Complete genome and description of *Corynebacterium incognita* sp. nov.: a new bacterium within the *Corynebacterium* genus

M. Boxberger[1,2], A. Antezack[1,2,3], S. Magnien[1,2], N. Cassir[1,2] and B. La Scola[1,2]

1) *Aix Marseille Université, IRD, AP-HM, MEPHI,* 2) *IHU-Méditerranée Infection and* 3) *Faculté des Sciences Médicales et Paramédicales, Ecole de Médecine Dentaire, Aix-Marseille Université, 27 boulevard Jean Moulin, 13385 Marseille, France*

Abstract

In 2020, as part of the diagnosis in IHU-Méditerranée Infection Institute in Marseille (France) we isolated the new bacterial strain Marseille-3630^T from a 7-year-old girl blood specimen (= CSUR: Q3630). Matrix-assisted desorption ionisation time−of-flight mass spectrometry failed to identify this isolate. Analysis of the 16S rRNA gene and genome-to-genome comparison suggested that this taxon belongs to a novel bacterial species within the family *Corynebacteriaceae* in the phylum *Actinobacteria*. We described here its main phenotypic characteristics, genome sequence and annotation of *Corynebacterium incognitum* strain Marseille-3630^T, a new member of the *Corynebacterium* genus, that we propose as type strain.

Keywords: Bacteria, *Corynebacterium*, genome, human, sp. nov., species, taxono-genomics

Original Submission: 20 January 2021; **Revised Submission:** 29 March 2021; **Accepted:** 28 April 2021

Article published online: 5 May 2021

Corresponding author: Bernard La Scola, Aix Marseille Université, Méditerranée-Infection, 19-21 Boulevard Jean Moulin, 13385, Marseille cedex 05, France

E-mail: bernard.la-scola@univ-amu.fr

Introduction

The genus *Corynebacterium* (Gr. fem. n. *korynê*, a club; L. neut. n. *bacterium*, a rod, and in biology a bacterium (so called because the first ones observed were rod-shaped); N.L. neut. n. *Corynebacterium*, a club bacterium) belongs to the large family *Corynebacteriaceae* and was first introduced into literature by Lehmann and Neumann in 1896 [1]. The genus *Corynebacterium* currently counts 177 species [2]; some of them are of medical, veterinary or biotechnological interest [3]. Species within this genus are ubiquitous and are potentially pathogens. As an illustration, toxin production by *C. diphtheriae* is involved in diphtheria, a dreadful historical disease in the 19th century and in the first half of the 20th century. This study contributes to the taxonomical and clinical knowledge of this genus by describing a novel species, strain Marseille-Q3630, isolated as part of a microbiological workup of a patient in IHU-Méditerranée Infection Institute in Marseille (France). We aimed at comparing strain Marseille-Q3630 to its closely related phylogenetic neighbours and proposed to establish for this strain the species name *Corynebacterium incognitum* sp. nov.

Materials and methods

Strain isolation

Corynebacterium incognitum strain Marseille-Q3630^T was isolated after 50 µL of liquid aerobic blood culture bottle (BACT/ALERT®, bioMerieux, Marcy l'Etoile, France) was seeded on Columbia Agar with 5% Sheep Blood media (Biomérieux, Marcy l'Etoile, France) and maintained à 37 °C for 24 hours. The strain is then routinely cultivated on Columbia Agar with 5% Sheep Blood media ((bioMerieux, Marcy l'Etoile, France) incubated in aerobiosis at 37 °C for 24 hours.

Strain identification

Matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) mass spectrometry (MS protein analysis was carried out using a Microflex spectrometer (Bruker Daltonics, Bremen, Germany) and spectra from strain Marseille-Q3630^T were imported into the MALDI BioTyper software (version 3.0, Bruker, Germany) and analysed by standard pattern matching (with default parameter settings). As MALDI-TOF analyse failed to identify this organism by analysing against our database [4] (https://www.mediterranee-infection.com/acces-ressources/base-de-donnees/urms-data-base/.), the genome was sequenced as described in the following section. BlastN tool from NCBI

https://doi.org/10.1016/j.nmni.2021.100893

were performed to compare the 16S best hits sequence from strain Marseille-Q3630 against the 16S database [5]. Phylogenetic trees for the 16s RNA and rpoB gene was obtained using the Maximum Likelihood method and Kimura 2-parameter within the MEGA 7 software [6].

Phenotypic tests

Different growth temperatures (20–56 °C), atmosphere conditions (anaerobic, aerobic and microaerophilic), using atmosphere generators (CampyGEN, Oxoid, USA) and pH (5.5–8.5) were tested. API ZYM, API Coryne and API 50 CH strips (BioMérieux, Marcy L'Etoile, France) were used to evaluate the biochemical properties of the strain in accordance with the manufacturer's instructions. Morphology was analysed by scanning electronic microscopy. Briefly, a colony was collected from agar and immersed into a 2.5 % glutaraldehyde fixative solution. The slide was gently washed in water; air-dried and examined to evaluate bacterial structure on a TM4000 microscope (Hitachi, Tokyo, Japan). Motility test was performed using the semisolid TCC media [7].

Genome sequencing

Genomic DNA (gDNA) of *Corynebacterium incognitum* strain Marseille-Q3630^T was extracted with the EZ1 biorobot (Qiagen) with EZ1 DNA tissues kit after a mechanical and chemical (lysozyme) pretreatment. gDNA was next sequenced on the MiSeq Technology (Illumina Inc, San Diego, CA, USA) with the paired end strategy and was prepared with the Nextera XT DNA sample prep kit (Illumina). Normalised libraries were pooled into a single library for sequencing on the MiSeq. Total information of 6.03 Gb was obtained from a 628 K/mm^2 cluster density with a cluster passing quality control filters of 96.38. The 12,164,746 paired end reads were filtered as per the read qualities. To improve the genome quality, an Oxford Nanopore approach was performed on 1D gDNA sequencing for the MinIon device using SQK-LSK109 kit. Library was constructed from 1 µg genomic DNA without fragmentation and end repair. About 753 actives pores were detected for the sequencing and the workflow WIMP was chosen for bioinformatic analysis in live. After 3.5 hours as run time and end life of the flow cell, 61.140 reads as raw data were generated.

Genome annotation and genome comparison

Genome assembly was proceeded using SPAdes software v3.10 [8]. Genome annotation was obtained through the NCBI Prokaryotic Genome Annotation Pipeline [9]. The TYGS online platform was used to determinate the closely related type strains. Determination of closest type strain genomes was done in two complementary ways: First, the Marseille-Q3630 genome was compared against all type strain genomes available in the TYGS database via the MASH algorithm, a fast approximation of intergenomic relatedness [10], and the ten type strains with the smallest MASH distances chosen per user genome. Second, an additional set of ten closely related type strains was determined via the 16S rDNA gene sequences. These were extracted from the user genomes using RNAmmer [11] and each sequence was subsequently BLASTed [5] against the 16S rDNA gene sequence of each of the currently 14,130 type strains available in the TYGS database. This was used as a proxy to find the best 50 matching type strains (in accordance with the bit score) for each user genome and to subsequently calculate precise distances using the Genome BLAST Distance Phylogeny approach under the algorithm 'coverage' and distance formula d5 [12]. These distances were finally used to determine the 10 closest type strain genomes for each of the user genomes. The Genome-to-Genome Distance Calculator Digital DDH values and confidence intervals were calculated using the recommended settings of the GGDC 2.1 The degree of genomic similarity based on the Orthologue group of genes of *Corynebacterium incognitum* strain Marseille-Q3630 with closely related species was estimated using the OrthoANI software [13].

Results

Strain identification and classification

Corynebacterium incognitum strain Marseille-Q3630^T was isolated from a 7-year-old girl that was consulting paediatric service for an acute febrile illness. Blood was collected and analysed in aerobic blood culture bottle (BACT/ALERT®, bioMerieux, Marcy l'Etoile, France). *Corynebacterium incognitum*

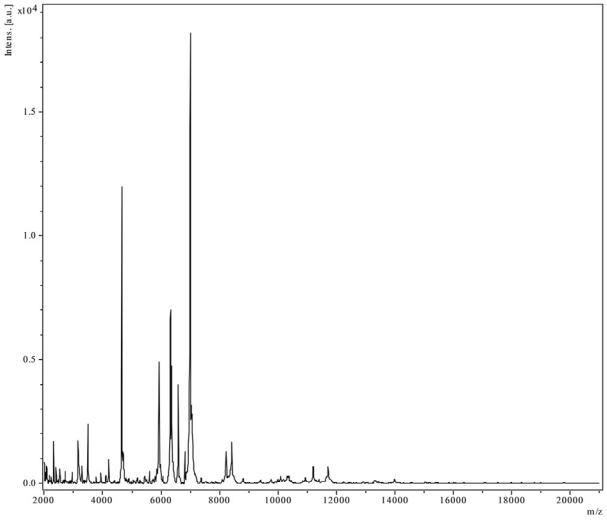

FIG. 1. MALDI-TOF MS reference mass spectrum. Spectra from 12 individual colonies of *Corynebacterium incognitum* strain Marseille-Q3630^T were compared and a reference spectrum was generated.

strain Marseille-Q3630^T failed to be identified by our systematic MALDI-TOF MS screening, suggesting that the corresponding species was not in the database (Fig. 1). Moreover, *Corynebacterium incognitum* strain Marseille-Q3630^T exhibited a 97.16% 16S rRNA sequence similarity with *Corynebacterium tuberculostearicum*^{T} (GenBank accession NR_028975.1), the phylogenetically closest bacterium standing in nomenclature and a 88.03% *rpoB* gene identity with *Corynebacterium aurimucosum* ATCC 700975 (extracted from the genome sequence CP001601.1) - shown to be more discriminant for *Corynebacterium* species, validating the <96.6% identity cutoff described by Khamis et al. [14]. (Fig. 2). Digital DNA–DNA hybridisation analysis between the novel organism and *Corynebacterium diphtheria* subsp. *lausannense* CHUV2995T exhibited value of 25.4%. Furthermore, genomic comparison using the OrthoANI parameter provided a value of 72.06% *Corynebacterium ureicelerivorans* strain IMMIB RIV-2301^T (Table 1 and Fig. S1).

Phenotypic characteristics

Colonies from strain Marseille-Q3630^T showed a white pigmentation and no haemolysis. Bacterial cells were gram-positive, non-motile, rod shaped, with a mean size of 0.5 × 0.9 µm determined by scanning electronic microscopy (Fig. 3). Strain Marseille-Q3630^T is facultatively aerobic. Optimal growth medium temperature, pH and NaCl concentration is comprised between 31–37 °C, 5.5–8.5 and 10–15 g/L, respectively. The sporulation test (20 minutes at 80 °C) is negative. Using API strips, positive reactions were shown for pyrazine-carboxamide, 2-naphthyl-phosphate, D-glucose, alkaline phosphatase, esterase lipase (C8), naphthol-AS-BI-phosphohydrolase, D-fructose. All other reactions tested were negative. In addition, this bacterium shows catalase positive and oxidase negative. These results are summarised in Table 2.

Genome properties

The assembly was achieved in a single contig with a coverage value 22.017x. The genome is 2.348.605 bp long with a 62.44% G + C content. The genome assembly of this strain was achieved on a single contig with a N_{50} value of 22.017. Of the 2.178 predicted genes, 2.083 were protein-coding genes and 63 were RNAs (3 16S rRNA, 4 5S rRNAs, 3 23S rRNAs, 50 tRNAs and 3 ncRNA) (Fig. S2) [15]. The genome properties and distribution of genes into COGs functional categories are detailed in Fig. S3. The *in silico* resistome of the strain Marseille-Q3630

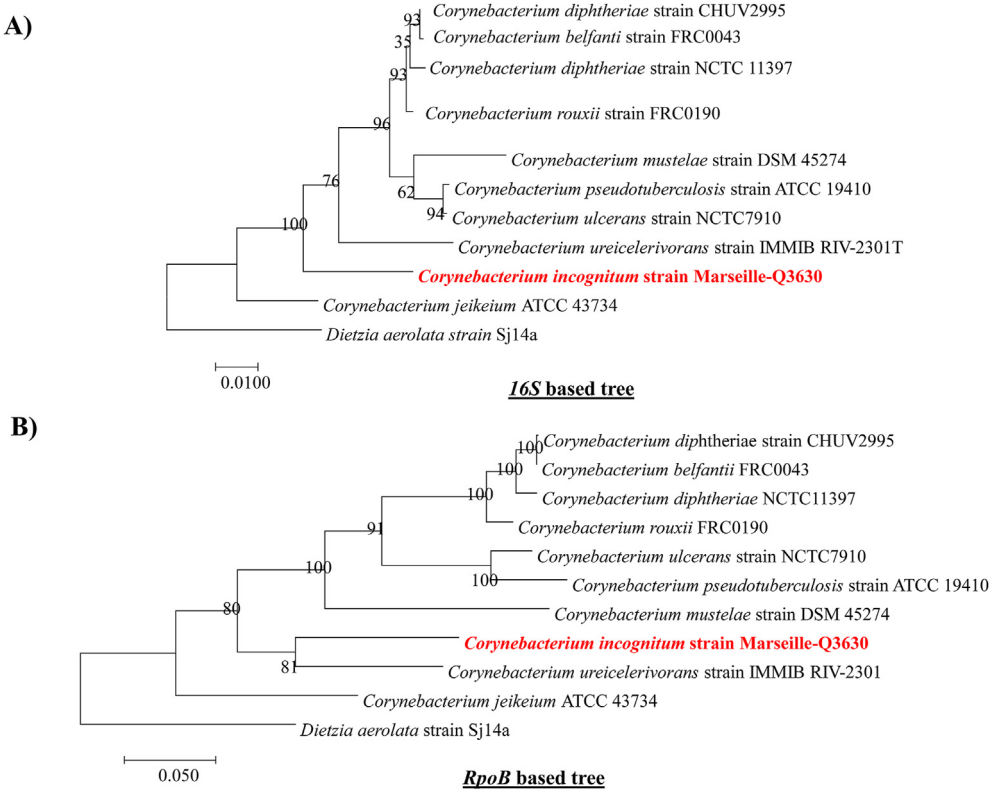

FIG. 2. A. *16S* rRNAgene (A) and *RpoB* gene (B) based phylogenetic trees highlighting the position of *Corynebacterium incognitum* sp., nov., strain Marseille-Q3630 (red) relative to other closely related bacterial taxa. Best hits of BlastN megablast was added to the relative pointed by the TYGS software. Sequences were aligned using Muscle v3.8.31 with default parameters and phylogenetic relationship inferred using the maximum likelihood method, with 1000 bootstrap replicates, within the MEGA software version 7.

TABLE 1. Digital DNA–DNA hybridisation values obtained by sequence comparison of all studied genomes using TYGS comparison server using the 2nd formula

Subject strain	OrthoANI value (Fig. S1)	dDDH (in %) with *Corynebacterium incognitum strain Marseille-Q3630T*	Confidence interval (in %)	G + C content difference (in %)
Corynebacterium diphtheriae subsp. lausannense CHUV2995	69.3094	**25,4**	[23.1 - 27.9]	8,49
Corynebacterium rouxii FRC0190 T	69.6315	25,1	[22.8 - 27.6]	9,21
Corynebacterium mustelae DSM 45274	68.4165	24,6	[22.3 - 27.1]	9,86
Corynebacterium diphtheriae NCTC 11397	69.5648	24,5	[22.2 - 27.0]	8,91
Corynebacterium belfantii FRC0043	69.4349	24,1	[21.8 - 26.6]	8,81
Corynebacterium pseudotuberculosis ATCC 19410	68.5154	23,8	[21.4 - 26.2]	10,25
Corynebacterium pseudotuberculosis DSM 20689	Not found in public database	23,8	[21.4 - 26.2]	10,25
Corynebacterium ureicelerivorans DSM 45051	**72.0657**	23,6	[21.3 - 26.0]	2,57
Corynebacterium ulcerans NCTC 7910	68.928	23,6	[21.3 - 26.1]	9,12
Corynebacterium jeikeium ATCC 43734	71.0532	23,6	[21.3 - 26.0]	0,8

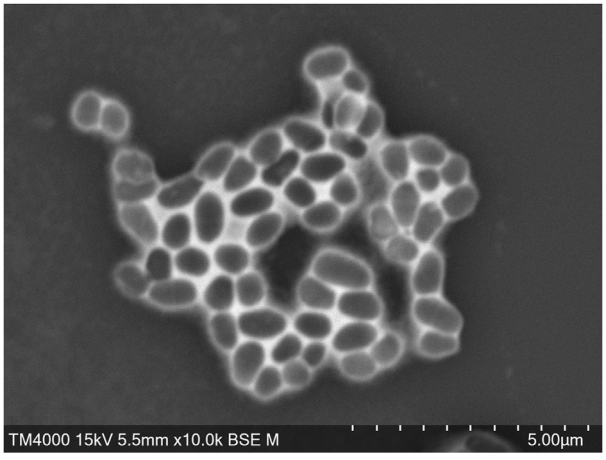

FIG. 3. Scanning electron microscopy of *Corynebacterium incognitum* sp. nov., strain Marseille-Q3630^T using a Tabletop microscope TM 4000 plus (Hitachi, Tokyo, Japan). The scale bar represents 5 µm.

shows *erm(X)_4* gene with 94.74% identity percentage, notably involved in erythromycin resistance.

Discussion

Using the taxono-genomics concept, i.e. the combination of the genomic and phenotypic properties of a putative new taxon [16], we have characterised a new bacterial species representing a new species within the family *Corynebacteriaceae* found in human. Although it was found in a blood culture of a 7-year-old girl, *Corynebacterium* spp are known to be associated with such contaminated samples. It is important to keep in mind that *Corynebacterium incognitun* could also be such skin contaminant [17]. This strain is most closely related to *Corynebacterium tuberculostearicum* ^{T} with a 16S rRNA sequence similarity value

TABLE 2. Differential characteristics of *Corynebacterium incognitum* strain Marseille- Q3630^T and closest species standing in nomenclature

Characteristics	*C. incognitum*	*C. tuberculostearicum*	*C. tuscaniense*	*C. macgninleyi*	*C. simulans*	*C. accolens*
Properties	Marseille-Q3630	Medalle X	ISS-5309	JCL-2	DSM 44415	CIP 104783T
Oxygen requirement	Facultative	Facultative	+	Facultative	Facultative	Facultative
Gram Strain	+	+	+	+	+	+
Motility	–	–	–	–	–	–
Endospore formation	–	–	–	na	–	–
Optimum temperature for growth (°C)	31–37 °C	na	na	37 °C	na	37 °C
Production of:						
Alkaline phosphatase	+	+	+	+	+	–
Catalase	+	+	+	+	+	+
Oxidase	–	–	na	–	na	–
α-Glucosidase	–	–	–	–	–	+
β-galactosidase	–	–	–	–	–	–
Acid from:						
N-Acetylglucosamine	–	+	–	–	–	+
L-arabinose	–	na	na	+	–	na
D-ribose	–	+	–	+	+	+
D-mannose	–	+	na	na	+	+
D-mannitol	–	–	–	+	–	–
D-glucose	+	+	+	+	+	+
D-fructose	+	+	na	+	+	+
D-maltose	–	–	+	–	–	–
D-lactose	–	–	–	–	–	–
Genome size	2,348,605	2,453,172	2,263,530	2,419,073	2,598,702	2,465,636
Isolation source	Human healthy skin	Clinical and food samples	Blood cultures of a patient with endocarditis	Bacterial flora of certain human body sites	Clinical samples	Clinical and food samples

TABLE 3. Description of *Corynebacterium incognitum* sp. nov. strain Marseille-Q3630^T

Type of description	New description
Species name	*incognitum*
Genus name	*Corynebacterium*
Specific epithet	*Corynebacterium*
Species Status	sp. nov.
Species etymology	Corynebacterium incognitum strain Marseille-Q3630T. Gr. fem. n. korynê, a club; L. neut. n. bacterium, a rod, and in biology a bacterium (so called because the first ones observed were rod-shaped); N.L. neut. n. Corynebacterium, a club bacterium. Incognitum, lat. incognitus « id. », der. cognitus, known.
Authors	Manon Boxberger, Angéline Antezack, Sibylle Magnien, Nadim Cassir, Bernard La Scola
Designation of the type strain	Marseille Q3630
Strain collection number	CSUR:3630
16S rRNA gene accession number	MT772002
Genome accession number	GCA_014217255.1
Genome status	Whole genome
Genome size	2,348,605-bp
GC%	62.44%
Country of origin	France
Date of isolation	2020
Source of isolation	Human healthy skin
Growth medium, incubation	Columbia Agar with 5% Sheep Blood media (Biomérieux, France), 37 °C, Aerobiosis, 1 day
Gram stain	Positive
Cell shape	Rod shaped
Cell size	0.9 × 0.5 µm
Motility	—
Sporulation	—
Colony morphology	White
Temperature range	21 °C–56 °C
Temperature optimum	37 °C
Relationship to O2	Facultative
O2 for strain testing	Anaerobiosis, microaerophilic, aerobiosis
Oxidase	Negative
Catalase	Positive

of 97.16%. Furthermore, genomic comparison using the OrthoANI parameter provided a value of 72.06% with *Corynebacterium ureicelerivorans*^{T} and a dDDH value of 25.4 % with *Corynebacterium diphtheria* subsp. lausannense CHUV2995T. Indeed, although they are not universal cutoff, a value lower than 70% of DDH and 95–96% of ANI are common indicators for a new species discovery, this new strain shows a *rpoB* gene identity with *Corynebacterium aurimucosum* of 88.03%, validating the <96.6% identity cutoff [14]. Taken altogether, these parameters have prompt us to propose *Corynebacterium incognitum* strain Marseille-Q3630^T as a new member of *Corynebacterium* genus. Gr. fem. n. *korynê*, a club; L. neut. n. *bacterium*, a rod, and in biology a bacterium (so called because the first ones observed were rod-*shaped*); N.L. neut. n. *Corynebacterium, a club bacterium. Incognitum, lat. incognitus « id. », der. cognitus,* known (Table 3).

Deposit in culture collections and sequences database

Corynebacterium incognitum strain Marseille-Q3630^T was deposited in CSUR collections under accession CSUR-Q3630. The 16S rRNA and genome sequences available in GenBank under accession numbers MT772002 and GCA_014217255.1, respectively.

Ethics committee

The study was validated by the ethics committee of the Institut Fédératif de Recherche IFR48.

Transparency declaration

None to declare. MB is PhD granted by the collaboration between M&L Laboratories and Aix Marseille University referenced PVM:2018-200. This study was supported by the French State managed by the National Research Agency under the "Investissements d'avenir (Investments for the Future)" program under the reference ANR-10-IAHU-03 (Méditerranée Infection) and by the Région Provence-Alpes-Côte d'Azur and the European funding FEDER PRIMI.

Acknowledgements

The authors acknowledge Dr Alexandre Toro and the clinical microbiology team of Martigues hospital for sending the sample to identify this strain. The authors are indebted to Ludivine Brechard for sequencing the genome and the platform of electron microscopy of IHU-MI for the electron micrographs.

Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.nmni.2021.100893.

Fig. S1: Heatmap generated with OrthoANI values calculated using the OAT software between *Corynebacterium incognitum* sp. nov., strain Marseille-Q3630^T and other closely related species standing in nomenclature.

Fig. S2: Graphical circular map of the genome from strain Marseille-Q3630^T, obtained by CG view tool [15].

Fig. S3: Distribution of functional classes of predicted genes according to the Clusters of Orthologous Groups of proteins of *C. incognitum* sp. nov. other closely related bacterial taxa and associated table.

Table S1: Distribution of functional classes of the predicted genes in *C. incognitum* strain Marseille-Q3630^T and closest species according to the clusters of orthologous groups of proteins.

References

[1] Lehmann Karl Bernhard, Royal College of Physicians of Edinburgh. Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik, vol. 1. Lehmann: Munchen; 1896.

[2] LPSN PAC. List of prokaryotic names with standing in nomenclature (bacterio.net), 20 years on. International Journal of Systematic and Evolutionary Microbiology 2018;68:1825–9. https://doi.org/10.1099/ijsem.0.002786.

[3] Oliveira A, Oliveira LC, Aburjaile F, Benevides L, Tiwari S, Jamal SB, et al. Insight of genus Corynebacterium: ascertaining the role of pathogenic and non-pathogenic species. Frontiers in Microbiology 2017;8:1937. https://doi.org/10.3389/fmicb.2017.01937.

[4] Grégory D, Chaudet H, Lagier J-C, Raoult D. How mass spectrometric approaches applied to bacterial identification have revolutionized the study of human gut microbiota. Expert Review of Proteomics 2018;15:217–29. https://doi.org/10.1080/14789450.2018.1429271.

[5] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K. BLAST+: architecture and applications. BMC Bioinformatics 2009;10. https://doi.org/10.1186/1471-2105-10-421.

[6] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 2016;33:1870–4. https://doi.org/10.1093/molbev/msw054.

[7] Tittsler RP, Sandholzer LA. The use of semi-solid agar for the detection of bacterial motility. Journal of bacteriology 1936;31:575–80.

[8] 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. Journal of Computational Biology 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021.

[9] Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Research 2016;44:6614–24. https://doi.org/10.1093/nar/gkw569.

[10] Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology 2016;17. https://doi.org/10.1186/s13059-016-0997-x.

[11] Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research 2007;35:3100–8. https://doi.org/10.1093/nar/gkm160.

[12] Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013;14. https://doi.org/10.1186/1471-2105-14-60.

[13] Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. International Journal of Systematic and Evolutionary Microbiology 2016;66:1100–3. https://doi.org/10.1099/ijsem.0.000760.

[14] Khamis A, Raoult D, La Scola B. rpoB gene sequencing for identification of Corynebacterium species. Journal of Clinical Microbiology 2004;42:3925–31. https://doi.org/10.1128/JCM.42.9.3925-3931.2004.

[15] Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Research 2008;36:W181–4. https://doi.org/10.1093/nar/gkn179.

[16] Abdallah RA, Beye M, Diop A, Bakour S, Raoult D, Fournier P-E. The impact of culturomics on taxonomy in clinical microbiology. Antonie Van Leeuwenhoek 2017;110:1327–37. https://doi.org/10.1007/s10482-017-0871-1.

[17] Doern GV, Carroll KC, Diekema DJ, Garey KW, Rupp ME, Weinstein MP, et al. Practical guidance for clinical microbiology Laboratories: a comprehensive update on the problem of blood culture contamination and a discussion of methods for addressing the problem. Clinical Microbiology Reviews 2019;33. https://doi.org/10.1128/CMR.00009-19. CMR.00009-19, e00009-19.