engineering proceedings

MDPI

Proceeding Paper

Pangenome Analysis and Physiological Characterization of *Gordonia alkanivorans* Strains Capable of Utilizing Persistent Organic Pollutants †

Ekaterina Frantsuzova [1], Alexander Bogun [1,2], Lidiya Shishkina [1,2], Anna Vetrova [1], Inna Solyanikova [1,3] and Yanina Delegan [1,*]

[1] Institute of Biochemistry and Physiology of Microorganisms, Federal Research Center "Pushchino Scientific Center for Biological Research of Russian Academy of Sciences" (FRC PSCBR RAS), 142290 Pushchino, Moscow Region, Russia; frantsuzova.ee@gmail.com (E.F.); bogun62@mail.ru (A.B.); kadnikova_lidiya@mail.ru (L.S.); phdvetrova@gmail.com (A.V.); innap_solyan@yahoo.com (I.S.)
[2] State Research Center for Applied Microbiology and Biotechnology, Obolensk 142279, Russia
[3] Belgorod State University, Belgorod 308015, Russia
* Correspondence: mewgia@ya.ru
† Presented at the 2nd International Electronic Conference on Processes: Process Engineering—Current State and Future Trends (ECP 2023), 17–31 May 2023; Available online: https://ecp2023.sciforum.net/.

Abstract: Members of many species of the genus *Gordonia* are known for their ability to utilize compounds of different structures. The aim of the work was to study the ability of nine *G. alkanivorans* strains to degrade persistent organic pollutants and to analyze the genomic peculiarities of these strains. The genomes of nine *Gordonia alkanivorans* strains were sequenced and assembled. Based on the unique genes in the genomes, the strains can be divided into two subgroups. The strains can be used in biotechnologies of the environmental treatment as alkane degraders. Additionally, they all utilize benzoate.

Keywords: *Gordonia alkanivorans*; genome sequencing; benzoate; alkanes; biodegradation

Citation: Frantsuzova, E.; Bogun, A.; Shishkina, L.; Vetrova, A.; Solyanikova, I.; Delegan, Y. Pangenome Analysis and Physiological Characterization of *Gordonia alkanivorans* Strains Capable of Utilizing Persistent Organic Pollutants. *Eng. Proc.* **2023**, *37*, x. https://doi.org/10.3390/xxxxx
Published: 17 May 2023

1. Introduction

Gordonia strains are known for their ability to adapt to various environmental conditions and their extensive metabolic capabilities due to the remarkable plasticity of the genomes of this genus [1–3]. It can be said that *Gordonia* strains are ubiquitous.

Members of many species of the genus *Gordonia* are known for their ability to utilize compounds of a different structure, including persistent organic pollutants. This makes *Gordonia* strains promising for application in the field of environmental biotechnology.

The species *Gordonia alkanivorans* was introduced in 1999 due to the isolation of a type strain of this species, HKI 0136T, from tar- and phenol-contaminated soil [4]. Currently, *G. alkanivorans* strains are known mainly as degraders of two types of compounds: thiophenes as a sulfur source [5,6] and alkanes as carbon and energy sources.

Previously, it was thought that thiophene catabolism in *G. alkanivorans* was only carried out using the operon *dsz* [5,7]. However, we have shown that there are *G. alkanivorans* strains that utilize thiophenes without *dsz* genes [8,9]. Alkane catabolism by *G. alkanivorans* strains is performed by P450 hydroxylases (CYP153); the absence of the *alk*B genetic system is characteristic of all the known representatives of this genus [8].

The aim of the work is to study the ability of nine *G. alkanivorans* strains to degrade persistent organic pollutants and to analyze the genomic peculiarities of these strains. The results will make it possible to assess the prospects for the application of these strains in biotechnologies for the remediation of contaminated soil ecosystems.

2. Materials and Methods

2.1. Bacterial Strains and Cultivation Conditions

We used 9 bacterial strains (Table 1) isolated from oil-contaminated soils and *G. alkanivorans* strain 135 [8,10] as a reference to investigate the catabolic properties.

Table 1. Information about the strains used in the study.

Strain	Collection Number	Isolation Source	Previously Identified as
96	IEGM 96	crude oil-contaminated soil, Ukraine	*G. rubripertincta*
144	IEGM 144	crude oil-contaminated soil, Lvov, Ukraine	*G. terrae*
129	IEGM 129	crude oil-contaminated soil, Ivano-Frankovsk, Ukraine	*G. rubripertincta*
132	IEGM 132	crude oil-contaminated soil, Ivano-Frankovsk, Ukraine	*G. rubripertincta*
133	IEGM 133	crude oil-contaminated soil, Ivano-Frankovsk, Ukraine	*G. rubripertincta*
134	IEGM 134	crude oil-contaminated soil, oilfield, Lvov, Ukraine	*G. rubripertincta*
142	IEGM 142	crude oil-contaminated soil, Ukraine	*G. rubripertincta*
12	-	oil-polluted soil, Moscow, Russia	*G. alkanivorans*
152	-	oil-polluted soil, Moscow, Russia	*G. alkanivorans*

The ability of the strains to utilize thiophenes as the only sulfur source was tested by the method described in [8]. The ability of the strains to grow on even alkanes (C_6–C_{20}) and aromatic compounds (naphthalene, phenol, benzoate, catechol) was tested using mineral medium of the following composition: K_2HPO_4—8.71 g/L, 5 M NH_4Cl solution—1 mL/L, 0.1 M Na_2SO_4 solution—1 mL/L, 62 mM $MgCl_2$ solution—1 mL/L, 1 mM $CaCl_2$ solution—1 mL/L, 0.005 mM of $(NH_4)_6Mo_7O_{24} \times 4H_2O$ solution, micronutrients—1 mL (micronutrient composition in g/L: ZnO—0.41 g, $FeCl_3 \times 6H_2O$—5.4 g, $MnCl_2 \times 4H_2O$—2 g, $CuCl_2 \times 2H_2O$—0.17 g, $CoCl_2 \times 6H_2O$—0.48 g, H_3BO_3—0.06 g), and pH 7.0. Alkanes were added at 7.5 mL/L, naphthalene, phenol, and benzoate at 1 g/L, and catechol at 0.1 g/L.

2.2. Genome Sequencing and Analysis

Genomic DNA of strains was isolated from a biomass grown on LB [11] agar using a DNeasy Blood & Tissue Kit (QIAGEN, 69506). Sequencing was performed on a MGI platform (DNBSEQ-G400) using the DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150) (2 × 150 bp). A paired-end library was prepared with the MGIEasy Universal DNA Library Prep Set. The information of the generated data is presented in Table 2.

Table 2. Number of sequencing data before and after filtration.

Strain	Read Pairs before Filtration	Read Pairs after Filtration	Read Pairs after Filtration (%)
96	2,962,813	2,835,521	95.70
144	5,382,881	5,150,561	95.68
129	3,147,953	3,003,848	95.42
132	4,354,967	4,162,088	95.57
133	5,192,473	5,019,737	96.67
134	5,815,565	5,558,107	95.57
142	3,330,918	3,171,608	95.22
12	9,985,963	8,692,673	87.05
152	6,985,286	4,928,247	70.55

The raw reads were filtered using Trimmomatic v. 0.39 [12] and assembled using SPAdes v. 3.15.4 [13]. Contigs shorter than 500 bp were removed (Table 3).

Table 3. Assembly metrics of the strains.

Strain	Genome Length, Mb	Number of Contigs	N50, bp
96	4.9	80	167,760
144	4.2	303	27,240
129	4.9	86	265,593
132	5.0	85	256,904
133	5.0	87	225,227
134	5.1	75	182,877
142	5.0	96	142,845
12	4.9	82	285,499
152	5.0	91	169,160

The Average Nucleotide Identity (ANI) value with the type strain *G. alkanivorans* NBRC16433 (BACI00000000.1) was calculated using the EzBioCloud ANI Calculator [14]. DNA-DNA hybridization (DDH) was calculated using the Genome-to-Genome Distance Calculator (GGDC) [15]. The genome was annotated with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.6 [16], Prokka [17] and RAST [18].

3. Results and Discussion

3.1. Identifiyng the Strains

Based on the results of whole genome sequencing data, several strains were reidentified (Table 4).

Table 4. Species identification of the strains.

Strain	ANI Value with the Type Strain of *G. rubripertincta*, %	DDH Value with the Type Strain of *G. rubripertincta*, %	ANI Value with the Type Strain of *G. alkanivorans*, %	DDH Value with the Type Strain of *G. alkanivorans*, %	Taxonomic Position of the Strain
96	92.58	76.20	98.45	89.70	*G. alkanivorans*
144	92.42	66.80	98.63	78.10	*G. alkanivorans*
129	92.45	76.00	98.42	88.60	*G. alkanivorans*
132	92.40	76.30	98.42	88.60	*G. alkanivorans*
133	92.40	76.10	98.37	88.50	*G. alkanivorans*
134	92.25	76.80	98.23	87.60	*G. alkanivorans*
142	92.54	76.30	98.23	91.00	*G. alkanivorans*
12	92.38	76.20	98.40	90.10	*G. alkanivorans*
152	92.48	76.60	98.34	88.00	*G. alkanivorans*

Thus, all the strains reliably belong to *Gordonia alkanivorans*. The difference in the DDH value between the studied strains and the type of strain *G. alkanivorans* NBRC16433 indicates some heterogeneity of the species, but all the strains pass the species threshold by both ANI (>96%) and DDH (>70%).

3.2. Physiological and Biochemical Characteristics of Strains

On agarized rich media (LB), all the strains form small round colonies of a pink-orange color. When growing on mineral media with alkanes as a carbon source or thiophenes as a sulfur source, lighter orange colonies are formed. The color change possibly indicates that when growing on mineral media with energy sources that are difficult to access, microorganisms expend energy for basic metabolic processes and substrate utilization, but not for the biosynthesis of secondary metabolites (carotenoids) which give cells their color.

Crude oil is a complex mixture of components, those being aliphatic and aromatic hydrocarbons, as well as their sulfur-, nitrogen-, and oxygen-containing derivatives. In this regard, we assumed that the strains isolated from areas contaminated with crude oil may be capable of utilizing compounds of each of these groups. All strains were capable of utilizing alkanes from C_{10} to C_{20} and benzoate; some strains are capable of using dibenzothiophene (DBT) as the sole source of sulfur (Table 5).

Table 5. Substrate specificity profile of strains.

Strain	Alkanes C_{10}–C_{16}	Alkanes C_{18}–C_{20}	Benzoate	Phenol	Naphthalene	Catechol	DBT
96	++	+	+	-	-	-	±
144	+	+	+	-	-	-	-
129	+	+	+	-	-	-	-
132	+	+	+	-	-	-	±
133	+	+	+	-	-	-	-
134	++	++	+	-	-	-	-
142	+	+	+	-	-	-	-
12	++	+	+	-	-	-	±
152	++	+	+	-	-	-	-
135	++	++	+	-	-	-	+

++ very good growth, + good growth, ± weak growth, - no growth.

Thus, there is still no strain of *G. alkanivorans* capable of utilizing PAHs (naphthalene) or their metabolites (catechol) known at this time.

3.3. Peculiarities of G. alkanivorans Genome Organization and Pangenome Analysis

Of the nine strains of *G. alkanivorans*, two (strains 142 and 152) have plasmid elements. The plasmid of strain 142 (p142) is 67,219 bp in length; the plasmid of strain 152 (p152) is 44,937 bp. We assume that they are circular, since *Gordonia* in general is not characterized by the maintenance of large linear plasmids as, for example, in rhodococci.

In total, 98% of the entire length of plasmid p142 is a 99% repeat of plasmids pCP89 (CP094666.1) of the *Gordonia amicalis* G2 strain (percent identity (PI) 99.70%) and pCP86 (CP096597.1) of the *G. amicalis* 6-1 strain (PI 99.98%) (Figure 1). It is interesting to note this relatedness between plasmids whose hosts are strains of different species.

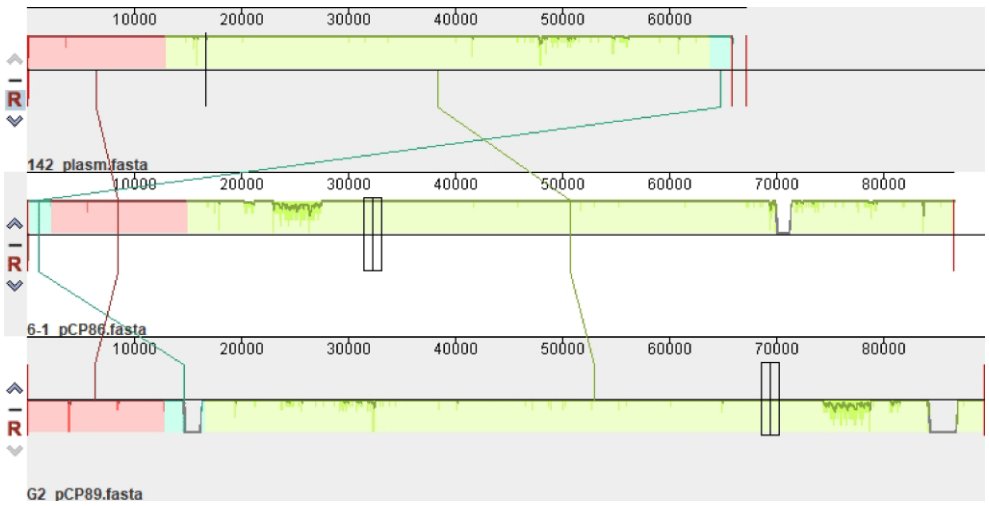

Figure 1. Mauve alignment demonstrating relatedness of the plasmids p142 (strain *G. alkanivorans* 142), pCP86 (*G. amicalis* 6-1) and pCP89 (*G. amicalis* G2). Vertical bars mark boundaries between elements.

The plasmid of strain 152 has a relatedness with the plasmid pG135. The ANI value between the plasmids p142 and p152 is 80.00%.

Based on the results of genome analysis and individual unique genes, the strains can be divided into two groups within the species. The representatives of the first group (strains 129, 144, 132, 133) are characterized by unique genes of tyrocidine and gramicidin biosynthesis. The representatives of the second group have a greater catabolic potential: the genomes of the strains contained (1) operons for the biosynthesis of steroid compounds, (2) additional copies of genes involved in dibenzothiophene catabolism, and (3) genes of aromatic compound catabolic process: cytochrome P450-pinF2 and phenol hydroxylase P5. All the studied *G. alkanivorans* strains lack *alk*B genes in their genomes; therefore, we assume that the ability to utilize alkanes in these strains is controlled by the CYP153 genetic system. The CYP153 hydroxylases have at least a 99% identity between strains and contain one amino acid substitution each: one hydroxylase has A/T variants at position 239, the other hydroxylase has A/S variants at position 7, and they are in the first and second strain groups, respectively.

4. Discussion

In general, we can say that the strains of *G. alkanivorans* species are similar in terms of their physiological properties and degradative potential. They utilize alkanes with different chain lengths, and, among the compounds with an aromatic structure, benzoate is available to them. Nevertheless, representatives of this species have no metabolic pathways of naphthalene degradation.

It is interesting to note that *Gordonia* plasmids do not appear to be species-specific. The function of plasmid p142 in strain 142 is currently unclear, but the fact that copies of the plasmid have been observed in members of another species (*G. amicalis*) may indicate the importance of this plasmid for the vital activity of *Gordonia*. We plan to obtain a plasmid-free eliminant of strain 142 in the future in order to obtain a better understanding of the functions of this plasmid. At this point, we can assume that plasmids such as p142 are required by strains for metal transport and resistance.

Based on the pangenome analysis, we were able to trace some regularities of strain distribution within the species. Nine representatives of *G. alkanivorans* can be divided into two subgroups, which are distinguished by unique genes. Considering the presence of a greater number of catabolic genes and operons in the representatives of the second subgroup (strains 96, 134, 142, 12, 152), we can assume that these strains can be promising in the field of biotechnologies for the purification of the environment from, for example, steroid compounds.

5. Conclusions

The genomes of nine strains of *Gordonia alkanivorans* isolated from oil-contaminated soils were sequenced and assembled. The genomes are about 5 Mb in size. Some of the strains contain plasmids, but the functions of these plasmids are currently not fully understood. A pangenome analysis of the strains has shown that, within the species, there are differences between the strains, allowing them to be conditionally divided into two subgroups according to the unique genes of each strain. The strains can be used in environmental treatment biotechnologies as alkane degraders. Additionally, all of the strains also utilize benzoate.

Author Contributions: Conceptualization, Y.D. and A.V.; methodology, E.F., A.B. and I.S.; software, Y.D.; validation, A.B., Y.D. and I.S.; formal analysis, A.V.; investigation, E.F., L.S. and A.V.; data curation, Y.D.; writing—original draft preparation, E.F. and Y.D.; writing—review and editing, A.B. and A.V.; visualization, E.F.; supervision, Y.D.; project administration, Y.D.; funding acquisition, Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the Russian Science Foundation, grant number 22-74-10082.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The whole-genome sequences of the strains can be found under BioProject number PRJNA955828, BioSamples SAMN34194808-SAMN34194816.

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

References

1. Arenskötter, M.; Bröker, D.; Steinbüchel, A. Biology of the Metabolically Diverse Genus *Gordonia*. *Appl. Environ. Microbiol.* **2004**, *70*, 3195–3204. https://doi.org/10.1128/AEM.70.6.3195-3204.2004.
2. Shen, F.-T.; Lu, H.-L.; Lin, J.-L.; Huang, W.-S.; Arun, A.B.; Young, C.-C. Phylogenetic analysis of members of the metabolically diverse genus *Gordonia* based on proteins encoding the *gyr*B gene. *Res. Microbiol.* **2006**, *157*, 367–375. https://doi.org/10.1016/j.resmic.2005.09.007.
3. Shen, F.-T.; Young, C.-C. Rapid detection and identification of the metabolically diverse genus *Gordonia* by 16S rRNA-gene-targeted genus-specific primers. *FEMS Microbiol. Lett.* **2005**, *250*, 221–227. https://doi.org/10.1016/j.femsle.2005.07.014.
4. Kummer, C.; Schumann, P.; Stackebrandt, E. *Gordonia alkanivorans* sp. nov., isolated from tar-contaminated soil. *Int. J. Syst. Evol. Microbiol.* **1999**, *49*, 1513–1522. https://doi.org/10.1099/00207713-49-4-1513.
5. Alves, L.; Marques, S.; Matos, J.; Tenreiro, R.; Gírio, F.M. Dibenzothiophene desulfurization by *Gordonia alkanivorans* strain 1B using recycled paper sludge hydrolyzate. *Chemosphere* **2008**, *70*, 967–973. https://doi.org/10.1016/j.chemosphere.2007.08.016.
6. Aminsefat, A.; Rasekh, B.; Ardakani, M.R. Biodesulfurization of dibenzothiophene by *Gordonia* sp. AHV-01 and optimization by using of response surface design procedure. *Microbiology* **2012**, *81*, 154–159. https://doi.org/10.1134/S0026261712020026.
7. Jaishankar, J.; Singh, P.; Srivastava, P. Draft Genome Sequence of a Biodesulfurizing Bacterium, *Gordonia* sp. Strain IITR100. *Genome Announc.* **2017**, *5*, e00230-17. https://doi.org/10.1128/genomeA.00230-17.
8. Delegan, Y.; Kocharovskaya, Y.; Frantsuzova, E.; Streletskii, R.; Vetrova, A. Characterization and genomic analysis of *Gordonia alkanivorans* 135, a promising dibenzothiophene-degrading strain. *Biotechnol. Rep.* **2021**, *29*, e00591. https://doi.org/10.1016/j.btre.2021.e00591.
9. Frantsuzova, E.; Delegan, Y.; Bogun, A.; Sokolova, D.; Nazina, T. Comparative Genomic Analysis of the Hydrocarbon-Oxidizing Dibenzothiophene-Desulfurizing *Gordonia* Strains. *Microorganisms* **2022**, *11*, 4. https://doi.org/10.3390/microorganisms11010004.
10. Delegan, Y.; Valentovich, L.; Vetrova, A.; Frantsuzova, E.; Kocharovskaya, Y. Complete Genome Sequence of *Gordonia* sp. 135, a Promising Dibenzothiophene- and Hydrocarbon-Degrading Strain. *Microbiol. Resour. Announc.* **2020**, *9*, e01450-19. https://doi.org/10.1128/MRA.01450-19.
11. Bertani, G. Studies on Lysogenesis I: The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **1951**, *62*, 293–300. https://doi.org/10.1128/jb.62.3.293-300.1951.
12. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
13. 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. *J. Comput. Biol.* **2012**, *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021.
14. Yoon, S.-H.; Ha, S.; Lim, J.; Kwon, S.; Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **2017**, *110*, 1281–1286. https://doi.org/10.1007/s10482-017-0844-4.
15. Meier-Kolthoff, J.P.; Auch, A.F.; Klenk, H.-P.; Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **2013**, *14*, 60. https://doi.org/10.1186/1471-2105-14-60.
16. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. https://doi.org/10.1093/nar/gkw569.
17. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
18. 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 Genom.* **2008**, *9*, 75. https://doi.org/10.1186/1471-2164-9-75.