Sequencing and description of the genome of a strain of *Stenotrophomonas geniculata* isolated from a patient infected with COVID-19 at Hospital Regional No.1 de Charo, Michoacán, México

Luis Jose Flores-Alvarez[1*], Irma Martínez-Flores[2*], Patricia Bustos[2*], Anel Gómez-García[1], Sergio Gutiérrez-Castellanos[3], Augusto César Poot-Hernández[4], Marina Arredondo-Santoyo[1§]

[1]Centro de Investigación Biomédica de Michoacán-IMSS. Morelia, Michoacán, México, Mexican Social Security Institute, Mexico City, Mexico City, Mexico

[2]Centro de Ciencias Genomicas, UNAM. Cuernavaca, Morelos, Universidad Nacional Autónoma de México, Mexico City, Mexico City, Mexico

[3]Centro de Investigación Biomédica de Michoacán-IMSS. Morelia, Michoacán, México. , Mexican Social Security Institute, Mexico City, Mexico City, Mexico

[4]Instituto de Fisiología Celular, UNAM. Cd de México, Universidad Nacional Autónoma de México, Mexico City, Mexico City, Mexico

[§]To whom correspondence should be addressed: mas_mina@hotmail.com

[*]These authors contributed equally.

Abstract

Stenotrophomonas is a bacterial genus that can be found in various environments, such as water, soil, and clinical samples. Due to their high genetic and phenotypic diversity, it is difficult to properly identify and classify all isolates. The COVID-19 pandemic caused an increase in nosocomial infections, which played a major role in the high mortality rate among patients in intensive care. This is the first report of the identification of *S. geniculata* as a nosocomial opportunistic pathogen isolated from a patient with COVID-19. Their genome was isolated, sequenced, and assembled, and it consists of 4,488,090 bp in 24 contigs, 4,103 coding sequences, and a G+C content of 66.58%.
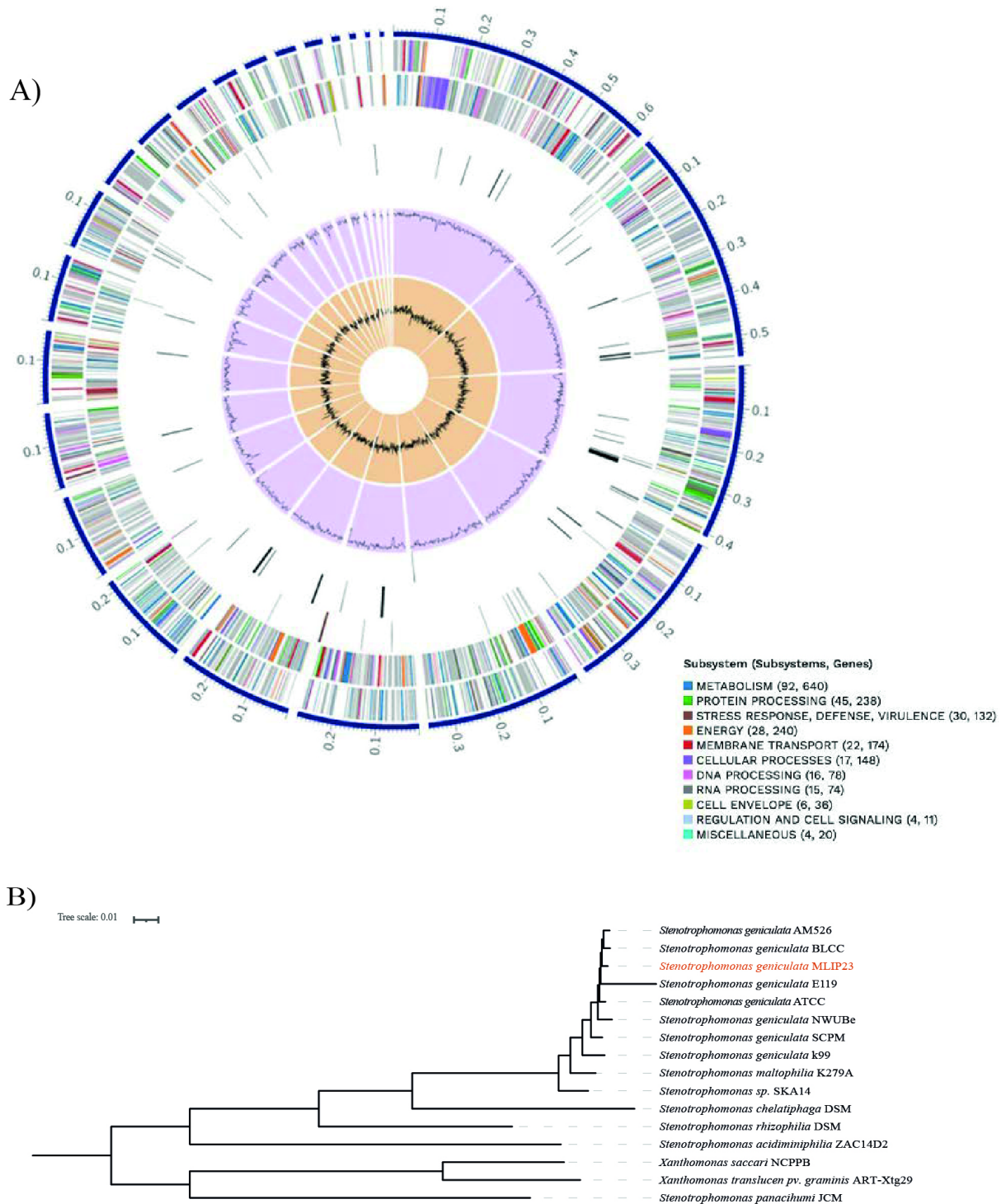
A)

Subsystem (Subsystems, Genes)
- METABOLISM (92, 640)
- PROTEIN PROCESSING (45, 238)
- STRESS RESPONSE, DEFENSE, VIRULENCE (30, 132)
- ENERGY (28, 240)
- MEMBRANE TRANSPORT (22, 174)
- CELLULAR PROCESSES (17, 148)
- DNA PROCESSING (16, 78)
- RNA PROCESSING (15, 74)
- CELL ENVELOPE (6, 36)
- REGULATION AND CELL SIGNALING (4, 11)
- MISCELLANEOUS (4, 20)

B)

Tree scale: 0.01

Stenotrophomonas geniculata AM526
Stenotrophomonas geniculata BLCC
Stenotrophomonas geniculata MLIP23
Stenotrophomonas geniculata E119
Stenotrophomonas geniculata ATCC
Stenotrophomonas geniculata NWUBe
Stenotrophomonas geniculata SCPM
Stenotrophomonas geniculata k99
Stenotrophomonas maltophilia K279A
Stenotrophomonas sp. SKA14
Stenotrophomonas chelatiphaga DSM
Stenotrophomonas rhizophilia DSM
Stenotrophomonas acidiminiphilia ZAC14D2
Xanthomonas saccari NCPPB
Xanthomonas translucen pv. graminis ART-Xtg29
Stenotrophomonas panacihumi JCM

Figure 1. Genome annotations graph and Phylogenetic tree based on the 1479 core genes of each genome:

A) Genome annotation graph. The graph shows the annotations of the assembled genome of *Stenotrophomonas* MLIP23 using the comprehensive genome analysis service of the Bacterial and Viral Bioinformatics Resource Center (BVBRC. https://www.bv-brc.org/) (Wattam et al., 2017). The analyses are performed considering annotation statistics and comparison with other genomes in BVBRC within this same species. The annotation graph shows, starting from the outer rings to the inner ones, 24 contigs, CDS in the forward strand, CDS in the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to known virulence factors, GC content (66.6%) and at the end the GC skew. The CDS colors on the forward and reverse strands indicate the subsystems to which these genes belong. B) Phylogenetic tree. Phylogenetic tree based on the 1479 core genes of families determined with BPGA (Chaudhari et al., 2016), corresponding to the most representative reference genomes, to generate the support in the tree and to determine the phylogenetic placement of MLIP23 genome; which was located in the *S. geniculata* branch.

Description

The coronavirus is a diverse group of viruses that infect different types of animals (Hu et al., 2021). At the end of December 2019, several health units in the city of Wuhan, China, reported a group of patients with pneumonia of unknown cause (Wu et al., 2020). The RNA sequencing of this virus resulted in a new betacoronavirus not previously identified (Wu et al., 2020; Zhu et al., 2020). This new, highly communicable disease became known as coronavirus disease 2019 (COVID-19) (Wu et al., 2020; Hui et al., 2020). Due to the COVID-19 pandemic, a percentage of these patients required care treatment in the intensive care unit (ICU) for invasive ventilation (Maes et al., 2021). Nosocomial infections were common in patients with very severe complications from COVID-19 who were in the ICU, and there seemed to be a connection due to very long stays in that unit (Bardi et al., 2021). Nosocomial diseases have a special place in hospital clinics due to their importance as one of the main causes of morbidity and mortality. The highest prevalence of nosocomial infections occurs in ICU, surgical, and orthopedic wards, being the most vulnerable population (Lisboa and Rello 2008). The nosocomial diseases in COVID-19 patients exhibited different groups of Gram-positive (*Clostridium sp, Staphylococcus aureus*) and Gram-negative bacteria (*Escherichia coli, Proteus, Klebsiella, Enterobacter, Serratia marcescens, Pseudomonas spp, Burkholderia cepacia, Stenotrofomonas maltofilia*) (Lisboa and Rello 2008).

Bacteria are classified and identified to distinguish between strains according to the criteria of interest to microbiologists and scientists, with species identification being the most important level of classification. Usually, the identification and verification of species involves biochemical, phenotypic criteria, and sometimes serological evaluations. However, currently, genetic criteria are necessary to define a species, particularly when dealing with atypical, rare, or recently described strains (Baron 1996).

The genus Stenotrophomonas is characterized by having at least 18 recognized species. These bacteria are found in various habitats such as soil, water, associated with plants, animals, and the human body. They are Gram-negative bacteria, morphologically uniform (individual cells or in pairs that are slightly elongated and curved), and they do not form endospores. Generally, they can have one or two polar flagella. Stenotrophomonas uses a limited number of carbon sources. Most species of Stenotrophomonas show positive results for oxide and nitrate reductase tests, as well as for citrate, malonate, hydrolysis of gelatin, and Tween 80 (Ghosh et al., 2020).

The species that have been found to infect humans have been sequenced and they are not highly virulent organisms, its virulence factors are strain-specific, e. g. the strain of *S. maltophilia* K279a possesses an unusual cell density-signaling pathway and mobile genetics elements such as hemagglutinin and hemolysins, this bacteria is truly an opportunistic microorganism for environmental adaptations (Crossman et al., 2008; Hu et al ., 2021).

Stenotrophomonas geniculata, also known as *Pseudomonas geniculata*, forms a lineage that is phylogenetically highly divergent with other strains of *Pseudomonas*. *S. geniculata* have been observed to be more closely related to *Stenotrophomonas* species (Saati-Santamaria et al., 2021). It is a Gram-negative bacillus that does not form spores, is mobile since it has a polar flagellum, and can form extracellular slime. In culture, pleomorphic forms may appear. On meat peptone agar, colonies are circular, convex, entire, smooth, and transparent. On agar slope, there is good growth (Lysenko 1960). The G+C genomic content the type of strain is 66.2%, and its genomic size is approximately 4.81 Mbp (Saati-Santamaría et al., 2021).

S.geniculata had not been described as a pathogen in humans, only *S. maltophili*a and *S. africana* were the species reported to date (Drancourt et al., 1997). *S. maltophilia* and *S. africana* are opportunistic pathogens with a very high potential to affect health, not only due to their ability to form biofilm (Pompilio et al., 2011; Brooke 2012; Flores-Treviño et al., 2014; Maes et al., 2021), antibiotic resistance genes, as well as potentially mobile genomic regions and coding for pili and fimbriae may contribute to greater resistance to different antimicrobials (Crossman et al., 2008).

The COVID-19 pandemic led to the emergence of various opportunistic organisms, some already documented in hospitals, and others, like the one described in this study, such as *S. geniculata*. This marks the first reported instance of this species affecting humans. When classification is done by colony morphology and biochemical test there can be inaccuracies that are only revealed when the organism is sequenced. Nowadays, molecular tools enable a more precise identification. In this study, we sequenced and assembled the genome of *Stenotrophomonas geniculata* isolated from a patient infected with COVID-19.

As a result, the genome had a length of 4,488,090 bp in 24 contigs, with 4,103 coding sequences, 66 tRNA, and a G+C content of 66.58%. No plasmids were identified. A total of 3,013 proteins with functional assignments were identified, mainly related to metabolism and protein processing. According to the PATRIC database, 42 antibiotic resistance genes and one virulence factor were identified. The resistance mechanisms included antibiotic inactivation enzymes (APH(3´)-II/APH(3´)-XV, L1 family), antibiotic target in susceptible species (Alr, Ddl, dxr, EF-G, EF-Tu, folA, Dfr, folP, gyrA, gyrB, Iso-tRNA, kasA, MurA, rho, rpoB, rpoC, S10p, S12p), and efflux pumps conferring antibiotic resistance (EmrAB-OMF, EmrAB-TolC, MacA, MacB, MdtABC-TolC, TolC/OpmH).

Methods

Strain isolation: The strain was isolated from a pharyngeal exudate of a patient with COVID-19, it was identified using the VITEK 2 Compact (Biomérieux) automated bacteria identification system.

Total DNA isolation: Total DNA was obtained from a pure culture which was grown in 5 mL of Luria Bertani broth overnight at 37°C. The pellet was recovered and bacterial genomic DNA extraction was performed using the ZymoBOIMICS DNA Miniprep Kit (ZymoReseach US Cat D4300) according to the manufacturer's instructions.

Genome sequencing: The genome sequencing service was carried out using the services of the Novogene company, according to its protocols, the genomic DNA was randomly fragmented by sonication, the DNA fragments were polished at the ends, A-sequences were added to the ends and ligated with the full-length adapters for Illumina sequencing. It was amplified by PCR with indexed P5 and P7 oligonucleotides. The PCR products as the final construction of the libraries were purified with the AMPure XP system. Libraries were then checked for size distribution using the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) and quantified using real-time PCR (to meet the 3 nM criteria). Qualified libraries are loaded onto Illumina sequencers after mixing according to their effective concentration and expected data volume. The result was a total raw data of 5.8G, raw reads 38767250, effective (%) 99.82, error (%) 0.03, Q20(%) 97.47, Q30(%)93.03, and GC(%) 65.42. The sequencing quality of the raw sequences obtained was analyzed using the FASTQC program (V.0.11.8).

Genome assembly and annotation. After trimming adapters of the reads with Trim Galore v0.6.4 tool, and to ensure the accuracy and reliability of the genome assembly, we assemble readings with two algorithms, MEGAHIT v1.2.8 and SPAdes v3.13.1. The assemblies were compared with the QUAST v5.0.2 tool; according to its metrics, the best assembler was SPAdes. Although both assemblers provide similar values in terms of the complete size of the assembly and percentage of GC, SPAdes showed longer reading contigs and higher N50 (360,976), suggesting higher assemble integrity, so Spades was the selected assembler; thus, the processed reads were assembled on a high-coverage (500 – 600 X), resulting in 24 contigs with a total length of 4,488,090 bp and an average G+C content of 66.58%. The pre-annotation of the genome was obtained from "The Bacterial and Viral Bioinformatics Resource Center" (BV-BRC, https://www.bv-brc.org/) using the default parameters. The website uses RAST tool to pre-annotate the antibiotic-resistance and virulence-related genes (Brettin et al., 2015). The genome contained 4,103 protein-coding sequences (CDS), 66 transfer RNA (tRNA) genes, and 5 ribosomal RNA (rRNA) genes (Fig. 1 A, genome annotations graph). Using the quality tools in BV-BRC, it was determined that the genome is 100% complete (with checkM program), with a Coarse consistency of 99.6 and Fine Consistency of 99.5.

The taxonomic classification. The genome annotation was carried out with proteins coming, on the one hand, from the reference genomes proposed by BV-BRC based on its phylogenetic analysis, and on the other, from the representative genomes obtained with the GTDBTK v2.1.1 program. Using GTDB-Tk version r207 (Pierre et al., 2022) assigning their classification according to their homology to the closest genomes. These genomes were obtained from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) and their proteins were annotated with Prokka v1.14.6 (prodigal). The information was used to estimate the phylogenetic classification of the studied genome.

Phylogenetic analysis. The BPGA v1.3 tool was used to conduct the phylogenetic analysis with 1479 core genes (Chaudhari et al., 2016). Briefly, BPGA uses clustering tools with default values, USEARCH ortholog clustering tool with a 50% identity cutoff for ortholog clustering of genomes, and with the proteins in fasta format, we were able to identify orthologous groups of functional genes. To visualize the phylogeny, inference of a phylogenetic tree (Fig. 1 B) using the maximum likelihood method was performed using the phylogenetic software package IQ-TREE v2.1.2 with the best parameters (1000 bootstrap replicates and LG+F+R7 model; Chernomor et al., 2016; Nguyen et al., 2015). Briefly, IQ-TREE automatically detects the sequence type and performs a rapid selection of the best-fitting replacement model via ModelFinder (Kalyaanamoorthy et at.,

2017), an effective tree search algorithm (Chernomor et al., 2016), and a Bootstrap approximation (Minh et al., 2014; Hoang et al., 2018). To draw and edit the phylogenetic trees, the Interactive Tree Of Life program (iTOL v6.6) was used (Letunic et al., 2019). The analysis showed that all strains used were closely related, and it could be estimated that local strain belongs to *Stenotrophomonas geniculata*, with a 98% identity and over 90% coverage when compared to clones from the same branch (Figure 1 B).

Reagents

This whole-genome shotgun project has been deposited at GenBank. IBID SUB13887264, Bioproject PRJNA1025098, BioSample SAMN37712844. Organism *Stenotrophomonas geniculata*. Sample: MLIP23.

References

Anzai Y, Kim H, Park JY, Wakabayashi H, Oyaizu H. 2000. Afiliación filogenética de las pseudomonas basada en la secuencia de ARNr 16S. Revista Internacional de Microbiología Sistemática y Evolutiva 50: 1563-1589. DOI: 10.1099/00207713-50-4-1563

Bardi T, Pintado V, Gómez-Rojo M, Escudero-Sanchez R, Azzam López A, Diez-Remesal Y, et al., Pestaña. 2021. Infecciones nosocomiales asociadas a COVID-19 en la unidad de cuidados intensivos: características clínicas y evolución. Revista europea de microbiología clínica y enfermedades infecciosas 40: 495-502. DOI: 10.1007/S10096-020-04142-W

Baron EJ. Classification. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 3. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8406/

Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ y otros, Xia. 2015. RASTtk: una implementación modular y extensible del algoritmo RAST para crear canales de anotación personalizados y anotar lotes de genomas. Informes científicos 5: 10.1038/srep08365. DOI: 10.1038/srep08365

Brooke JS. 2012. Stenotrophomonas maltophilia: un patógeno oportunista global emergente. Reseñas de microbiología clínica 25: 2-41. DOI: 10.1128/CMR.00019-11

Chaudhari NM, Gupta VK, Dutta C. 2016. BPGA: un proceso de análisis pangenómico ultrarrápido. Informes científicos 6: 10.1038/srep24373. DOI: 10.1038/srep24373

Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: clasificación amigable con la memoria con la base de datos de taxonomía del genoma. Bioinformática 38: 5315-5316. DOI: 10.1093/bioinformatics/btac672

Chernomor O, von Haeseler A, Minh BQ. 2016. Estructura de datos consciente de Terrace para inferencia filogenómica a partir de supermatrices. Biología Sistemática 65: 997-1008. DOI: 10.1093/sysbio/syw037

Crossman LC, Gould VC, Dow JM, Vernikos GS, Okazaki A, Sebaihia M, et al., Avison. 2008. El genoma completo, el análisis comparativo y funcional de Stenotrophomonas maltophilia revela un organismo fuertemente protegido por determinantes de resistencia a los medicamentos. Biología del genoma 9: R74. DOI: 10.1186/gb-2008-9-4-r74

Drancourt M, Bollet C, Raoult D. 1997. Stenotrophomonas africana sp. nov., un patógeno humano oportunista en África. Revista Internacional de Bacteriología Sistemática 47: 606-606. DOI: 10.1099/00207713-47-2-606

Flores-Treviño S, Gutiérrez-Ferman JL, Morfín-Otero R, Rodríguez-Noriega E, Estrada-Rivadeneyra D, Rivas-Morales C, et al., Garza-González. 2014. Stenotrophomonas maltophilia en México: resistencia antimicrobiana, formación de biopelículas y diversidad clonal. Revista de Microbiología Médica 63: 1524-1530. DOI: 10.1099/jmm.0.074385-0

Ghosh R, Chatterjee S, Mandal NC. 2020. Estenotrofomonas. Microbios beneficiosos en agroecología: 427-442. DOI: 10.1016/B978-0-12-823414-3.00020-4

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2017. UFBoot2: Mejora de la aproximación de Bootstrap ultrarrápida. Biología Molecular y Evolución 35: 518-522. DOI: 10.1093/molbev/msx281

Hu B, Guo H, Zhou P, Shi ZL. 2020. Características del SARS-CoV-2 y COVID-19. Nature Reviews Microbiología 19: 141-154. DOI: 10.1038/s41579-020-00459-7

Hu M, Li C, Xue Y, Hu A, Chen S, Chen Y, et al., Zhou. 2021. Aislamiento, caracterización e investigación genómica de una cepa fitopatógena de *Stenotrophomonas maltophilia* . Fitopatología® 111: 2088-2099. DOI: 10.1094/PHYTO-11-20-0501-R

Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al., Petersen. 2020. La continua amenaza epidémica de 2019-nCoV que suponen los nuevos coronavirus para la salud mundial: el último brote de nuevo coronavirus de 2019 en Wuhan,

microPublication
BIOLOGY μP

China. Revista Internacional de Enfermedades Infecciosas 91: 264-266. DOI: 10.1016/j.ijid.2020.01.009

Janda JM, Abbott SL. 2002. Identificación bacteriana para publicación: ¿Cuándo es suficiente? Revista de Microbiología Clínica 40: 1887-1891. DOI: 10.1128/JCM.40.6.1887-1891.2002

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: selección rápida de modelos para estimaciones filogenéticas precisas. Métodos de la naturaleza 14: 587-589. DOI: 10.1038/nmeth.4285

Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: actualizaciones recientes y nuevos desarrollos. Investigación de ácidos nucleicos 47: W256-W259. DOI: 10.1093/nar/gkz239

Lisboa T, Rello J. 2008. Prevención de infecciones nosocomiales: estrategias para mejorar la seguridad de los pacientes en la Unidad de Cuidados Intensivos. Medicina Intensiva 32: 248-252. DOI: 10.1016/s0210-5691(08)70947-1 Cite

Lysenko O. 1961. Pseudomonas: un intento de realizar una clasificación general. Revista de Microbiología General 25: 379-408. DOI: 10.1099/00221287-25-3-379

Maes M, Higginson E, Pereira-Dias J, Curran MD, Parmar S, Khokhar F, et al., Conway Morris. 2021. Neumonía asociada a ventilador en pacientes críticos con COVID-19. Cuidados críticos 25: 10.1186/s13054-021-03460-5. DOI: 10.1186/S13054-021-03460-5

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Aproximación ultrarrápida para Bootstrap filogenético. Biología Molecular y Evolución 30: 1188-1195. DOI: 10.1093/molbev/mst024

Munita JM, Arias CA. 2016. Mecanismos de resistencia a los antibióticos. Espectro de microbiología 4: 10.1128/microbiolspec.vmbf-0016-2015. DOI: 10.1128/microbiolspec.VMBF-0016-2015

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: un algoritmo estocástico rápido y eficaz para estimar filogenias de máxima verosimilitud. Biología Molecular y Evolución 32: 268-274. DOI: 10.1093/molbev/msu300

Patil PP, Midha S, Kumar S, Patil PB. 2016. Secuencia del genoma de cepas tipo del género Stenotrophomonas. Fronteras en Microbiología 7: 10.3389/fmicb.2016.00309. DOI: 10.3389/fmicb.2016.00309

Pompilio A, Crocetta V, Ghosh D, Chakrabarti M, Gherardi G, Vitali LA, Fiscarelli E, Di Bonaventura G. 2016. Stenotrophomonas maltophilia Diversidad fenotípica y genotípica durante una colonización de 10 años en los pulmones de un paciente con fibrosis quística. Fronteras en Microbiología 7: 10.3389/fmicb.2016.01551. DOI: 10.3389/fmicb.2016.01551

Van Den Mooter M, Swings J. 1990. Análisis numérico de 295 características fenotípicas de 266 cepas de Xanthomonas y cepas relacionadas y una taxonomía mejorada del género. Revista Internacional de Bacteriología Sistemática 40: 348-369. DOI: 10.1099/00207713-40-4-348

Wu JT, Leung K, Leung GM. 2020. Previsión inmediata y previsión de la posible propagación nacional e internacional del brote de 2019-nCoV originado en Wuhan, China: un estudio de modelado. The Lancet 395: 689-697. DOI: 10.1016/S0140-6736(20)30260-9

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al., Tan. 2020. Un nuevo coronavirus procedente de pacientes con neumonía en China, 2019. New England Journal of Medicine 382: 727-733. DOI: 10.1056/NEJMOA2001017

Wattam AR, Davis JJ, Assaf R, Boisvert Sb, Brettin T, Bun C, et al., Stevens. 2016. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Research 45: D535-D542. DOI: 10.1093/nar/gkw1017