frontiers
in Bioengineering and Biotechnology

The Complete Genome Sequence and Structure of the Oleaginous *Rhodococcus opacus* Strain PD630 Through Nanopore Technology

Andrea Firrincieli[1†], Beatrice Grigoriev[1†], Hana Dostálová[2] and Martina Cappelletti[1]*

[1]Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy, [2]Institute of Microbiology of the CAS, Prague, Czechia

INTRODUCTION

Rhodococcus bacterial strains are characterized by wide metabolic versatility and extraordinary resistance to environmental stresses (de Carvalho et al., 2014; Orro et al., 2015; Cappelletti et al., 2016, 2019; Pátek et al., 2021). The high versatility and adaptability of *Rhodococcus* strains is partly related with large and complex genomes (up to 10.1 Mbp) including high genetic redundancy and the presence of several circular and linear (mega)plasmids, which harbour peculiar catabolic and biosynthetic genes (Cappelletti et al., 2019). Within *Rhodococcus* genus, *R. opacus* strain PD630 is considered a model oleaginous strain for its ability to produce and accumulate lipids (mostly triacyglycerols, TAGs) using different carbon sources, including low-cost and renewable resources such as lignocellulose (Alvarez et al., 1996; Anthony et al., 2019; Cappelletti et al., 2020; Alvarez et al., 2021; Donini et al., 2021). Notably, under specific growth conditions, this strain is capable of accumulating up to 80% of its cellular dry weight in TAGs (Alvarez and Steinbüchel, 2002); that is a rare feature in the prokaryotic and eukaryotic kingdoms. Multi-omic approaches have been applied to obtain system-level information about metabolic and regulatory pathways involved in these biosynthetic processes. Novel molecular tools for genome editing (CRISPR/Cas9 and recombineering) have been recently developed highlighting the possible utilization of *R. opacus* PD630 as synthetic biology platform for lipids production (Donini et al., 2021; Liang and Yu, 2021).

A first assembly of the *R. opacus* PD630 genome was submitted by the Broad Institute in 2011 (Holder et al., 2011) and included 491 contigs. Later in 2014, a "complete" version of the PD630 genome was submitted by the Institute of Biophysics of the Chinese Academy of Sciences (hereafter IBP_PD630) (Chen et al., 2013) that was recently indicated as "Anomalous assembly" and "contaminated" and therefore deleted from NCBI RefSeq. IBP_PD630 reported the PD630 genome to be composed by one chromosome and nine plasmids (two circular and seven linear plasmids). This result was divergent from the typical number of extrachromosomal elements reported for genomes of this genus, five being the maximum number of extrachromosomal elements described in a single *Rhodococcus* strain (Cappelletti et al., 2019). Despite the assembly-related issues, IBP_PD630 has been used as reference genome in many works involving -omics analyses for the detection of genetic determinants involved in aromatics tolerance and conversion into lipids (DeLorenzo et al., 2018; Kim et al., 2019; Chen et al., 2013; Yoneda et al., 2016; Henson et al., 2018).

In this study, we combined Oxford Nanopore sequencing with Illumina high quality data to solve the architecture of the *R. opacus* PD630 genome and to obtain its whole sequence. Here, in addition to the high-quality complete genome sequence, we demonstrate that this strain does not possess nine

plasmids as previously stated, but instead harbours one chromosome and three (one linear and two circular) plasmids. Further, by solving the final structure of *R. opacus* PD630, we found that the large linear plasmid included extrachromosomal genes involved in lipid and xenobiotics' metabolism. This work provides the correct PD630 genomic framework for the development of genetic engineering strategies to boost the application of this strain as microbial cell factory for lipid production.

MATERIALS AND METHODS

Bacterial Growth and Genomic DNA Extraction

R. opacus PD630 (DSM 44 193) was purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). For genomic DNA extraction, PD630 was cultivated in 50 ml Luria Bertani (LB) broth at 30°C for 24 h under shaking conditions at 150 rpm. The culture was centrifuged at 4°C for 10 min at 5,000 rpm and the whole cell pellet was subjected to the procedure indicated by Cappelletti et al. (Cappelletti et al., 2011) with slight modifications that involved the utilization of a 10% sodium dodecyl sulfate (SDS)-based lysis buffer to disrupt the cells instead of the mechanical treatment (mediated by bead beater). The genomic DNA was quantified via Qubit dsDNA BR assay kit with the Qubit 4.0 fluorometer (Life Technologies).

Oxford Nanopore Whole Genome Sequencing

The genomic DNA was fragmented for 10 s via sonication and the integrity of the DNA after fragmentation was checked via electrophoresis gel. The sequencing library was performed with the Oxford Nanopore Ligation Sequencing Kit (SQK-LSK110), according to manufacturer's instructions. The genomic library was loaded on a FLO-MIN106D (chemistry R9.4.1) flow cell and sequenced with the MinION Mk1C device (Oxford Nanopore Technology, ONT). The sequencing run was performed until 3.5 Gb of bases were obtained that corresponded to a total of 694,680 reads. Base calling of the FAST5 data from MinION was carried out with Guppy GPU 5.0 in high-accuracy mode and with default parameters (chunks per runner 256; chunk size 2000; minimum qscore 7). Evaluation of sequencing quality of the basecalled FAST5 data was performed using PycoQC v2.5.2. Sequencing reads were deposited in the Sequence Read Archive (SRA) SRX12606194.

Sequencing Data Assembly and Annotation

The genome was assembled using Canu assembler v1.2 (Koren et al., 2017) and the draft assembly was corrected using Illumina short reads [SRX875494 (Janet, 2014)] (Illumina HiSeq 2500) with a single round of Pilon v. 1.24 (Walker et al., 2014). The draft assembly was finally circularized using Circlator (Hunt et al., 2015) indicating the *dnaA* gene of *R. opacus* PD630 (OPAG_07542) as the chromosomal sequence start. A final round of polishing was performed with Berokka to trim overhanging ends from plasmids. Finally, genome completeness was assessed via BUSCO (Manni et al., 2021) and annotated *de novo* using the standalone version of Prokaryotic Genome Annotation Pipeline (PGAP) (Li et al., 2020).

Comparative Analysis With Other PD630 Genome Assembly Versions

The synteny analysis between the overall organization of PD630 Chen's assembly (Chen et al., 2013) and the genome presented in this work was performed by Sibelia (Minkin et al., 2013). Liftoff was used to identify the PD630 genes from Chen's assembly that did not map against our genome due to the presence of miscalled nucleotides, insertion/deletions (indels), and structural variants. A gene was considered to successfully map when the alignment coverage and sequence identity was equal to 100%.

Functional Annotation of Protein Coding Genes in Plasmids of *R. opacus* Strains

Proteins harbored by plasmids of *R. opacus* strains and *R. jostii* RHA1 were downloaded from RefSeq and annotated using the functional database KEGG with KofamKOALA (Aramaki et al., 2019) (**Supplementary Table S1**). According to the Genome Taxonomy Database (Chaumeil et al., 2019), we included in the analysis the only *R. opacus* strains with a complete genome currently available in RefSeq (Accessed: December 10th, 2021), i.e., *R. opacus* B4, *R. opacus* DSM 44 186, *R. opacus* 1CP, and *R. opacus* KT112-7.

INTERPRETATION OF DATA SET

Genome Assembly and Annotation

After base calling a total of 667,679 reads with a median Phred score of 13.43, and a median length of 3.1 Kb were generated (**Supplementary Figures S1A,B**). The draft assembly generated by Canu consisted of four contigs with a total length of 9.16 Mbp. After one polishing round in Pilon using Illumina HiSeq paired-end reads, the size of the Canu-assembled contig slightly increased mostly because of single and/or di-nucleotide insertions (**Supplementary File 1**). Large variants (>100 bp) were manually checked via blastn. As a result, all of them perfectly matched with sequences of *R. opacus* PD630 (**Supplementary File 1**). A change in the final size (>10 Kb) of the contigs RoPD630, pRoPD630_2, and pRoPD630_3 was observed after the circularization and trimming steps (Circlator + Berokka) (**Table 1**), due to the removal of duplicated sequences at their ends. The presence of start-end overlaps in contigs is a well-known behaviour of Canu assembler which tends to generate contigs with a contiguity above 100% (Wick and Holt, 2021).

The final assembly (obtained after circularization and trimming) provided a genome with an overall size of 9.16 Mb, including one circular chromosome (RoPD630) of 8.37 Mb and three plasmids, one linear (pRoPD630_1) and two circular (pRoPD630_2 and pRoPD630_3). The average G + C content

Predicted topology	RoPD630 (CP080954)	pRoPD630_1 (CP080955)	pRoPD630_2 (CP080956)	pRoPD630_3 (CP080957)
	Circular	Linear	Circular	Circular
Canu	8,411,399	538,757	210,131	121,266
Pilon	8,412,872	538,909	210,169	121,283
Circlator	8,374,353	538,909	210,169	85,051
Berokka	8,374,353	538,909	67,985	85,051
Final bp - Canu bp[a]	−37,046	+152	−42,145	−36,215

[a]*Difference in the number of bp after all the steps with respect to the initial Canu output.*

of the circular chromosome was 67.38% whereas plasmids pRoPD630_1, pRoPD630_2, and pRoPD630_3 showed lower G + C values, i.e., 65.27, 65.06, and 63.91%, respectively (**Table 1**). The genome completeness was checked via BUSCO, resulting in 100% completeness and 1.4% duplication. According to the PGAP, FABIT_PD630 genome contained 8,280 genes, of which 7,972 protein coding genes, 66 RNA genes, and 242 pseudo-genes.
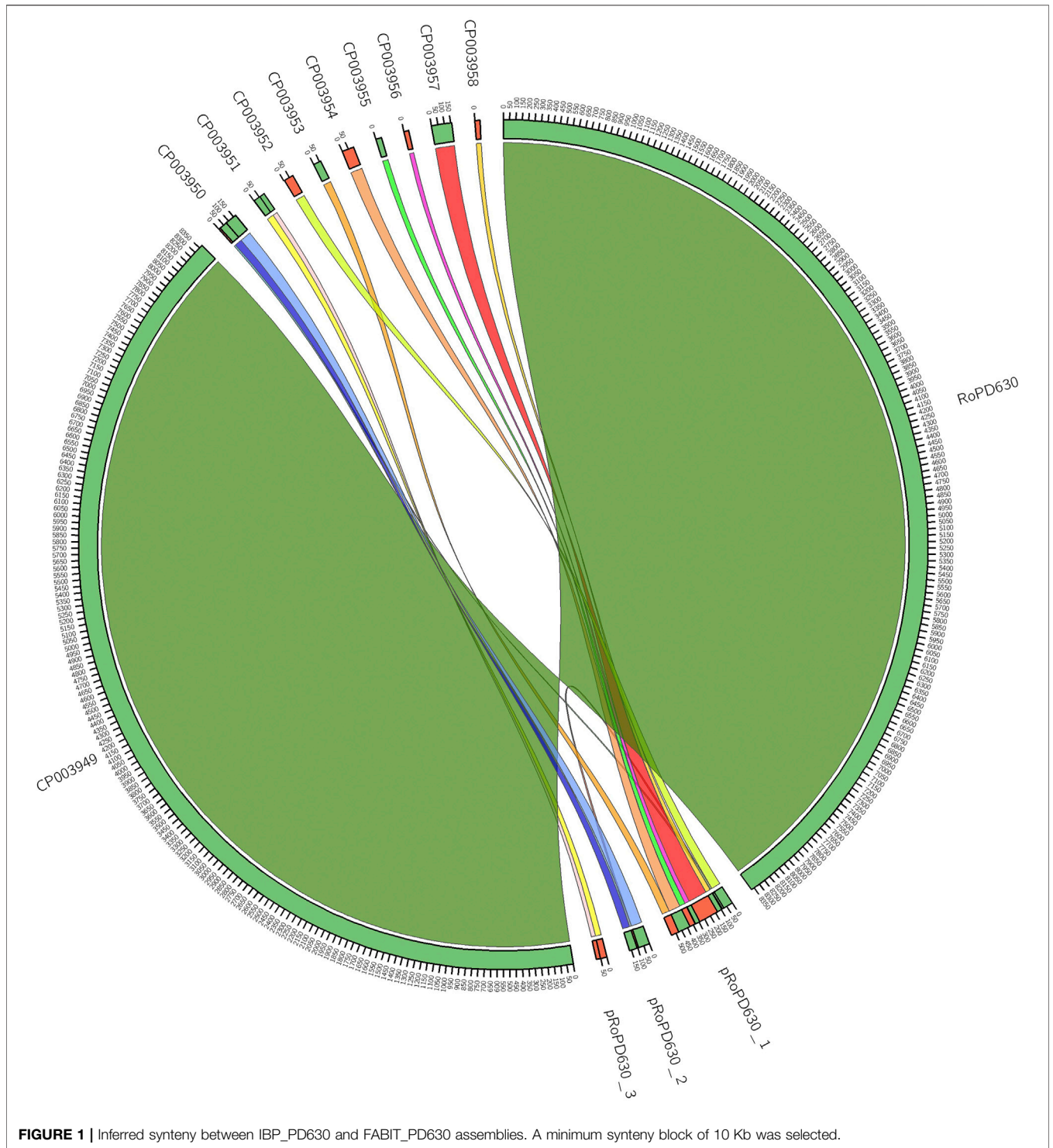
Genome Assembly Comparison Between the IBP_PD630 and FABIT_PD630 Genomes

We specifically performed comparative analysis between the IBP_PD630 and FABIT_PD630 genomes to analyse their differences at structural and genetic levels. As a result, FABIT_PD630 possesses a lower number of genes as compared with IBP_PD630, i.e., 8,280 against the 9,005 genes detected in IBP_PD630 assembly. The different number of genes can be attributed to the distinct annotation pipelines that were used in the two works and/or to actual differences in the genome sequence (i.e., sequence variants). To get deep into the reason on this discrepancy, we investigated the possibility that missing genes were due to structural and nucleotide variants. Therefore, we only focused on the identification of the IBP_PD630 genes that did not map with a perfect alignment (i.e., alignment coverage and sequence identity of 100%) with FABIT_PD630 assembly. As a result, only 238 IBP_PD630 genes (around 30% of the number of genes missing in FABIT_PD630 as compared to IBP_PD630) did not align with a perfect match. Furthermore, the number of genes not mapping to our genome was reduced to 84 (<1% of all the genes of Chen's assembly) when a coverage and sequence identity of 90% was imposed (**Supplementary File 2**). These results indicate that the differences observed between FABIT_PD630 and IBP_PD630 in terms of identified gene number were mainly due to the annotation pipeline. Annotation pipelines are known to have different sensitivity to misannotations and therefore to the detection of open reading frames (Monnahan et al., 2020).

Syntenic analysis of IBP_PD630 plasmids showed near-perfect match between the seven linear plasmids (CP03952-CP03958) of IBP_PD630 and the linear plasmid pRoPD630_1 of FABIT_PD630. This indicates that the sequences CP03952-CP03958 do not correspond to distinct plasmids but represent the fragments of a single linear megaplasmid present in *R. opacus* PD630. Similarly, a nearly-perfect match was observed between each

of the circular plasmids CP03950 and CP03951 and the circular plasmids pRoPD630_2 and pRoPD630_3, respectively (**Figure 1**). Therefore, these results demonstrate that *R. opacus* PD630 possesses only three plasmids instead of nine. The difference between the genetic structure of the two genomes can be attributed to the distinct sequencing and library preparation technologies used in the two works to assemble BIP_PD630 and FABIT_PD630. Indeed, Chen et al. (Chen et al., 2013) carried out a primary assembly using 454 pyrosequencing reads, followed by scaffolding step using Illumina mate-pair library with insert size of 3,000 bp. Although Illumina mate-pair libraries can be successfully used to improve contiguity and therefore close gaps between contigs (Wetzel et al., 2011), the utilization of a single mate-pair library could have been not able to correctly solve repetitive regions of different sizes. Indeed, the capacity of finding links between contigs is limited by the length of the gap itself (Wetzel et al., 2011). An additional downside of Illumina mate-pair reads that might have hampered a correct assembly is the uneven sequencing depth introduced during PCR amplification step which is prone to GC content related bias (Sohn and Nam, 2016). Conversely, long-read sequencing technologies, including the ONT sequencing used in our work, are not typically affected by these issues as they apply library preparation strategies that are PCR-free and therefore less biased towards regions with high AT/GC content. On the other hand, in order to cope with the error rate limits of the ONT technology (Laver et al., 2015), we used high quality Illumina sequencing data (available online under the accession number SRX875494) to correct the possible sequencing errors introduced during long-reads sequence generation.
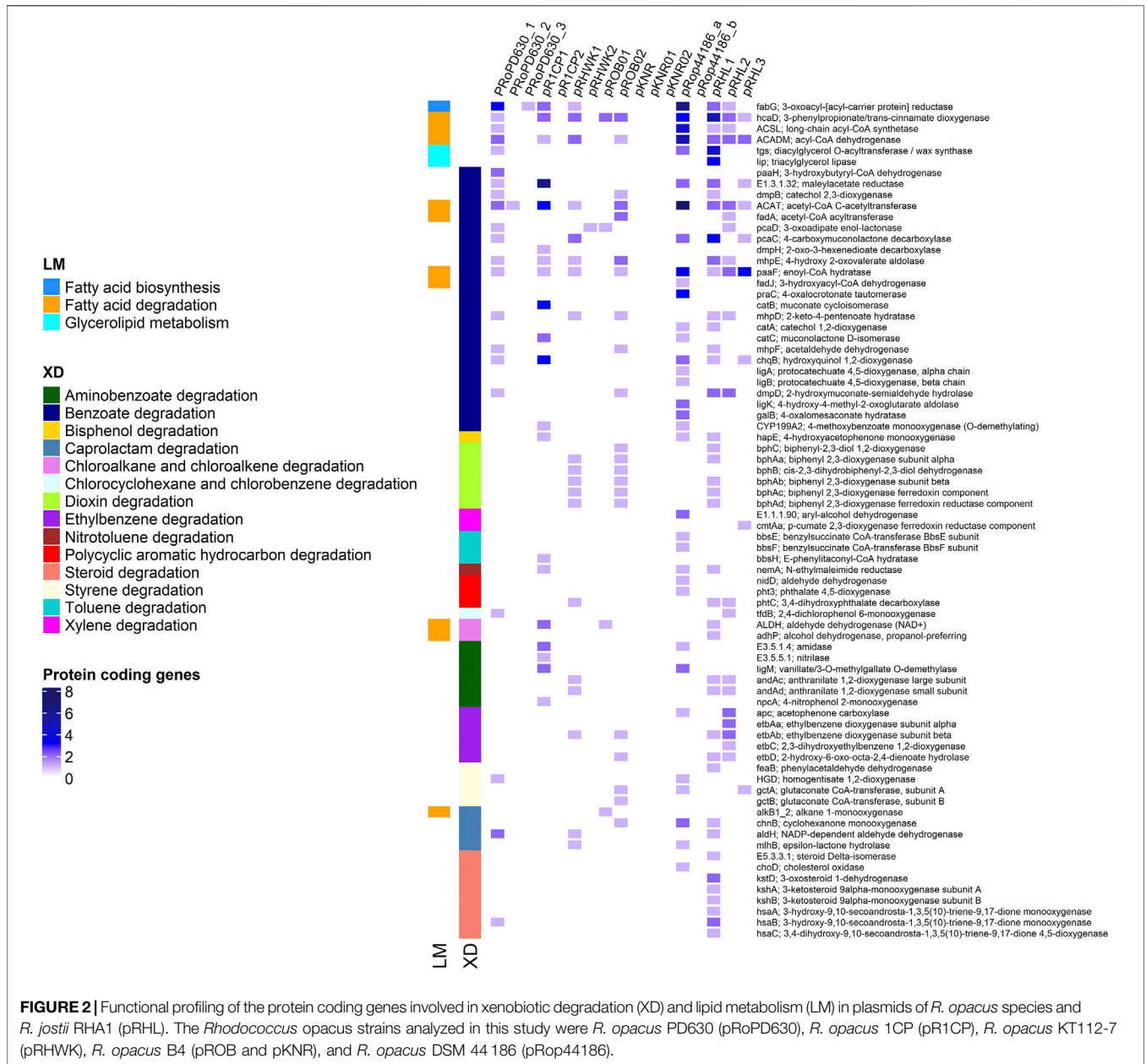
Our work provides the definitive structure of the plasmids of PD630 that have drawn attention in relation with the aromatics bioconversion capacity of this strain. In this regard, Henson et al. (Henson et al., 2018) showed that Adaptive Laboratory Evolution (ALE) of PD630 strain that was sequentially cultured on different aromatics led to multiple plasmid loss events. The selective loss of extrachromosomal elements seemed to benefit the PD630 growth under specific conditions by leading to the deletion of superfluous genes and saving in large plasmid replication costs (DeLorenzo et al., 2018). Furthermore, the PD630 plasmids were found to harbour several genes encoding enzymes involved in the catabolism of hetero- and poly-cyclic aromatic compounds as well as multiple uncharacterized mono and dioxygenases. This observation pointed out the interest of these plasmids as potential targets for genetic engineering strategies and bacterial strain

FIGURE 1 | Inferred synteny between IBP_PD630 and FABIT_PD630 assemblies. A minimum synteny block of 10 Kb was selected.

improvement (Cappelletti et al., 2019; Anthony et al., 2019; DeLorenzo et al., 2018). The knowledge of the correct PD630 genome structure and plasmid organization provided by our work is therefore crucial to assess the molecular and genetic bases of the PD630 peculiar metabolic capacities and to correctly design strain improvement strategies.

Functional Profiles of the *R. opacus* Species Plasmids

To gain a better understanding on the repertoire of catabolic potential harbored by each of the three *R. opacus* PD630 plasmids defined in this study, their KEGG functional annotation was performed. Additionally, a comparative analysis of all the

FIGURE 2 | Functional profiling of the protein coding genes involved in xenobiotic degradation (XD) and lipid metabolism (LM) in plasmids of *R. opacus* species and *R. jostii* RHA1 (pRHL). The *Rhodococcus opacus* strains analyzed in this study were *R. opacus* PD630 (pRoPD630), *R. opacus* 1CP (pR1CP), *R. opacus* KT112-7 (pRHWK), *R. opacus* B4 (pROB and pKNR), and *R. opacus* DSM 44 186 (pRop44186).

plasmids of *R. opacus* strains available in the database (with a complete genome and unanimously identified as belonging to this species) was conducted. *R. jostii* RHA1 was also included in the analysis as the model strain of *Rhodococcus* genus. As a result, the percentage of KEGG orthologues detected in PD630 was 30.8% in pRoPD630_1, 18.7% in pRoPD630_2, and 11.8% in pRoPD630_3. All the three PD630 plasmids harbored genes associated with lipid metabolism, biosynthesis of vitamin, and secondary metabolites, whereas only pRoPD630_1 and pRoPD630_2 included also genes related to carbon, energy, amino acid, and nucleotide metabolism and xenobiotic degradation (**Supplementary Table S1**). However, pRoPD630_1 included a significantly higher number of genes associated to all of these metabolisms as compared to pRoPD630_2 and pRoPD630_3. In particular, pRoPD630_1

showed a high number of genes associated with xenobiotics degradation and fatty acid catabolism/anabolism that are the metabolic capacities mostly studied for this strain and in general for *R. opacus* species (Anthony et al., 2019; Donini et al., 2021). Conversely, in the PD630 genome version of (Chen et al., 2013), these genes were distributed over the seven scaffolds that were erroneously identified as separate plasmids. These gene functions were also found to be mostly co-localized on one of the several plasmids carried by the other *R. opacus* strains under analysis, while in *R. jostii* RHA1 they were scattered (**Figure 2**). By resolving the final structure of all the PD630 replicons, we have now the full comprehension of the organization of biotechnologically relevant genes also associated with its plasmids.

DATA AVAILABILITY STATEMENT

AUTHOR CONTRIBUTIONS

AF conceived the bioinformatic analysis and together with BG conducted the in silico analyses and wrote the first draft of the manuscript. HD carried out the PD630 culture preparation, DNA extraction, and together with MC prepared the library and carried out the experiment on the MinION Mk1C. MC conceived and supervised the study, acquired the funding, edited the manuscript and shaped the last version of it.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.810571/full#supplementary-material

Supplementary File 1 | Pilon polishing round results.

Supplementary File 2 | Liftoff analysis results of IBP_PD630 genes mapped over FABIT_PD630 assembly using increasing coverage and identity thresholds: (i) 100% coverage and identity, (ii) 95% coverage and identity, and (iii) 90% coverage and identity. Each folder includes two files, "mapped" and "unmapped." The "mapped" file is a GFF (General Feature Format) file including all IBP_PD630 genes successfully mapped against FABIT_PD630 and their coordinates. Conversely, the "unmapped" file is a text file reporting the IBP_PD630 genes that do not satisfy the minimum coverage and identity values specified in the Liftoff analysis. GFF files provide the framework to identify the IBP_PD630 genes in the FABIT_PD630 assembly Finally, a Sequence Alignment Map (SAM) file is provided to inspect the alignment of IBP_PD630 genes over FABIT_PD630 genome.

REFERENCES

Alvarez, H. M., and Steinbüchel, A. (2002). Triacylglycerols in Prokaryotic Microorganisms. *Appl. Microbiol. Biot.* 60, 367–376. doi:10.1007/s00253-002-1135-0

Alvarez, H. M., Hernández, M. A., Lanfranconi, M. P., Silva, R. A., and Villalba, M. S. (2021). Rhodococcus as Biofactories for Microbial Oil Production. *Molecules* 26, 4871. doi:10.3390/molecules26164871

Alvarez, H. M., Mayer, F., Fabritius, D., and Steinbüchel, A. (1996). Formation of Intracytoplasmic Lipid Inclusions by Rhodococcus opacus Strain PD630. *Arch. Microbiol.* 165, 377–386. doi:10.1007/s002030050341

Anthony, W. E., Carr, R. R., DeLorenzo, D. M., Campbell, T. P., Shang, Z., Foston, M., et al. (2019). Development of Rhodococcus opacus as a Chassis for Lignin Valorization and Bioproduction of High-Value Compounds. *Biotechnol. Biofuels* 12, 1–14. doi:10.1186/s13068-019-1535-3

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2019). KofamKOALA: Kegg Ortholog Assignment Based on Profile HMM and Adaptive Score Threshold. *Method. Biochem. Anal.* 36, 2251–2252. doi:10.1093/bioinformatics/btz859

Cappelletti, M., Fedi, S., Frascari, D., Ohtake, H., Turner, R. J., and Zannoni, D. (2011). Analyses of Both the alkB Gene Transcriptional Start Site and alkB Promoter-Inducing Properties of Rhodococcus sp. Strain BCP1 Grown on N -alkanes. *Appl. Environ. Microbiol.* 77, 1619–1627. doi:10.1128/aem.01987-10

Cappelletti, M., Fedi, S., Zampolli, J., Di Canito, A., D'Ursi, P., Orro, A., et al. (2016). Phenotype Microarray Analysis May Unravel Genetic Determinants of the Stress Response by Rhodococcus aetherivorans BCP1 and Rhodococcus opacus R7. *Res. Microbiol.* 167, 766–773. doi:10.1016/j.resmic.2016.06.008

Cappelletti, M., Presentato, A., Piacenza, E., Firrincieli, A., Turner, R. J., and Zannoni, D. (2020). Biotechnology of Rhodococcus for the Production of Valuable Compounds. *Appl. Microbiol. Biotechnol.* 104, 8567–8594. doi:10.1007/s00253-020-10861-z

Cappelletti, M., Zampolli, J., Di Gennaro, P., and Zannoni, D. (2019). "Genomics of Rhodococcus," in *Biology of Rhodococcus* (Springer International Publishing), 23–60. doi:10.1007/978-3-030-11461-9_2

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database. *Method. Biochem. Anal.* 36, 1925–1927. doi:10.1093/bioinformatics/btz848

Chen, Y., Ding, Y., Yang, L., Yu, J., Liu, G., Wang, X., et al. (2013). Integrated Omics Study Delineates the Dynamics of Lipid Droplets in Rhodococcus opacus PD630. *Nucleic Acids Res.* 42, 1052–1064. doi:10.1093/nar/gkt932

de Carvalho, C. C. C. R., Marques, M. P. C., Hachicho, N., and Heipieper, H. J. (2014). Rapid Adaptation of Rhodococcus erythropolis Cells to Salt Stress by Synthesizing Polyunsaturated Fatty Acids. *Appl. Microbiol. Biotechnol.* 98, 5599–5606. doi:10.1007/s00253-014-5549-2

DeLorenzo, D. M., Rottinghaus, A. G., Henson, W. R., and Moon, T. S. (2018). Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630. *ACS Synth. Biol.* 7, 727–738. doi:10.1021/acssynbio.7b00416

Donini, E., Firrincieli, A., and Cappelletti, M. (2021). Systems Biology and Metabolic Engineering of Rhodococcus for Bioconversion and Biosynthesis Processes. *Folia Microbiol.* 66, 701–713. doi:10.1007/s12223-021-00892-y

Henson, W. R., Campbell, T., DeLorenzo, D. M., Gao, Y., Berla, B., Kim, S. J., et al. (2018). Multi-omic Elucidation of Aromatic Catabolism in Adaptively Evolved Rhodococcus opacus. *Metab. Eng.* 49, 69–83. doi:10.1016/j.ymben.2018.06.009

Holder, J. W., Ulrich, J. C., DeBono, A. C., Godfrey, P. A., Desjardins, C. A., Zucker, J., et al. (2011). Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development. *Plos Genet.* 7, e1002219. doi:10.1371/journal.pgen.1002219

Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., and Harris, S. R. (2015). Circlator: Automated Circularization of Genome Assemblies Using Long Sequencing Reads. *Genome Biol.* 16, 1–10. doi:10.1186/s13059-015-0849-0

Janet, D. (2014). Whole Genome Sequencing of *Rhodococcus opacus* PD630. Short Read Archive Available at: https://www.ncbi.nlm.nih.gov/sra/SRX875494.

Kim, H. M., Chae, T. U., Choi, S. Y., Kim, W. J., and Lee, S. Y. (2019). Engineering of an Oleaginous Bacterium for the Production of Fatty Acids and Fuels. *Nat. Chem. Biol.* 15, 721–729. doi:10.1038/s41589-019-0295-5

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-Mer Weighting and Repeat Separation. *Genome Res.* 27, 722–736. doi:10. 1101/gr.215087.116

Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. (2015). Assessing the Performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quantification* 3, 1–8. doi:10.1016/j.bdq.2015.02.001

Li, W., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., Badretdin, A., et al. (2020). RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline Reach with Protein Family Model Curation. *Nucleic Acids Res.* 49, D1020–D1028. doi:10.1093/nar/gkaa1105

Liang, Y., and Yu, H. (2021). Genetic Toolkits for Engineering Rhodococcus Species with Versatile Applications. *Biotechnol. Adv.* 49, 107748. doi:10.1016/j. biotechadv.2021.107748

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi:10.1093/molbev/msab199

Minkin, I., Patel, A., Kolmogorov, M., Vyahhi, N., and Pham, S. (2013). Sibelia: A Scalable and Comprehensive Synteny Block Generation Tool for Closely Related Microbial Genomes. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8126 LNBI, 215–229. doi:10.1007/978-3-642-40453-5_17

Monnahan, P. J., Michno, J. M., O'Connor, C., Brohammer, A. B., Springer, N. M., McGaugh, S. E., et al. (2020). Using Multiple Reference Genomes to Identify and Resolve Annotation Inconsistencies. *BMC* 21. doi:10.1186/s12864-020-6696-8

Orro, A., Cappelletti, M., D'Ursi, P., Milanesi, L., Di Canito, A., Zampolli, J., et al. (2015). Genome and Phenotype Microarray Analyses of Rhodococcus sp. BCP1 and Rhodococcus opacus R7: Genetic Determinants and Metabolic Abilities with Environmental Relevance. *PLoS ONE* 10, e0139467. doi:10.1371/journal. pone.0139467

Pátek, M., Grulich, M., and Nešvera, J. (2021). Stress Response in Rhodococcus Strains. *Biotechnol. Adv.* 53, 107698. doi:10.1016/j. biotechadv.2021.107698

Sohn, J.-i., and Nam, J.-W. (2016). The Present and Future of de Novo Whole-Genome. *Brief Bioinform* 19, bbw096. doi:10.1093/bib/bbw096

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* 9, e112963. doi:10.1371/journal.pone.0112963

Wetzel, J., Kingsford, C., and Pop, M. (2011). Assessing the Benefits of Using Mate-Pairs to Resolve Repeats in De Novo Short-Read Prokaryotic Assemblies. *BMC Bioinformatics* 12, 1–14. doi:10.1186/1471-2105-12-95

Wick, R. R., and Holt, K. E. (2021). Benchmarking of Long-Read Assemblers for Prokaryote Whole Genome Sequencing. *F1000Res* 8, 2138. doi:10.12688/f1000research.21782.4

Yoneda, A., Henson, W. R., Goldner, N. K., Park, K. J., Forsberg, K. J., Kim, S. J., et al. (2016). Comparative Transcriptomics Elucidates Adaptive Phenol Tolerance and Utilization in Lipid-accumulating Rhodococcus opacus PD630. *Nucleic Acids Res.* 44, 2240–2254. doi:10.1093/nar/gkw055

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.