Current Research in Biotechnology 4 (2022) 152–166

Contents lists available at ScienceDirect

Current Research in Biotechnology

journal homepage: www.elsevier.com/locate/crbiot

Identification of putative producers of rhamnolipids/glycolipids and their transporters using genome mining

Meryam Magri [a], Ahmad M. Abdel-Mawgoud [a,b,*]

[a] Institute of Integrative Biology and Systems, Laval University, 1030 Ave. de la Médecine, Quebec, QC G1V 0A6, Canada
[b] Department of Biochemistry, Microbiology and Bioinformatics, Faculty of Science and Engineering, Laval University, 1045 Ave. de la Médecine, Quebec, QC G1V 0A6, Canada

ARTICLE INFO

Keywords:
Rhamnolipids
RhlABC orthologs
Genome mining
Thin layer chromatography
Rhamnolipid transport

ABSTRACT

Rhamnolipids (RLs) are microbial glycolipids (GLs) with interesting structure-dependent bioactivities and physicochemical properties making them suitable for diverse medical and industrial applications. The discovery of RLs with more interesting bioactivities and properties has relied on laborious screening of new RL producers isolated from the environment, and has resulted in the redundant identification of already known RL producers and structures. Here, we present a genome mining approach that enabled the identification of 80 RL-producing species (including the two reference species), 71 of which were previously unreported. Distance trees of two of their RL biosynthetic enzymes, RhlAB, allowed for the identification of 11 distinct clades. Preliminary experimental validation with thin layer chromatography on one non-pathogenic RL/GL producer, *Nevskia soli*, confirmed its putative production of RLs. Additionally, this study led to the discovery of the putative RL transport mechanism involving three transmembrane proteins whose coding genes are highly conserved and clustered with one of the RL biosynthetic gene clusters in most RL/GL producers identified in this study.

Introduction

Rhamnolipids (RLs) are extracellular microbial glycolipids (GLs) produced by certain bacterial species, the best known of which are *Pseudomonas* and *Burkholderia* species (Abdel-Mawgoud et al., 2010). RLs are amphiphilic molecules with hydrophobic and hydrophilic structural moieties conferring interesting surfactant properties (Abdel-Mawgoud et al., 2010). The hydrophobic moiety is a 3-hydroxyalkanoyloxyalkanoate (HAA), a dimer of two ester-linked *R*-3-hydroxyalkanoates, which is associated with a monomeric or dimeric rhamnose unit, the hydrophilic moiety, via *O*-glycosidic bonds, thus creating mono-RLs or di-RLs congeners (Fig. 1) (Abdel-Mawgoud et al., 2010). In addition to these congeners, RLs show a diversity of homologs differing in fatty acid length, branching and chain saturation (Abdel-Mawgoud et al., 2010). For simplicity, mono-RLs and di-RLs with two 10-carbon fatty acid chains are denoted RC10C10 and RRC10C10, respectively (RCxCy and RRCxCy, where R denotes rhamnose units, and x and y indicate the numbers of carbon atoms of the acyl chains). RLs are produced as mixtures of congeners and homologs, whose composition differs across microbial species (Tiso et al., 2015) and even in the same species at different genetic backgrounds or cultivation conditions (Abdel-Mawgoud et al., 2014b).

RLs have a wide range of biological activities, such as antibacterial (Haba et al., 2003), antifungal (Haba et al., 2003), antiviral (Remichkova et al., 2008), antiparasitic (Silva et al., 2015), immunomodulatory (Andrä et al., 2006; Luzuriaga-Loaiza et al., 2018) and anti-adherent activities, as extensively reviewed in several articles (Abdel-Mawgoud and Stephanopoulos, 2018; Rodrigues et al., 2006; Thakur et al., 2021). Furthermore, RLs are excellent detergents having high biodegradability and low toxicity. These characteristics have enabled RL to be used in bioremediation, agricultural and cosmeceutical industries (Abalos et al., 2001). Interestingly, the nature and extent of RL's bioactivities and physicochemical properties vary significantly across RL homologs and congeners. For example, the surface activity of RLs, measured in terms of the critical micelle concentration (CMC), is structure dependent, such that mono-RLs (RC10C10, Fig. 1**A**) have CMCs of approximately 40 mg/L, whereas di-RLs (RRC10C10, Fig. 1**B**) are more surface active, with CMCs as low as 5 mg/L (Tiso et al., 2015). This surface activity decreases in di-RLs with longer fatty acid chain length, such as di-RLs with C14-fatty acids (RRC14C14, Fig. 1**B**), whose CMC is 25 mg/L (Costa

* Corresponding author at: Pavillon Charles-Eugène-Marchand, 1030, avenue de la Médecine, Room 3167 (Office), Québec (Québec) G1V 0A6, Canada.
E-mail address: ahmad.saleh@bcm.ulaval.ca (A.M. Abdel-Mawgoud).

https://doi.org/10.1016/j.crbiot.2022.02.002
Received 14 October 2021; Revised 8 February 2022; Accepted 11 February 2022

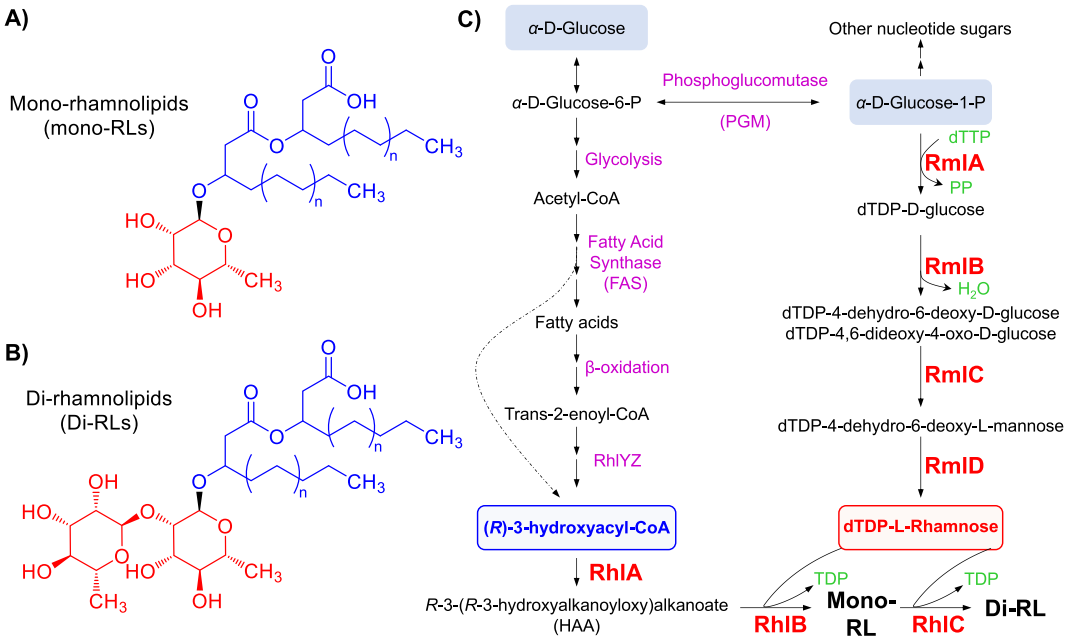


Fig. 1. Chemical structures and biosynthetic pathway of rhamnolipids. **(A)** Mono-rhamnolipids (mono-RLs, RC10C10). **(B)** Di-rhamnolipids (di-RLs, RRC10CC10). **(C)** Biosynthesis of RLs. Fatty acid precursors of RLs are channeled mainly from β-oxidation, although supply from FAS II is not excluded (dotted line).

et al., 2011). RL mixtures containing RLs with unsaturated fatty acid chains (e.g., RC12:1C10) and/or lacking one of the two fatty acid chains (e.g., RC10) have lower surface activities, where their CMCs are as high as 234 mg/L (Abalos et al., 2001). This structure-dependent surface activity might be attributable to the different three-dimensional conformations taken by different homologs and congeners of RLs that affect their aggregation behaviors and consequently their CMC values (Abalos et al., 2001).

More interestingly, the bioactivities of RLs are also structure dependent. For example, the long-chain di-RL, RRC14C14, but not the medium-chain RL, RRC10C10, has immunomodulatory activity on human leukocytes (Van Gennip et al., 2009). Furthermore, the immunomodulatory activity of RLs was reported to be fatty acid chain length-dependent, as demonstrated by the RL-mediated immunity of *Arabidopsis* against plant pathogens (Luzuriaga-Loaiza et al., 2018).

The structure dependency of RL bioactivities and physicochemical properties has prompted the search for new producers of RLs with novel structures, which might have more interesting bioactivities and physicochemical properties. Traditionally, the search for new RL/GL producers has relied on classical screening of environmental microorganisms. The tensioactive properties of RLs were the primary basis for classical screening of RL/GL producers based on the detection of the surface or interfacial tension lowering effects of RLs in the culture supernatants of various microorganisms (Walter et al., 2010). These surface tension measurements are conducted on culture supernatants with methods including Du-Nouy-Ring tensiometry, drop collapse tests, oil spread tests and the measurement of emulsification index (Walter et al., 2010). Another indirect method for screening for RL producers relying on another physical property of RLs include the bacterial adhesion to hydrocarbon (BATH) and similar assays that are based on the RL-mediated increase of cell surface hydrophobicity (Walter et al., 2010). The chemical properties of RLs were also the basis for other methods for screening for RL producers, such as the cetyltrimethylammonium bromide (CTAB) assay that relies on the interaction of anionic RLs with cationic surfactants (e.g. CTAB) and cationic methylene blue. Anionic molecules, including RLs, form an insoluble ion pair that appears as a dark blue halo around RL-producing colonies on mineral agar medium containing CTAB and methylene blue (Siegmund and Wagner, 1991). The biological effects of RLs have also been used to indirectly screen for RL producers, such as in the hemolysis test. This test is based on the hemolytic effect of RLs that form a clear hemolysis zones around RL-producing colonies grown on blood agar. Finally, RLs are screened for through the direct chemical analysis of target RL molecules using liquid chromatography coupled to mass spectrometry (MS) (Abdel-Mawgoud et al., 2014a; Déziel et al., 2000; Gauthier et al., 2019). Apart from the MS-based direct chemical analysis of RLs, all RL screening methods are indirect, non-specific and laborious, and they are subject to substantial interference from many confounding factors (Walter et al., 2010). Despite being the only direct and precise methods for screening RL/GL producers, MS-based methods are expensive and low throughput, and they rely on prior knowledge of the chemical structure or at least the exact mass/molecular formula of the produced RLs. Moreover, these previously described approaches have largely resulted in redundant identification of already known RL structures and producers (Haba et al., 2000). This could be partly attributed to the fact that these direct and indirect screening approaches all require the cultivation of potential RL producers; however, non-cultivable or scarcely cultivable organisms cannot be efficiently screened. Given the limiting biological factors in RL production and discovery, chemical synthesis of RLs was sought to bypass or decrease the reliance on biological production of RLs, to potentially enable the overproduction of typical RLs or accelerate the discovery of novel ones (Bauer et al., 2006; Cloutier et al., 2021; Compton et al., 2020; Nott et al., 2013; Palos Pacheco et al., 2017). Although chemical synthesis of RLs has yielded both typical and novel RLs, these methods are largely tedious, expensive, ecologically unfriendly, and less sustainable.

Bioinformatic tools can be used for easy and fast discovery of homologs of RL biosynthetic enzymes producing typical or novel RLs. This approach requires prior knowledge of the biosyntehtic enzymes of RLs. Briefly, RLs are biosynthesized from rhamnose and fatty acid precursors in three sequential reactions catalyzed by RhlA, RhlB and RhlC (denoted RhlABC for simplicity) (Déziel et al., 2003). RhlA is an acyl transferase that synthesizes HAA from two *R*-3-hydroxy fatty acids (Abdel-Mawgoud et al., 2013) mainly supplied from β-oxidation (Abdel-Mawgoud et al., 2014b). RhlB and RhlC are

rhamnosyl transferases that use dTDP-L-rhamnose to rhamnosylate HAA and form mono-RLs (Ochsner et al., 1994) and di-RLs (Rahim et al., 2001), respectively (Fig. 1**C**). RhlABC is expressed from *rhlABC*, in which *rhlAB* are often clustered together (in Pseudomonadales and Burkholderiales and sometimes in Enterobacteriales), with (e.g. in Burkholderiales) or without (e.g. in Pseudomonadales) *rhlC*. The rhamnosyl donor itself, dTDP-L-rhamnose, is biosynthesized from glucose-1-phosphate in four metabolic steps catalyzed by RmlB, RmlD, RmlA and RmlC (denoted RmlBDAC for simplicity) (Abdel-Mawgoud et al., 2011; Rahim et al., 2000) (Fig. 1**C**). With the aim of discovering new RhlABC orthologs, we developed and applied a feature-guided genome mining workflow that led to the identification of 80 RL-producing species (including the reference strains), 71 of which were previously unreported. We then tentatively confirmed RL production in one of the identified strains. Moreover, this study led to the discovery of the putative RL transport mechanism.

Material and methods

Computational analysis

Genome mining

Genome mining for orthologs of RhlA, RhlB and RhlC (denoted RhlABC for simplicity) relied on the amino acid sequences of RhlABC of *P. aeruginosa* UCBPP-PA14 (Lee et al., 2006) and *Burkholderia thailandensis* E264 (Brett et al., 1998) as lead proteins (Table 1). Four genome databases were explored: the non-redundant protein (NCBI), genus-specific, Orthologous Matrix (OMA) and Annotree databases (Fig. 2). Searching in the non-redundant protein database (NCBI) was performed with the NCBI BLASTp tool and the amino acid sequences of lead proteins. BLASTp (NCBI) parameters were set at their default values, except that the maximum number of hits was increased from the default 100 to 20,000, and the following taxonomic orders were excluded: Pseudomonadales (TaxID:72274) and Burkholderiales (TaxID:80840). We filtered hits to keep only those with E-values $\leq 1 \times 10^{-5}$, percentage coverage $\geq$ 46% and percentage identity $\geq$ 23% (corresponding to an overall alignment score of $\geq$ 50). Afterward, hits referring to organisms not identified by both genus and species names were eliminated, because they referred primarily to unconfirmed or sometimes redundant genomes. Searching in the genus-specific genome databases of *Pseudomonas* (https://www.pseudomonas.com/) and *Burkholderia* (https://www.burkholderia.com/) genera was performed with the BLASTp tool with the same cutoff values described above.

Searching in the genus-specific genome databases of *Pseudomonas* (https://www.pseudomonas.com/) and *Burkholderia* (https://www.burkholderia.com/) genera was performed with the BLASTp tool set at the same cutoff values described above.

Searching of the OMA database (Altenhoff et al., 2020) was performed with the UniProt identifiers of lead RhlABC (Table 1).

Finally, searching the Annotree database (Mendler et al., 2019) was conducted using the KEGG orthology (KO) identifiers of lead RhlABC (Table 1), at default parameters, namely, percentage identity $\geq$ 30% (minimum set), E value $\leq 1 \times 10^{-5}$ and minimum percentage coverage $\geq$ 70%. Two rounds of search were conducted in Annotree database, one using both KO identifiers of RhlA and RhlB, and the second using the identifiers of RhlA, RhlB and RhlC

Curation of raw strain lists

Three curation steps were performed to generate one curated hit list of unique species harboring RhlAB or RhlABC orthologs. The first curation involved hits referring to several strains belonging to the same species, of which only the one with the best identity and coverage scores was retained. The second curation involved the elimination of strain hits whose genomes did not harbor genes encoding both RhlA and RhlB, whether clustered together or not. The strain lists including orthologs of any RhlABC of *P. aeruginosa* (denoted $RhlABC_{PA}$ for simplicity) and any RhlABC of *B. thailandensis* (denoted $RhlABC_{BT}$ for simplicity) were examined by drawing Venn diagrams with the Jvenn online tool (Bardou et al., 2014). Only strains harboring orthologs of both RhlA and RhlB (RhlAB) or all of RhlA, RhlB and RhlC (RhlABC) were maintained. The third curation involved assigning yes/no scores to strain hits according to the clustering of *rhlABC* genes in the respective genomes. A score of "yes" was assigned to hits whose genomes showed clustering of at least *rhlAB* (i.e., those located within several thousand base pairs of each other), and a score of "no" was assigned to those with no clustering of *rhlABC* genes (i.e., those located within more than 10,000 base pairs of each other). Relative genomic locations of genes were determined by comparison of the start and stop positions of open reading frames of *rhlABC*, as retrieved from the NCBI Genome Database of respective strains (**Table S1**). The annotated genomes of two species in our curated list (Table 3), *Burkholderia catarinensis*

Table 1
General and strain-specific identifiers of lead proteins associated with rhamnolipid biosynthesis and transport in the two reference strains used in this genome mining study.

Protein	KO[a]	*P. aeruginosa* UCBPP-PA14			*B. thailandensis* E264[c]		
		Locus Tag	UniProt	GI[b]	Locus Tag	UniProt	GI[b]
RhlA	K18100	PA14_19100	A0A0H2ZE72	ABJ12734.1	BTH_II1075, BTH_II1881	Q2T424	ABC34569.1, ABC34824.1
RhlB	K18101	PA14_19110	A0A0H2ZD76	ABJ12733.1	BTH_II1076, BTH_II1880	Q2T425	ABC34429.1, ABC35399.1
RhlC	K12990	PA14_49760	A0A0H2Z832	ABJ10298.1	BTH_II1079, BTH_II1877	Q2T428	ABC35897.1, ABC36027.1
RmlB	K01710	PA14_68170	A0A0H2ZIN4	ABJ14544.1	BTH_I1469	Q2SYI4	ABC37040.1
RmlD	K00067	PA14_68190	A0A0H2ZHP4	ABJ14545.1	BTH_I1472	Q2SYI1	ABC37745.1
RmlA	K00973	PA14_68200	A0A0H2ZIM7	ABJ14546.1	BTH_I1470	G3LYM7	ABC36359.1
RmlC	K01790	PA14_68210	A0A0H2ZIY2	ABJ14547.1	BTH_I1471	Q2SYI2	ABC38476.1
RhlD	K03446	PA14_68140	A0A0H2ZIY1	ABJ14543.1	BTH_II1077 BTH_II1879	Q2T426	ABC34473.1, ABC34833.1
RhlE	K03543	PA14_68130	A0A0H2ZIX9	ABJ14542.1	BTH_II1081, BTH_II1875	Q2T6C1	ABC35318.1
RhlF	–	PA14_68120	A0A0H2ZIM3	ABJ14541.1	BTH_II1080, BTH_II1876	Q2T429	ABC35408.1, ABC35493.1

[a] KEGG Orthology identifiers of the lead proteins RhlA/B/C, RmlBDAC and putative rhamnolipid transporters RhlDEF used in this study.
[b] GeneInfo identifiers
[c] The genome of *B. thailandensis* contains two identical copies of the each of the *rhlA, rhlB, rhlC* genes, all located on one of its two chromosomes.

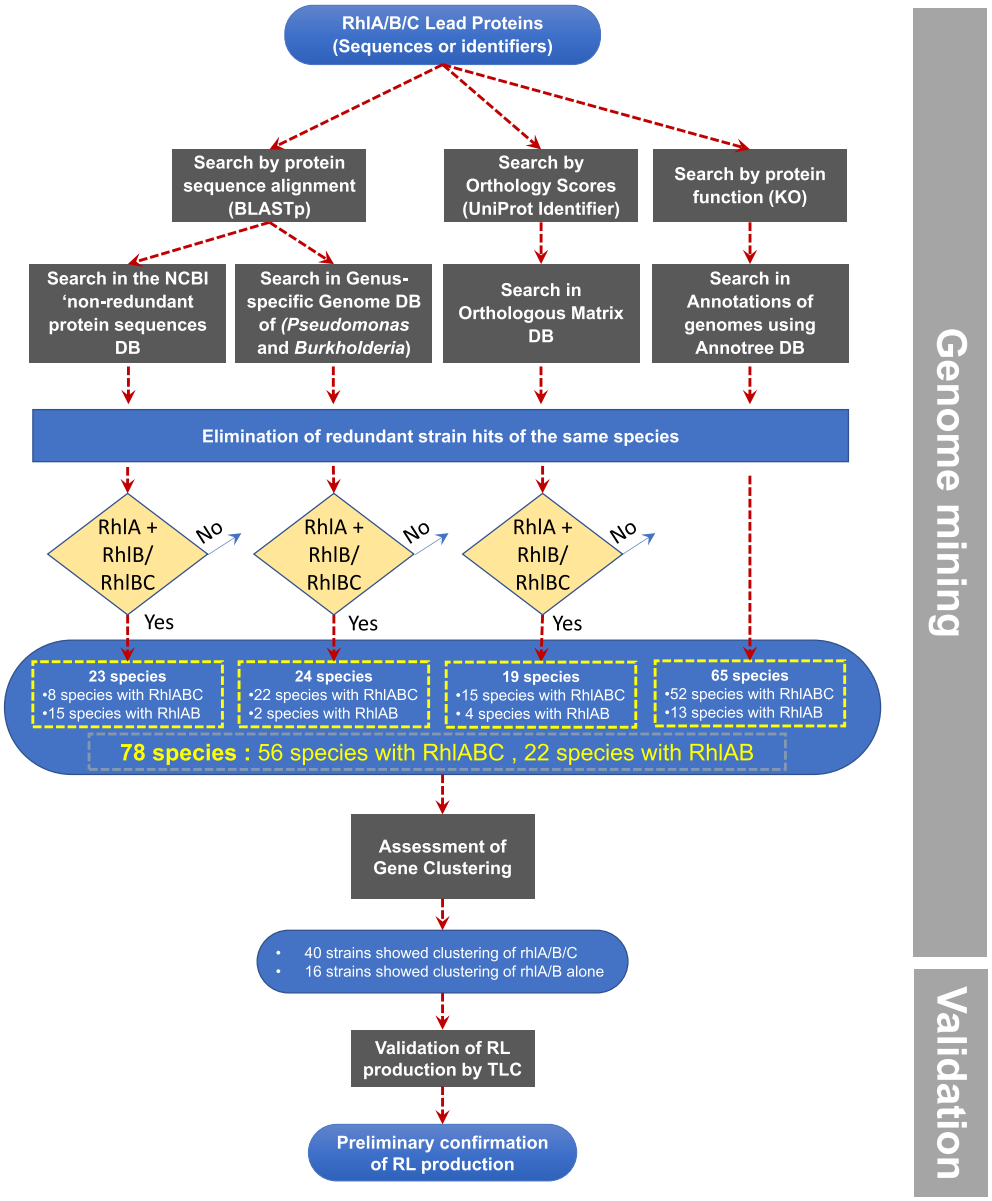


Fig. 2. Genome mining workflow used to identify putative RhlABC orthologs and potential RL/GL producers, followed by the experimental validation of RL production.

and *Burkholderia mesoacidophila*, were not available in the NCBI database. These two strains were therefore not included in this analysis. Finally, a consolidated list was generated including all curated strain hits harboring orthologs of both $RhlABC_{PA}$ and of $RhlABC_{BT}$.

Construction of distance trees of RhlABC orthologs

The amino acid sequences of the RhlABC orthologs of all hits were retrieved in FASTA format with the Batch Entrez tool of NCBI and subjected to multiple sequence alignment with the Clustal Omega function of Jalview software (Waterhouse et al., 2009). Distance trees were constructed in MEGA CC software (Kumar et al., 2018) with the following settings: the maximum likelihood (ML) method and Jones-Taylor-Thornton model, with the ML heuristic method set to nearest-neighbor-interchange, wherein the initial tree for ML was automatically constructed with neighbor-joining/BioNJ and 1000 bootstrap replications. Individual trees were constructed for each of the RhlA, RhlB and RhlC orthologs. In addition, a fourth tree was constructed using concatenated RhlAB sequences. The concatenation of aligned FASTA files for RhlA and RhlB was performed in Phylosuite software (Zhang et al., 2020). Trees were constructed on the computational platform of the Institute of Integrative and Systems Biology, Laval University, QC, Canada.

Construction of taxonomy tree for putative RL-producing strains

A phylum-level taxonomy tree was constructed with the Annotree web tool and the respective NCBI taxonomy IDs of the strains (Mendler et al., 2019). The NCBI taxonomy IDs (NCBI:TaxID) of strains were retrieved from the NCBI TaxIdentifier web tool.

Protein function predictions

The functions of putative RL/GL transporters (RhlDEF) were predicted with the following tools and approaches: the functional annotations in the respective annotated genome, the PHMMER online tool

"Biosequence analysis using profile hidden Markov Models" (Potter et al., 2018) and Interpro tools (Blum et al., 2021).

Experimental validation

Strains, media and culture conditions

Nevskia soli DSM 19509 (DSMZ) and *B. thailandensis* E264 (Brett et al., 1998) were cultivated under rotation (70 rpm) at 28 °C in test tubes containing 3 ml of defined medium whose composition per liter of 25 mM sodium phosphate buffer (pH 7.0) was as follows: glucose (20 g), KNO_3, (2 g), $MgSO_4{\cdot}7H_2O$ (0.4 g), $CaCl_2{\cdot}2H_2O$ (0.1 g), tryptone (1 g) and trace element solution (2 ml). *N. soli* additionally required casamino acids (5 g) and yeast extract (5 g). The trace element solution contained the following per liter of water: $FeSO_4{\cdot}7H_2O$ (2 g), $MnSO_4{\cdot}H_2O$ (1.5 g) and $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$ (0.6 g).

RL extraction and detection

The supernatant of 2-day-old liquid cultures of *N. soli* and *B. thailandensis* was collected after centrifugation at 14,000 rpm for 10 min, then extracted by vortexing with two volumes of chloroform/methanol (2:1, v/v). The chloroform layer was collected and evaporated to obtain a residue of RL extract. The dried residue of RL extract of *N. soli* was dissolved in 100 μL methanol. RL extract of *B. thailandensis*, and standard RLs (R90-50G, Millipore Sigma, originally from a Pseudomonas aeruginosa source) were used as reference long-chain and medium-chain RLs, respectively, and were prepared at 100 mg/L in methanol. A normal phase thin layer chromatography (TLC) plate (Silica Gel 60 F254, Millipore Sigma) was loaded with 5 μL of *N. soli* and reference RLs, developed with chloroform:methanol:water (65:25:3, v/v/v), dried and stained with orcinol reagent (per liter of water: 2 g orcinol, 114 ml concentrated H_2SO_4) with heating at 100 °C for 10 min. RL spots appeared violet to brownish.

Results

Genome mining for strains harboring RhlABC orthologs

To identify new RL/GL producers harboring RhlABC orthologs, we used the amino acid sequences of RhlABC of the well-studied RL/GL producers *P. aeruginosa* UCBPP-PA14 (Abdel-Mawgoud et al., 2014b) and *B. thailandensis* E264 (Dubeau et al., 2009) as lead proteins (Table 1). Our genome mining approach included screening for orthologs of RhlABC in four genome databases: the NCBI non-redundant protein, genus-specific, OMA and Annotree databases (Fig. 2). Our search for RhlABC orthologs in the non-redundant protein database (NCBI) aimed at identifying RL/GL producers other than the common ones belonging to the *Pseudomonas* and *Burkholderia* genera. Therefore, we adjusted the search parameters to exclude hits belonging to the orders *Pseudomonadales* and *Burkholderiales*, because RL/GL producers belonging to these orders were searched for in the respective genus-specific databases, as described later. Hits were filtered according to the cutoffs described in the Materials and methods. A significantly lower number of ortholog hits was obtained for RhlA than for each of RhlB and RhlC (**Table S2**) suggesting that the RhlA enzyme (an acyl transferase) is the most distinctive enzyme in the RL biosynthetic pathway, whereas many of the obtained RhlBC homologs are probably involved in other glycosyl transferase reactions.

We then searched for new RL/GL producers belonging to the *Pseudomonas* and *Burkholderia* genera with BLASTp tools in the respective genus-specific genome databases. These two databases contain more curated and well-annotated genomes of different species and strains of respective genera. The numbers of RhlA, RhlB and RhlC ortholog hits were approximately 33, 10 and 3, respectively for the *Pseudomonas* database, and 246, 268 and 260, respectively, for the *Burkholderia* database (**Table S2**). We also searched the OMA database, which contains an orthology matrix of more than 2400 genomes (Altenhoff et al., 2020), for RL/GL producers harboring RhlABC orthologs. The numbers of RhlA, RhlB and RhlC ortholog hits obtained from OMA database were approximately 40, 370 and 640, respectively (**Table S2**). The raw strain lists obtained from searching in the NCBI, gene-specific and OMA databases were curated as explained below.

Finally, we searched the Annotree database for new RL/GL producers harboring RhlABC orthologs. This database allowed for searching for RhlABC orthologs with universal protein function, KO, identifiers rather than protein-specific identifiers or homology-based searches, and provided strain hits along with their taxonomical information represented in a phylogenetic tree.

Our search in Annotree yielded a total of 65 species, 52 of which harbored RhlABC orthologs and belonged to six genera (*Pseudomonas*, *Burkholderia*, *Dickeya*, *Lonsdalea*, *Paraburkholderia* and *Caballeronia*), and 13 of which harbored only RhlAB and belonged to seven genera (*Pseudomonas*, *Nevskia*, *Pantoea*, *Lonsdalea*, *Robbsia*, *Dickeya* and *Serratia*) (Table 3, **Figure S2**).

Curation of strain lists harboring putative RhlABC orthologs

We next sought to curate the different raw strain lists obtained from genome mining of different databases to generate one consolidated curated hit list of unique species harboring RhlAB or RhlABC orthologs.

The first step of curation aimed at eliminating redundant species where hits referring to different strains were filtered so that to keep one strain of that species having the best identity and coverage scores. The second step of curation aimed at eliminating strain hits harboring solitary instances of RhlA or RhlB or RhlC, or harboring RhlBC alone. Theoretically, an RL/GL producer should harbor at least both RhlA and RhlB (RhlAB for simplicity) to be able to produce mono-RLs and should harbor RhlABC to produce di-RL (Fig. 1). With the help of Venn diagrams (**Figure S1**), we identified strains harboring RhlAB and those harboring RhlABC (Table 2). This curation step was not required for the strain hits obtained from Annotree database, because the approach we used for searching this database resulted in hits that all harbored either RhlAB or RhlABC.

For hits obtained from the NCBI non-redundant protein sequence database, 18 of the 23 strains were shared between the hit list obtained using $RhlABC_{PA}$ and that using $RhlABC_{BT}$ lead proteins. Only three species were unique in the group of orthologs to $RhlABC_{PA}$ (*Gordonia amicalis*, *Enterobacter cloacae* and *Granulicella arctica*), and 2 species were unique in the group of orthologs to $RhlABC_{BT}$ (*Serratia ficaria* and *Serratia plymuthica*) (**Figure S1.a, Figure S1.b**).

For hits obtained from the *Pseudomonas* genome database, two species harbored RhlAB orthologs, and two other species harbored RhlABC orthologs (**Figure S1.c**). Twenty species were obtained from the *Burkholderia* genome database, all of which harbored RhlABC (**Figure S1.d**).

Hits for species obtained from the OMA database were all already obtained from other databases except for one species, *Opitutus terrae*, which was unique to the OMA database (**Figure S1.e, Figure S1.f**). Interestingly, this species is an obligate anaerobe and thus is the first obligate anaerobic RL producer reported to date in the literature.

Overall, a total of 78 species (or 80 when including the reference species), 19 of which had RhlAB orthologs and 59 of which had RhlABC, were recovered from our search in all databases (Table 2 **and** Table 3).

The third step of curation was assigning yes/no scores to hits according to the clustering (co-localization)/lack of clustering of the coding genes of their RhlAB/RhlABC orthologs in the genomes of respective species. Typically, coding genes of RhlABC clustered as *rhlAB* without *rhlC* or as *rhlABC* in the genomes of the known RL/GL producers *P. aeruginosa* and *B. thailandensis*, respectively. In the latter species, the intergenic distances in the *rhlABC* clusters ranged from

Table 2
Summary of numbers of curated hits of strains harboring RhlAB or RhlABC orthologs as obtained from different database searches.

Database[a]	Total [b]		RhlAB[c]			RhlABC[c]		
	$\sum_{NR}$	$\sum_{Uq}$	PA	BT	$\sum_{NR}$	PA	BT	$\sum_{NR}$
NC	23	9	13	14	15	8	6	8
GS	24	3	2	0	2	2	20	22
OM	19	1	4	0	4	10	15	15
AN	65	30	–	–	21	–	–	47
$\sum_{Total}$	78	–	–	–	19	–	–	59
$\sum_{Clustering}$	56				16			40
$\sum_{Rhamnose}$	76							
$\sum_{RhlDEF}$	75							
RG1[d]	54							
RG2[d]	22							
RG3[d]	1							

[a] NC: NCBI non-redundant protein sequence database; GS: Genus-specific database; OM: OMA database; AN: Annotree database.
[b] Sums ($\sum$) are expressed as: $\sum_{NR}$ for total nun-redundant hits and $\sum_{Uq}$ for total hits that are unique to the specific database (i.e. not shared across other databases)
[c] PA and BT are number of strains carrying orthologs to RhlABC of *P. aeruginosa* and *B. thailandensis*, respectively.
[d] RG: Risk Group. RG1, 2 and 3, require containment level 1, 2 and 3, respectively.

2 bp to less than 300 bp. Strain hits whose genomes showed such clustering were suggested in this study to have a higher probability of being RL/GL producers than those whose genomes did not show such clustering. On the basis of the start and stop positions of open reading frames of *rhlABC* in putative RL/GL producers, 56 strains had a similar *rhlABC* genetic arrangement to that of *P. aeruginosa* and *B. thailandensis* (Table 2), whereas 20 strains had atypical genetic arrangements in which *rhlABC* genes were significantly spaced in the genome (Fig. 3).

The names of the 76 putative RL-producing strains, their biosafety levels, the protein identifiers of their RhlABC orthologs and their homology scores, as compared with those of *P. aeruginosa* UCBPP-PA14 and *B. thailandensis* E264, are tabulated in **Table S3**.

We verified the ability of putative RL/GL producers identified to date to furnish the dTDP-L-rhamnose precursors required for RL biosynthesis by searching for orthologs of RmlBDAC of *P. aeruginosa* ($RmlBDAC_{PA}$). Interestingly, all 76 strains possessed the four genes *rmlBDAC*. Moreover, these genes clustered together except in three strains: *Robbsia andropogonis* LMG 2129, *Gordonia amicalis* DSM 44,461 and *Granulicella arctica* X4EP2, in which one of the four enzymes was distantly located from the *rmlBDAC* cluster (Fig. 3). Overall, these findings indicate that all putative RL/GL producers should be able to provide the dTDP-L-rhamnose precursors required for RL biosynthesis.

Finally, a single consolidated curated strain list was created that included all 78 curated strains (including the reference strains) harboring orthologs of $RhlABC_{PA/BT}$ lead proteins (Table 3**, Supplementary Excel file**).

Phylogenetic trees of strains and RhlABC orthologs

The identified putative RL/GL producers belonged to 16 genera: *Pseudomonas* (n = 11), *Burkholderia* (n = 28), *Paraburkholderia* (n = 11), *Robbsia* (n = 1), *Caballeronia* (n = 2), *Streptococcus* (n = 1), *Stenotrophomonas* (n = 1), *Dickeya* (n = 9), *Gordonia* (n = 1), *Pantoea* (n = 3), *Granulicella* (n = 1), *Lonsdalea* (n = 3), *Serratia* (n = 3), *Nevskia* (n = 1), *Opitutus* (n = 1) and *Enterobacter* (n = 1) (Table 3). These genera belonged to four phyla: *Proteobacteria* (n = 74), *Acidobacteria* (n = 1), *Actinobacteria* (n = 1), *Firmicutes* (n = 1) and *Verrucomicrobia* (n = 1) (Table 3**, Figure S3**).

The pathogenicity of RL/GL producers is a matter of concern in the industrial production of RLs. Therefore, we examined the biosafety classifications of the potential RL-producing strains identified and assigned the most stringent biosafety classification found in the following databases: Canadian Epathogen (https://health.canada.ca/en/epathogen), DSMZ (https://www.dsmz.de/) and ATCC (https://www.atcc.org/). Approximately 70% of species were classified as RG1 (requiring level 1 containment), 28% of the strains were classified as RG2 (requiring level 2 containment), and only one strain was classified as RG3 (Table 3).

Next, we analyzed the phylogenetic distance between the amino acid sequences of the RhlABC orthologs of the 76 identified strains. Individual distance trees were constructed for each of the identified RhlA, RhlB and RhlC orthologs (**Figure S4**). To represent both RhlA and RhlB in a single distance tree, we constructed a distance tree based on the concatenated sequences of both proteins (Fig. 3).

We identified 11 RhlAB clades, which we denoted by the name of a member species. The identified clades and the number of their member species are as follows: *Pseudomonas aeruginosa* clade (nine species), *Pseudomonas mediterranea* clade (five species), *Nevskia soli* clade (two species), *Dickeya dadantii* clade (nine species), *Lonsdalea quercina* clade (three species), *Pantoea ananatis* clade (six species), *Gordonia amicalis* clade (three species), *Caballeronia mineralivorans* clade (six species), *Paraburkholderia aspalathi* clade (five species), *Burkholderia ambifaria* clade (19 species) and *B. thailandensis* clade (11 species) (Fig. 3). The RhlABC identifiers of the representative species of each clade are presented in **Table S4**.

Discovery of the putative RL transporters

We observed that the *rhlABC* clusters of nearly all putative RL/GL producers of the genera *Burkholderia*, *Paraburkholderia* and *Caballeronia* co-clustered with three genes predicted to encode transport proteins (Fig. 3), namely, DHA2 family efflux major facilitator superfamily (MFS) transporter permease inner membrane subunit (which we named RhlD); efflux RND transporter periplasmic adaptor subunit (which we named RhlE); and efflux transporter outer membrane subunit (which we named RhlF) (**Figures 3, S5**). In these genera, *rhlDEF* co-clustered with *rhlABC* in the following order: ***rhlABDCFE***. Because of the conserved clustering of *rhlDEF* with the *rhlABC* in all putative RL-producing strains of the *Burkholderia*, *Paraburkholderia* and *Caballeronia* genera, we examined their presence in the genomes of other putative RL/GL producers. To do so, we used the amino acid sequences of RhlDEF of *B. thailandensis* E264 as lead proteins (Table 1) in our search for orthologs in the genomes of putative RL/GL producers identified herein.

Unexpectedly, we identified RhlDEF orthologs in the genomes of most putative RL/GL producers identified in this study. The only exceptions were three strains: *Stenotrophomonas rhizophila*, in which

Table 3
Identified bacterial species that harbor RhlABC/RhlAB from different search methods.

Genus & Species[a]	Name of database[b]				RhlAB, RhlABC	BSL[c]	Ref.[d]
	NC	GS	OM	AN			
Phylum: Proteobacteria; Order: Pseudomonadales							
Pseudomonas aeruginosa	+	+	+	+	RhlABC	RG2	[1]
Pseudomonas batumici	–	–	–	+	RhlABC	RG2	–
Pseudomonas brassicacearum	–	–	–	+	RhlABC	RG1	–
Pseudomonas cedrina	–	–	–	+	RhlABC	RG1	–
Pseudomonas corrugata	–	–	–	+	RhlABC	RG1	–
Pseudomonas kilonensis	–	–	–	+	RhlABC	RG1	–
Pseudomonas mediterranea	–	–	–	+	RhlABC	RG1	–
Pseudomonas mucidolens	–	–	–	+	RhlABC	RG1	–
Pseudomonas denitrificans	–	+	–	–	RhlABC	RG1	–
Pseudomonas otitidis	–	+	–	–	RhlABC	RG2	–
Pseudomonas fluorescens [a]	–	+	–	+	RhlAB	RG1	–
Pseudomonas lactis	–	+	–	–	RhlAB	RG1	–
Phylum: Proteobacteria; Order: Burkholderiales							
Burkholderia thailandensis	+	+	+	+	RhlABC	RG1	[2]
Burkholderia ambifaria	–	+	+	+	RhlABC	RG1	–
Burkholderia anthina	–	+		+	RhlABC	RG1	–
Burkholderia cenocepacia	–	+	+	+	RhlABC	RG2	–
Burkholderia cepacia	–	+	–	+	RhlABC	RG2	–
Burkholderia contaminans	–	+	–	+	RhlABC	RG2	–
Burkholderia diffusa	–	+	–	+	RhlABC	RG2	–
Burkholderia gladioli	–	+	+	+	RhlABC	RG2	–
Burkholderia glumae	–	+	+	+	RhlABC	RG1	[3]
Burkholderia lata	–	+	+	+	RhlABC	RG1	–
Burkholderia latens	–	+	–	+	RhlABC	RG2	–
Burkholderia oklahomensis	–	+	–	+	RhlABC	RG2	–
Burkholderia plantarii	–	+	–	+	RhlABC	RG1	[4]
Burkholderia pseudomallei	–	+	+	+	RhlABC	RG3	[5]
Burkholderia pseudomultivorans	–	+	–	+	RhlABC	RG2	–
Burkholderia pyrrocinia	–	+	–	+	RhlABC	RG1	–
Burkholderia seminalis	–	+	–	+	RhlABC	RG2	[6]
Burkholderia stagnalis	–	+	–	+	RhlABC	RG2	–
Burkholderia territorii	–	+	–	+	RhlABC	RG2	–
Burkholderia ubonensis	–	+	–	+	RhlABC	RG1	–
Burkholderia catarinensis	–	–	–	+	RhlABC	RG1	–
Burkholderia humptydooensis	–	–	–	+	RhlABC	RG2	–
Burkholderia mallei	–	+	–	+	RhlABC	RG3	–
Burkholderia mesoacidophila	–	–	–	+	RhlABC	NA	–
Burkholderia metallica	–	–	–	+	RhlABC	RG2	–
Burkholderia paludis	–	–	–	+	RhlABC	RG2	–
Burkholderia puraquae	–	–	–	+	RhlABC	RG1	–
Burkholderia singularis	–	–	–	+	RhlABC	RG2	–
Burkholderia stabilis	–	–	–	+	RhlABC	RG2	–
Paraburkholderia acidophila	–	–	–	+	RhlABC	RG1	–
Paraburkholderia aromaticivorans	–	–	–	+	RhlABC	RG1	–
Paraburkholderia aspalathi	–	–	–	+	RhlABC	RG1	–
Paraburkholderia bryophila	–	–	–	+	RhlABC	RG1	–
Paraburkholderia fungorum	–	–	–	+	RhlABC	RG1	–
Paraburkholderia ginsengiterrae	–	–	–	+	RhlABC	RG1	–
Paraburkholderia megapolitana	–	–	–	+	RhlABC	RG1	–
Paraburkholderia phenazinium	–	–	–	+	RhlABC	RG1	–
Paraburkholderia sprentiae	–	–	–	+	RhlABC	RG1	–
Paraburkholderia bannensis	–	–	–	+	RhlABC	NA	–
Paraburkholderia tropica	–	–	–	+	RhlABC	RG1	–
Caballeronia udeis	–	–	–	+	RhlABC	RG1	–
Caballeronia mineralivorans	–	–	–	+	RhlABC	RG1	–
Robbsia andropogonis	–	–	–	+	RhlAB	RG1	–
Phylum: Proteobacteria; Order: Enterobacterales							
Lonsdalea iberica	+	–	–	+	RhlABC	RG1	–
Lonsdalea quercina	+	–	–	+	RhlAB	RG1	–
Lonsdalea britannica	+	–	–	+	RhlAB	RG1	–
Dickeya dadantii	+	–	+	+	RhlABC	RG1	–
Dickeya dianthicola	+	–	–	+	RhlABC	RG1	–
Dickeya chrysanthemi	+	–	+	+	RhlABC	RG1	–
Dickeya fangzhongdai	+	–	–	+	RhlAB	RG1	–
Dickeya paradisiaca	+	–	+	+	RhlAB	RG1	–
Dickeya solani	+	–	–	+	RhlAB	RG1	–
Dickeya zeae	+	–	+	+	RhlAB	RG1	–
Dickeya lacustris	+	–	–	–	RhlAB	RG1	–
Dickeya undicola	+	–	–	–	RhlAB	RG1	–
Pantoea allii	–	–	–	+	RhlAB	RG1	–
Pantoea ananatis	+	–	+	+	RhlAB	RG1	[7]

Table 3 (*continued*)

Genus & Species[a]	Name of database[b]				RhlAB, RhlABC	BSL[c]	Ref.[d]
	NC	GS	OM	AN			
Pantoea stewartii	+	–	–	+	RhlAB	RG1	[8]
Serratia rubidaea	+	–	–	+	RhlAB	RG2	[9]
Serratia ficaria	+	–	–	–	RhlAB	RG1	–
Serratia plymuthica	+	–	–	–	RhlAB	RG1	–
Enterobacter cloacae[a]	+	–	–	–	RhlABC	RG2	–
Phylum: Proteobacteria; Order: Xanthomonadales							
Stenotrophomonas rhizophila	+	–	–	–	RhlABC	RG1	–
Phylum: Proteobacteria; Order: Nevskiales							
Nevskia soli	+	–	–	+	RhlAB	RG1	–
Phylum: Acidobacteria; Order: Acidobacteriales							
Granulicella arctica	+	–	–	–	RhlAB	RG1	–
Phylum: Actinobacteria; Order: Corynebacteriales							
Gordonia amicalis	+	–	–	–	RhlAB	RG1	–
Phylum: Firmicutes; Order: Lactobacillales							
Streptococcus dysgalactiae subsp. *Equisimilis*[a]	+	–	–	–	RhlAB	RG2	–
Phylum: Verrucomicrobia; Order: Opitutales							
Opitutus terrae	–	–	+	–	RhlAB	RG1	–

References: **[1]** (Abdel-Mawgoud et al., 2014b); **[2]** (Dubeau et al., 2009); **[3]** (Costa et al., 2011); **[4]** (Hörmann et al., 2010); **[5]** (Häussler et al., 1998); **[6]** (Araújo et al., 2017); **[7]** (Smith et al., 2016); **[8]** (Rooney et al., 2009); **[9]** (Nalini and Parthasarathi, 2013)

[a] Some species harbor RhlABC orthologs in only one specific strain of that species. The names of these specific strains are provided in the Figure 3 and in the supplementary information. The total number of species in this list is 80, including the two reference strains shown in bold.

[b] NC: NCBI non-redundant protein sequence database; GS: genus-specific database; OM: OMA database; AN: Annotree database.

[c] BSL: biosafety level.

[d] Only references to studies in which the organisms were identified by sequencing, and in which RL/GL production was definitively confirmed with precise methods, are mentioned (see Discussion).

we found only RhlEF, and *Dickeya dadantii* and *Gordonia amicalis*, in which we found only RhlD (Fig. 3). Interestingly, *rhlDEF* clustered with the *rhlABC* or *rmlBDAC* genes in most of the putative RL producers identified in this study. The *rhlD*, *rhlE* and *rhlF* genes were clustered within the *rhlABC* cluster in the form of *rhlAB**D**C**FE*** in the species of *Burkholderia* (n = 27), *Paraburkholderia* (n = 11) and *Caballeronia* (n = 2), or in the form of *rhl**EF**AB**D**C* in *Nevskia soli* (Fig. 3). In addition, *rhlDEF* clustered upstream of *rmlBDAC* in *P. aeruginosa* (in the form of *rhl**F**DE-rmlBDAC*) and in *P. otitidis* (in the form of *rhlDE**F**-rmlBDAC*), or clustered downstream of *rmlBDAC* in *P. denitrificans*, *Streptococcus dysgalactiae* subsp. *Equisimilis* (both in the form of *rmlBDAC-rhlDFE*) and *Enterobacter cloacae* (in the form of *rmlBDAC-rhlDEF*). In other putative RL/GL producers (n = 16), the *rhlDEF* cluster was distant from the *rhlABC* or *rmlBDAC* clusters, namely in *L. quercina*, *L. britannica*, *L. iberica*, *Pantoea allii*, *Pantoea sewartii*, *Pantoea ananatis*, *P. cedrina*, *P. mucidolens*, *P. lactis*, *P. fluorescens*, *P. corrugata*, *P. batumici*, *P. kilonensis*, *P. mediterranea*, *P. brassicacearum* and *Robbsia andropogonis*. Finally, the *rhlDEF* genes were also found non-clustered (n = 11, in *Granulicella arctica*, *Serratia plymuthica*, *Opitutus terrae* and the eight *Dickeya* species) (Fig. 3).

Overall, our data demonstrated the presence of *rhlDEF* in the genomes of 75 of the 78 putative producers (96%), and these genes clustered (n = 46; 58%) with *rhlABC* (n = 41; 53%) or *rmlBDAC* (n = 5; 4%); or clustered alone (n = 16; 20%); or did not cluster at all (n = 11; 14%)

Tentative validation of RL/GL production

In the scope of this study, we used TLC to tentatively validate the presence of RLs in one putative RL-producing species identified herein. To do so, we selected one of the non-pathogenic RL producers identified in this study, *Nevskia soli* DSM 19509.

TLC results indicated that the total lipid extract of *N. soli* stained positively for glycolipids and showed di-RL spots comparable to di-RL spots of standard RLs in *P. aeruginosa* and *B. thailandensis* (Fig. 4). Moreover, the migration distance of the GL spots in *N. soli* was slightly shorter than that of the di-RL spots in *B. thailandensis* and slightly longer than that in *P. aeruginosa* (Fig. 4). Therefore, the GL of *N. soli* is most probably a di-GL whose lipid chain lengths are shorter than C14 that is present in di-RL of *B. thailandensis*, and longer than C10 that is present in the RL of *P. aeruginosa;* thus GL of *N. soli* is most probaby carrying fatty acids that are 12 carbon atoms long (C12). With respect to the sugar moiety of GL of *N. soli*, we emphasize that the orcinol reagent used to stain RL/GL relies on its reaction with a reducing sugar unit that could be rhamnose, glucose or another reducing sugar. Thus, the nature of the sugar could not be confirmed by orcinol staining. Our preliminary high-resolution MS analyses indicated that the putative GL of *N. soli* is not any of the known RLs (data not shown). Interestingly, the RhlABC orthologs of *N. soli* had several notable features. We found that *rhlA* clustered with two tandem *rhlB* orthologs, i.e., as an *rhlAB$_1$B$_2$* cluster. RhlB$_1$ and RhlB$_2$ presented high similarities to RhlB of both *P. aeruginosa* (51.08% identity and 97% coverage) and *B. thailandensis* (55.39% identity and 97% coverage) (**Table S3**). In contrast, RhlB$_2$ showed very low similarity to RhlC in *P. aeruginosa* (50% identity, 6% coverage) or *B. thailandensis* (42.11% identity, 23% coverage) (**Table S3**). RhlB$_1$ and RhlB$_2$ showed moderate similarity relative to each other (47.47% identity, 97% coverage). Because *N. soli* appeared to produce di-GLs, on the basis of our TLC results, we suggest that RhlB$_1$ and RhlaB$_2$ should add the first and second glycosyl units of the produced di-GL, respectively. The lack of similarity between RhlB$_2$ and RhlC might indicate that, in contrast to RhlC, which adds a second glycosyl on top of the first glycosyl unit via *O*-glycosidic linkage, RhlB$_2$ could be adding the second glycosyl via an ester linkage with the free carboxylic group of the distal 3-hydroxy fatty acid of GL. This structure has been reported in other microbial GLs, such as trehalolipids and mannosylerythritol lipids (Abdel-Mawgoud and Stephanopoulos, 2018).

Discussion

RLs are microbial GLs with diverse structure-dependent biological activities and physicochemical properties that can be adapted for

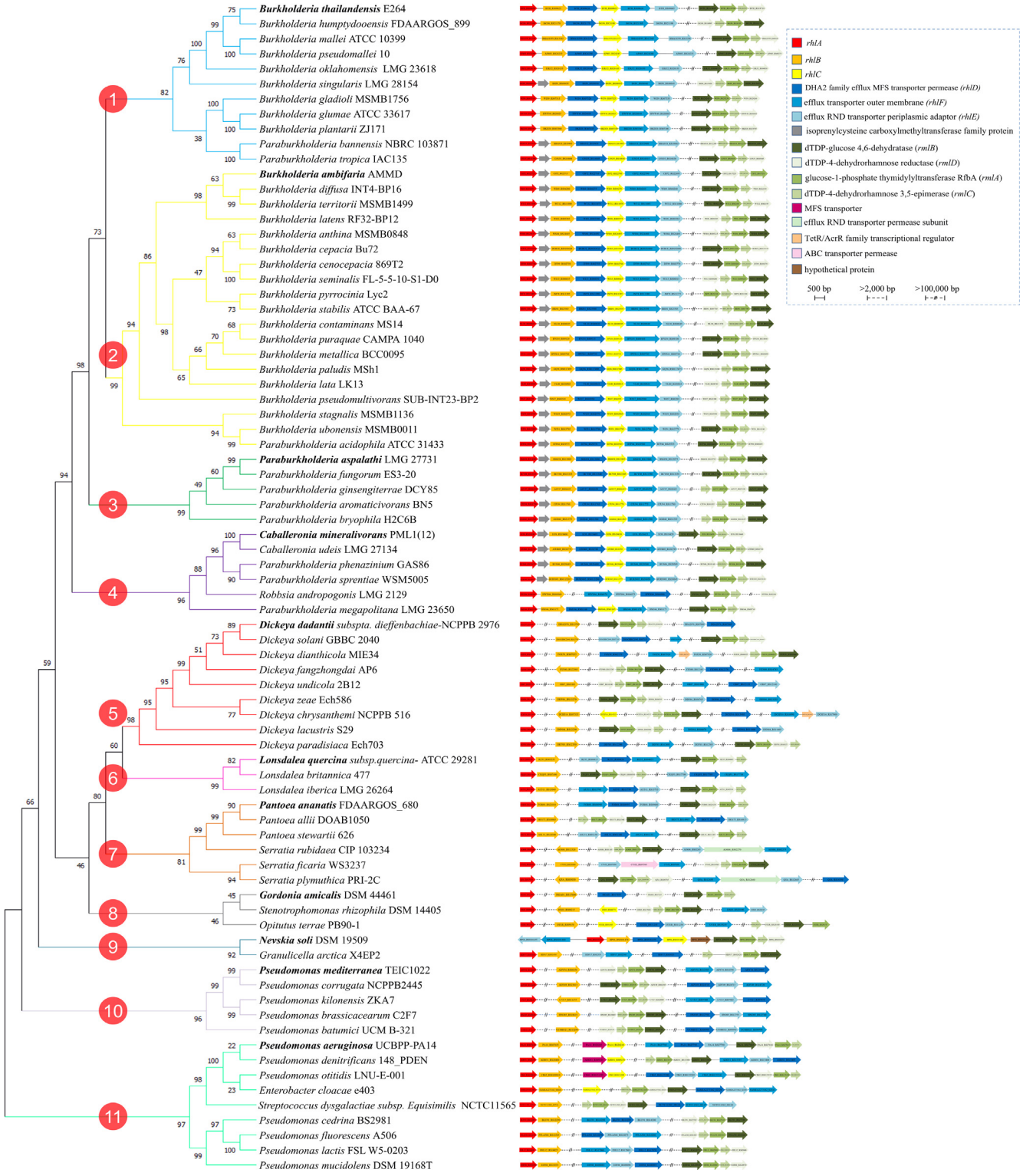


Fig. 3. Maximum likelihood-based distance tree based on RhlAB orthologs and the genetic arrangement of *rhl*ABC, *rmlBDAC* and *rhlDEF* gene clusters. Main color coding: red (*rhlA*), orange (*rhlB*), yellow (*rhlC*), blue scale (*rhlD/E/F* implicated in RL transport) and green scale (*rmlBDAC*, rhamnose biosynthesis). The tree was generated by using the concatenated sequences of RhlA and RhlB. Bootstrap values were calculated from 1000 replications. The 11 clades are numbered and colored differently. Each clade is denoted by the name of one member species, shown in bold. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

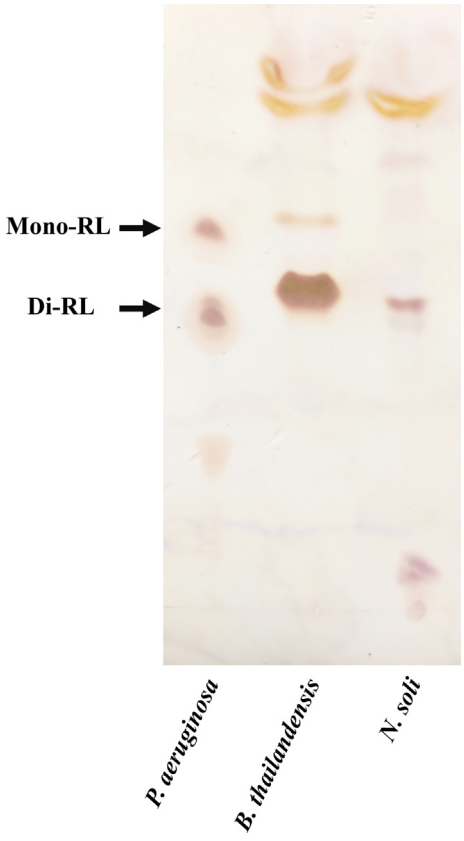


Fig. 4. Thin layer chromatography of the extracts of *Nevskia soli* DSM 19509, demonstrating putative production of glycolipids. The RLs of *P. aeruginosa* and *B. thailandensis* were used as standards. RLs/GLs appear as brown spots after spraying with orcinol reagent.

potential medical, pharmaceutical, industrial and environmental applications (Abalos et al., 2001; Abdel-Mawgoud and Stephanopoulos, 2018; Hošková et al., 2015; Kiran et al., 2016). The structure-dependent diversity of bioactivities and the physical properties of RLs have prompted the discovery of new RL/GL producers of novel RLs that might have more interesting activities. Classical screening of environmental microorganisms to identify new RL producers, in addition to being laborious and costly, has largely resulted in the redundant discovery of previously identified RLs/GLs (Abdel-Mawgoud et al., 2007; Haba et al., 2000).

We used a genome mining approach that led to the discovery of 71 previously unreported potential RL/GL producers out of 78 total identified strains (or 80 when including the two reference species).

In this study, four databases were mined for orthologs of the RL biosynthetic enzymes, RhlABC. These databases vary in the number of genomes contained, the degree of curation and the level of annotation. Although many strain hits were shared across the four databases, each database yielded unique hits (Table 2), thus indicating the importance of using multiple approaches in parallel to improve depth and coverage in genome mining. Notably, the Annotree database resulted in the highest number of total non-redundant hits (n = 65) and most importantly the highest number of unique hits (n = 30) not recovered by the other three databases (Table 2). The Annotree database contains protein-function annotations of more than 27,000 bacterial and 1500 archaeal genomes derived from the Genome Taxonomy DataBase, which is relatively standardized with respect to taxonomic nomenclature and phylogeny, and even includes a large number of novel prokaryotic genomes derived from metagenomic sources (Mendler et al., 2019).

Among the RhlABC biosynthetic enzymes of RLs, we found that RhlA, the acyl transferase synthesizing the lipid backbone of RL, is the most distinctive of the three, as deduced from the number of hits obtained for each of the three enzymes (**Table S2**), compared to the widespread presence of rhamnosyl transferases in prokaryotes and eukaryotes (Kiran et al., 2016). This finding implies that RhlA could be used alone as a reliable marker for initial screening of putative RL/GL producers in genome and metagenome mining studies, hits of which could subsequently be curated on the basis of the co-presence and clustering of RhlB/C with RhlA. A recent study aimed at discovering and testing known and new RhlA orthologs has identified interesting putative HAA-producing species (Germer et al., 2020). However, many of these putative HAA-producing species were not in our list of putative RL/GL producers. After investigating these species individually, we found that they indeed do not harbor any RhlB/RhlC orthologs.

The only species belonging to the genus *Pseudomonas* that has been definitively reported in literature to produce RLs is *P. aeruginosa* (Abdel-Mawgoud et al., 2014b), excluding the many reports that have claimed the production of RL by other *Pseudomonas* species without confirming RL production with reliable methods using advanced analytical techniques. In this study, we discovered, however, 11 additional species of *Pseudomonas* putatively producing RL (Table 3). Similarly, *Burkholderia thailandensis* (Dubeau et al., 2009), *B. glumae* (Costa et al., 2011), *B. plantarii* (Hörmann et al., 2010), *B. pseudomallei* (Häussler et al., 1998) and *B. seminalis* (Araújo et al., 2017) are the only *Burkholderia* species reported in the literature to produce RLs. Here, we report 38 additional putative RL-producing species of *Burkholderia* (n = 24), *Paraburkholderia* (n = 11), *Caballeronia* (n = 2) and *Robbsia* (n = 1), all belonging to the order Burkholderiales (Table 3). Only three organisms belonging to the order Enterbacterales reported in the literature as RL /GL producers are genetically supported according to our study: *Pantoea ananatis* (Smith et al., 2016), *Pantoea stewartii* (Rooney et al., 2009) and *Serratia rubidaea* (Nalini and Parthasarathi, 2013). Of note, a previously claimed RL producing Enterbacterales, *Enterobacter asburiae* (Rooney et al., 2009), showed not to be genetically supported, because this species does not harbor any orthologs to RhlAB (data not shown). Nonetheless, the same team has reported another RL-producing Enterbacterales, *Enterbacter hormaechei* (Rooney et al., 2009) that, despite harboring orthologs to RhlABC of *P. aerugninosa*, has a weak score for its RhlB, whose percentage coverage was only 22%—a value less than half the set coverage threshold in our study. This discrepancy explains why these two Enterbacterales were not maintained in our screening, although we do not exclude a potential RL-producing ability of *Enterbacter hormaechei*. Here, we report 16 additional putative RL-producing species belonging to the Enterbacterales, including species of *Lonsdalea*, *Dickeya*, *Pantoea*, *Serratia* and *Enterobacter*. Interestingly, six additional taxonomical orders, each with one putative RL producer, are first reported here. The most notable species belonging to these orders is *Opitutus terrae*, the first obligate anaerobic native RL/GL producer reported to date. We note that although the known RL producer, *Pseudomonas aeruginosa,* is an aerobic organism, it can also grow anaerobically in the presence of terminal electron acceptors, such as nitrate, or when *L*-arginine is a substrate for growth (Wu et al., 2005). A medium has been optimized for the anaerobic production of RL by *P. aeruginosa* using glycerol and nitrate as carbon and nitrogen sources, respectively (Zhao et al., 2021). Comparing RL production by *Opitutus terrae* and investigating whether this organism might help avoid the bioreactor-associated foaming problems usually encountered during cultivation of aerobic RL producers is interesting.

Unexpectedly, of all identified species, three harbored chromosomal *rhlABC* orthologs in a strain-specific manner (*rhlABC* was absent in all other strains of the same species): *Streptococcus dysgalactiae* subsp. *Equisimilis* NCTC11565, *Pseudomonas fluorescens* A506 and *Enterobacter cloacae* e403. Except for these specific strains, *S. dysgalac-*

tiae, *P. fluorescens* and *E. cloacae* strains all lack *rhlABC* genes. We suggest that these strain-specific non-conserved *rhlABC*$^{+}$ genotypes might have been acquired through horizontal gene transfer events. This possibility is corroborated by the reported origin of *P. fluorescens* A506 being a multi-antibiotic resistant spontaneous mutant (Lindow et al., 1996; Stockwell et al., 2010).

Interest in finding non-pathogenic RL/GL producers that could be used for industrial production of RLs is increasing. Currently, five non-pathogenic RL/GL-producing species have been reported in the literature: *Burkholderia glumae*, *B. thailandensis*, *B. plantari*, *Pantoea ananatis* and *Pantoea stewartii*. Of these species, the first two and last two produce long- chain RL (RRC14C14) (Costa et al., 2011; Dubeau et al., 2009; Hörmann et al., 2010) and cyclic glucolipids (GC10C10) (Gauthier et al., 2019; Rooney et al., 2009; Smith et al., 2016), respectively. Nonetheless, non-pathogenic strains producing medium-chain mono- and di-RLs (RC10C10 and RRC10C10), such as those produced by pathogenic *P. aeruginosa* (Abdel-Mawgoud et al., 2014b), have not been reported to date. More than 50 RL/GL producers with non-pathogenic potential identified in this study should provide a rich resource to be explored that might lead to the discovery of non-pathogenic producers of medium-chain-RLs. One of these is *Nevskia soli* DSM 19509, which attracted our attention for being phylogenetically distant from all other RL producers (Figure S2). Our analysis showed that it might produce medium-chain RL/GL (RRC12C12/GGC12C12) (Fig. 4), yet this strain scarcely grows to optical densities of 0.5 after 4 days of incubation and thus may not be well suited for industrial scale applications. Overall, this study provides a comprehensive list of putative RL/GL producers that have been genetically validated. Moreover, this study casts doubt on previous claims of RL producers that are not supported genetically (Kiran et al., 2016).

The clustered feature of RL biosynthetic genes, *rhlABC* or at least *rhlAB*, was observed in 56 of the 78 putative RL/GL producers identified in this study (Table 2, Fig. 3). Although we retained strains with non-clustered *rhlABC*, we think that strains with clustered *rhlAB* are more likely to be RL producers, and that non-clustered RhlB and RhlC could still be of rhamnosyl/glycosyl transferases yet implicated in other metabolic pathways.

Regulation of *rhlA/B/C* genes and RL production in the 22 strains that lacked *rhlABC* clustering is intriguing. Mostly, individual *rhlABC* genes are reported to be parts of operons, although many of these need to be revisted. The clustered *rhlAB* genes of *P. aeruginosa* have been reported to be part of one operon (Pearson et al., 1997), although this finding is not supported by bioinformatic analysis, because transcriptional terminators are predicted between *rhlA* and *rhlB* according to the *Pseudomonas* Genome Database (Winsor et al., 2016). On the other hand, *rhlC* is distantly located in the genome of *P. aeruginosa* and is part of an operon with an upstream gene of an unknown function (Winsor et al., 2016). In *Burkholderia*, *rhlA* is not predicted to be part of an operon at all, yet *rhlBC* are parts of one operon together with transporter genes, as predicted in the *Burkholderia* Genome Database (Winsor et al., 2008).

The consolidated tree of identified RhlAB orthologs revealed 11 clades, which we named according to one constituent species. The largest of these was the clade of *B. ambifaria* (Table 4, Fig. 3). In contrast to the RhlAB tree, the RhlC tree is hardly structured into two main clades, yet with low bootstrap scores (**Figure S4.c**).

We previously suggested that the diversity of simple glycolipids stems from the diversity of different acyl and glycosyl transferases (Abdel-Mawgoud and Stephanopoulos, 2018). Likewise, we suggest that different RhlAB clades may biosynthesize different RL or GL structures, originating from the different substrate specificities of different RhlA and RhlB orthologs. Although many of the newly identified clades do not yet have associated RL/GL structures, the different RL/GL structures attributed to different clades according to the literature corroborate the hypothesis that each RhlAB clade is associated with a distinct RL/GL structural class (Table 4). For example, *P. aeruginosa* clade 1 should produce medium-chain mono-($RC_{10\pm2}C_{10\pm2}$) and di-RLs ($RRC_{10\pm2}C_{10\pm2}$) (Abdel-Mawgoud et al., 2014b), whereas *B. thailandensis* clade 11 might produce long- chain di-RLs ($RRC_{14\pm2}C_{14\pm2}$) (Dubeau et al., 2009). The largest but least studied clade, *B. ambifaria* clade 10, should produce long-chain di-RL with one odd chain fatty acid (RRC15C14) (Araújo et al., 2017). Clade 6 of *Pantoae ananatis* produces a cyclic 15-membered macrodilactone mono-glucolipid (GC10C10) containing glucose instead of rhamnose as the sugar moiety, whose C1 and C6 hydroxyl groups are lactonized with the free 3-hydroxyl and the carboxyl groups of the C10C10 HAA unit, respectively (Gauthier et al., 2019). Further investigations are required to determine the RL/GL structures of the remaining clades to aid in determining the relationship between RhlABC sequence and the resultant RL/GL structure.

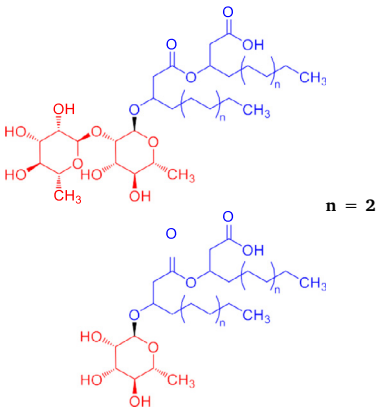
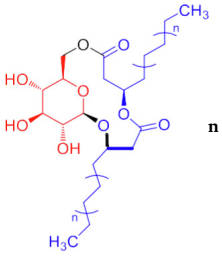
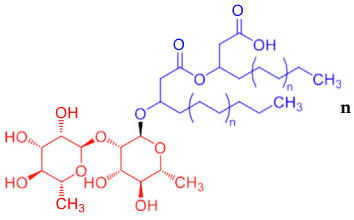
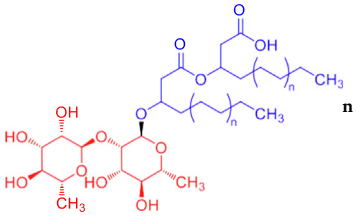
A recent study aiming at diversifying HAA, the precursor of RLs (Fig. 1), has screened RhlA orthologs and characterized the nature of the fatty acids of corresponding HAA, as determined by the heterologous expression of different RhlA orthologs in *E. coli* (Germer et al., 2020). Although the results of Germer et al. (2020) may help predict the chain length specificity of different RhlA orthologs and their ultimate RL structures, the findings should be interpreted cautiously, because these data were obtained through heterologous expression in a non-native host, *E. coli* (Germer et al., 2020). The fatty acid pool of this heterologous host is necessarily different from that of the native organisms, thus potentially affecting the nature of the produced HAA fatty acid chain lengths. Each RhlA ortholog has a certain tolerance to a range of fatty acid chain lengths rather than to one specific fatty acid chain length; e.g., RhlA of *P. aeruginosa* is specific to C8–C12 (with the C10 homolog being the most preferred substrate), whereas that of *B. thailandensis* is specific to C12–C16 (with the C14 homolog being the most preferred substrate) (Abdel-Mawgoud et al., 2010). Therefore, the expression of RhlA of *B. thailandensis* in *E. coli* could result in HAA that does not necessarily reflect the real specificity of that RhlA, because *E. coli* might be limited in long chain fatty acids that are instead available in *B. thailandensis*.

Moreover, Germer et al. (2020) have described seven species reported in the literature as RL producers, which were also identified in our study (Table 3), except for two strains that we excluded: *Pseudomonas chlororaphis* and *Dietzia maris*. *Pseudomonas chlororaphis* and *Dietzia maris* have been "experimentally" reported to produce mono-RL of RC10C12 type (Gunther et al., 2005) and RL of RRC10C10 (Wang et al., 2014), respectively.

We excluded *Pseudomonas chlororaphis* because, with the Diamond BLASTP tool of the *Pseudomonas* genome database (Winsor et al., 2016), we found that the 39 curated genomes of *P. chlororaphis* are all devoid of RhlB or RhlC orthologs, although they harbor RhlA orthologs. The same was found according to the genome database of NCBI, which includes 59 completely sequenced genomes of *P. chlororaphis*. The genome of the specific strain reported to be an RL producer by Gunther et al. (2005), *P. chlororaphis* strain NRRL B-30761, has not been sequenced. However, the partial sequence of *rhlB* of the NRRL B-30761 strain has been deposited under GenBank ID: JN415770.1. A BLASTN search on this DNA sequence indicated that it is nearly identical to that of *P. fluorescens* A506, with a very high alignment score of 2331, percentage coverage of 100%, E-value of 0 and percentage identity of 99.6% with zero gaps over the 1275 nucleotides. Therefore, we concluded that the reported RL/GL producer by Gunther et al. (2005) might actually be *P. fluorescens* A506 or a close strain to it. Consequently, we can conclude that the work of Gunther et al. (2005) was most probably based on a mal-identified strain that was *P. fluorescens* and not *P. chlororaphis*.

We also excluded *Dietzia maris* As-13-3, reported by Wang et al. (2014), because a BLASTP search of $RhlA_{PA\ or\ BT}$ orthologs in the genome database of NCBI against the genomes of all species of the genus *Dietzia* (taxid:37914) indicated that all *Dietzia* species are devoid of any RhlA orthologs, except for a unique hit corresponding to the

Table 4
Chemical structures of the most abundant RLs produced by each clade.

	RhlAB Clade name	RL/GL acronyms[a,b]	RL/GL structures	Ref.
1	*Pseudomonas aeruginosa*	RC10C10, RRC10C10	 **n = 2**	(Abdel-Mawgoud et al., 2014a)
2	*Pseudomonas mediterranea*	NA	NA	NA
3	*Nevskia soli*	NA	NA	NA
4	*Dickeya dadantii*	NA	NA	NA
5	*Lonsdalea quercina*	NA	NA	NA
6	*Pantoea ananatis* [c]	GC10C10	 **n = 2 (Ananatoside A)**	(Gauthier et al., 2019)
7	*Gordonia amicalis*	NA	NA	NA
8	*Caballeronia mineralovorans*	NA	NA	NA
9	*Paraburkholderia aspalathi*	NA	NA	NA
10	*Burkholderia ambifaria*	RRC15C14	 **n = 4**	(Araújo et al., 2017)
11	*Burkholderia thailandensis*	RRC14C14	 **n = 4**	(Costa et al., 2011; Häussler et al., 1998; Hörmann et al., 2010; Toribio et al., 2010)

[a] Acronym key: R and G indicate rhamnose and glucose units, respectively. Each C indicates an *R*-3-hydroxy acyl chain suffixed by a number indicating the chain length.

[b] The indicated acyl chain length represents the most abundant homolog; less abundant homologs are also naturally co-produced with slightly shorter or longer chains, mostly varying by ± 2 carbons.

[c] Other rhamnolipids, e.g., RC10C10, were reported for this clade, specifically in *Pantoae stewartii*; however, according to the mass spectra, the signal was very weak and appeared to be attributable to cross-sample carryover.

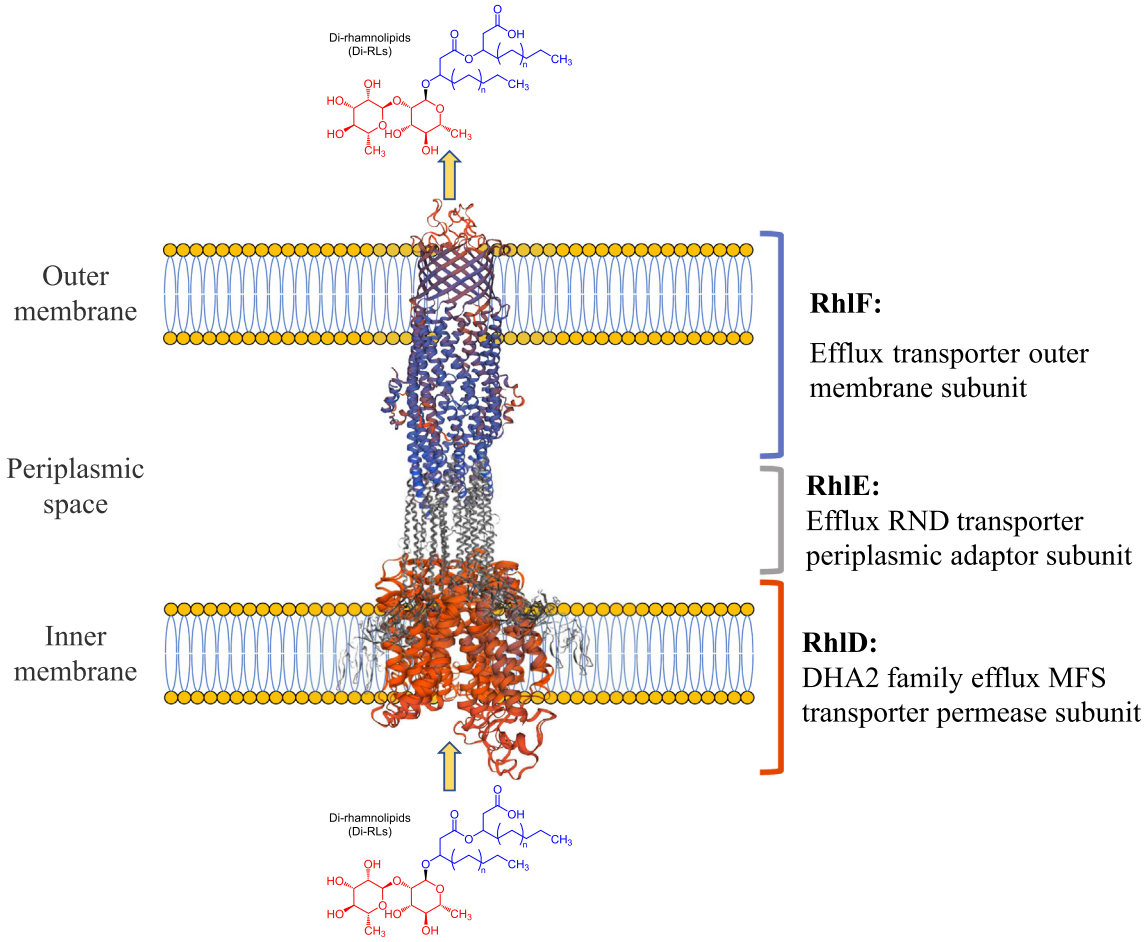


Fig. 5. Schematic representation of predicted RL transporter mechanisms. The predicted transport mechanism consisted of three components: DHA2 family efflux MFS transporter permease subunit (orange), efflux RND transporter periplasmic adaptor subunit (gray) and efflux transporter outer membrane subunit (blue). The modeled structures were constructed by using the homology modeling method with the sequences of *B. thailandensis* (the three models were built with the SWISS-MODEL webserver). These structures are only for demonstration and were not pre-processed or refined. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

HAA synthase "RhlA" (GenBank: AJD81708.1) of *Dietzia maris* As-13-3, submitted by Wang et al. (2014) themselves. Alignment of the DNA coding sequence of this submission against all genomes of the genome database of NCBI revealed that the sequence submitted by Wang et al. (2014) is 100% identical to that of *Burkholderia cenocepacia*. Therefore, *Dietzia maris* As-13-3 reported by Wang et al. (2014) is most probably *Burkholderia cenocepacia*.

Burkholderia kururiensis and *Paraburkholderia kururiensis* were reported to produce RLs of type RRC10C10 (Tavares et al., 2012), yet this claim is genetically unsupported, because none of the sequenced genomes (NCBI) of *B. kururiensis* or *Paraburkholderia kururiensis* harbor any orthologs of RhlAB of *P. aeruginosa* or *B. thailandensis*. Moreover, *Burkholderia/Paraburkholderia* species are not expected to produce medium chain RLs; they instead produce long-chain RLs of C14 and above.

In conclusion, only RL/GL producers that were validated taxonomically (well-identified), analytically (with reliable methods of RLs/GLs analysis) or genetically (harboring RL- /GL-encoding genes) are cited in Table 3. These criteria explain why many reported RL/GL producers, even those described in our previous review (Abdel-Mawgoud et al., 2010) and elsewhere in the literature, were not included in this study.

The mechanism of RL transport is poorly understood in the literature where two hypotheses have been proposed. The first hypothesis is that a gene upstream to *rhlC*, PA1131 in *P. aeruginosa*, predicted to encode an MFS protein, might be implicated in RL transport (Rahim et al., 2001). This possibility was refuted by a later study showing mono-RL production by a recombinant *P. putida* strain harboring heterologous *rhlABC* yet lacking any PA1131 homolog (Wittgens et al., 2017). The second hypothesis is that other putative transport genes upstream and downstream to *rhlC* in *B. thailandensis* E264 might be implicated in RL transport (Dubeau et al., 2009).

Indeed, except for those of three strains, the genomes of all RL/GL producers identified in this study harbor orthologs of putative RL transporters of *B. thailandensis* E264, which we named RhlD, RhlE and RhlF, for the inner membrane, periplasmic and outer membrane units, respectively (Fig. 5). In terms of genomic localization, *rhlDEF* primarily clustered with *rhlABC* or *rmlBDAC* clusters, or was otherwise dispersed in the genome. This finding incites future investigation of the exact roles of individual tranposrt proteins, RhlDEF, and their biotechnological potential. We believe that modulation of the RhlDEF putative RL transporters might aid in controlling RL/GL production and might have potential applications in medicine and industry.

Conclusion and perspectives

Our genome mining approach led to the discovery of a wealth of putative RL/GL producers. This approach could be iteratively used for the discovery of novel producers of rhamnolipids and glycolipids, as long as the number of annotated genomes is continually increasing. This process should ultimately help discover novel rhamnolipids/glycolipids with more potent or even novel activities, and could support

structure activity relationship studies on RLs. Moreover, the findings of this study could ultimately help predict rhamnolipid and glycolipid structures from the amino acid sequences of RL-biosynthetic enzymes by using artificial intelligence. The discovery of putative RL transporters is perhaps the most interesting finding of this study. This finding, with further molecular studies, should allow for better understanding of the transport of rhamnolipids, a phenomenon that has long remained obscure. Control of the discovered RL transport mechanism might have implications in the control of rhamnolipid production for medical and industrial applications.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was funded from the startup funds (Laval University, Québec ,Canada) provided to Prof. Ahmad M. Abdel-Mawgoud. We thank Prof. Eric Déziel (INRS-Armand-Frappier Santé Biotechnologie Research Centre, QC, Canada) for gifting us *B. thailandensis* E264. Meryam Magri was a visiting MSc student from the Medical Biotechnology Laboratory (MedBiotech), Faculty of Medicine and Pharmacy, Mohammed V University in Rabat, Morocco. We also thank Halim Maroufi for the valuable discussions at the begnning of this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crbiot.2022.02.002.

References

Abalos, A., Pinazo, A., Infante, M.R., Casals, M., Garcia, F., Manresa, A., 2001. Physicochemical and antimicrobial properties of new rhamnolipids produced by *Pseudomonas aeruginosa* AT10 from soybean oil refinery wastes. Langmuir 17 (5), 1367–1371.

Abdel-Mawgoud, A.M., Aboulwafa, M.M., Hassouna, N.-A.-H., 2007. Microbial production of surfactants: Screening and identification of two promising isolates and their biosurfactants. Egypt. J. Biotechnol. 27, 166–185.

Abdel-Mawgoud, A.M., Hausmann, R., Lépine, F., Müller, M.M., Déziel, E., 2011. Rhamnolipids: Detection, analysis, biosynthesis, genetic regulation & bioengineering of production. In: Soberón-Chávez, G. (Ed.), Biosurfactants: From Genes to Applications (Microbiology Monographs). first ed. Springer, Münster, Germany, pp. 13–55.

Abdel-Mawgoud, A.M., Lépine, F., Déziel, E., 2013. A chiral high-performance liquid chromatography–tandem mass spectrometry method for the stereospecific analysis of enoyl-coenzyme A hydratases/isomerases. J. Chromatogr. A 1306, 37–43.

Abdel-Mawgoud, A.M., Lépine, F., Déziel, E., 2014a. Liquid chromatography/mass spectrometry for the identification and quantification of rhamnolipids. In: Filloux, A., Ramos, J.-L. (Eds.), Pseudomonas Methods and Protocols. Springer New York, pp. 359–373.

Abdel-Mawgoud, A.M., Lépine, F., Déziel, E., 2010. Rhamnolipids: Diversity of structures, microbial origins and roles. Appl. Microbiol. Biotechnol. 86 (5), 1323–1336.

Abdel-Mawgoud, A.M., Lépine, F., Déziel, E., 2014b. A stereospecific pathway diverts β-oxidation intermediates to the biosynthesis of rhamnolipid biosurfactants. Chem. Biol. 21 (1), 156–164.

Abdel-Mawgoud, A.M., Stephanopoulos, G., 2018. Simple glycolipids of microbes: Chemistry, biological activity and metabolic engineering. Synth. Syst. Biotechnol. 3 (1), 3–19.

Altenhoff, A.M., Train, C.-M., Gilbert, K.J., Mediratta, I., Mendes de Farias, T., Moi, D., Nevers, Y., Radoykova, H.-S., Rossier, V., Warwick Vesztrocy, A., Glover, N.M., Dessimoz, C., 2020. OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. Nucleic Acids Res. 49 (D1), D373–D379.

Andrä, J., Rademann, J., Howe, J., Koch, M.H.J., Heine, H., Zahringer, U., Brandenburg, K., 2006. Endotoxin-like properties of a rhamnolipid exotoxin from *Burkholderia* (*Pseudomonas*) *plantarii*: immune cell stimulation and biophysical characterization. Biol. Chem. 387 (3), 301–310.

Araújo, F.D., Araújo, W.L., Eberlin, M.N., 2017. Potential of *Burkholderia seminalis* TC3.4.2R3 as biocontrol agent against *Fusarium oxysporum* evaluated by mass spectrometry imaging. J Am Soc Mass Spectrom 28 (5), 901–907.

Bauer, J., Brandenburg, K., Zähringer, U., Rademann, J., 2006. Chemical synthesis of a glycolipid library by a solid-phase strategy allows elucidation of the structural specificity of immunostimulation by rhamnolipids. Chem. – A Eur. J. 12 (27), 7116–7124.

Bardou, P., Mariette, J., Escudie, F., Djemiel, C., & Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. BMC Bioinformatics, 15, 293. doi:10.1186/1471-2105–15-293.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., Finn, R. D., 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49 (D1), D344–D354.

Brett, P.J., DeShazer, D., Woods, D.E., 1998. *Burkholderia thailandensis* sp. nov., a *Burkholderia pseudomallei*-like species. Int. J. Syst. Bacteriol. 48 Pt 1, 317–320.

Cloutier, M., Prevost, M.J., Lavoie, S., Feroldi, T., Piochon, M., Groleau, M.C., Legault, J., Villaume, S., Crouzet, J., Dorey, S., Diaz De Rienzo, M.A., Deziel, E., Gauthier, C., 2021. Total synthesis, isolation, surfactant properties, and biological evaluation of ananatosides and related macrodilactone-containing rhamnolipids. Chem. Sci. 12 (21), 7533–7546.

Compton, A.A., Deodhar, B.S., Fathi, A., Pemberton, J.E., 2020. Optimization of a chemical synthesis for single-chain rhamnolipids. ACS Sustainable Chem. Eng. 8 (24), 8918–8927.

Costa, S.G.V.A.O., Déziel, E., Lépine, F., 2011. Characterization of rhamnolipid production by *Burkholderia glumae*. Lett. Appl. Microbiol. 53 (6), 620–627.

Déziel, E., Lépine, F., Milot, S., Villemur, R., 2000. Mass spectrometry monitoring of rhamnolipids from a growing culture of *Pseudomonas aeruginosa* strain 57RP. Biochim. Biophys. Acta-Mol. Cell Biol. Lipids 1485 (2–3), 145–152.

Déziel, E., Lépine, F., Milot, S., Villemur, R., 2003. *rhlA* is required for the production of a novel biosurfactant promoting swarming motility in *Pseudomonas aeruginosa*: 3-(3-hydroxyalkanoyloxy)alkanoic acids (HAAs), the precursors of rhamnolipids. Microbiol.-SGM 149, 2005–2013.

Dubeau, D., Déziel, E., Woods, D., Lépine, F., 2009. *Burkholderia thailandensis* harbors two identical *rhl* gene clusters responsible for the biosynthesis of rhamnolipids. BMC Microbiol. 9 (1), 263.

Gauthier, C., Lavoie, S., Piochon, M., Martinez, S., Milot, S., Déziel, E., 2019. Structural determination of ananatoside A: An unprecedented 15-membered macrodilactone-containing glycolipid from *Pantoea ananatis*. Carbohydr. Res. 471, 13–18.

Germer, A., Tiso, T., Müller, C., Behrens, B., Vosse, C., Scholz, K., Froning, M., Hayen, H., Blank, L.M., Drake, H.L., 2020. Exploiting the natural diversity of RhlA acyltransferases for the synthesis of the rhamnolipid precursor 3-(3-Hydroxyalkanoyloxy)alkanoic acid. Appl. Environ. Microbiol. 86 (6). https://doi.org/10.1128/AEM.02317-19.

Gunther, N.W., Nunez, A., Fett, W., Solaiman, D.K.Y., 2005. Production of rhamnolipids by *Pseudomonas chlororaphis*, a nonpathogenic bacterium. Appl. Environ. Microbiol. 71 (5), 2288–2293.

Haba, E., Abalos, A., Jauregui, O., Espuny, M.J., Manresa, A., 2003. Use of liquid chromatography-mass spectroscopy for studying the composition and properties of rhamnolipids produced by different strains of *Pseudomonas aeruginosa*. J. Surfactants Deterg. 6 (2), 155–161.

Haba, E., Espuny, M.J., Busquets, M., Manresa, A., 2000. Screening and production of rhamnolipids by *Pseudomonas aeruginosa* 47T2 NCIB 40044 from waste frying oils. J. Appl. Microbiol. 88 (3), 379–387.

Häussler, S., Nimtz, M., Domke, T., Wray, V., Steinmetz, I., 1998. Purification and characterization of a cytotoxic exolipid of *Burkholderia pseudomallei*. Infect. Immun. 66 (4), 1588–1593.

Hörmann, B., Müller, M.M., Syldatk, C., Hausmann, R., 2010. Rhamnolipid production by *Burkholderia plantarii* DSM 9509^T. Eur. J. Lipid Sci. Technol. 112 (6), 674–680.

Hošková, M., Ježdík, R., Schreiberová, O., Chudoba, J., Šír, M., Čejková, A., Masák, J., Jirků, V., Řezanka, T., 2015. Structural and physiochemical characterization of rhamnolipids produced by *Acinetobacter calcoaceticus, Enterobacter asburiae* and *Pseudomonas aeruginosa* in single strain and mixed cultures. J. Biotechnol. 193, 45–51.

Kiran, G.S., Ninawe, A.S., Lipton, A.N., Pandian, V., Selvin, J., 2016. Rhamnolipid biosurfactants: evolutionary implications, applications and future prospects from untapped marine resource. Crit. Rev. Biotechnol. 36 (3), 399–415.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35 (6), 1547–1549.

Lee, D.G., Urbach, J.M., Wu, G., Liberati, N.T., Feinbaum, R.L., Miyata, S., Diggins, L.T., He, J.X., Saucier, M., Déziel, E., Friedman, L., Li, L., Grills, G., Montgomery, K., Kucherlapati, R., Rahme, L.G., Ausubel, F.M., 2006. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. Genome Biol. 7 (10).

Lindow, S.E., McGourty, G., Elkins, R., 1996. Interactions of antibiotics with *Pseudomonas fluorescens* strain A506 in the control of fire blight and frost injury to pear. Phytopathology 86, 841–848.

Luzuriaga-Loaiza, W.P., Schellenberger, R., De Gaetano, Y., Obounou Akong, F., Villaume, S., Crouzet, J., Haudrechy, A., Baillieul, F., Clément, C., Lins, L., Allais, F., Ongena, M., Bouquillon, S., Deleu, M., Dorey, S., 2018. Synthetic rhamnolipid bolaforms trigger an innate immune response in *Arabidopsis thaliana*. Sci. Rep. 8 (1), 8534.

Mendler, K., Chen, H., Parks, D.H., Lobb, B., Hug, L.A., Doxey, A.C., 2019. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Res. 47 (9), 4442–4448.

Nalini, S., Parthasarathi, R., 2013. Biosurfactant production by *Serratia rubidaea* SNAU02 isolated from hydrocarbon contaminated soil and its physico-chemical characterization. Bioresour. Technol. 147, 619–622.

Nott, K., Richard, G., Laurent, P., Jérôme, C., Blecker, C., Wathelet, J.-P., Paquot, M., Deleu, M., 2013. Enzymatic synthesis and surface properties of novel rhamnolipids. Process Biochem. 48 (1), 133–143.

Ochsner, U.A., Fiechter, A., Reiser, J., 1994. Isolation, characterization, and expression in *Escherichia coli* of the *Pseudomonas aeruginosa rhlAB* genes encoding a rhamnosyltransferase involved in rhamnolipid biosurfactant synthesis. J. Biol. Chem. 269 (31), 19787–19795.

Palos Pacheco, R., Eismin, R.J., Coss, C.S., Wang, H., Maier, R.M., Polt, R., Pemberton, J. E., 2017. Synthesis and characterization of four diastereomers of monorhamnolipids. J. Am. Chem. Soc. 139 (14), 5125–5132.

Pearson, J.P., Pesci, E.C., Iglewski, B.H., 1997. Roles of *Pseudomonas aeruginosa las* and *rhl* quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes. J. Bacteriol. 179 (18), 5756–5767.

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., Finn, R.D., 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46 (W1), W200–W204.

Rahim, R., Burrows, L.L., Monteiro, M.A., Perry, M.B., Lam, J.S., 2000. Involvement of the *rml* locus in core oligosaccharide and O polysaccharide assembly in *Pseudomonas aeruginosa*. Microbiol.-SGM 146, 2803–2814.

Rahim, R., Ochsner, U.A., Olvera, C., Graninger, M., Messner, P., Lam, J.S., Soberón-Chávez, G., 2001. Cloning and functional characterization of the *Pseudomonas aeruginosa rhlC* gene that encodes rhamnosyltransferase 2, an enzyme responsible for di-rhamnolipid biosynthesis. Mol. Microbiol. 40 (3), 708–718.

Remichkova, M., Galabova, D., Roeva, I., Karpenko, E., Shulga, A., Galabov, A.S., 2008. Anti-herpesvirus activities of *Pseudomonas* sp S-17 rhamnolipid and its complex with alginate. Zeitschrift Fur Naturforschung Section C-a J. Biosci. 63 (1–2), 75–81.

Rodrigues, L., Banat, I.M., Teixeira, J., Oliveira, R., 2006. Biosurfactants: potential applications in medicine. J. Antimicrob. Chemother. 57 (4), 609–618.

Rooney, A.P., Price, N.P., Ray, K.J., Kuo, T.M., 2009. Isolation and characterization of rhamnolipid-producing bacterial strains from a biodiesel facility. FEMS Microbiol. Lett. 295 (1), 82–87.

Siegmund, I., Wagner, F., 1991. New method for detecting rhamnolipids excreted by *Pseudomonas* species during growth on mineral agar. Biotechnol. Tech. 5 (4), 265–268.

Silva, V.L., Lovaglio, R.B., Von Zuben, C.J., Contiero, J., 2015. Rhamnolipids: solution against *Aedes aegypti*? Front. Microbiol. 6, 88.

Smith, D.D.N., Nickzad, A., Déziel, E., Stavrinides, J., McMahon, K., 2016. A novel glycolipid biosurfactant confers grazing resistance upon *Pantoea ananatis* BRT175 against the social amoeba *Dictyostelium discoideum*. mSphere 1 (1). https://doi.org/10.1128/mSphere.00075-15.

Stockwell, V.O., Johnson, K.B., Sugar, D., Loper, J.E., 2010. Control of fire blight by *Pseudomonas fluorescens* A506 and *Pantoea vagans* C9–1 applied as single strains and mixed inocula. Phytopathology 100 (12), 1330–1339.

Tavares, L.F.D., Silva, P.M., Junqueira, M., Mariano, D.C.O., Nogueira, F.C.S., Domont, G.B., Freire, D.M.G., Neves, B.C., 2012. Characterization of rhamnolipids produced by wild-type and engineered *Burkholderia kururiensis*. Appl. Microbiol. Biotechnol. 97 (5), 1909–1921.

Thakur, P., Saini, N.K., Thakur, V.K., Gupta, V.K., Saini, R.V., Saini, A.K., 2021. Rhamnolipid the glycolipid biosurfactant: Emerging trends and promising strategies in the field of biotechnology and biomedicine. Microb. Cell Fact. 20 (1), 1.

Tiso, T., Germer, A., Küpper, B., Wichmann, R., Blank, L., 2015. Methods for recombinant rhamnolipid production. Humana Press, 1–30.

Toribio, J., Escalante, A.E., Soberon-Chavez, G., 2010. Rhamnolipids: Production in bacteria other than Pseudomonas aeruginosa. Eur. J. Lipid Sci. Technol. 112 (10), 1082–1087.

Van Gennip, M., Christensen, L.D., Alhede, M., Phipps, R., Jensen, P.O., Christophersen, L., Pamp, S.J., Moser, C., Mikkelsen, P.J., Koh, A.Y., Tolker-Nielsen, T., Pier, G.B., Hoiby, N., Givskov, M., Bjarnsholt, T., 2009. Inactivation of the *rhlA* gene in *Pseudomonas aeruginosa* prevents rhamnolipid production, disabling the protection against polymorphonuclear leukocytes. Apmis 117 (7), 537–546.

Walter, V., Syldatk, C., Hausmann, R., 2010. Screening concepts for the isolation of biosurfactant producing microorganisms. Biosurfactants 672, 1–13.

Wang, W., Cai, B., Shao, Z., 2014. Oil degradation and biosurfactant production by the deep sea bacterium *Dietzia maris* As-13-3. Front. Microbiol. 5 (711).

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25 (9), 1189–1191.

Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A., Brinkman, F.S., 2016. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. Nucleic Acids Res. 44 (D1), D646–D653.

Winsor, G.L., Khaira, B., Van Rossum, T., Lo, R., Whiteside, M.D., Brinkman, F.S., 2008. The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. Bioinformatics 24 (23), 2803–2804.

Wittgens, A., Kovacic, F., Müller, M.M., Gerlitzki, M., Santiago-Schübel, B., Hofmann, D., Tiso, T., Blank, L.M., Henkel, M., Hausmann, R., Syldatk, C., Wilhelm, S., Rosenau, F., 2017. Novel insights into biosynthesis and uptake of rhamnolipids and their precursors. Appl. Microbiol. Biotechnol. 101 (7), 2865–2878.

Wu, M., Guina, T., Brittnacher, M., Nguyen, H., Eng, J., Miller, S.I., 2005. The *Pseudomonas aeruginosa* proteome during anaerobic growth. J. Bacteriol. 187 (23), 8185–8190.

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W.X., Wang, G.T., 2020. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol. Ecol. Resour. 20 (1), 348–355.

Zhao, F., Wu, Y., Wang, Q., Zheng, M., Cui, Q., 2021. Glycerol or crude glycerol as substrates make *Pseudomonas aeruginosa* achieve anaerobic production of rhamnolipids. Microb. Cell Fact. 20 (1), 185.