



Application of the Whole Genome-Based Bacterial Identification System, TrueBac ID, Using Clinical Isolates That Were Not Identified With Three Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) Systems

Sung-Min Ha , M.Sc.^{1,2*}, Chang Ki Kim , M.D., Ph.D.^{3,4*}, Juhye Roh , M.D.⁵, Jung-Hyun Byun , M.D.^{4,5,6}, Seung-Jo Yang , Ph.D.¹, Seon-Bin Choi , M.Sc.¹, Jongsik Chun , Ph.D.^{1,2}, and Dongeun Yong , M.D., Ph.D.^{4,5,7}

¹ChunLab, Inc., Seoul, Korea; ²School of Biological Sciences & Institute of Molecular Biology and Genetics, Seoul National University, Seoul, Korea; ³Seoul Clinical Laboratories, Yongin, Korea; ⁴Research Institute of Bacterial Resistance, Yonsei University College of Medicine, Seoul, Korea; ⁵Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul, Korea; ⁶Department of Laboratory Medicine, Gyeongsang National University Hospital, and Gyeongsang National University College of Medicine, Jinju, Korea; ⁷Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, Korea

Background: Next-generation sequencing is increasingly used for taxonomic identification of pathogenic bacterial isolates. We evaluated the performance of a newly introduced whole genome-based bacterial identification system, TrueBac ID (ChunLab Inc., Seoul, Korea), using clinical isolates that were not identified by three matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) systems and 16S rRNA gene sequencing.

Methods: Thirty-six bacterial isolates were selected from a university-affiliated hospital and a commercial clinical laboratory. Species was identified by three MALDI-TOF MS systems: Bruker Biotyper MS (Bruker Daltonics, Billerica, MA, USA), VITEK MS (bioMérieux, Marcy l'Étoile, France), and ASTA MicroIDSys (ASTA Inc., Suwon, Korea). Whole genome sequencing was conducted using the Illumina MiSeq system (Illumina, San Diego, CA, USA), and genome-based identification was performed using the TrueBac ID cloud system (www.truebacid.com).

Results: TrueBac ID assigned 94% (34/36) of the isolates to known (N=25) or novel (N=4) species, genomospecies (N=3), or species group (N=2). The remaining two were identified at the genus level.

Conclusions: TrueBac ID successfully identified the majority of isolates that MALDI-TOF MS failed to identify. Genome-based identification can be a useful tool in clinical laboratories, with its superior accuracy and database-driven operations.

Key Words: Next generation sequencing, Genome-based identification, TrueBac ID, Performance

Received: October 14, 2018

Revision received: January 8, 2019

Accepted: May 22, 2019

Corresponding author:

Dongeun Yong, M.D., Ph.D.
Department of Laboratory Medicine and
Research Institute of Bacterial Resistance,
Yonsei University College of Medicine,
50-1 Yonsei-ro, Seodaemun-gu, Seoul
03722, Korea
Tel: +82-2-2228-2802
Fax: +82-2-364-1583
E-mail: deyong@yuhs.ac

*These authors contributed equally to this work.



© Korean Society for Laboratory Medicine

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Primary and nosocomial bacterial infections are significant causes of morbidity and mortality worldwide [1]. Identification of bacte-

rial isolates at the species level is the first and crucial step in routine clinical laboratories, as it provides essential guidance regarding treatment. Although conventional biochemical testing is still used, whole-cell matrix-assisted laser desorption ionization

time-of-flight mass spectrometry (MALDI-TOF MS) has been widely adopted for the routine identification of pathogenic bacteria [2]. MALDI-TOF MS can rapidly identify isolates by comparing the proteomic profiles of highly conserved and abundant proteins with an already compiled profile database of reference strains. Therefore, the accuracy of MALDI-TOF MS identification is heavily dependent on the software and spectral database/libraries [3]. MALDI-TOF MS is particularly useful for identifying frequently isolated pathogenic species because of better coverage of the spectral database. However, its ability to identify infrequently isolated species is questionable.

16S ribosomal RNA (16S rRNA) gene sequencing has been primarily used for cases, in which routine, conventional methods fail to identify the isolates. The 16S rRNA gene is a phylogenetic marker that is present in all bacteria and has played an essential role in the development of bacterial phylogeny and classification [4, 5]. Recently, the similarity cut-off for 16S rRNA gene sequence (98.7%) was proposed as a boundary for bacterial species [6]. However, a 16S rRNA gene sequence similarity of $\geq 98.7\%$ does not guarantee that the test isolate is a member of the species. Almost identical 16S rRNA gene sequences have been found in different species [6, 7].

Unlike 16S rRNA gene sequencing, whole genome sequencing (WGS) provides a clear-cut criterion for bacterial classification. Furthermore, bacterial species is now defined by the relatedness of genome sequences [8]. A category of algorithms, named the overall genome-related index [4], has been devised to calculate the genomic similarity for taxonomic purposes, and a general guideline for bacterial identification using WGS data has been published [8]. WGS has considerable potential in clinical diagnostics as it could provide accurate identification of the species, and resolution can be achieved up to the strain [9]. As the cost of WGS is continuously decreasing, its use as a routine test has been validated in large hospitals [10, 11].

If the genome sequences of the type strains representing all the known bacterial species are available, any isolate could be identified with high confidence. However, until a few years ago, the availability of such data was not satisfactory [4, 12]. The utility of genome-based methods for clinical diagnostics requires re-evaluation in light of the recent expansion in the bacterial genome sequence database. This is the first study to evaluate TrueBac ID (ChunLab Inc., Seoul, Korea), which is the first commercial whole genome-based bacterial identification system. Its database contains highly curated and taxonomically validated genome data of type and reference strains. We evaluated the performance of TrueBac ID in identifying clinical bacterial isolates

that could not be identified using commercial MALDI-TOF MS systems and 16S rRNA gene sequencing.

METHODS

Bacterial isolates

In this retrospective study, a total of 36 clinical isolates that were either unidentified or identified with low confidence by three commercial MALDI-TOF MS systems were collected from two institutes in Korea. Fifteen isolates were chosen from Severance hospital, Seoul, Korea, and 21 isolates were from Seoul Clinical Laboratories, Yongin, Korea. The isolates were recovered from clinical specimens (blood, pus, sputum, tracheal aspirate, urine, and wounds) from April 2017 to January 2018. Since this study focuses on the identification of the isolates, an approval from the Institutional Review Board was not required, and the demographic data of the patients were not included.

Initially, we used Bruker Biotyper (Bruker Daltonics, Billerica, MA, USA) for species identification at both institutes. 16S rRNA gene PCR and sequencing were carried out for the isolates that showed no possible identification (score value: < 1.70) or low confidence identification (score value: $1.70-1.99$) MALDI-TOF MS results. For isolates showing uncertain identification results, we employed two additional MALDI-TOF identification systems: the VITEK MS system (bioMérieux, Marcy l'Étoile, France) and ASTA MicroIDSys (ASTA Inc., Suwon, Korea). A colony grown on sheep blood agar was smeared and dried on the target plates of each instruments. Matrix solution (α -cyano-4-hydroxycinnamic acid) and 70% formic acid (Sigma-Aldrich, St. Louis, MO, USA) provided by the manufacturer were overlaid on the spot, and the peptide profile was acquired using Microflex with Biotyper Software 3.1 (Bruker), VITEK MS V3.0 (bioMérieux), and ASTA MicroIDSys 3.0.4 (ASTA Inc.). Mass spectra were analyzed according to the manufacturers' instructions.

Genome-based identification

Genomic DNA was extracted from the isolates using the FastDNA SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA, USA), and 550-bp long fragments were generated using the M220 Focused-ultrasonicator (Covaris Ltd, Brighton, UK). The sequencing library was constructed using the TruSeq DNA Library LT kit (Illumina, San Diego, CA, USA), according to the manufacturer's protocols. WGS was performed on an Illumina MiSeq system (Illumina) with a 300 bp paired-end reads sequencing kit (MiSeq Reagent Kit v3; Illumina).

The raw data from the MiSeq instrument in the FASTQ format

were directly uploaded to the TrueBac ID cloud system (www.truebacid.com) and analyzed with the TrueBac ID-Genome system. The current version of the system uses trimmomatic for filtering low-quality reads [13]. The genome assembly was then carried out using the SPAdes software [14], as well as proprietary software specifically designed for the assembly of the 16S rRNA gene from the raw data.

The main section of the TrueBac ID-Genome system consists of (1) the proprietary reference database, named the TrueBac database, which is curated to hold up-to-date nomenclature, 16S rRNA gene, and genome sequences of type/reference strains, and (2) the optimized bioinformatics pipeline that provides the identification of a query genome sequence using the average nucleotide identity (ANI) [4, 8, 15]. We used TrueBac database version 2018-08, which contains 10,439 genomes representing 10,152 species and 287 subspecies (7,702 with valid names, 261 with invalid names, 138 with Candidatus names [16], and 2,338 genomospecies). Genomospecies is defined as a hitherto unknown species that is supported by its genome sequences [17-19]. The database also contains 18,476 16S rRNA gene sequences representing each species/subspecies.

The algorithmic identification scheme using WGS was slightly modified from that of Yoon, *et al.* [5]. First, the most phylogenetically closely related pool of taxa was identified using a search of three genes—16S rRNA, *recA*, and *rpIC*—which were extracted from the whole genome assembly [5]. The latter two genes were a part of the 92 recently defined bacterial core genes [20]. The taxonomically meaningful similarity of 16S rRNA gene sequences was calculated as previously described [21]. In addition to the gene-based searches, we used the Mash tool (<https://github.com/marbl/mash>) for additional fast whole-genome based searches [22]. The top-hits of the above four searches were then pooled, and the ANI was calculated using the MUMmer tool (<http://mummer.sourceforge.net/>) [15].

The algorithmic cut-off for species-level identification was set at 95% ANI [8, 15]. If the closely related taxa in a 16S rRNA gene comparison did not have the corresponding genome sequences in the database, the species assignment was made when the 16S rRNA gene sequence similarity to the best hit taxon was $\geq 99\%$ with $>0.8\%$ separation between species [23]. Using these criteria, a genome sequence could be assigned to a species held in the TrueBac database, identified to the genus level (e.g. *Bacillus* sp.), identified as a novel species (e.g., *Chryseobacterium* sp. nov.), or regarded as unidentifiable.

In some cases, two or more species belonging to the same species were not yet formally reclassified. For isolates assigned

to these species, the TrueBac ID system generated the final decision as a “species group” instead of individual species.

RESULTS

Of the 36 isolates, TrueBac ID successfully identified 25 isolates as known species (Table 1). Four isolates were new species that had not been previously recognized. Three genomospecies, labeled CP015506_s, BBQM_s, and JHEL_s, were assigned. Detailed taxonomic information on these genomospecies is available at www.ezbiocloud.net [5]. Two isolates were identified as “species group” (*Shewanella algae* group and *Tsukamurella tyrosinosolvens* group). The remaining two isolates were identified at only the genus level because of the lack of relevant reference genome sequences. Isolate YUMC P471 was most closely related to *Bacillus beringensis*, with 98.91% 16S rRNA gene similarity, which is lower than the cut-off (99%) we used. Isolate YUMC R2593 was found to be closely related to *Chryseobacterium bernardetii* and *Chryseobacterium vietnamense* with 99.65% and 98.92% 16S rRNA gene similarity, respectively. As the difference in 16S rRNA gene similarity between two *Chryseobacterium* species was only 0.73%, which is lower than the cut-off (0.8%) set for the 16S rRNA gene-based identification scheme [23], the identification was made to only the genus level. Overall, TrueBac ID identified 94% (34 of 36) of the isolates at the species level or as a novel species.

Of the 34 isolates that were conclusively identified at the species level, 26 were assigned to known species using the ANI calculation against the type strain genomes in the database, yielding a true or definitive identification. The remaining eight isolates were identified by 16S rRNA gene similarity, according to the CLSI guidelines [23].

Of the 25 isolates identified as known species by TrueBac ID, all three MALDI-TOF systems failed to identify 17 isolates. The MALDI Biotyper System identified nine (eight matched with TrueBac ID and one mismatched), the VITEK MS identified seven (three matched with TrueBac ID and four mismatched), and the ASTA system identified seven (four matched with TrueBac ID and three mismatched). The detailed identification results with the genome assembly and gene sequences we reported are available at <https://www.truebacid.com/genome/demo/clinical/korea>.

DISCUSSION

Overall, TrueBac ID performs well for isolates that MALDI-TOF

Table 1. Identification results of TrueBac ID and three MALDI-TOF MS systems

Isolate name	Specimen	TrueBac ID	Evidence used by TrueBac ID (%)	MALDI-TOF MS ID systems		
				MALDI Biotyper	Vitek MS	ASTA MicroIDSys
SCL P5	Pus	<i>Porphyromonas pogoanae</i>	16S rRNA gene (100)	-	-	-
SCL U17	Urine	<i>Myroides injenensis</i>	ANI (99.81)	-	-	-
SCL P33	Pus	<i>Corynebacterium provencense</i>	ANI (99.18)	-	-	-
SCL O41	Other	<i>Brachybacterium nesterenkovii</i>	ANI (97.15)	<i>Brachybacterium nesterenkovii</i>	-	-
SCL P44	Superficial wound	<i>Brevibacterium otitidis</i>	16S rRNA gene (99.85)	<i>Brevibacterium luteolum</i>	-	-
SCL B45	Blood	<i>Ignavigranum ruoffiae</i>	ANI (99.47)	-	-	-
SCL B46	Blood	<i>Oceanobacillus sojiae</i>	16S rRNA gene (99.93)	-	-	-
SCL B79	Blood	<i>Shewanella algae</i> group	ANI (98.19)	<i>Shewanella algae</i>	<i>Shewanella putrefaciens</i>	
SCL B80	Blood	<i>Roseomonas mucosa</i>	ANI (99.83)	<i>Roseomonas mucosa</i>	<i>Roseomonas gilardii</i>	
SCL C89	CSF	<i>Brevibacterium otitidis</i>	16S rRNA gene (99.93)	-	<i>Brevibacterium luteolum</i>	-
SCL P96	Pus	<i>Brevundimonas albigilva</i>	ANI (99.49)	-	-	-
SCL P122	Pus	<i>Nocardia nova</i>	ANI (97.36)	<i>Nocardia nova</i>	-	-
SCL C146	CAPD	<i>Sphingobacterium hotanense</i>	16S rRNA gene (99.65)	-	-	-
SCL B151	Blood	<i>Chryseobacterium</i> sp. nov.	ANI (94.72)	-	-	-
SCL U158	Urine	<i>Enterococcus thailandicus</i>	ANI (99.52)	<i>Enterococcus thailandicus</i>	-	<i>Enterococcus thailandicus</i>
SCL P174	Pus	<i>Gemella massiliensis</i>	ANI (99.53)	<i>Gemella bergeri</i>	<i>Gemella bergeri</i>	<i>Gemella morbillorum</i>
SCL T176	Lung biopsy	<i>Nocardia farcinica</i>	ANI (99.13)	<i>Nocardia farcinica</i>	<i>Nocardia farcinica</i>	-
SCL B191	Blood	<i>Haematomicrobium sanguinis</i>	ANI (95.87)	-	-	-
SCL P193	Brain abscess	<i>Nocardia nova</i>	ANI (97.61)	-	-	-
SCL C196	Body fluid	<i>Myroides injenensis</i>	ANI (99.84)	-	-	<i>Myroides odoratimimus</i>
SCL B197	Blood	<i>Rothia dentocariosa</i>	ANI (96.28)	<i>Rothia dentocariosa</i>	-	<i>Rothia dentocariosa</i>
YUMC B10632	Blood	<i>Herbaspirillum huttiense</i>	ANI (97.44)	<i>Herbaspirillum huttiense</i>	-	-
YUMC B11605	Blood	BBQM_s (<i>Cupriavidus</i>)	ANI (98.66)	-	-	<i>Cupriavidus basilensis</i>
YUMC B12492	Blood	<i>Tsukamurella tyrosinosolvens</i> group	ANI (98.11)	<i>Tsukamurella paurometabola</i>	-	<i>Dialister pneumosintes</i>
YUMC B1400	Blood	<i>Dysgonomonas</i> sp. nov.	16S rRNA gene (98.34)	-	-	-
YUMC B14785	Blood	<i>Cohnella</i> sp. nov.	16S rRNA gene (94.98)	-	-	-
YUMC B1698	Blood	<i>Brucella melitensis</i>	ANI (99.85)	-	-	-
YUMC B2387	Blood	<i>Pseudoxanthomonas</i> sp. nov.	16S rRNA gene (98.36)	-	-	-
YUMC B413	Blood	<i>Pandoraea apista</i>	ANI (99.26)	<i>Pandoraea apista</i>	-	-
YUMC P471	Pus	<i>Bacillus</i> sp.	16S rRNA gene (99.45)	-	-	-
YUMC P647	Pus	CP015506_s (<i>Bacillus</i>)	ANI (99.04)	-	<i>Bacillus firmus</i>	-
YUMC P721	Pus	JHEL_s (<i>Kocuria</i>)	ANI (99.38)	-	-	-
YUMC P800	Pus	<i>Kerstersia gyiorum</i>	ANI (98.77)	<i>Kerstersia gyiorum</i>	-	<i>Kerstersia gyiorum</i>
YUMC P891	Pus	<i>Tsukamurella inchonensis</i>	ANI (98.88)	-	-	<i>Tsukamurella inchonensis</i>
YUMC R2593	Sputum	<i>Chryseobacterium</i> sp.	16S rRNA gene (99.65)	<i>Chryseobacterium gleum</i>	-	<i>Chryseobacterium gleum</i>
YUMC T569	Tip	<i>Ornithinibacillus contaminans</i>	ANI (98.77)	-	-	-

Abbreviations: MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; rRNA, ribosomal RNA; CSF, cerebrospinal fluid; CAPD, continuous ambulatory peritoneal dialysis; ANI, average nucleotide identity; -, no available identification result.

MS systems and 16S rRNA gene sequencing fail to identify. The ability to identify rare species is largely influenced by database coverage. The TrueBac ID system contains >10,000 species, whereas commercially available MALDI-TOF MS systems contain only ~2,500 species [24].

Because of advances in DNA sequencing technologies and the introduction of genomics into bacterial taxonomy, numerous species have been newly described. On average, approximately 100 new species were described every month in 2017 (data from www.ezbiocloud.net). The TrueBac ID system reference database is updated every month, enabling detection of recently described species. For example, isolate SCL P33 was identified as "*Corynebacterium provencense*," which was recently discovered in a human fecal specimen [25] and also associated with otitis in a cat [26]. Similarly, isolate SCL P174 was assigned to "*Gemella massiliensis*," which was isolated from the sputum of a healthy individual in France [27]. Neither species could be identified by MALDI-TOF Biotyper previously [25, 27]. We could not identify isolate SCL P33 using any of the MALDI-TOF MS systems employed, whereas isolate SCL P174 was misidentified as *Gemella bergeri* using the MALDI Biotyper and VITEK MS, and as *Gemella morbillorum* using the ASTA system.

One of the major benefits of whole genome-based identification is that it can provide a scientifically sound decision for the recognition of novel species. We confirmed four novel species and three genomospecies based on 16S rRNA gene or genomic evidence. Isolates YUMC P721, YUMC P647, and YUMC B11605 were identified as genomospecies JHEL_s, CP015506_s, and BBQM_s, respectively. These species were never officially proposed and only tentatively named by the EzBioCloud database [5], so they can be considered novel species. As genomospecies represent previously isolated species, the use of this concept can provide further insights into species ecology. For example, the genomospecies CP015506_s included in this study is a species of the genus *Bacillus*, and there are three genome sequences in the EzBioCloud database from three different sources: an oral swab of a patient (USA), seawater (Korea), and soil (India) (https://www.ezbiocloud.net/genome/list?tn=CP015506_s). This additional information implies that the species is widespread in nature and may be associated with human diseases.

Isolate SCL B79 showed high ANI values to *Shewanella upe- nei* (98.19%), as well as to *Shewanella algae* (98.15%); both ANI values are clearly higher than the species boundary (95–96%) [8]. The ANI value between the type strains of the two *Shewanella* species is 98.1%, indicating that they taxonomically represent a heterotypic synonym. Similarly, isolate YUMC B12492

isolated from blood was assigned as "*Tsukamurella carboxydivorans* group," which consists of *Tsukamurella tyrosinosolvens* and *Tsukamurella carboxydivorans*; these two species share high genome sequence similarity (98.9% ANI). These potential synonyms are treated as a "species group" in the TrueBac ID system to avoid possible confusion in species-level identification. We expect that all species groups we reported will eventually be combined to meet the currently accepted taxonomic scheme [8].

Overall, TrueBac ID could identify the species level for >90% of the isolates. Moreover, it demonstrated the ability to recognize new species with high confidence. This is a significant advantage of genome-based ID over other methods, including MALDI-TOF MS and biochemical tests. In addition to its superior accuracy, WGS is not influenced by media and growth conditions, in contrast to other methods based on phenotypes including MALDI-TOF MS [28].

Although 16S rRNA gene sequencing has been widely used as the gold standard for bacterial identification [29], this method is not feasible for some clinically important species with highly similar 16S rRNA gene sequences [3]. We demonstrated that WGS exhibited sufficient taxonomic coverage to be employed as a scientifically sound gold standard when any new diagnostic method or commercial system is evaluated.

This study has some limitations. We collected only those isolates that were not properly identified by MALDI-TOF MS. However, the proportion of those isolates would be low in most laboratories. In addition, the clinical significance of the isolates was not clearly defined. We assume that not all the isolates are true pathogens. Lastly, we did not examine how accurately identifying the isolates can improve patient care.

In conclusion, TrueBac ID successfully identified the majority of clinical bacterial isolates that were not identified by commercially available MALDI-TOF MS systems or 16S rRNA gene sequencing. TrueBac ID was more useful than other conventional diagnostic methods in recognizing new species. As the coverage of type strain-genome sequence database continues to grow and the cost of DNA sequencing continues to decrease, genome-based identification can be a useful tool for diagnostic laboratories, with its superior accuracy and database-driven operations.

Authors' Disclosures of Potential Conflicts of Interest

SMH, SJY, SBC, and JC are employees of ChunLab, Inc., the creator of the TrueBac ID system. DY owns shares in ChunLab, Inc.

Acknowledgment

This work was supported by the BioNano Health Guard Research Center, funded by the Ministry of Science, ICT and Future Planning (MSIP) of Korea as a Global Frontier Project (Grant Number H-GUARD_2014M3A6B2060509); by the Nano Material Technology Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No.2017M3A7B4039936); and by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Korea (grant No. HI17C1807). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ORCID

Sung-Min Ha <https://orcid.org/0000-0003-3945-8329>
 Chang Ki Kim <https://orcid.org/0000-0003-4729-3114>
 Juhye Roh <https://orcid.org/0000-0003-0078-1145>
 Jung-Hyun Byun <https://orcid.org/0000-0001-5909-5807>
 Seung-Jo Yang <https://orcid.org/0000-0001-7084-5816>
 Seon-Bin Choi <https://orcid.org/0000-0001-6995-8039>
 Jongsik Chun <https://orcid.org/0000-0003-3385-5171>
 Dongeun Yong <https://orcid.org/0000-0002-1225-8477>

REFERENCES

1. Vouga M and Greub G. Emerging bacterial pathogens: the past and beyond. *Clin Microbiol Infect* 2016;22:12-21.
2. Croxatto A, Prod'homme G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiol Rev* 2012;36:380-407.
3. Van den Abeele AM, Vogelaers D, Vandamme P, Vanlaere E, Houf K. Filling the gaps in clinical proteomics: a do-it-yourself guide for the identification of the emerging pathogen *Arcobacter* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Microbiol Methods* 2018;152:92-7.
4. Chun J and Rainey FA. Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*. *Int J Syst Evol Microbiol* 2014;64:316-24.
5. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing Ez-BioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 2017;67:1613-7.
6. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 2014;64:346-51.
7. Fox GE, Wisotzkey JD, Jurtshuk PJ Jr. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* 1992;42:166-70.
8. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018;68:461-6.
9. Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogstraat DR, Cookson BT. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *J Clin Microbiol* 2015;53:1072-9.
10. Roach DJ, Burton JN, Lee C, Stackhouse B, Butler-Wu SM, Cookson BT, et al. A year of infection in the Intensive Care Unit: prospective whole genome sequencing of bacterial clinical isolates reveals cryptic transmissions and novel microbiota. *PLoS Genet* 2015;11:e1005413.
11. Mellmann A, Bletz S, Böking T, Kipp F, Becker K, Schultes A, et al. Real-Time Genome sequencing of resistant bacteria provides precision infection control in an institutional setting. *J Clin Microbiol* 2016;54:2874-81.
12. Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Göker M, Parker CT, et al. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biol* 2014;12:e1001920.
13. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114-20.
14. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455-77.
15. Richter M and Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009;106:19126-31.
16. Murray RG and Stackebrandt E. Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described prokaryotes. *Int J Syst Bacteriol* 1995;45:186-7.
17. Fischer S, Mayer-Scholl A, Imholt C, Spierling NG, Heuser E, Schmidt S, et al. *Leptospira* genomospecies and sequence type prevalence in small mammal populations in Germany. *Vector Borne Zoonotic Dis* 2018;18:188-99.
18. Patil PP, Kumar S, Midha S, Gautam V, Patil PB. Taxonogenomics reveal multiple novel genomospecies associated with clinical isolates of *Stenotrophomonas maltophilia*. *Microb Genom* 2018;4:e000207.
19. Salipante SJ, Kalapila A, Pottinger PS, Hoogstraat DR, Cummings L, Duchin JS, et al. Characterization of a multidrug-resistant, novel *Bacteroides* genomospecies. *Emerg Infect Dis* 2015;21:95-8.
20. Na SI, Kim YO, Yoon SH, Ha SM, Baek I, Chun J. UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J Microbiol* 2018;56:280-5.
21. Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, et al. Introducing Ez-Taxon-e: a prokaryotic 16S rRNA gene sequence database with phylogenotypes that represent uncultured species. *Int J Syst Evol Microbiol* 2012;62:716-21.
22. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016;17:132.
23. CLSI. Interpretive criteria for identification of bacteria and fungi by DNA target sequencing. Approved guideline MM18-A. Wayne, PA: Clinical and Laboratory Standards Institute, 2008.
24. Jang KS and Kim YH. Rapid and robust MALDI-TOF MS techniques for microbial identification: a brief overview of their diverse applications. *J Microbiol* 2018;56:209-16.
25. Ndongo S, Andrieu C, Fournier PE, Lagier JC, Raoult D. '*Actinomyces provencensis*' sp. nov., '*Corynebacterium bouchesdurhonense*' sp. nov., '*Corynebacterium provencense*' sp. nov. and '*Xanthomonas massiliensis*' sp. nov., 4 new species isolated from fresh stools of obese French patients. *New Microbes New Infect* 2017;18:24-7.

26. Kittl S, Brodard I, Rychener L, Jores J, Roosje P, Gobeli Brawand S. Otitis in a cat associated with *Corynebacterium provencense*. BMC Vet Res 2018;14:200.
27. Fonkou MDM, Bilen M, Cadoret F, Fournier PE, Dubourg G, Raoult D. '*Enterococcus timonensis*' sp. nov., '*Actinomyces marseillensis*' sp. nov., '*Leptotrichia massiliensis*' sp. nov., '*Actinomyces pacaensis*' sp. nov., '*Actinomyces oralis*' sp. nov., '*Actinomyces culturomici*' sp. nov. and '*Gemella massiliensis*' sp. nov., new bacterial species isolated from the human respiratory microbiome. New Microbes New Infect 2017;22:37-43.
28. Martiny D, Visscher A, Catry B, Chatellier S, Vandenberg O. Optimization of *Campylobacter* growth conditions for further identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). J Microbiol Methods 2013;94:221-3.
29. Mellmann A, Cloud J, Maier T, Keckevoet U, Ramminger I, Iwen P, et al. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. J Clin Microbiol 2008; 46:1946-54.