Canadian Journal of Microbiology

Identification of plasmids from Brazilian *Chromobacterium violaceum* strains

Journal:	Canadian Journal of Microbiology
Manuscript ID	cjm-2021-0197.R2
Manuscript Type:	Article
Date Submitted by the Author:	15-Sep-2021
Complete List of Authors:	Lima, Daniel; Instituto Federal de Educação Ciência e Tecnologia do Rio Grande do Norte; Universidade Federal do Rio Grande do Norte Medeiros, Inácio Gomes; Universidade Federal do Rio Grande do Norte, ; Universidade Federal do Rio Grande do Norte de Cássia Silva-Portela, Rita; Universidade Federal do Rio Grande do Norte da Silva Junior, Francisco Carlos; Universidade Federal do Rio Grande do Norte, Fassarella Agnez-Lima, Lucymara; Universidade Federal do Rio Grande do Norte; Universidade Federal do Rio Grande do Norte de Souza, Jorge Estefano; Universidade Federal do Rio Grande do Norte; Universidade Federal do Rio Grande do Norte Batistuzzo de Medeiros, Silvia R; Universidade Federal do Rio Grande do Norte
Keyword:	plasmids, Biotechnology, Genomics, Microbial phylogenetics, Bacteriophages
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

Identification of plasmids from Brazilian *Chromobacterium violaceum* strains

Daniel Chaves de Lima (chavesbio@gmail.com)^{1,2}; Inácio Gomes Medeiros(inaciogmedeiros@gmail.com)^{2,4,5}; Rita de Cássia Silva-Portela(silvaportela.rcb@gmail.com)²; Francisco Carlos da Silva Junior(carlosjsilva3@gmail.com)²; Lucymara Fassarela Agnez-Lima(lfagnez@ufrnet.br)^{2,4}; Jorge Estefano Santana de Souza(jorge@imd.ufrn.br)^{3,4}; Silvia R Batistuzzo de Medeiros*(sbatistu@gmail.com)²

¹ Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, CEP: 59015-000, Natal, Brazil;

² Laboratório de Biologia Molecular e Genômica, Universidade Federal do Rio Grande do Norte, CEP 59078-970, Natal, Brazil;

³ Bioinformatics Multidisciplinary Environment, Instituto Metrópole Digital, Universidade Federal do Rio Grande do Norte, CEP 59078-970, Natal, Brazil;

⁴ Programa de Pós-Graduação em Bioinformática, Instituto Metrópole Digital, Universidade Federal do Rio Grande do Norte, CEP 59078-970, Natal, Brazil;

⁵ Instituto do Cérebro, Universidade Federal do Rio Grande do Norte, CEP 59056-450, Natal, Brazil;

* Correspondence:

Silvia Regina Batistuzzo de Medeiros
Departamento de Biologia Celular e Genética
Centro de Biociências – UFRN
Campus Universitário, Lagoa Nova.
Natal – RN – Brazil
CEP 59072-970
Fone: +55-84-3211-9209
Fax: +55-84-3215-3346
E-mail: sbatistu@cb.ufr.br

Abstract

Chromobacterium violaceum is an opportunistic pathogen found in tropical and subtropical regions worldwide. *C. violaceum* infections are difficult to treat, and many strains are resistant to antibiotics. Recently, a novel plasmid (pChV1) was discovered in the type strain ATCC 12472, suggesting that other *C. violaceum* strains may harbor extra-chromosomal DNA. The aim of the present study was to detect and compare new plasmids in Brazilian strains of *C. violaceum* using next generation sequencing techniques. We obtained draft genomes of six plasmids from strains isolated from the Amazon region and aligned them to pChV1. At least three plasmids, CVAC05, CVACO2, and CVT8, were similar to pChV1. Phylogenetic analysis suggested that these new extra-chromosomal DNA sequences have a common origin to pChV1, but have diverged. Many of the ORFs detected were related to plasmid segregation/maintenance, viral structural proteins, and proteins with unknown functions. These findings may enable better genetic manipulation of *C. violaceum*, which will enhance our ability to exploit this valuable microorganism in industrial and clinical applications.

Keywords: plasmids; biotechnology; genomics; microbial phylogenetics; bacteriophages; antibiotic resistance

1 Introduction

Chromobacterium violaceum is a gram-negative beta-proteobacteria, which is known for its ability to produce a purple pigment named violacein. First described in the late 1800s (Boisbaudran 1882), it is a free-living, widely distributed bacillus that is found in soil and water in tropical and subtropical regions. It can infect

wounded individuals via environmental exposure (Bachti et al. 2021), and is difficult to treat in immune-suppressed individuals (Yang and Li 2011).

Since its discovery, investigations into *C. violaceum* biology have expanded significantly, leading to the exploitation of its biotechnological and pharmacological capacities (Kothari et al. 2017). One of the most remarkable characteristics of *C. violaceum* is the purple pigment violacein, produced by the *vio* operon, which has anticancer and antibiotic activity (Choi et al. 2015). The production of violacein is mediated by quorum sensing, and because the pigment is easily quantified, *C. violaceum* has been used as a model organism for studying the mechanisms by which this pigment exerts its beneficial effects (Durán et al. 2016; Kothari et al. 2017; Mion et al. 2021; Ahmed et al. 2021). These properties have resulted in *C. violaceum* being studied globally; however, as this species is resistant to transfection/conjugation or maintenance of plasmids, the development of new tools enabling the easy and reproducible engineering of the *C. violaceum* genome are urgently needed.

Few studies have described genetic manipulation of the *C. violaceum* genome to date. Dr. P. Williams' group obtained a white mutant, named CV026, by Tn5 transposon mutagenesis (McClean et al. 1997; Morohoshi et al. 2008). Stable transformation of *C. violaceum* was demonstrated using high-voltage electroporation (Broetto et al., 2006). A previous study described VioS, a repressor of the violacein operon (Devescovi et al. 2017), and episomal vectors were recently developed for *C. violaceum* and other non-model gram-negative bacteria (Liow et al. 2020). Curiously, the CV026 strain, which was used as an indicator of quorum sensing activity, has recently been reclassified as a member of the *C. subtsugae* species (Harrison and Soby 2020).

The complete ATCC 12472 *C. violaceum* genome revealed a single chromosome with more than 4.7 M bases. Approximately 40% of its open reading frames (ORFs) lack predicted functions (Vasconcelos and et al. 2003). Currently, this number has decreased to 18% due to conserved homology with other ORFs in other bacteria (personal communication). More than 10 years after the establishment of the *C. violaceum* genome, Lima et al. discovered a plasmid in this strain comprising a phage with DNA sequences to ensure plasmid partitioning, as well as hypothetical protein-encoding genes. In the same study, using restriction pattern analysis, the authors demonstrated that this plasmid may also be present in other strains isolated from Brazil (Lima et al. 2018). The presence of plasmids in several strains of *C. violaceum* might be useful for the further development of tools to facilitate the genetic manipulation of this organism, and thus to enhance our understanding of quorum sensing mechanisms and bioremediation processes.

The aim of the present study was to isolate, sequence, and compare plasmids from several Brazilian strains of *C. violaceum*. We obtained the sequences of six plasmids from strains isolated from the Amazon region and compared them with that of the reference plasmid pCh1V. We used phylogenetic analysis to trace the likely origins of these plasmids. The identification of novel plasmids could supply the means to effectively exploit *C. violaceum* in industrial microbiology applications.

2 Material and methods

2.1 Plasmids preparation

Plasmids were isolated from six Brazilian strains of *C. violaceum* (CVACO2, CVACO5, CVT8, CVT19, CVT24, and CVRP5), as described previously (Lima et

al. 2018). For each strain, three independent colonies were inoculated into three Erlenmeyer flasks containing 400 mL of LB medium for 16-18 hours. Cultures were centrifuged at 4°C for 5 min at 12,400 rcf. Pellets were resuspended in 25 mL of resuspension buffer (50 mM Tris-HCl, 10 mM EDTA, RNase 100 µg/mL, pH 8.0). Then, 25 mL of lysis buffer (SDS 1%, 0.2 M NaOH) was added, and the mixture was incubated for 5 min at 25°C. Plasmid DNA was precipitated by adding 25 mL of 3 M potassium acetate (pH 5.5), followed by centrifugation at 23,940 rcf for 10 min at 4°C. The supernatant was transferred to a new tube and 0.7 volumes of isopropanol was added. After an additional centrifugation at 23,940 rcf for 10 min at 4°C, the resulting pellets were resuspended in 1 mL of TE buffer and one volume of phenol:chloroform was added. After centrifugation at 20,850 rcf for 10 min at 4°C, the aqueous phase was transferred to a fresh tube, and the DNA was precipitated with one volume of isopropanol. Finally, the DNA was washed with 80% ethanol and four independent preparations were resuspended in 500 µL of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The plasmids were resolved electrophoretically on a 0.6% agarose gel at 50 V (Supplementary Data 1). The plasmid samples were then gel-purified using β-agarase (NEB, Ipswich, Massachusetts, USA, catalog #M0392S) according to the manufacturer's instructions.

2.2 Next generation sequencing (NGS) and assembly

Plasmids from the Brazilian strains were analyzed by NGS using an Ion Torrent Personal Genome Machine™ (PGM™) System (Life Technologies, Carlsbad, CA, USA). The integrity and size of the DNA were checked by agarose gel electrophoresis and quantified using a Qubit Fluorometer (Invitrogen). A library

was then generated using an Ion Xpress Plus Fragment Library Kit (Life Technologies) according to the manufacturer's instructions. DNA was fragmented by enzymatic lysis using Ion Shear Plus Reagent, linked to adapter P1 and Ion Xpress barcode adapter, and concomitantly nick repaired to connect the adapter and the DNA insert. Size selection of the library was performed using an E-Gel Size Select 2% agarose gel (Invitrogen, Carlsbad, CA, USA). An amplification reaction was set up in a final volume of 130 μ L with the following cycling profile: 95°C for 5 min, followed by 10 cycles of 95°C for 15 s, 58°C for 15 s, and 70°C for 60 s. After each step, purification was performed using magnetic beads with an Agencourt® AMPure® XP Kit (Beckman Coulter). The template was prepared using an Ion PGM™ Template OT2 Kit (Life Technologies). The libraries were diluted, and 25 pM was added to the aqueous master mix containing ion sphere particles (ISPs) at proportions specified by the manufacturer. An emulsion was created using an Ion OneTouch™ 2 instrument. After emulsion polymerase chain reaction (PCR), DNA-positive ISPs were recovered and enriched according to standard protocols. A sequencing primer was annealed to the DNA-positive ISPs, and the sequencing polymerase was bound prior to loading the ISPs on an Ion 318 sequencing chip. Sequencing of the samples was performed according to the Ion PGM 400 sequencing kit protocol. The run was programmed to include 850 nucleotide flows, to deliver 400 base read lengths on average. The libraries were sequenced using an Ion Torrent PGM™.

The sequenced libraries were assembled with SPAdes v3.10.1 (Bankevich et al. 2012) and grouped by each strain as single-end inputs with default parameters. The possible remaining gaps from the obtained scaffolds were closed with Redundans v. 10/2014 (Pryszcz and Gabaldón 2016), which was configured to join

contigs with a minimum of three pairs. Gap-free scaffolds were submitted to CAP3 software in 08/2013 (Huang and Madan 1999) for analysis with default parameters. New scaffolds obtained from CAP3 execution were aligned against the pChV1 reference genome with BWA-SW 0.7.12-r1039 (Li and Durbin 2010) with default parameters, and then manually reordered based on the obtained alignments. Annotated open reading frames (ORFs) of *C. violaceum* plasmid obtained from the NCBI database, accession number NC_005085.1, were mapped to assembled scaffolds via BLAST, version 2.6.0+ (Altschul et al. 1990), with a minimum coverage and identity cutoff of 30%. A synteny map of the location and the order of appearance of *C. violaceum* plasmid ORFs in each scaffold (Figure 1) was built using in-house developed Python (<https://www.python.org>) scripts.

As plasmids from strains CVT8, CVT19, CVT24, and CVRP5 were more dissimilar to pChV1, CVACO2, and CVACO5, we conducted *de novo* genome assembly to access the remaining sequences from these plasmids. To this end, we removed the reads that had been previously aligned with pChV1 and the *C. violaceum* ATCC 12472 reference genome and proceeded with the assembly process by initial assembly with SPAdes v3.10.1, followed by gap closing with Redundans v. 10/2014 using a minimum of three pairs. The resulting contigs were submitted for CAP3 analysis in 08/2013. The final contigs were annotated with D-FAST v. 1.1.6 (Tanizawa et al. 2018). All tools were used with the default parameters.

2.3 PCR conditions

PCR was performed using the Platinum Taq DNA polymerase (Invitrogen). One microliter of each plasmid was added to the reaction mix containing 0.75×

Buffer, 1 mM Mg, 5% dimethyl sulfoxide, 0.2 mM dNTPs, 0.3 mM of each primer, and 1 U DNA polymerase in a final volume of 25 μ L, as per the manufacturer's instructions. PCR conditions for amplification of region 1 comprised of an initial denaturation at 94°C for 5 min, 40 cycles of 94°C for 1 min, 65°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 5 min. PCR for region 2 comprised of an initial denaturation at 94°C for 5 min, 40 cycles of 94°C for 1 min, 67.9°C for 30 s, 72°C for 30 s, and a final extension at 72 °C for 5 min. PCR for region 3 comprised of an initial denaturation at 94°C for 5 min, 40 cycles of 94°C for 1 min, 67°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 5 min. Reactions with water instead of DNA were used as the negative controls. To verify the size of the amplicons, 5 μ L of the PCR reaction was resolved on a 0.8% agarose TAE gel electrophoretically and stained with ethidium bromide. A 1 kb (Promega) DNA ladder was used as the reference. The primer sequences used are shown in Supplementary Data 2.

2.4 Phylogenetic analysis

To evaluate the evolution of the plasmids, phylogenetic trees were constructed based on two different ORFs. We selected a phage tail ORF and hypothetical ORF 04, both of which were detected in all strains (Figure 1). First, we performed local alignment of each ORF using the blastx tool (Altschul et al. 1990) to search for similar sequences in the NCBI database. We then used ClustalW to align the sequences (Larkin et al. 2007). Trees were built using MEGA X, and evolutionary history was inferred based on the maximum likelihood method (Jones et al. 1992; Kumar et al. 2018).

We also conducted phylogenetic analysis to measure evolutionary relatedness between plasmids from Brazilian strains and pChV1 using *parA*, *parB* and *repA* genes. After ClustalW alignment we calculated the best model using MEGA X (Neighbor-joining tree and Maximum Likelihood). Trees were built using Maximum Likelihood method and Tamura-Nei model (Tamura and Nei 1993) and Phylogeny was tested using Bootstrap method (500 replications) with General Time Reversible model. Evolutionary analyses were conducted in MEGA X.

3 Results

In the present study, we describe the draft genomes of plasmids isolated from Brazilian strains of *C. violaceum*. All six sequenced plasmids had >260,000 reads of varying sizes, which were aligned with the previously published plasmid pChV1 (Table 1). Strains CVACO5 and CVACO2 appeared to harbor the same plasmid based on the number of reads aligned with pChV1, horizontal genome coverage, number of ORFs from pChV1, and the previously described restriction profile of the plasmids (Lima et al. 2018).

Figure 1 shows synteny analysis comparing the assemblies of plasmids isolated from Brazilian strains with pChV1 with all the ORF positions annotated in each strain (Figure 1). The majority of the plasmids had ORFs coding for the ParA and RepA plasmid replication proteins, which are involved in copy number regulation and partitioning. ORFs coding for phage proteins, such as phage tail, tail sheath, and baseplate, were also detected. We also identified two non-specific endonucleases that were not annotated when Lima *et al.* first described pChV1 (Lima et al. 2018). Detailed references of the ORFs found in all strains are listed in

Supplementary Data 3, which shows the results of a BLAST search against the non-redundant NCBI database.

This sequencing study corroborates the results of Lima *et al.* (Lima *et al.* 2018) and strongly suggests that the *C. violaceum* strains CVACO2, CVACO5, and CVT8 harbor a plasmid similar to that of the ATCC 12472 strain (pChV1). Conversely, strains CVT19, CVT24, and CVRP5 may contain different plasmids than ATCC 12472. We were unable to assemble any contigs with ORFs for the plasmid from strain CVRP5. Nevertheless, *de novo* genome assembly of this plasmid (and plasmids from CVT8, CVT19, and CVT24) produced assembled sequences with some ORFs (Table 2). We were unable to produce a proper *de novo* assembly for the plasmid from strain CVT19. Table 2 shows the new ORFs detected in plasmids from CVT8, CVT24 and CVRP5 strains. Most of the detected ORFs correspond to hypothetical proteins without any conserved domains. However, plasmids from CVT8 and CVT24 have ORFs that encode proteins for plasmid maintenance. The CVT8 plasmid contains ORFs coding for phage proteins (ORFs 06–09). Interestingly, plasmids from all strains had ORFs coding for transposases, such as IS5 and IS21 transposases, which are also involved in plasmid maintenance and distribution (Figure 1; Table 2). Supplementary Data 4 shows the amount and size of each contig produced after genome assembly in all six Brazilian strains of *C. violaceum*. It also describes the location of these ORFs in each contig.

PCR was used to confirm that some strains had the same plasmid as strain ATCC 12472, although some regions of the plasmids were not sequenced. Three pairs of primers were used to amplify three distinct regions of the plasmids. One region (region 1) was sequenced in all the Brazilian strains, whereas the other two

regions (regions 2 and 3) were not amplified in some strains (Figure 2). As the sequencing was partial, we performed PCR to confirm that plasmids from CVT19 and CVT24 were different from the other plasmids (Figure 2). Region 1 was amplified in plasmids isolated from all Brazilian strains. This region has already been sequenced for most strains (Figure 1). In contrast, region 2 was absent in plasmids from CVT19 and CVRP5, whereas region 3 was not present in CVT8 alone. This provides further evidence that the two strains (CVACO2 and CVACO5) have the same extra-chromosomal DNA as ATCC 12472, whereas the other plasmids diverged from pChV1 during evolution. Despite some similarities in partial alignment (Figure 1), the PCR results clearly show the absence of some regions (Figure 2), corroborating the previously published restriction pattern analysis (Lima et al. 2018).

We next constructed phylogenetic trees based on the two different ORFs that were detected in plasmids from all strains (except for CVRP5). Figure 3A shows a tree based on ORF_04, which codes for a hypothetical protein without a conserved domain. Figure 3B depicts a tree based on an ORF that codes for a phage tail protein. Phylogenetic analysis revealed that plasmids from strains CVACO2 and CVACO5 were more closely related to pChV1 than the other sequenced strains, at least with respect to ORF_04.

To further measure evolutionary relatedness between these plasmids we built three phylogenetic trees from *parA*, *parB* and *repA* genes (Figure 4). These genes were detected in strains CVACO2, CVACO5, CVT8 and CVT19, although in some of them we were not able to provide the entire ORF (Figure 1). Accordingly to the trees, plasmid from CVT8 is more distantly related to other plasmids while CVACO2, CVACO5 and CVT19 form a more uniform group with pChV1. The low

reliability values of some branches can be explained due the presence of partial ORFs in this analysis. Nevertheless, the phylogenetic analysis together with sequencing data reinforces our assumption that pChV1 is more similar to plasmids isolated from CVACO2 and CVACO5 strains.

4 Discussion

In the present study, we sequenced plasmids isolated from six Brazilian strains of *C. violaceum* (CVACO2, CVACO5, CVT8, CVT19, CVT24, and CVRP5). All six strains were previously analyzed by Lima *et al.* (Lima et al. 2018) when they isolated an undiscovered plasmid from *C. violaceum* ATCC 12472, who used restriction enzyme analysis to show that some Brazilian strains may have a plasmid similar to pChV1. The Illumina® Platform was used by Lima *et al.* to sequence plasmid pChV1 (Lima et al. 2018), whereas we used Ion Torrent NGS technology in the present study.

Alignment of 6 plasmids to the pChV1 genome reference revealed that strains CVACO2, CVACO5, and CVT8 closely matched the previously published sequence, suggesting that these strains likely harbored a plasmid similar to pChV1. However, comparison of restriction patterns with the sequencing data suggested that CVCA5 and CVCO2 possess the same plasmid, whereas CVT8 was different from pChV1 (Lima et al. 2018). In agreement with Lima et al. (2018), in the present study, plasmids from strains CVT19, CVT24, and CVRP5 had minimal alignment with pChV1, confirming the presence of a different plasmid. *De novo* genome assembly confirmed this, based on the presence of plasmids containing ORFs that are not present in pChV1. The position of RPA changed in

these plasmids, suggesting a reordering of the genome in plasmids from strains CVACO2 and CVT8.

Many ORFs from plasmids derived from CVT8, CVT24, and CVRP5 strains were annotated as sequences belonging to *C. violaceum*. When its genome was first sequenced (Vasconcelos et al. 2003), pChV1 was not detected. New genome sequencing of *C. violaceum* first revealed the presence of the pChV1 sequence, although Lima *et al.* (2018) were the first to describe this genetic element. Thus, the matching of some ORFs described in the present study to sequences belonging to *C. violaceum* reflects the fact that several strains of this bacterium probably harbor one or more plasmids.

Both PCR and sequencing results strongly suggested that CVACO2 and CVACO5 contain a plasmid that is similar to pChV1. However, we were unsuccessful in obtaining 100% horizontal coverage of any of the other strains. Several studies describing the differences between available NGS platforms have been conducted (Li and Durbin 2010; Goodwin et al. 2016; Shendure et al. 2017). For example, Quail *et al.* (2012) compared ion torrent technology with two other platforms, and demonstrated that in complex genomes with a significant number of repetitive regions, such as *Plasmodium falciparum*, the performance of ion torrent was worse than that of other platforms, enabling only 65% coverage of the genome. The genome of pChV1 has both repetitive sequences and homopolymer regions, which may have hampered the sequencing of some regions of the plasmids in the present study.

Five of the Brazilian strains contained ORFs related to plasmid replication and maintenance, including RepA, RepB, and ParA proteins (Supplementary data 3). Plasmids from CVACO5, CVACO2, and CVT8 strains presented 33, 31, and 21

ORFs from pChV1, respectively, which reinforces our view that these three strains most likely harbor the same or a very similar plasmid to pChV1. Conversely, the other 3 strains may have a plasmid that is more divergent from pChV1. Nevertheless, except for CVRP5, all strains had some ORFs that were present in the reference plasmid, suggesting that all of them share a common evolutionary origin (Figure 3 and Figure 4). It is important that several ORFs encoding transposase genes were detected in all six plasmids analyzed. Transposases are enzymes that are involved in the translocation of DNA transposons to another location in the plasmid and/or genome (Harmer et al. 2020). In fact, more than one transposase gene can be found in a single plasmid, likely due to the heteromeric structure of some of these enzymes (Peters et al. 2014; Wang et al. 2020).

The presence of several ORFs coding for phage proteins, such as phage tail, phage sheath, and baseplate, suggests that this extra-chromosomal DNA is a defective P1-like bacteriophage that infected a distant ancestor of the current *C. violaceum* and diverged across different strains. The presence of both plasmid maintenance and bacteriophage-related genes has also been documented in N15 and SSU5, and the term 'phage-plasmids' (P-Ps) was proposed for these mobile genetic elements (Kim et al. 2012; Ravin 2015). Recently, Pfeifer et al. screened over 15,000 plasmid and bacteriophage sequences, and identified 780 P-Ps (Pfeifer et al. 2021). These authors suggest that P-Ps are not just chimeric mobile genetic elements, but have an ancient evolutionary origin.

In the present study, we analyzed the sequences of plasmids from six different strains of *C. violaceum* from the Amazon region. Plasmids in at least two of these strains (CVACO2 and CVACO5) are identical to pChV1. The other strains likely harbor one or more extra-chromosomal genetic elements that are different

from pChV1. We identified ORFs coding for hypothetical proteins, plasmid-maintaining proteins, and transposases. The discovery of different plasmids in several strains of *C. violaceum* may be useful for the further development of tools for easy, accurate, and reproducible modification of its genome. To achieve this, more research on these plasmids should be conducted to fully understand the role of these extra-chromosomal elements in *C. violaceum*, including the function of the newly identified ORFs.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We would like to thank Spartaco Astolfi Filho, who supplied all the Brazilian strains of *C. violaceum*.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

Author contribution

D.C.L. and R.C.S. isolated the plasmids and conducted the NGS sequencing; I.G.M. conducted the assembly and annotation of the genomes; F.C.S.J. performed PCR analysis and interpreted the results; D.C.L., I.G.M., L.F.A., J.E.S.S. and S.R.B. worked in designing, writing and editing of the manuscript.

Data availability

Raw data were submitted to NCBI (BioProject ID PRJNA755587).

5 References

- Ahmed, S.O., Zedan, H.H., and Ibrahim, Y.M. 2021. Quorum sensing inhibitory effect of bergamot oil and aspidosperma extract against *Chromobacterium violaceum* and *Pseudomonas aeruginosa*. Arch. Microbiol. doi:10.1007/S00203-021-02455-8.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2.
- Bachti, A., Josephine, D., Evan, S., and Guntur, D. 2021. *Chromobacterium violaceum*: A Review of an Unexpected Scourge. Int. J. Gen. Med. **14**: 3259–3270. Int J Gen Med. doi:10.2147/IJGM.S272193.
- 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., Pyshkin, A. V, Sirotkin, A. V, Vyahhi, N., Tesler, G., Alekseyev, M.A., and Pevzner, P.A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. **19**(5): 455–77. doi:10.1089/cmb.2012.0021.
- Boisbaudran, P.-É.L. de. 1882. Matière colorante se formant dans la colle de farine. C. R. Hebd. Seances Acad. Sci.
- Broetto, L., Cecagno, R., Sant’Anna, F.H., Weber, S., and Schrank, I.S. 2006. Stable transformation of *Chromobacterium violaceum* with a broad-host-range plasmid. Appl. Microbiol. Biotechnol. doi:10.1007/s00253-005-0140-5.
- Choi, S.Y., Yoon, K.H., Lee, J. II, and Mitchell, R.J. 2015. Violacein: Properties

- and production of a versatile bacterial pigment. doi:10.1155/2015/465056.
- Devescovi, G., Kojic, M., Covaceuszach, S., Cámara, M., Williams, P., Bertani, I., Subramoni, S., and Venturi, V. 2017. Negative regulation of violacein biosynthesis in *Chromobacterium violaceum*. Front. Microbiol. doi:10.3389/fmicb.2017.00349.
- Durán, N., Justo, G.Z., Durán, M., Brocchi, M., Cordi, L., Tasic, L., Castro, G.R., and Nakazato, G. 2016. Advances in *Chromobacterium violaceum* and properties of violacein-Its main secondary metabolite: A review. doi:10.1016/j.biotechadv.2016.06.003.
- Goodwin, S., McPherson, J.D., and McCombie, W.R. 2016. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet **17**: 333–351. doi:10.1038/nrg.2016.49.
- Harmer, C.J., Pong, C.H., and Hall, R.M. 2020. Structures bounded by directly-oriented members of the IS26 family are pseudo-compound transposons. Plasmid **111**: 102530. doi:10.1016/j.plasmid.2020.102530.
- Harrison, A.M., and Soby, S.D. 2020. Reclassification of *Chromobacterium violaceum* ATCC 31532 and its quorum biosensor mutant CV026 to *Chromobacterium subtsugae*. AMB Express. doi:10.1186/s13568-020-01140-1.
- Huang, X., and Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Res. **9**(9): 868–77. doi:10.1101/gr.9.9.868.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. The rapid generation of mutation data matrices. Bioinformatics **8**(3): 275–82.
- Kim, M., Kim, S., and Ryu, S. 2012. Complete Genome Sequence of Bacteriophage SSU5 Specific for *Salmonella enterica* serovar Typhimurium

- Rough Strains. J. Virol. doi:10.1128/jvi.01796-12.
- Kothari, V., Sharma, S., and Padia, D. 2017. Recent research advances on *Chromobacterium violaceum*. doi:10.1016/j.apjtm.2017.07.022.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35**(6): 1547–1549. doi:10.1093/molbev/msy096.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**(21): 2947–8. doi:10.1093/bioinformatics/btm404.
- Li, H., and Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26**(5): 589–95. doi:10.1093/bioinformatics/btp698.
- Lima, D.C., Nyberg, L.K., Westerlund, F., and De Medeiros, S.R.B. 2018. Identification and DNA annotation of a plasmid isolated from *Chromobacterium violaceum*. Sci. Rep. **8**(1). doi:10.1038/s41598-018-23708-5.
- Liow, L.T., Go, M.K., Chang, M.W., and Yew, W.S. 2020. Toolkit Development for Cyanogenic and Gold Biorecovery Chassis *Chromobacterium violaceum*. ACS Synth. Biol. doi:10.1021/acssynbio.0c00064.
- McClean, K.H., Winson, M.K., Fish, L., Taylor, A., Chhabra, S.R., Camara, M., Daykin, M., Lamb, J.H., Swift, S., Bycroft, B.W., Stewart, G.S.A.B., and Williams, P. 1997. Quorum sensing and *Chromobacterium violaceum*: Exploitation of violacein production and inhibition for the detection of N-acylhomoserine lactones. Microbiology. doi:10.1099/00221287-143-12-3703.

- Mion, S., Carriot, N., Lopez, J., Plener, L., Ortalo-Magné, A., Chabrière, E., Culioli, G., and Daudé, D. 2021. Disrupting quorum sensing alters social interactions in *Chromobacterium violaceum*. *npj Biofilms Microbiomes*. doi:10.1038/s41522-021-00211-w.
- Morohoshi, T., Kato, M., Fukamachi, K., Kato, N., and Ikeda, T. 2008. N-acylhomoserine lactone regulates violacein production in *Chromobacterium violaceum* type strain ATCC 12472. *FEMS Microbiol. Lett.* **279**(1): 124–30. doi:10.1111/j.1574-6968.2007.01016.x.
- Peters, J.E., Fricker, A.D., Kapili, B.J., and Petassi, M.T. 2014. Heteromeric transposase elements: Generators of genomic Islands across diverse bacteria. *Mol. Microbiol.* **93**(6): 1084–92. doi:10.1111/mmi.12740.
- Pfeifer, E., Moura De Sousa, J.A., Touchon, M., and Rocha, E.P.C. 2021. Bacteria have numerous distinctive groups of phage-plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res.* doi:10.1093/nar/gkab064.
- Pryszcz, L.P., and Gabaldón, T. 2016. Redundans: An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**(12): e113. doi:10.1093/nar/gkw294.
- Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., and Gu, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**(1): 341. BioMed Central.
- Ravin, N. V. 2015. Replication and Maintenance of Linear Phage-Plasmid N15. *Microbiol. Spectr.* doi:10.1128/microbiolspec.plas-0032-2014.
- Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J.,

- Schloss, J.A., and Waterston, R.H. 2017. DNA sequencing at 40: Past, present and future. *Nature* **550**: 345–353. doi:10.1038/nature24286.
- Tamura, K., and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* doi:10.1093/oxfordjournals.molbev.a040023.
- Tanizawa, Y., Fujisawa, T., and Nakamura, Y. 2018. DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* **34**(6): 1037–1039. doi:10.1093/bioinformatics/btx713.
- Vasconcelos, A.T.R. de, Almeida, D.F. de, Hungria, M., Guimarães, C.T., Antônio, R.V., Almeida, F.C., Almeida, L.G.P. de, Almeida, R. de, Alves-Gomes, J.A., Andrade, E.M., Araripe, J., Araújo, M.F.F. de, Astolfi-Filho, S., Azevedo, V., Baptista, A.J., Bataus, L.A.M., Batista, J. da S., Beló, A., Berg, C. van den, Bogó, M., Bonatto, S., Bordignon, J., Brigido, M.M., Brito, C.A., Brocchi, M., Burity, H.A., Camargo, A.A., Cardoso, D. das D. de P., Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C. de M., Cavada, B.S., Chueire, L.M.O., Creczynski-Pasa, T.B., Cunha-Junior, N.C. da, Fagundes, N., Falcão, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S., Ferrari, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R., Freitas, N.S.A. de, Furlan, L.R., Gazzinelli, R.T., Gomes, E.A., Gonçalves, P.R., Grangeiro, T.B., Grattapaglia, D., Grisard, E.C., Hanna, E.S., Jardim, S.N., Laurino, J., Leoi, L.C.T., Lima, L.F.A., Loureiro, M. de F., Lyra, M. do C.C.P. de, Madeira, H.M.F., Manfio, G.P., Maranhão, A.Q., Martins, W.S., Mauro, S.M.Z. di, Medeiros, S.R.B. de, Meissner, R. de V., Moreira, M.A.M., Nascimento, F.F. do, Nicolás, M.F., Oliveira, J.G., Oliveira, S.C., Paixão, R.F.C., Parente, J.A., Pedrosa, F. de O., Pena, S.D.J., Pereira, J.O.,

Pereira, M., Pinto, L.S.R.C., Pinto, L. da S., Porto, J.I.R., Potrich, D.P.,
 Ramalho-Neto, C.E., Reis, A.M.M., Rigo, L.U., Rondinelli, E., Santos, E.B.P.
 do, Santos, F.R., Schneider, M.P.C., Seuanez, H.N., Silva, A.M.R., Silva,
 A.L. da C. da, Silva, D.W., Silva, R., Simões, I. de C., Simon, D., Soares,
 C.M. de A., Soares, R. de B.A., Souza, E.M., Souza, K.R.L. de, Souza, R.C.,
 Steffens, M.B.R., Steindel, M., Teixeira, S.R., Urmenyi, T., Vettore, A.,
 Wassem, R., Simpson, A.Z., and Andrew John George. 2003. The Complete
 Genome Sequence of *Chromobacterium violaceum* Reveals Remarkable
 and Exploitable Bacterial Adaptability. *Proc. Natl. Acad. Sci. U. S. A.*
100(20): 11660–11665. doi:10.1073/pnas.1832124100.

Vasconcelos, A.T.R., and et al. 2003. The complete genome sequence of
Chromobacterium violaceum reveals remarkable and exploitable bacterial
 adaptability. *Proc. Natl. Acad. Sci. U. S. A.* **100**(20): 11660–5.
 doi:10.1073/pnas.1832124100.

Wang, H.C., Lin, S.J., Mohapatra, A., Kumar, R., and Wang, H.C. 2020. A review
 of the functional annotations of important genes in the ahpnd-causing pva1
 plasmid. *Microorganisms* **8**(7): 996. doi:10.3390/microorganisms8070996.

Yang, C.H., and Li, Y.H. 2011. *Chromobacterium violaceum* infection: A clinical
 review of an important but neglected infection. *J. Chinese Med. Assoc.*
74(10): 435–441. doi:10.1016/j.jcma.2011.08.013.

Strain	Number of reads	Reads aligned with pChV1	Genome coverage (%)	ORFs	Number of Contigs
CVACO5	312611	212006	67	33	12
CVACO2	306254	58905	50	31	9
CVT8	286690	14041	41	21	10
CVT19	268146	492	30	16	10
CVT24	377791	86	4	2	2
CVRP5	331100	89	2	0	1

Table 1

Draft

ORF NUMBER	ANNOTATION	LENGTH
	CVT8	
ORF_01	OmpA_C-like super family and Na_H_Exchanger super family domains [<i>Chromobacterium violaceum</i>]	606
ORF_02	MotB super family and MotB super family domains [<i>Chromobacterium violaceum</i>]	327
ORF_03	hypothetical protein - No Conserved Domains [<i>Chromobacterium violaceum</i>]	246
ORF_04	hypothetical protein - No Conserved Domains [<i>Chromobacterium violaceum</i>]	104
ORF_05	hypothetical protein - No Conserved Domains [<i>Chromobacterium violaceum</i>]	247
ORF_06	phage minor tail protein L [<i>Chromobacterium vaccinii</i>]	230
ORF_07	phage tail protein [<i>Chromobacterium vaccinii</i>]	124
ORF_08	phage tail sheath family protein [<i>Xenophilus sp.</i> AP218F]	510
ORF_09	phage tail protein [<i>Chromobacterium violaceum</i>]	148
ORF_10	hypothetical protein - No Conserved Domains [<i>Chromobacterium violaceum</i>]	140
ORF_11	IS5 family transposase [<i>Chromobacterium phragmitis</i>]	310
ORF_12	hypothetical protein - No Conserved Domains [<i>Chromobacterium violaceum</i>]	101
ORF_13	AAA family ATPase [<i>Chromobacterium violaceum</i>]	398
ORF_14	hypothetical protein - No Conserved Domains [<i>Chromobacterium violaceum</i>]	135
ORF_15	hypothetical protein - No Conserved Domains [<i>Chromobacterium violaceum</i>]	308

CVT24

ORF_01	transposase [<i>Pseudogulbenkiania</i> sp. MAI-1]	176
ORF_02	RepB family DNA primase [<i>Escherichia coli</i>]	314
ORF_03	Initiator Replication protein [<i>Acidithiobacillus thiooxidans</i>]	301
ORF_04	BrnA antitoxin family protein [<i>Xylella fastidiosa</i>]	101
ORF_05	BrnT family toxin [<i>Acidithiobacillus thiooxidans</i>]	102
ORF_06	replication initiation protein [<i>Aquitalea magnusonii</i>]	252
ORF_07	DNA cytosine methyltransferase [<i>Extensimonas vulgaris</i>]	318
ORF_08	NgoMIV family type II restriction endonuclease [<i>Pseudomonas formosensis</i>]	287
ORF_09	ATPase [<i>Chromobacterium amazonense</i>]	203
ORF_10	IS21-like element helper ATPase IstB [<i>Chromobacterium paludis</i>]	243
ORF_11	IS21 family transposase [<i>Chromobacterium paludis</i>]	496
ORF_12	IS21 family transposase, partial [<i>Chromobacterium amazonense</i>]	337
ORF_13	hypothetical protein - No Conserved Domain [<i>Burkholderia</i> sp. BCC0398]	503
ORF_14	hypothetical protein - No Conserved Domain [<i>Azospirillum brasilense</i>]	140
ORF_15	transposase [<i>Burkholderia</i>]	153
ORF_16	transposase [<i>Burkholderia cenocepacia</i>]	126
ORF_17	hypothetical protein - No Conserved Domain [<i>Burkholderia ubonensis</i>]	83
ORF_18	hypothetical protein - No Conserved Domain [<i>Burkholderia gladioli</i>]	132

ORF_19	IS256 family transposase [<i>Paludibacterium paludis</i>]	378
ORF_20	Tn3 family transposase, partial [<i>Xanthomonas arboricola</i>]	171
ORF_21	hypothetical protein	82
ORF_22	hypothetical protein - Saf-pilin pilus formation protein SafA [<i>Kosakonia arachidis</i>]	172
ORF_23	hypothetical protein - No Conserved Domain [<i>Chromobacterium vaccinii</i>]	95
ORF_24	TraY domain-containing protein [<i>Burkholderiales bacterium</i>]	76
ORF_25	type II toxin-antitoxin system RelE/ParE family toxin [<i>Janthinobacterium sp.</i> B9-8]	89
ORF_26	hypothetical protein (plasmid) - Rep_3 super family [uncultured prokaryote]	304
ORF_27	IS5 family transposase [<i>Chromobacterium piscinae</i>]	310
ORF_28	IS5 family transposase [<i>Chromobacterium haemolyticum</i>]	167
ORF_29	IS5 family transposase [<i>Chromobacterium vaccinii</i>]	366
ORF_30	hypothetical protein - No Conserved Domain [<i>Pseudomonas hussainii</i>]	90
ORF_31	transposase [<i>Chromobacterium amazonense</i>]	279
ORF_32	transposase [<i>Chromobacterium haemolyticum</i>]	91

CVRP5

ORF_01	hypothetical protein - No Conserved Domain [<i>Chromobacterium violaceum</i>]	67
ORF_02	hypothetical protein - No Conserved Domain [<i>Chromobacterium violaceum</i>]	395
ORF_03	hypothetical protein - No Conserved Domain [<i>Chromobacterium violaceum</i>]	186

ORF_04	IS5 family transposase [<i>Chromobacterium violaceum</i>]	378
ORF_05	transposase, partial [<i>Pseudomonas veronii</i>]	62
ORF_06	lytic transglycosylase domain-containing protein [<i>Chromobacterium violaceum</i>]	145
ORF_07	prepilin peptidase [<i>Chromobacterium violaceum</i>]	260

Table 2

Legends

Figure 1

Comparison of the six plasmids isolated from Brazilian strains of *C. violaceum* with that of pChV1, which was used as the reference sequence in this assembly. Synteny analysis suggests reordering of the RPA ORF in plasmids from strains CVACO2 and CVT8, and non-specific endonucleases in CVT8 and CVT19. The numbers in 'primers' are the regions amplified by PCR in all plasmids. White dots inside some ORFs represent partial ORFs, probably due to low sequencing coverage of these regions. Arrows representing each ORF were drawn in scale.

Figure 2

PCR of three different regions from *C. violaceum* strains. The regions from pChV1 were chosen to assess if these regions were also present in the plasmids from other Brazilian strains.

Figure 3

Phylogenetic analysis of ORF_04 (A) and the ORF coding for phage tail protein (B) based on the maximum likelihood method. The sequence from CVT19 was removed from the analysis in A, as its annotated ORF_04 was partial. The evolution inference in A corroborates the hypothesis that plasmids from CVACO2 and CVACO5 strains are more similar to pChV1. Black circles represent

Chromobacterium spp., and white circles represent sequences from Brazilian strains of *C. violaceum*.

Figure 4

Phylogenetic analysis of *parA* (Figure 4A), *parB* (Figure 4B) and *repA* (Figure 4C) present in plasmids isolated in CVACO2, CVACO5, CVT8 and CVT19 strains and in pChV1. Trees with the highest log likelihood are shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value.

Table 1

Summary of the sequenced plasmids from Brazilian strains.

Table 2

ORFs found in plasmids from CVT8, CVT24, and CVRP5 strains after *de novo* assembly. The species names in brackets refer to the organism whose sequence had the best blastp search result. To perform *de novo* assembly, we removed the reads that aligned previously with pChV1 and *C. violaceum* genome. The ORFs in

this table are new putative genes that are present in plasmids from *C. violaceum* Brazilian strains.

Draft

Figure 1

Draft

Figure 2

Draft

Figure 3

Figure 4

© The Author(s) or their Institution(s)

Figure 2 PCR reactions of three different regions from *C. violaceum* strains. The regions from pChV1 were chosen to assess if they were also present in the Brazilian strains.

225x127mm (300 x 300 DPI)

Figure 3 Phylogenetic analysis of ORF_04 (A) and the ORF coding for phage tail protein (B) based on Maximum Likelihood method. Sequence from CVT19 was removed from analysis in A as the annotated ORF_04 in this strain was partial. The evolution inference in A corroborates that plasmids from strains CVACO2, CVACO5, and CVT8 are more similar to pChV1. Black circles represent *Chromobacterium* spp. whereas white circles represent sequences from Brazilian strains of *C. violaceum*.

225x127mm (300 x 300 DPI)

Phylogenetic analysis of parA (Figure 4A), parB (Figure 4B) and repA (Figure 4C) present in plasmids isolated in CVACO2, CVACO5, CVT8 and CVT19 strains and in pChV1. Trees with the highest log likelihood are shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value.

60x72mm (300 x 300 DPI)