *L. goodfellowii* strain JCM16774 and *L. sp.* W10393 (**Figure 3D**). However, neither of these two genomes carried type III *cas* genes.

The present analysis revealed that the type VI-A CRISPR-*cas* system, which carries the single effector protein Cas13a (formerly C2c2), is broadly distributed throughout *Leptotrichia* species (**Table 1**). Eleven type VI-A CRISPR-*cas* loci were identified in five known and one unclassified *Leptotrichia* species. As shown in **Figure 3E**, the type VI-A loci of *Leptotrichia* were further classified into four subtypes, VI-A-Lept1 to VI-A-Lept4. Cas13a is a crRNA-guided RNA-targeting effector, which has two distinct RNase motifs (Shmakov et al., 2015; O'Connell, 2019). In addition to the single RNA-guided ssRNA cleavage activity, once activated by the cognate target RNA, Cas13a becomes a promiscuous RNase that can non-sequence-specifically cleave host cell RNA (namely collateral RNA cleavage), leading to host cell death or cellular dormancy (Abudayyeh et al., 2016; East-Seletsky et al., 2016). Previously, Cas13a loci were identified and characterized in *L. wadei* F0279, *L. buccalis* C-1031-b and *L. shahii* DSM 19757, which were named as LwaCas13a, LbuCas13a and LshCas13a, respectively (Abudayyeh et al., 2016, 2017; East-Seletsky et al., 2016). In addition, Cas13a loci were identified in three *Leptotrichia* species in the current study, *L. massiliensis, L. trevisanii* and the unclassified *Leptotrichia* strain F0557. Genomic analysis showed that, except for the cases of *L. wadei* JMUB3934 and *L. massiliensis* F581, of which *cas13a* genes were truncated, all of the VI-A CRISPR Cas13a systems identified in this study were thought to have a biological function, as based on the gene structures of the three characterized systems. Moreover, the CRISPR-associated genes *cas1* and *cas2*, encoding an adaptation module of CRISPR-Cas13a, were usually located downstream of Cas13a in subtypes VI-A-Lep3 and VI-A-Lep4. This was also the case for the VI-A-Lep1 CRISPR-*cas* system of *L. wadei* F0279. It has been suggested that the Cas13a of VI-A-Lep3 and 4, including *L. shahii* JCM16776 and *L. wadei* JMUB3933, may be more efficient than the other subtypes because the *cas13a*, *cas1* and *cas2* genes make up an operon

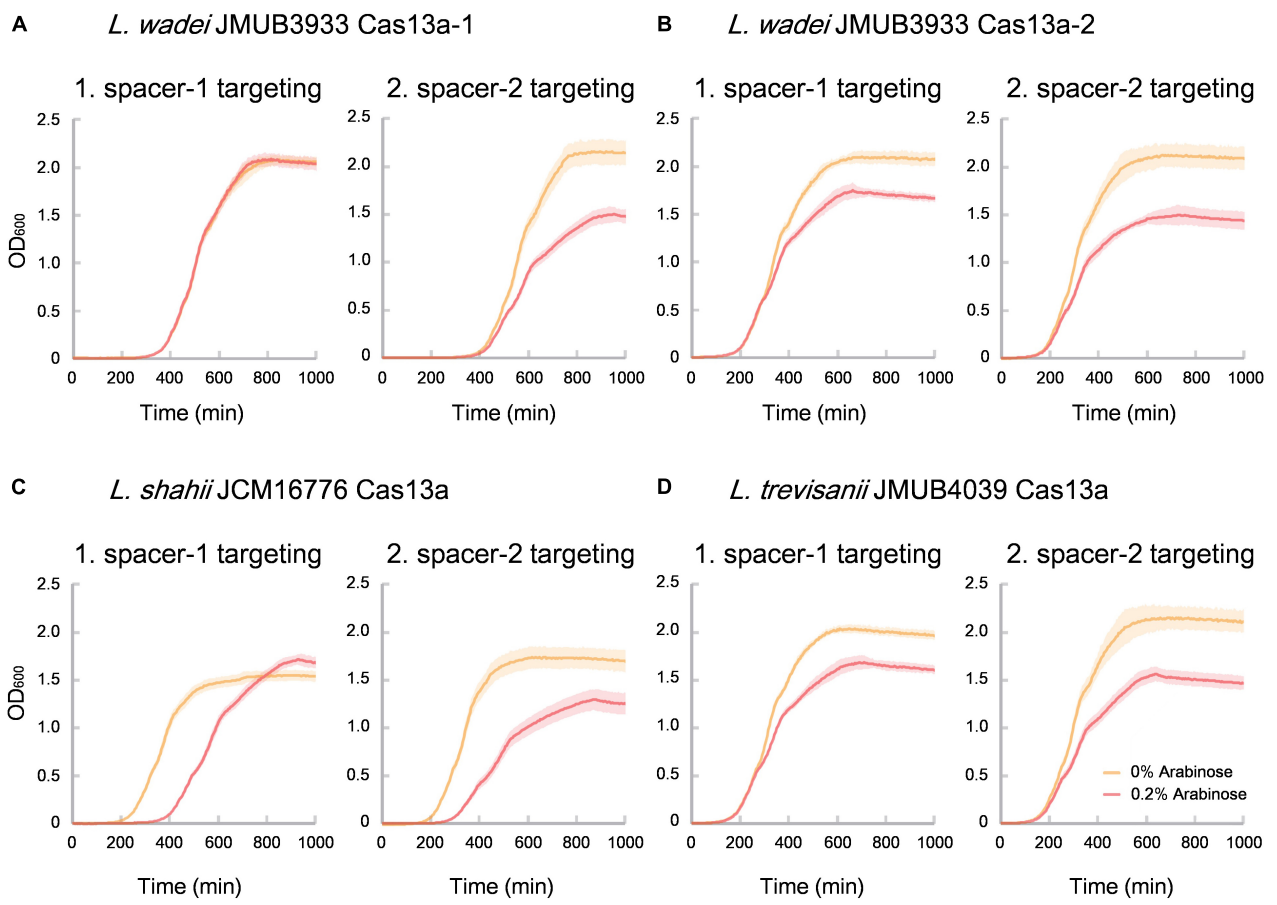


FIGURE 4 | Collateral growth inhibitory activities of **(A)** LwaCas13a-1, **(B)** LwaCas13a-2, **(C)** LshCas13a, and **(D)** LtrCas13a. Native CRISPR-Cas13a expression vector carrying kanamycin-resistant gene and arabinose-inducible target RNA (protospacer) expression vector which carries chloramphenicol-resistant gene were co-transformed into *E. coli* MC1061 and incubated at 37°C on LB plates containing chloramphenicol and kanamycin until colonies appeared. The independent colonies were transferred to LB medium containing chloramphenicol, kanamycin and 0.2% arabinose (inducer of target RNAs) and incubated with gentle shaking at 25 rpm at 37°C. Optical density at 600 nm (OD 600) were measured every 5 min and the growth curves were obtained. Spacer sequences used here did not match against the *E. coli* genome with >80% identity.

with a structure that usually provides good spatial regulation during compensation.

Three strains of *Leptotrichia* have multiple loci of same types of CRISPR-Cas systems. Strain JMUB3933 and *L. massiliensis* P3007 carried two loci of the type VI-A system as well as *L. wadei* JMUB3936 had two loci of the type III-A system. In all cases of the three strains, only one of the two loci possessed *cas1* and *cas2* genes encoding spacer acquisition module. Therefore, it is suggested that the multiple CRISPR-Cas systems share the adaptation modules to acquire a spacer sequence into each CRISPR array. Such communal modules and CRISPR arrays are also observed in co-habiting type I systems (Makarova et al., 2011). Moreover, *Leptotrichia* strains frequently possess multiple CRISPR-Cas systems of different types. Our genome analysis revealed that ten of twenty nine *Leptotrichia* species carry more than two of types VI-A, I-B,

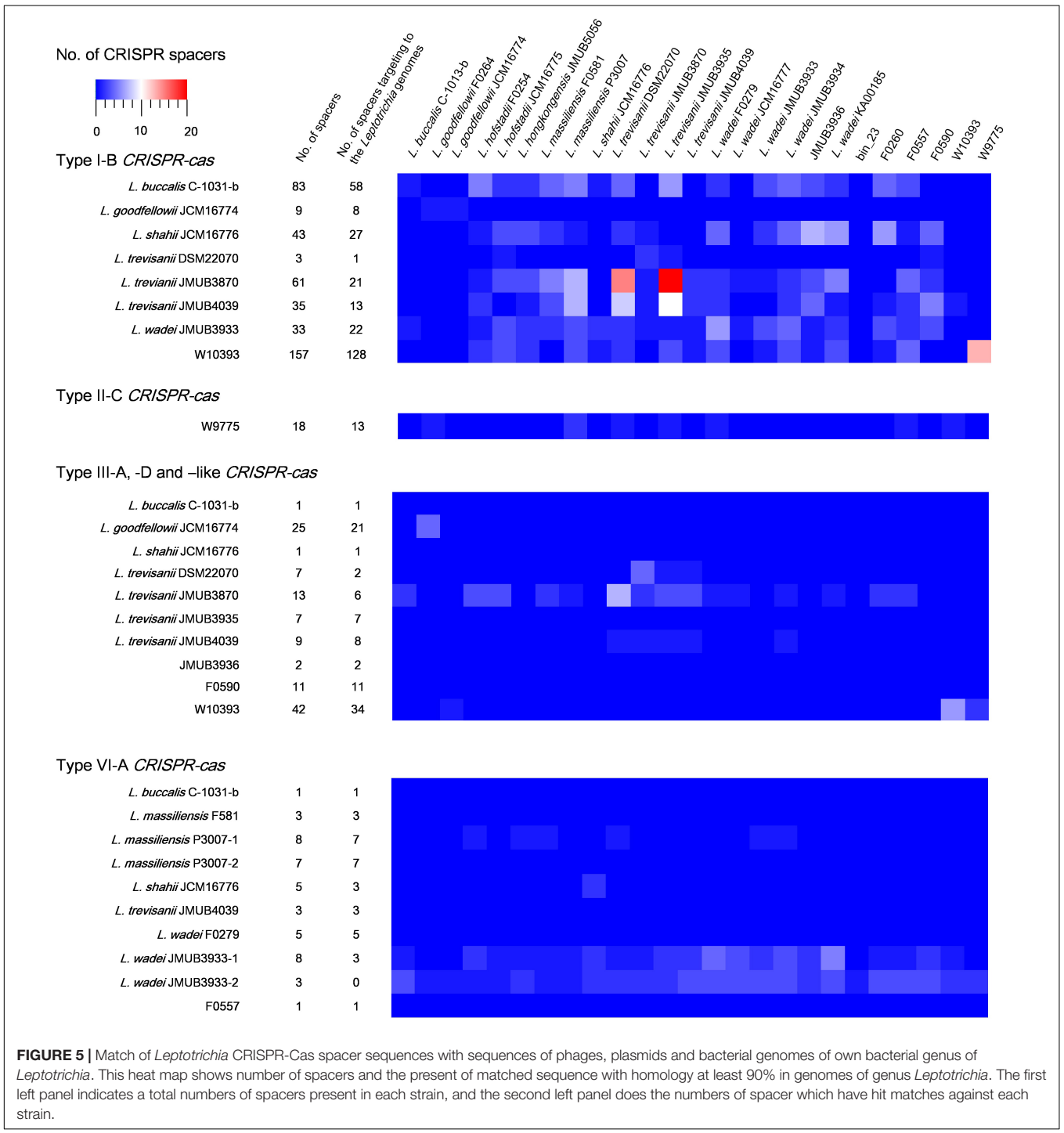


FIGURE 5 | Match of *Leptotrichia* CRISPR-Cas spacer sequences with sequences of phages, plasmids and bacterial genomes of own bacterial genus of *Leptotrichia*. This heat map shows number of spacers and the present of matched sequence with homology at least 90% in genomes of genus *Leptotrichia*. The first left panel indicates a total numbers of spacers present in each strain, and the second left panel does the numbers of spacer which have hit matches against each strain.

III-A/D and III-like systems on the genomes. Similar to the cases of same type systems, different types of CRISPR-Cas systems in single *Leptotrichia* genome also share the adaptation modules. For example, *L. trevisanii* JMUB4039 carries three CRISPR-*cas* loci; types I-B, III-D and VI-A. Since only the type I-B locus includes adaptation module genes *cas1* and *cas2*, the other types III-D and VI-A could utilize the Cas1 and Cas2 of type I-B. These multiple CRISPR-Cas systems might be the basis for an effective defense against foreign mobile genetic elements. In addition to their canonical function in defending foreign nucleic acids, new biological functions of CRISPR-Cas systems have been recognized. In *Pseudomonas aeruginosa*, CRISPR-Cas system is known to be required for bacteriophage DMS3-dependent inhibition of biofilm formation (Cady and O'Toole, 2011). While in *Porphyromonas gingivalis*, CRISPR-Cas can affect efficiency of IS transposition and genome recombination (Watanabe et al., 2013). Relationship between CRISPR-Cas and bacterial virulence has also been described in *Campylobacter jejuni*, *Neisseria meningitides,* and *Legionella pneumoniae* (Gunderson and Cianciotto, 2013; Louwen et al., 2013; Sampson et al., 2013).

Of the identified CRISPR-Cas13a systems, the Cas13a effectors were highly divergent with amino acid sequence similarities ranging from 61% to 90% to that of *L. shahii* JCM16776 (**Supplementary Table S4**), which is the most intensively studied for collateral RNA cleavage activity (Abudayyeh et al., 2016). In the present study, the two most distanced CRISPR-Cas13a systems, LwaJMUB3933Cas13a-1 and LwaJMUB3933Cas13a-2, were identified in *L. wadei* strain JMUB3933. The amino acid sequence similarities of LwaJMUB3933Cas13a-1 and LwaJMUB3933Cas13a-2 to that of *L. shahii* JCM16776 (LshCas13a) were 61% and 87%, respectively. In addition, Cas13a was firstly identified in *L. trevisanii* genome (*L. trevisanii* JMUB4039, LtrCas13a). The cytotoxic activities dependent on collateral ssRNA cleavage activity of the four systems (LwaJMUB3933Cas13a-1 and -2, LtrCas13a, LshCas13a) were compared with the use of a co-expression assay, where two different vectors carrying the CRISPR-Cas13a with spacer sequence and target sequences were co-transformed into competent *E. coli* cells, and then the growth curves were observed with or without 0.2% arabinose inducing mRNA which contain the target sequences. As shown in **Figure 4**, except for spacer-1 targeting of LwaJMUB3933Cas13a-1, the four Cas13a systems could significantly affect the growth of the host cells, indicating that the collateral cellular dormancy as a result of ssRNA cleavage were maintained among those strains. Here we demonstrated the collateral cytotoxic activities of four Cas13a using representative native spacers, a comprehensive study is necessary to draw precise conclusions regarding enzymatic activities.

CRISPR-Cas spacers represent a sequential achievement of former intruder encounters, and the retained spacers reflect evolutionary phylogeny or relatedness of the strains. In the present study, the CRISPR spacers of all strains were analyzed, as well as the protospacers of the strains with complete whole genomes. A protospacer is a spacer homolog existing in the plasmid, phage or chromosomal DNA that matches the spacers of CRISPR elements. The number and distribution of spacers and the corresponding protospacers are summarized in **Supplementary Table S5** and **Figure 5**. As shown in **Figure 5**, the number of spacers had varied among the strains, ranging from 1 to 157, which reflects the biological diversity of *Leptotrichia* species that acquired a CRISPR-Cas system for mediation of immunity. Type I-B, II-C and III-like CRISPR-Cas systems generally contains higher number of spacers than type III-A, III-D, and VI-A. The protospacer distribution showed that cross-targeting events within the species were very rare, but often occurred inter-species. Notably, some spacers matched against protospacers existed in the same genome. For example, spacer-1 (ATAATGTCTCAACTTAGTTTCAGTGCACTCTA) and spacer-2 (TTTTTAAACTCATCAGCCAACCAGTCAATGA) targeting of LwaJMUB3933Cas13a-2 completely matched against 50S ribosomal protein L33 gene *rpmG* and chaperone gene *dnaK* of strain JMUB3933, respectively. The CRISPR Cas13a system of LwaJMUB3933Cas13a-2 and the natural spacers could significantly affect the growth of the host *E. coli* cells (**Figure 4D**). As a possible interpretation, CRISPR-Cas systems may be regulated by unknown mechanisms that degrade their own function, as self-targeting results in persistence or cell death. In fact, some bacteria have intact CRISPR-*cas* genes and spacers, targeting their native chromosomal loci (Sampson and Weiss, 2014). Aklujkar et al. reported that a *Pelobacter carbinolicus* strain had type I CRISPR-Cas system with a spacer targeting the histidyl-tRNA synthetase gene, thereby regulating its expression (Aklujkar and Lovley, 2010). Sampson et al. (2013) also described that type II CRISPR-Cas system in *Francisella novicida* regulated the expression of an endogenous transcript encoding a bacterial lipoprotein. Hence, we postulated that CRISPR-Cas systems in *Leptotrichia* may play a role in more than just self-defense immunity.

CONCLUSION

In this study, we overviewed CRISPR-Cas systems in genus *Leptotrichia* and the comparative genomic analysis show that types VI-A as well as types I-B and III-A/D systems are predominantly distributed in *Leptotrichia*. Since most genus *Leptotrichia* have been found from bacteria composed of oral and gut flora in human and animal, and RNA type bacteriophages are infrequently found in gut, the RNA-targeting type VI-A system (CRISPR-Cas13a) might play an important role for bacterial surviving in the niche of oral cavity and gastrointestinal tract. However, this study showed that the VI-A system of *Leptotrichia* could get activated by natural spacers, including self-targeting spacers, to impede host cell growth, therefore, we postulated that CRISPR-Cas system has a more significant role other than immunity in the biology of *Leptotrichia* species.

DATA AVAILABILITY STATEMENT

The genome sequence has been deposited at DDBJ/GenBank: PRJDB7856 (Bioproject) and DRA007809 (Raw data).

ETHICS STATEMENT

Ethics approval and consent to participate were not required, following the ethical guidelines for medical and health research involving human subjects by Ministry of Health, Labor and Welfare, Japan since this study analyzed bacteria which were isolated as a clinical specimen and patients' personal health information could not be accessed (http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hokabunya/kenkyujigyou/i-kenkyu/index.html). Bacterial isolates were taken from hospitals as part of the standard patient care and used anonymously.

AUTHOR CONTRIBUTIONS

SW, KK, and LC designed the study. SW, BC, KK, YA, X-ET, YS, MK, TB, KT, YT, F-YL, AA, and TS performed the experiments and analyzed the data. YN coordinated the clinical sample collection. SW, BC, and KK prepared the figures and drafted the manuscript. SW, BC, KK, X-ET, and LC edited and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by JSPS KAKENHI (Grant Nos. 15H05654 and 19K08960 to SW, 18K15149 to KK, 17K15691 to YS, 19K15740 to MK, and 17K19570 to LC), the Takeda Science Foundation (SW and LC), the JSPS International Research Fellow (Grant No. 17F17713 to LC), and the Japan Agency for Medical Research and Development J-PRIDE (Grant Nos. JP17fm0208028, JP18fm0208028, and JP19fm0208028 to LC). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank Dr. Toyoko Oguri at the Juntendo University; Dr. Katsuko Okuzumi at the Dokkyo University Hospital; Dr. Aki Kawashima, Dr. Haruki Sawamura, and Dr. Akitaka Yokoyama at the Ichinomiyanishi Hospital; Dr. Noriyuki Abe at the Tenri Hospital; and Dr. Eri Morita, Dr. Yuji Hatano, and Dr. Satoru Hayashi at the Japanese Red Cross Gifu Hospital for kindly providing the *Leptotrichia* strains. We also thank Dr. Feng Zhang at the Massachusetts Institute of Technology and Dr. Masato Suzuki for kindly gifting the plasmids.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02838/full#supplementary-material

REFERENCES

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA targeting with CRISPR-Cas13. *Nature* 10, 280–284. doi: 10.1038/nature24049

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353:aaf5573. doi: 10.1126/science.aaf5573

Aklujkar, M., and Lovley, D. R. (2010). Interference with histidyl-tRNA synthetase by a CRISPR spacer sequence as a factor in the evolution of Pelobacter carbinolicus. *BMC Evol. Biol.* 10:230. doi: 10.1186/1471-2148-10-230

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 07, W16–W21. doi: 10.1093/nar/gkw387

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* 45, 273–297. doi: 10.1146/annurev-genet-110410-132430

Bik, E. M., Long, C. D., Armitage, G. C., Loomer, P., Emerson, J., Mongodin, E. F., et al. (2010). Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J.* 4, 962–974. doi: 10.1038/ismej.2010.30

Bolivar, I., Whiteson, K., Stadelmann, B., Baratti-Mayer, D., Gizard, Y., Mombelli, A., et al. (2012). Bacterial diversity in oral samples of children in niger with acute noma, acute necrotizing gingivitis, and healthy controls. *PLoS Negl Trop Dis.* 6:e1556. doi: 10.1371/journal.pntd.0001556

Cady, K. C., and O'Toole, G. A. (2011). Non-Identity-Mediated CRISPR-Bacteriophage Interaction Mediated via the Csy and Cas3 Proteins. *J. Bacteriol.* 193, 3433–3445. doi: 10.1128/JB.01411-10

Clarridge, J. E. III (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.* 17, 840–862. doi: 10.1128/CMR.17.4.840-862.2004

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. *Science* 11, 1019–1027. doi: 10.1126/science.aaq0180

East-Seletsky, A., O'Connell, M. R., Burstein, D., Knott, G. J., and Doudna, J. A. (2017). RNA targeting by functionally orthogonal type VI-A CRISPR-Cas enzymes. *Mol. Cell* 66, 373–383. doi: 10.1016/j.molcel.2017.04.008

East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R., et al. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* 538, 270–273. doi: 10.1038/nature19802

Eribe, E. R., and Olsen, I. (2008). Leptotrichia species in human infections. *Anaerobe* 14, 131–137. doi: 10.1016/j.anaerobe.2008.04.004

Eribe, E. R., Paster, B. J., Caugant, D. A., Dewhirst, F. E., Stromberg, V. K., Lacy, G. H., et al. (2004). Genetic diversity of *Leptotrichia* and description of *Leptotrichia goodfellowii* sp. nov., *Leptotrichia hofstadii* sp. nov., *Leptotrichia shahii* sp. nov. and *Leptotrichia wadei* sp. nov. *Int. J. Syst. Evol. Microbiol.* 54, 583–592. doi: 10.1099/ijs.0.02819-0

Eribe, E. R. K., and Olsen, I. (2017). species in human infections II. *J. Oral. Microbiol.* 9:1368848. doi: 10.1080/20002297.2017.1368848

Foster, K., Kalter, J., Woodside, W., Terns, R. M., and Terns, M. P. (2019). The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA Biol.* 16, 449–460. doi: 10.1080/15476286.2018.1493334

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31, 2877–2878. doi: 10.1093/bioinformatics/btv271

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., et al. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356, 438–442. doi: 10.1126/science.aam9321

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship

to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Gunderson, F. F., and Cianciotto, N. P. (2013). The CRISPR-associated gene cas2 of *Legionella pneumophila* is required for intracellular infection of amoebae. *mBio* 4:e00074-13. doi: 10.1128/mBio.00074-13

Gupta, R. S., and Sethi, M. (2014). Phylogeny and molecular signatures for the phylum Fusobacteria and its distinct subclades. *Anaerobe* 28, 182–198. doi: 10.1016/j.anaerobe.2014.06.007

Haft, D. H., Selengut, J., Mongodin, E. F., and Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* 1:e60. doi: 10.1371/journal.pcbi.0010060

Han, W., Pan, S., López-Méndez, B., Montoya, G., and She, Q. (2017). Allosteric regulation of Csx1, a type IIIB-associated CARF domain ribonuclease by RNAs carrying a tetraadenylate tail. *Nucleic Acids Res.* 45, 10740–10750. doi: 10.1093/nar/gkx726

Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., and Harris, S. R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* 16:294. doi: 10.1186/s13059-015-0849-0

Ivanova, N., Gronow, S., Lapidus, A., Copeland, A., Glavina Del Rio, T., Nolan, M., et al. (2009). Complete genome sequence of *Leptotrichia buccalis* type strain (C-1013-b). *Stand. Genomic Sci.* 1, 126–132. doi: 10.4056/sigs.1854

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Kong, H. H., Oh, J., Deming, C., Conlan, S., Grice, E. A., Beatson, M. A., et al. (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res.* 22, 850–859. doi: 10.1101/gr.131029.111

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444

Louwen, R., Horst-Kreft, D., de Boer, A. G., van der Graaf, L., de Knegt, G., Hamersma, M., et al. (2013). A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain-Barre syndrome. *Eur. J. Clin. Microbiol. Infect. Dis.* 32, 207–226. doi: 10.1007/s10096-012-1733-4

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2018). Classification and nomenclature of CRISPR-cas systems: where from here? *CRISPR J.* 1, 325–336. doi: 10.1089/crispr.2018.0033

Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61. doi: 10.1038/nature15386

Munson, M. A., Banerjee, A., Watson, T. F., and Wade, W. G. (2004). Molecular analysis of the microflora associated with dental caries. *J. Clin. Microbiol.* 42, 3023–3029. doi: 10.1128/JCM.42.7.3023-3029.2004

Niewoehner, O., and Jinek, M. (2016). Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA.* 22, 318–329. doi: 10.1261/rna.054098.115

O'Connell, M. R. (2019). Molecular mechanisms of RNA targeting by Cas13-containing Type VI CRISPR-Cas systems. *J. Mol. Biol.* 431, 66–87. doi: 10.1016/j.jmb.2018.06.029

Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L., and Weiss, D. S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* 497, 254–257. doi: 10.1038/nature12048

Sampson, T. R., and Weiss, D. S. (2014). CRISPR-Cas systems: new players in gene regulation and bacterial physiology. *Front. Cell. Infect. Microbiol.* 4:37. doi: 10.3389/fcimb.2014.00037

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shah, S. A., Alkhnbashi, O. S., Behler, J., Han, W., She, Q., Hess, W. R., et al. (2019). Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families. *RNA Biol.* 16, 530–542. doi: 10.1080/15476286.2018.1483685

Sheppard, N. F., Glover, C. V., Terns, R. M., and Terns, M. P. (2016). The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease. *RNA* 22, 216–224. doi: 10.1261/rna.039842.113

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., et al. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* 60, 385–397. doi: 10.1016/j.molcel.2015.10.008

Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5307–E5316. doi: 10.1073/pnas.1803440115

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Täubel, M., Rintala, H., Pitkäranta, M., Paulin, L., Laitinen, S., Pekkanen, J., et al. (2009). The occupant as a source of house dust bacteria. *J. Allergy Clin. Immunol.* 124, 834–840. doi: 10.1016/j.jaci.2009.07.045

Vestergaard, G., Garrett, R. A., and Shah, S. A. (2014). CRISPR adaptive immune systems of *Archaea*. *RNA Biol.* 11, 156–167. doi: 10.4161/rna.27990

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Watanabe, S., Aiba, Y., Tan, X. E., Li, F. Y., Boonsiri, T., Thitiananpakorn, K., et al. (2018). Complete genome sequencing of three human clinical isolates of Staphylococcus caprae reveals virulence factors similar to those of *S. epidermidis* and *S. capitis*. *BMC Genomics* 19:810. doi: 10.1186/s12864-018-5185-9

Watanabe, S., Sasahara, T., Arai, N., Sasaki, K., Aiba, Y., Sato'o, Y., et al. (2016). Complete genome sequence of *Streptococcus pyogenes* strain JMUB1235 isolated from an acute *Phlegmonous gastritis* patient. *Genome Announc.* 4:e1133-16. doi: 10.1128/genomeA.01133-16

Watanabe, T., Nozawa, T., Aikawa, C., Amano, A., Maruyama, F., and Nakagawa, I. (2013). CRISPR regulation of intraspecies diversification by limiting IS transposition and intercellular recombination. *Genome Biol. Evol.* 5, 1099–1114. doi: 10.1093/gbe/evt075

Woo, P. C., Wong, S. S., Teng, J. L., Leung, K. W., Ngan, A. H., Zhao, D. Q., et al. (2010). Leptotrichia hongkongensis sp. nov., a novel *Leptotrichia* species with the oral cavity as its natural reservoir. *J Zhejiang Univ. Sci. B.* 11, 391–401. doi: 10.1631/jzus.B1000056

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Watanabe, Cui, Kiga, Aiba, Tan, Sato'o, Kawauchi, Boonsiri, Thitiananpakorn, Taki, Li, Azam, Nakada, Sasahara and Cui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.