A Scientometric Overview of Bioinformatics Tools in the *Pseudomonas Putida* Genome Study

Alex Batista Trentin(✉), Beatriz Thomas Metzner, Glecio Oliveira Barros, and Deborah Catharine de Assis Leite

Federal University of Technology – Paraná, Dois Vizinhos, Brazil
a.trentinx@gmail.com

Abstract. *Pseudomonas putida* is a microorganism widely used in environmental science due to its high degradation power of recalcitrant compounds. This study aimed to perform a scientometric analysis of the global panorama of publications on the *P. putida* genome, in addition to the bioinformatics tools mainly used. The growth of publications on the *P. putida* genome is continuous until 2020, being France, Spain and USA the main countries with publications on the subject. Illumina was the main sequencing platform used and, in this set of articles, 120 genomes were sequenced, 106 complete and 14 drafts. The main assembly software was SPAdes, comprising 23.3% of the articles, and NCBI PGAP was the main genome annotation tool, in 25% of the documents. Thus, this study allowed the visualization of the main bioinformatics tools used for the analysis of the *P. putida* genome, besides presenting the advance of the research on this subject, and also supporting next studies with *P. putida.*

Keywords: Bioinformatic tools · Scientometric analysis · *Pseudomonas putida*

1 Introduction

The planet is fighting against all types of environmental pollution, with the soil being one of the places most affected by environmental degradation. There is, therefore, a need to control soil pollution in order to maintain fertility as well as productivity [1, 2]. In this context, the development of technologies to remedy these degraded environments becomes indispensable [3].

Pseudomonas genera, belonging to the *Gammaproteobacteria* class, have been extensively studied over time due to their great potential for degradation and biotransformation of xenobiotic and recalcitrant compounds, and are likely to be used in various biotechnological processes of environmental recovery [4]. *Pseudomonas putida* is an ubiquitous bacterium, mainly in soil, classified as chemoorganotrophic, capable of metabolizing large carbon chains, as well as several recalcitrant pollutants [5].

From the sequencing of the genome, as well as from bioinformatics analyses, it is possible to know the proteins and metabolic pathways of *Pseudomonas putida* applied to environmental remediation and degradation of xenobiotic compounds [6]. Thus,

J. C. Setubal and W. M. Silva (Eds.): BSB 2020, LNBI 12558, pp. 162–167, 2020.
https://doi.org/10.1007/978-3-030-65775-8_15

using molecular level analysis it is possible to study new techniques and environmental applications for this organism [7].

This work aimed to perform a systematic analysis, by means of the scientometric methodology, of the global publications on the genome of *Pseudomonas putida*, analyzing mainly the bioinformatics tools used for the assembly and annotation of the sequenced genomes.

2 Methods

The searches for performing the scientometric analysis were performed in the Web of Science database using the terms "Genome" AND "*Pseudomonas putida*" AND "soil". Initially 159 documents were found on the researched terms, however, a manually filtering was performed in order to find only those documents that reported the sequencing process of the microorganism. After the filtering, 29 papers were left, being 100% journal articles, containing the sequencing of the genome.

After filtering, the data was extracted to *Microsoft Excel* and *Citespace* software in order to produce graphs of data and connections on: main countries with publications on the subject, knowledge areas applied, keywords most used, number of publications and citation per year. Also, all the documents were read, in order to extract the quantity of sequenced genomes, the sequencing platforms, besides the genome annotation and assembly softwares, these data were tabulated and had graphs generated in Excel.

3 Results and Discussion

The first publication studying the genomic sequencing of *Pseudomonas putida* bacterium associated with soil, within the database, was in 2002, followed by a gap of 11 years, only starting the publications again in 2014, however, this time with a continuous growth of publications until 2020. The rate of citations on the subject grew continuously from 2002, totaling 1129 citations until August 2020, as shown in the Fig. 1. The growth in genomic studies of this soil-related organism is due to the extremely versatile metabolism of *P. putida*, it's great capacity of adaptation in several environments, the resistance to physical-chemical stresses, and the genes associated with degradation of recalcitrant compounds. This way, studies using this organism in biological remediation processes have maintained constant growth [8, 9]. Also, the top 5 of the main countries that published on the subject are: Spain, Japan, USA, Germany and France. The number of publications by country each year is also shown in Fig. 1.

3.1 Knowledge Areas

From the analysis performed at CiteSpace Software it is possible to see that Microbiology, Environmental Sciences and Biotechnology are the main areas of knowledge associated with the theme, fact based on the search direction of the research, aiming at the application of *P. putida* in soils, due to the high diversity of metabolisms of this microorganism associated with bioremediation [10]. The Fig. 2 presents the network of

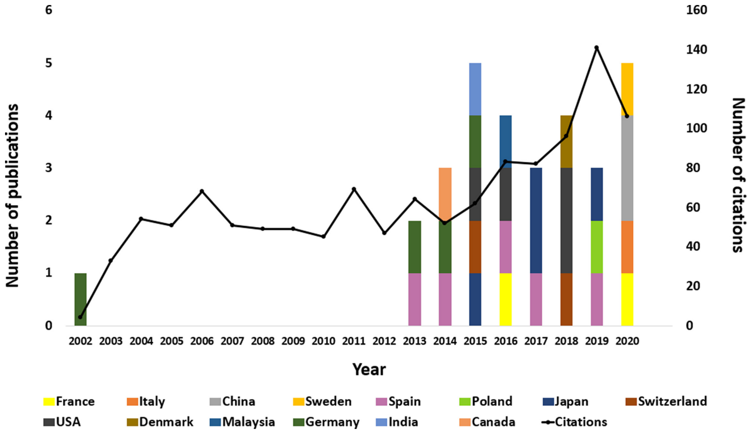

Fig. 1. Number of publications by country and year, and the number of citations per year

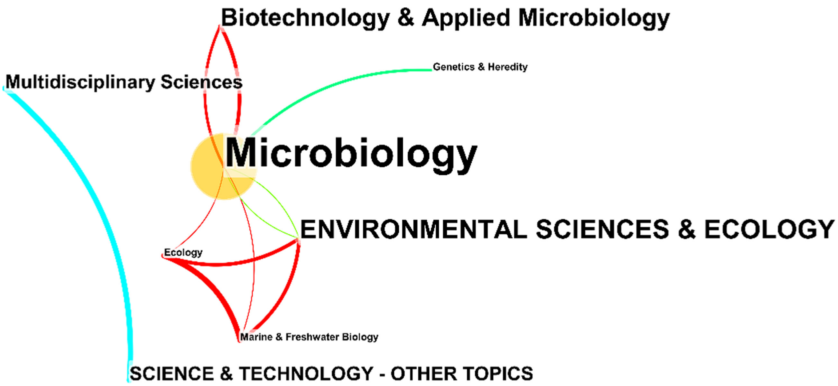

Fig. 2. Network of connections over knowledge areas. Yellow circles represent the centrality. (Color figure online)

connections related to the areas of study, and the higher the font of the letter, the more frequently this area is used in these researches.

Also, the yellow circle represents the centrality, a factor that indicates the amount of connections that the theme performs [11], thus, it is visualized that Microbiology is the only area that presents a significant centrality, due to the fact that this set of data approaches the use of a microorganism, therefore, even the most specific researches are still interconnected with microbiology, making microbiology coherent as a central area of knowledge [12].

3.2 Main Journals

In this data set 12 journals published the articles, most of them related to microbiology, environmental science or genomics. The main journal was Microbiology Resource

Announcements, with 9 articles published, 31% of the 29 documents. Table 1 presents all journals, along with the number of documents and the impact factor.

Table 1. Journals and number of published articles

Journal	Papers	Impact factor
Microbiology Resource Announcements	9	0.89
Environmental Microbiology	3	4.93
Environmental Microbiology Reports	3	2.97
Amb Express	2	2.49
Applied And Environmental Microbiology	2	4.016
Current Microbiology	2	1.73
Environmental Science And Pollution Research	2	3.30
Plos One	2	2.74
Biotechnology And Bioengineering	1	4.00
Microbial Ecology	1	3.86
Scientific Reports	1	3.99
Standards In Genomic Sciences	1	1.44

3.3 Sequencing Data

In total, 117 genomes of *Pseudomonas putida* were sequenced in the 26 documents, 106 of them complete and 14 drafts. Four sequencing platforms were used for these works, and some of these documents used more than one platform, the main one being Illumina, in 17 papers; followed by Roche 54 GS FLX, in 5 papers; Pac Bio, in 4 of the 26 documents; and Ion Torrent, in 1 document. Figure 3A and B presents the percentages of whole/draft genome and the main platforms.

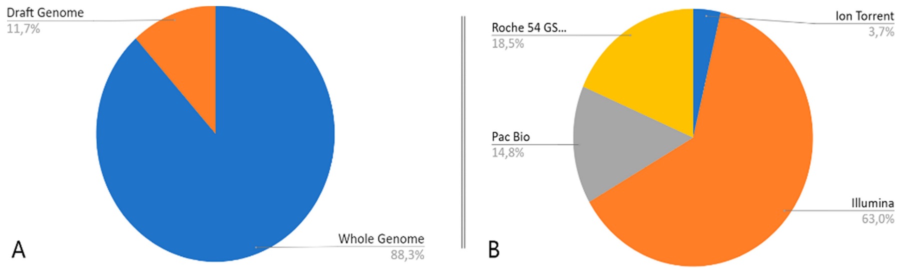

Fig. 3. (A) Percentage of 120 whole/draft genome. (B) Percentage of the main sequencing platforms used.

3.4 Assembly

In total, 9 genome assembly softwares were used, the most used being SPAdes, one of the main bacterial genome assembly softwares [13]. Other software such as CLC Genomics Workbench, Newbler and HGAP were used, as shown in Fig. 4. However, many papers (6 out of 29) did not inform the assembly software.

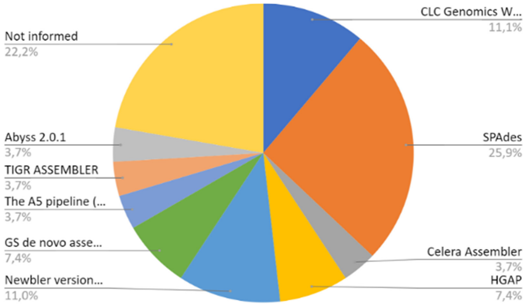

Fig. 4. Genome assembly softwares used

3.5 Annotation

In the same way as in the assembly, most of the articles did not inform the genome annotation software (7 of the 26 documents). There were 5 softwares used in the 26 papers, the main one being the NCBI Prokaryotic Genome Annotation Pipeline, a tool in constant modification and evolution for the annotation process [14]. Figure 5 shows the software used along with the percentage within the 26 papers.

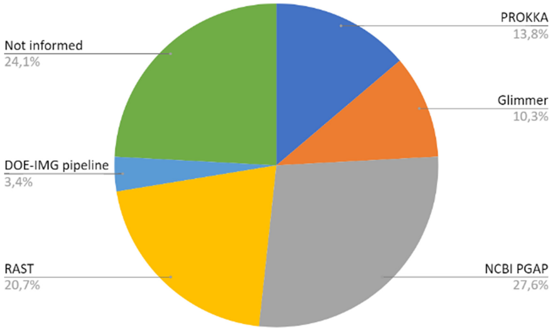

Fig. 5. Genome annotation softwares used in the data set

4 Conclusions

Pseudomonas putida is a widely studied organism with an extreme potential in the area of bioremediation of affected environments. In this context, the genomic study of this microorganism is of extreme relevance, since from the analyses at molecular level it is possible to know the main metabolic pathways, as well as the genes and proteins used by *P. putida* to perform the degradation processes of recalcitrant compounds.

In this way, the bioinformatics tools are able to analyze these molecular data and allow the development of new technologies using this microorganism, besides allowing the dissemination of genomic data. Thus, this systematic study allowed the visualization of the main bioinformatics tools used for the analysis of the *P. putida* genome, besides presenting the advance of the research on the subject, helping possible future studies with *P. putida* with application in soil environmental science and other environments.

References

1. Ashraf, A.M., Maah, J.M., Yusoff, I.: Soil contamination, risk assessment and remediation, vol. 1, no. 1, pp. 3–5. Intech (2014)
2. Ranga, P., Sharma, D., Saharan, S.B.: Bioremediation of azo dye and textile effluents using Pseudomonas putida MTCC 244. Asian J. Microbiol. Biotechnol. Environ. Sci. **22**(2), 88–94 (2019)
3. Horemans, B., Breugelmans, P., Saeys, W., Springael, D.: A soil-bacterium compatibility model as a decision-making tool for soil bioremediation. Environ. Sci. Technol. **51**(3), 1605–1615 (2017)
4. Loh, K.C., Cao, B.: Paradigm in biodegradation using Pseudomonas putida—A review of proteomics studies. Enzyme Microbial Technol. **43**(1), 1–12 (2008)
5. Tsirinirindravo, H.L., et al.: Bioremediation of soils polluted by petroleum hydrocarbons by Pseudomonas putida. Int. J. Innov. Eng. Sci. Res. **2**(5), 9–18 (2018)
6. Iyer, R., Iken, B., Damania, A., Krieger, J.: Whole genome analysis of six organophosphate-degrading rhizobacteria reveals putative agrochemical degradation enzymes with broad substrate specificity. Environ. Sci. Pollut. Res. **25**(14), 13660–13675 (2018)
7. Nelson, K.E., et al.: Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. Environ. Microbiol. **4**(12), 799–808 (2002)
8. Crovadore, J., et al.: Whole-genome sequence of Pseudomonas putida strain 1312, a potential biostimulant developed for agriculture. Microbiol. Resour. Announc. **10**(7), 1–2 (2018)
9. Weimer, A., et al.: Industrial biotechnology of Pseudomonas putida: advances and prospects. Appl. Microbiol. Biotechnol. **104**(1), 7745–7766 (2020)
10. Abyar, H., et al.: The role of Pseudomonas putida in bioremediation of naphthalene and copper. World J. Fish Marine Sci. **3**(5), 444–449 (2011)
11. Ping, Q., He, J., Chen, C.: How many ways to use CiteSpace? A study of user interactive events over 14 months. J. Assoc. Inf. Sci. Technol. **68**(5), 1234–1256 (2017)
12. Chen, C.: Manual do CiteSpace, 1st edn. Drexel University, Philadelphia (2014)
13. Prjibelski, A., et al.: Using SPAdes de novo assembler. Curr. Prot. Bioinform. **70**(1), 1–29 (2020)
14. Tatusova, T., et al.: NCBI prokaryotic genome annotation pipeline. Nucl. Acids Res. Adv. **1**(1), 1–11 (2016)