 doi:10.20944/preprints202209.0098.v1

Article

Comparison of Two Bacterial Characterization Techniques for the Genomic Analysis of Aquatic Environments

Roberto Bonnelly [1,], Victor V. Calderón [1], Confesora Pinales [1,], Irene Ortiz [2,4], Argeny Ovando [2,4], Willy Lara [3], Santo E. Mateo-Perez [3],Oscar Cardenas-Alegria [5], Rommel T. Ramos[5], Edian F. Franco [1,2,3‡]* and Luis O. Maroto [1,2‡]*

[1] Instituto Tecnologico de Santo Domingo (INTEC), Santo Domingo, Dominican Republic
[2] Instituto Superior de Formación Docente Salomé Ureña (ISFODOSU), Santo Domingo, Dominican Republic
[3] Instituto de Biotecnologia e Industria (IIBI) Santo Domingo, Dominican Republic
[4] Universidad ISA (UNISA), Santiago, Dominican Republic
[5] Federal University of Para, Para, Brazil
* Correspondence: edian.franco@intec.edu.do (E.F); luis.maroto@intec.edu.do (L.M)
‡ These authors contributed equally to this work.

Abstract: Our research team compared the performance of matrix-assisted laser desorption/ionization followed by a time of flight (MALDI-TOF) mass spectrometry and genomic DNA extraction followed by sequencing, assembly and alignment for phylogenetic assessment. We performed these comparisons to determine our methodology's overall efficacy and accuracy for environmental bacteria. In addition, we collected samples from various contaminated rivers in the Dominican Republic. For both methods, we analyzed these results and reported the main differences between each method.

Keywords: MALDI-TOF; DNA sequencing; environment; bacteria; microbiome; bioinformatics; Dominican Republic

1. Introduction

The Matrix-Assisted Laser Desorption Ionization-Time of Flight mass spectrometry (MALDI-TOF MS) has been used to profile bacterial proteins from cell extracts, and the procedure provides a unique spectral mass fingerprint of the microorganisms [1]. There are MALDI-TOF MS results reported to have over 95% accuracy. For this reason, the rapid process and the lower cost per sample it offers became a common methodology in clinical microbiology for species identification [2] [3]. Bacterial identification is essential in fields like the clinical to develop effective antimicrobial decisions, as well as the environment to process greater amounts of samples at lower cost and time, as stated before, which is a major advantage when it comes to the characterization of the microbiomes [3].

Genomic analysis of complex environmental samples is becoming an essential tool for understanding the evolutionary history and functional and ecological biodiversity, its only downside being a high-cost and long-time procedure [4,5]. On the other hand, with the widespread use of DNA sequencing in the last decades, it started to play a pivotal role in accurately identifying bacterial isolates and discovering novel bacteria. This methodology is essential, particularly in bacteria with unusual phenotypic profiles, slow-growing, uncultivable bacteria, and culture-negative infections. DNA barcoding employs standardized species-specific genomic regions (DNA barcodes) to generate vast DNA libraries to identify unknown specimens. An excellent example of this in bacteria is the 16S ribosomal RNA.

Our focus for this article will be to compare the efficacy of the MALDI-TOF method with the whole genome sequencing and assembly method in aquatic environmental samples. The objective is to determine how close the accuracy of mass spectrometry is to the accuracy of a high-precision and high-cost method in DNA sequencing.

 doi:10.20944/preprints202209.0098.v1

2. Materials and Methods

2.1. Sample Collection

The samples of water were collected from four rivers in the Dominican Republic. These samples provided information for twenty-nine genomes, thirteen from the La Isabela River, seven from the Ozama River, three from Yaque del Norte River, and six from Yaque del Sur River.

2.2. Bacteria Isolation.

Aliquots of 1, 10, and 50 mL were filtered from the samples using the standardized membrane filtration method utilizing 0.22μm diameter cellulose filter (Millipore). The membranes were placed on 2 MacConkey agar media (Oxoid, UK), one with 4 g/mL of imipenem and the other with 8 g/mL of cefotaxime, incubated at 37°C for 24-48 hours, following the recommendation of [6]. The individual colonies were purified on the same media and stored in 25% glycerol at -70°C. Then, the individual colonies from this process were isolated in chromogenic culture media (ChromoAgar) and stored in 25% glycerol at -70°C.

2.3. Identification of isolates by MALDI-TOF

As recommended, the isolates were characterized with the MALDI-TOF technique (matrix-assisted laser desorption ionization-time of flight) [7]. For identification, isolated colonies were cultured on blood agar for 24 hours at 35°C, using the whole cell transfer protocol. In this protocol, approximately 0.1mg of the culture is transferred to the sample carrier, covered with 1 µL of the matrix solution (10 mg mL-1 α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 2.5% trichloroacetic acid) and let it dry at room temperature for 20 minutes. To check the reproducibility of the spectra, each colony was grown and analyzed in triplicate.

Mass analysis was carried out on a MALDI-TOF AutoFlex spectrometer using the Flex Software Control version 3.4 program and the recommended settings [7]. The voltage of ion source 1: was 20kV; ion source voltage 2: was 19kV; lens voltage: was 6.5kV; mass range: 2-20kDa and the final spectrum was the sum of 10 individual spectra, each obtained from 200 runs of the laser at random locations on the plate as recommended by the manufacturer.

2.4. DNA extraction

2.4.1. Method 1

For samples from La Isabela and Ozama rivers total DNA purification, the QIAamp® DNA commercial system from the QIAGEN commercial house was used with slight modifications as follows: the bacterial pellet was resuspended in 420 μL of modified lysis buffer (20 μL proteinase K, 200 μL of TSB and 100 μL of ATL buffer (Qiagen). DNA extraction was performed starting from centrifuging 2 mL of colonies grown in TSB liquid medium for 24 hours at 35°C. The mixture was incubated for 10 minutes at 56°C. Then, 50 μL of absolute ethanol was added and incubated for 3 minutes at room temperature. From this point on, the protocol continues according to the manufacturer's recommendations using silica columns to purify bacterial genetic material. The DNA was resuspended in 50 μL of TE buffer (Tris-HCL 10mM, EDTA 1mM, pH 8.0). The concentration and quality of the extracted DNA were evaluated on a 1% agarose gel stained with SYBR Green and run at 100V for 60 minutes.

2.4.2. Method 2

For total DNA purification samples from Yaque del Norte and Yaque del Sur rivers, the DNeasy Blood and Tissue commercial system from QIAGEN commercial house was used with no modifications. The extraction was performed by centrifuging 2 mL of colonies grown in a TSB liquid medium for 24 hours at 35ºC. The mixture was incubated for 10 minutes at 56ºC. Then 200 μL of absolute ethanol were added and mixed through the vortex. Afterward, the sample was pipetted into a DNeasy Mini spin column, placed in a

 doi:10.20944/preprints202209.0098.v1

2 mL collection tube, and centrifuged according to the manufacturers instructions. Two washes were done with the buffers AW1 and AW2, respectively, then the column was transferred into a new 2mL collection tube. Finally, 200 μL of Buffer AE from the kit were added to elute the DNA. It was incubated for 1 minute at room temperature (15ºC-25ºC) and centrifuged for one last time.

2.5. MALDI-TOF analysis

A Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) was performed according to the CLSI guidelines [7–9] for the bacterial identification. For the correct realization of this procedure, the BioTyper® 3.1 software (Bruker Daltonics, Germany) was used, equipped with MBT 6903 MPS library (released in 2019), MALDI BioTyper Preprocessing Standard Method, and the MALDI Biotyper MSP Identification Standard Method adjusted by the manufacturer. The isolated colonies were cultured on blood agar for 24 hours at 35ºC. Approximately 0.1mg of each new culture was inoculated in a sample carrier with the complete cell transfer protocol. Samples were coated with 1 μL of matrix solution (10 mg/mL) consisting of α-Cyano-4-hydroxycinnamic acid in 50% acetonitrile and 2.5% trichloroacetic acid and left to dry at 25°C for 20 minutes. Identification was made in triplicates.

2.6. Genome Sequencing, Assembly, and Analysis

For the construction of sequencing libraries, (I) the genomic DNA was randomly fragmented by sonication; (II) DNA fragments were ended polished, A-tailed, and ligated with the full-length adapters of Illumina sequencing, and followed by further PCR amplification with P5 and indexed P7 oligos; and (III) the PCR products as the final construction of the libraries were purified with AMPure XP system (Beckman Coulter Inc., Indianapolis, IN, USA). Sequencing library size distribution quality control was performed with an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) and quantified by real-time PCR (to meet the criteria of 3 nM). Whole genomes were sequenced using Illumina NovaSeq 600 using the PE 150 strategy at the America Novogene Bioinformatics Technology Co., Ltd.

Genomes were assembled using the Assembly HiSeq Pipeline, a SnakeMake pipeline to assemble sequencing data produced by Illumina [6]. The pipeline integrates different quality control tools like FastQC [10] to analyze and visualize read quality, AdapterRemoval v2 [11] for removing sequencing adapters, and KmerStream [12] for computing k-mer distribution. For the genome graph construction, two leading assemblers were used: Edena V3 [13] and Spades 3.9.1 [14]; CD-HIT [15] and Unicycler [16] were used to optimize and integrate the assemblies previously produced. Whole-genome annotation was performed with RAST [17] and Prokka [18]. To predict and reconstruct individual plasmid sequences in the genome assemblies, we used MOB-recon [19]. Finally, QUAST [20] computed assembly quality metrics, and each genome phylogenetic affiliation was confirmed through JSpeciesWS web tools [21] using the contigs generated by the assemblies. All genomes shotgun projects have been deposited to DDBJ/ENA/GenBank.

3. Results

Table 1 shows each sample with their results for the MALDI-TOF and DNA Sequencing methods. The results that differed between both methods were highlighted, and the reasons for these discordances were discussed. In general, most of the results coincided as expected. However, the not coinciding results ratio was more significant than the expected under 10%.

 doi:10.20944/preprints202209.0098.v1

Table 1. In this table, we show the samples' Identification Codes with their respective results for both methods. Accession numbers were added if available.

Sample Code	MALDI-TOF Result	JSpeciesWS Result	Accession	Size	GC%
INTEC BC5 1.1	*Enterobacter cloacae*	*Enterobacter cloacae subsp. Cloacae SMART_901*	JACSEN000000000	5,169,767	54.9
INTEC BI4 1.1	*Enterobacter kobei*	*Enterobacter kobei 35730*	JACSEP000000000	5,761,677	54.1
INTEC AC6 1.1	*Escherichia coli*	*Escherichia coli SQ2203*	JACSHL000000000	5,069,210	50.9
INTEC BI10 1.1	*Escherichia coli*	*Escherichia coli 50816743*	JACSEO000000000	6,128,811	50.6
INTEC BC4	*Escherichia coli*	*Escherichia coli KOEGE 40 (102a)*	JACZEG000000000	4,965,451	50.9
INTEC BC8	*Escherichia coli*	*Escherichia coli O32:H37 str. P4*	JACZEH000000000	6,104,323	50.6
INTEC AI11 1.1	*Acinetobacter baumannii*	*Acinetobacter baumannii ABBL129*	JACSHM000000000	4,205,625	38.9
INTEC BI5	*Acinetobacter baumannii*	*Acinetobacter baumannii BR097*	JACXKJ000000000	4,461,496	38.8
INTEC BI9	*Acinetobacter baumannii*	*Acinetobacter baumannii NIPH 67*	JACXLE000000000	3,993,158	38.8
INTEC AI6	*Acinetobacter baumannii*	*Acinetobacter baumannii UH6507*	SAMN16287478	4,690,858	38.8
INTEC AI12	*Acinetobacter baumannii*	*Acinetobacter baumannii BR097*	JACZEF000000000	3,995,472	38.8
INTEC AI10	*Acinetobacter baumannii*	*Acinetobacter baumannii BR097*	JACZEE000000000	4,491,903	38.8
INTEC BI15	*Acinetobacter baumannii*	*Acinetobacter baumannii NIPH 67*	JACZEI000000000	4,465,154	38.8
DC2	***Escherichia coli***	***Pseudomonas monteilii***	**-**	**14,844,902**	**60.3**
DC8	*Escherichia coli*	*Escherichia coli*	-	4,710,092	50.7
DC10	***Acinetobacter pitii***	***Achromobacter xylosoxidans***	**-**	**33,600,749**	**61.7**
EC4	*Klebsiella pneumoniae*	*Klebsiella pneumoniae*	-	5,527,705	57.2
EC7	*Klebsiella pneumoniae*	*Klebsiella pneumoniae*	-	5,424,764	57.3
FC5	*Acinetobacter baumanii*	*Acinetobacter baumanii*	-	3,809,530	38.9
FC7	*Acinetobacter baumanii*	*Acinetobacter baumanii*	-	3,841,875	38.9
YNP1-2	***Bacillus licheniformis***	***Acinetobacter pittii***	**-**	**3,964,892**	**39.1**
YNP2-2	*Klebsiella pneumoniae*	*Klebsiella pneumoniae*	-	5,493,818	57.3
YNP5-3	*Pseudomonas aeruginosa*	*Pseudomonas aeruginosa*	-	11,657,323	62.3
YSP2-1	***Enterobacter bugadensis***	***Enterobacter mori***	**-**	**15,341,733**	**54.3**
YSP3-2	***Serratia marcescens***	***Salmonella enterica***	**-**	**5,775,726**	**49.9**
YSP4	*Raoultella ornithinolytica*	*Raoultella ornithinolytica*	-	5,788,965	55.7
YSP4-2	***Salmonella spp.***	***Klebsiella pneumoniae***	**-**	**11,182,713**	**48.3**
YSP5	***Escherichia coli***	***Klebsiella pneumoniae***	**-**	**7,744,840**	**45.0**
YSP6-2	***Klebsiella variicola***	***Salmonella enterica***	**-**	**5,321,435**	**52.6**

 doi:10.20944/preprints202209.0098.v1

Table 2. In this table we show the costs and time taken for each technique, as well as mention their limitations.

Technique	Cost per Sample	Time per Sample	Limitations
Whole Genome Sequencing	$95.00USD	Seven work days	CB; SS
MALDI-TOF MS	$7.50USD	Two work days	CB; SS
Shotgun Metagenomics	$180.00USD	Fourteen work days	-
Differential Culture Media	$10.00USD	Two work days	CB; SS

CB = Culture-based technique SS = Single Sample

4. Discussion

We conducted this research for three years. We used different techniques to identify the bacteria. We required culture-based (MALDI-TOF), shotgun methodologies (Shotgun Sequencing), and Whole genome sequencing (WGS). It is important to mention the number of organisms that each technique covers per sample: MALDI-TOF does one per sample, WGS does one per sample, and shotgun does the whole microbiome per the sample pierce as WGS; hence MALDI-TOF is a more reliable and cost-effective option for bacteria identification compared to WGS, many new packages allow you to also scan for particular antimicrobial resistance proteins inside bacteria using MALDI-TOF. This is very limited at the moment since MALDI-TOF is still a culture-based technique and requires at least eighteen hours for microbe isolation plus another eighteen to twenty-four hours for growth to be able to produce enough proteins to be read by MALDI-TOF, but this technology will eventually allow fast identification of ARGs based on expressed proteins in the cytoplasm.

The MALDI-TOF method results were found to have a 72.41% coincidence with the results obtained from the whole genomes phylogenetic analysis from JspeciesWS, unlike the ones described by [2]. It was observed that the eight results that differed between them were caused by contamination in the sequences. It was stated this way because the data analyzed by the JspeciesWS tool showed that the similitude percentages from the updated sequences were below the cutoff stated as acceptable (98.9%), as well as the sizes and GC% being too large. This meant filtering of the sequences was needed to divide the different sequences and identify the contamination taxonomically. Also, it is important to mention that the high cost of the genomic methodologies justifies themselves with their high specificity.

The contamination in the sequences might have been a manipulation mistake. After the filtering with MaxBin 2.0 [22] and Autometa 1.0 [23], we could see that the results coincided with the MALDI-TOF, and the filtered sequences had a 99.99% similitude with the database sequences. In addition, these post-filtering results meant a 100% match between the results from both methods, proving the high efficacy of the MALDI-TOF method in bacterial identification.

5. Conclusions

The samples were retrieved and processed successfully. The bacterial isolation was done for the MALDI-TOF procedure and the bacteria were identified. After that process, the only results to be obtained were the ones from the DNA Sequencing. On that topic, the DNA from each sample was extracted and purified to be sequenced. The fastq files were obtained and analyzed through the corresponding bioinformatic's tools.

A contamination problem that might be related to a manipulation mistake was detected in the phylogenetics assessments, causing these results to differ from the MALDI-TOF results. It was possible to detect the contamination due to the bioinformatic tool giving a range of acceptable similitude percentage between the submitted fasta files (the samples) and the sequences from the database (reference sequences), as well as features like the size and the GC% from the files being too high.

It was determined that, after the filtering process, bacterial identification results from both methods, MALDI-TOF and DNA Sequencing matched on the 100% of the results, exceeding the hypothesis of the coincidence rate being over 90% and proving how well

implemented this technique currently is, as well as why it is commonly used in clinical laboratories for microbiology purposes. In addition, it's important to state that there are more uses being standarized for this technique, like the identification of proteins related to antibiotic resistance, expanding even more it's scope in microbiology.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/1010000/s1, File S1: Reports of results of the samples processed by MALDI-TOF techniques.

Author Contributions: Conceptualization, E.F. and L.M.; methodology, E.F, O.C, R.R, L.M.; software, V.C. and R.B.; Investigation R.B., W.L, I.O., A.G. S.M. and C.P; validation, A.G., I.O, R.B, V.C and O.C; formal analysis, R.B and V.C; resources, E.F and L.M.; data curation, R.B and V. C.; writing—original draft preparation, E.F., R.B. and V.C .; writing—review and editing, E.F and R.R and L.M .; visualization, R.B and V.C; supervision, R.R., O.C, E.F and L,M ; project administration, E.F and L.M.; funding acquisition, E.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fondo Nacional de Innovación y Desarrollo Científico y Tecnológico (FONDOCYT) of Miniterio de Eduacion Superior Ciencia y Tecnología (MESCyT) grant number 2018-2019-2B4-157.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Acknowledgments:

This research project was successfully conducted thanks to the support provided by the Research Vice-Rectory and the Deanship of Basic and Environmental Sciences at Instituto Tecnologico de Santo Domingo.

The Federal University of Para team was support by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, Pró-reitoria de Pesquisa e Pós-graduação(PROPESP)-UFPA and Pró-Reitora de Relações Internacionais (PROINTER)-UFPA.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carbonnelle, E.; Mesquita, C.; Bille, E.; Day, N.; Dauphin, B.; Beretti, J.L.; Ferroni, A.; Gutmann, L.; Nassif, X. MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clinical biochemistry* **2011**, *44*, 104–109.
2. Carbonnelle, E.; Grohs, P.; Jacquier, H.; Day, N.; Tenza, S.; Dewailly, A.; Vissouarn, O.; Rottman, M.; Herrmann, J.L.; Podglajen, I.; et al. Robustness of two MALDI-TOF mass spectrometry systems for bacterial identification. *Journal of microbiological methods* **2012**, *89*, 133–136.
3. Shannon, S.; Kronemann, D.; Patel, R.; Schuetz, A.N. Routine use of MALDI-TOF MS for anaerobic bacterial identification in clinical microbiology. *Anaerobe* **2018**, *54*, 191–196.
4. Shokralla, S.; Spall, J.L.; Gibson, J.F.; Hajibabaei, M. Next-generation sequencing technologies for environmental DNA research. *Molecular ecology* **2012**, *21*, 1794–1805.
5. Hassan, F.M.; Salih, W.Y.; Al-Haideri, H.H. Next-Generation Sequencing Technologies for Environmental DNA as an Efficient Bio Indicator for Bacterial Biodiversity in Tigris River, Iraq. *Systematic Reviews in Pharmacy* **2020**, *11*, 1107–1114.
6. Freitas, D.Y.; Araújo, S.; Folador, A.R.; Ramos, R.T.; Azevedo, J.S.; Tacão, M.; Silva, A.; Henriques, I.; Baraúna, R.A. Extended spectrum beta-lactamase-producing gram-negative bacteria recovered from an Amazonian lake near the city of Belém, Brazil. *Frontiers in microbiology* **2019**, *10*, 364.
7. Strejcek, M.; Smrhova, T.; Junkova, P.; Uhlik, O. Whole-cell MALDI-TOF MS versus 16S rRNA gene analysis for identification and dereplication of recurrent bacterial isolates. *Frontiers in microbiology* **2018**, *9*, 1294.
8. Weinstein, M.P.; Lewis, J.S. The clinical and laboratory standards institute subcommittee on antimicrobial susceptibility testing: background, organization, functions, and processes. *Journal of clinical microbiology* **2020**, *58*, e01864–19.
9. Humphries, R.; Bobenchik, A.M.; Hindler, J.A.; Schuetz, A.N. Overview of changes to the clinical and laboratory standards institute performance standards for antimicrobial susceptibility testing, M100. *Journal of clinical microbiology* **2021**, *59*, e00213–21.
10. Andrews, S.; et al. FastQC: a quality control tool for high throughput sequence data, 2010.
11. Schubert, M.; Lindgreen, S.; Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes* **2016**, *9*, 1–7.

12. Melsted, P.; Halldórsson, B.V. KmerStream: streaming algorithms for k-mer abundance estimation. *Bioinformatics* **2014**, *30*, 3541–3547.
13. Hernandez, D.; François, P.; Farinelli, L.; Østerås, M.; Schrenzel, J. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome research* **2008**, *18*, 802–809.
14. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* **2012**, *19*, 455–477.
15. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152.
16. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS computational biology* **2017**, *13*, e1005595.
17. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **2008**, *9*, 1–15.
18. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069.
19. Robertson, J.; Nash, J.H. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microbial genomics* **2018**, *4*.
20. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075.
21. Richter, M.; Rosselló-Móra, R.; Oliver Glöckner, F.; Peplies, J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **2016**, *32*, 929–931.
22. Wu, Y.W.; Simmons, B.A.; Singer, S.W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **2016**, *32*, 605–607.
23. Miller, I.J.; Rees, E.R.; Ross, J.; Miller, I.; Baxa, J.; Lopera, J.; Kerby, R.L.; Rey, F.E.; Kwan, J.C. Autometa: automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic acids research* **2019**, *47*, e57–e57.