RESEARCH ARTICLE

The draft genome of Andean *Rhodopseudomonas* sp. strain AZUL predicts genome plasticity and adaptation to chemical homeostasis

Aisha E. Guardia[1†], Agustín Wagner[2†], Juan P. Busalmen[1], Cecilia Di Capua[3], Néstor Cortéz[3^] and María V. Beligni[4*]

Abstract

The genus *Rhodopseudomonas* comprises purple non-sulfur bacteria with extremely versatile metabolisms. Characterization of several strains revealed that each is a distinct ecotype highly adapted to its specific micro-habitat. Here we present the sequencing, genomic comparison and functional annotation of AZUL, a *Rhodopseudomonas* strain isolated from a high altitude Andean lagoon dominated by extreme conditions and fluctuating levels of chemicals. Average nucleotide identity (ANI) analysis of 39 strains of this genus showed that the genome of AZUL is 96.2% identical to that of strain AAP120, which suggests that they belong to the same species. ANI values also show clear separation at the species level with the rest of the strains, being more closely related to *R. palustris*. Pangenomic analyses revealed that the genus *Rhodopseudomonas* has an open pangenome and that its core genome represents roughly 5 to 12% of the total gene repertoire of the genus. Functional annotation showed that AZUL has genes that participate in conferring genome plasticity and that, in addition to sharing the basal metabolic complexity of the genus, it is also specialized in metal and multidrug resistance and in responding to nutrient limitation. Our results also indicate that AZUL might have evolved to use some of the mechanisms involved in resistance as redox reactions for bioenergetic purposes. Most of those features are shared with strain AAP120, and mainly involve the presence of additional orthologs responsible for the mentioned processes. Altogether, our results suggest that AZUL, one of the few bacteria from its habitat with a sequenced genome, is highly adapted to the extreme and changing conditions that constitute its niche.

Keywords: Chemical resistance, High-altitude Andean lakes, Pangenomic analysis, Purple non-sulfur bacteria, *Rhodopseudomonas*

†Aisha E. Guardia and Agustín Wagner share first authorship.

Néstor Cortéz deceased.

*Correspondence: mvbeligni@mdp.edu.ar

⁴ Instituto de Investigaciones Biológicas (IIB-CONICET-UNMdP), Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina
Full list of author information is available at the end of the article

Introduction

The genus *Rhodopseudomonas* is composed of gram-negative, purple non-sulfur photosynthetic bacteria from the Alphaproteobacteria class. Members of this genus are widespread in nature, evidenced by the fact that numerous strains have been isolated from very diverse environments [1, 2]. They have extremely versatile metabolisms: they can use carbon dioxide or organic compounds as carbon sources, and light, inorganic or organic

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 2 of 19

compounds as energy sources. They can grow with or without oxygen, fix nitrogen and degrade a great variety of organic compounds [3]. Within this genus, *Rhodopseudomonas palustris* is a model organism for the study of anoxygenic photosynthesis [4, 5]. The sequencing of the first *R. palustris* strain, CGA009, provided the basis for attributing a great deal of the metabolic versatility of this genus to specific genes and pathways [3]. As more strains were sequenced and analyzed phenotypically [1, 2, 6] the remarkable physiological differences between strains became apparent. Comparison of complete genome sequences showed that, although a core set of physiological processes is shared among strains, each behaves as a distinct ecotype that is highly adapted to sediment microenvironments within its natural habitat [2].

In this work, we report the sequencing, genomic comparison and functional annotation of a novel *Rhodopseudomonas* strain, named AZUL, which we previously characterized electrochemically [7]. This strain was isolated from Laguna Azul, an endorheic basin that is part of a number of water bodies collectively known as high-altitude Andean lakes (HAALs), located at the Central Andes region in South America [8]. These habitats experience exposure to severe environmental parameters, including high salinity, osmolarity, UV radiation, barometric pressure, pH and temperature and, as such, are paramount examples of extreme environments [9]. As a consequence, indigenous microbial communities have adapted to combinations of adverse chemical concentrations and physical stress in each particular niche. Several microorganisms isolated from HAALs have been taxonomically identified and characterized, which has shown that the organisms living in these habitats are an important reserve of molecular traits for resistance to environmental conditions such as arsenic and osmotic or UV stress [9–19]. Although metagenomic analyses have begun to tap into the diversity of species within HAALs [20, 21], very few genomes from organisms living in these habitats have been sequenced [8, 9, 22–25].

This work is the first report on the genome sequence of a *Rhodopseudomonas* from HAALs. Our analyses show that strain AZUL features the main characteristics of the genus, reported previously [2, 3] and is also furnished with remarkable specific traits, particularly related to membrane transport, resistance to toxic compounds and responses to nutrient limitation. Altogether, our findings suggest that this strain is highly specialized in maintaining cell homeostasis in a hostile habitat.

Materials and methods
Strain and culture conditions
Rhodopseudomonas sp. strain AZUL was isolated from a water sample obtained from the high altitude lagoon called Laguna Azul (27°34′ 17.3″ S 68°32′ 19.6″ W) kindly provided by María Eugenia Farías (PROIMI-CONICET). The water sample was filtered for microorganism enrichment, filters were incubated in Peptone Yeast (PY) medium (1% p/v bactopeptone, 0,05% (w/v) yeast extract, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 45 μM $FeSO_4$) at 28º C under anaerobic illuminated conditions. When brown–red coloration was detected (after 2–3 days), cultures were streaked onto PY-agar plates to select for colonies capable of growing in light, anaerobic conditions.

DNA isolation, sequencing and genome assembly
Rhodopseudomonas sp. AZUL cells were grown in PY medium and harvested by centrifugation at $5,000 \times g$. Genomic DNA was extracted using the Wizard Genomic DNA Purification kit (Promega). Library preparation, sequencing and data analysis was done at Instituto Nacional de Agrobiotecnología de Rosario (INDEAR) using an Illumina HiSeq 1500 system, according to the manufacturer's instructions. Briefly, two micrograms of purified DNA were resuspended in 50 μL of TE buffer. DNA fragmentation by nebulization, repair and end-adenylation, adapter ligation, gel purification and enrichment (amplification) were done according to the protocol described in TruSeq® DNA Sample Preparation Guide, Illumina. One microgram of the DNA library was run in a 2100 Bioanalyzer (Agilent Technologies) using the High Sensitivity DNA kit, quantified by qPCR (Light Cycler 480 Roche) using the Kapa Library Quantification kit and normalized to 2 nM. One equimolar pool of the libraries from the same sequencing lane was prepared by mixing 10 μl of each. The pool was used for the generation of clusters in a single lane of the sequencing cell. Paired-end (PE) sequencing was done (2×100 bp). Genome assembly starting from paired-end reads was done using the a5pipeline v20140113. Quality of the assembly (completeness, contamination and annotation self-consistency) was estimated using the genome-quality analysis service 'EvalG-EvalCon' within the Genome Annotation service at PATRIC [26]. Preliminary genome annotation was done using the RAST tool kit [27]. The *Rhodopseudomonas* sp. AZUL genome sequence is available at the RAST server (https://rast.nmpdr.org/), using the Login: guest and the Password: guest and also at NCBI GenBank, with the assembly accession GCA_024330085.1.

Average Nucleotide Identity (ANI) analysis
In addition to AZUL, the genome sequences of 38 *Rhodopseudomonas* strains were retrieved from NCBI genomes (https://www.ncbi.nlm.nih.gov/genome/), all available by January 2022. Genome sequences were compared using OAU, a command line tool for

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 3 of 19

calculating OrthoANI values using the USEARCH algorithm (https://help.ezbiocloud.net/oau-manual/) [28].

Pangenomic analyses

Genome quality was analyzed using PATRIC [26], similarly to what was previously done with AZUL. Only genomes that had values of completion and consistency higher than 90% and contamination lower than 10% were included in the pangenomic analyses (Additional File 2: Table S1). Clustering and pangenomic analyses were done using the softwares Roary and GET_HOMOLOGUES (GH). GH builds upon orthology-calling approaches based on heuristic pairwise best-match methods, as described [29]. Within the GH package, BLAST results were clustered with the bidirectional best-hit (BDBH) [30], the COGtriangles [31] and the OrthoMCL (OMCL) [32] clustering algorithms. The size of the pangenome was not determined using BDBH since it uses a single reference genome and, thus, cannot track genes not present in the reference. The exponential [33] and binomial mixture models [34] were fitted to the data obtained with OMCL within the GH package to estimate theoretical core and pangenome sizes. The main script get_homologues.pl was called under default settings, with a minimal sequence identity of 70%. Homologous clusters were computed reporting all clusters (-t) and excluding paralogs. Stringency was added to the analysis by scanning Pfam domain composition of the clusters using hmmscan from HMMER3 package (-D). Exponential decay or binomial mixture models were fitted to the core-genome cluster data by calling the auxiliary script plot_pancore_matrix.pl. For genome composition analysis, the -c flag was used under the exponential model to obtain tables of re-sampled core and pangenome sizes with the same settings as the previous GH analyses. The auxiliary script parse_pangenome_matrix.pl was used to analyze the structure of the pangenome, computing the strict core, relaxed core, shell, and cloud components.

In Roary [35], coding regions from.gff3 files produced by PROKKA [36] were used. The Roary pipeline extracted and converted them to protein sequences, filtered to remove partial sequences and iteratively pre-clustered with CD-HIT [37]. An all-against-all comparison was performed with BLASTP on the reduced sequences with a defined percentage sequence identity of 70%, with all the rest of the parameters set to default. The software then clustered the sequences with MCL and merged them with the pre-clustered results. For pangenomic analysis, the option -s was used in order not to split paralogs into different groups. Phylogenetic reconstruction using the core genome was done using Roary, as part of its default pipeline [35].

Functional annotation

The output of Roary annotation was used in conjunction with the results of pangenomic analysis to estimate the number of paralogs in protein families, using both default and -s options (splitting and not splitting paralogs). Genes resulting unique or rare in the Roary output were subjected to HMMER clustering analyses for confirmation (using nhmmer) and compared with the remaining strains to determine whether they were absent in other strains or possessed homologs that clustered separately. In addition, the total number of genes from different protein families, classes and subsystems were determined for the 31 strains with the combined outputs of Roary and PATRIC. In the PATRIC website, we used the Metabolomics, Comparative Pathway tool (https://www.patricbrc.org/app/ComparativePathway). Operon visualization was done using Geneious Prime® version 2020.0.5 (https://www.geneious.com/). Further operon editing for Fig. 5 was done using Inkscape 0.92 (https://www.inkscape.org).

Inductively coupled plasma mass spectrometry (ICP-MS)

The water sample from Laguna Azul was analyzed by Inductively Coupled Plasma Mass Spectrometry (ICP-MS), using the TotalQuant method, for the semi-quantitative estimation of element content. The analysis was done in a Perkin Elmer NexION 350X equipment (CEFOBI- CONICET, Rosario, Argentina) following the manufacturer's instructions and a single calibration standard containing elements distributed across the mass range, used to create a response table. The complete mass spectrum was determined and interpreted using the Syngistix software. After a total spectrum evaluation for each element, the resulting final isotope intensity counts were summed for each element and were then compared with the stored response table. Each determination was done in duplicates and averaged. With this method, accuracies are within $\pm 50\%$.

Results and discussion

General features of the *Rhodopseudomonas* sp. AZUL genome

In view of the great adaptability of the *Rhodopseudomonas* genus and the characteristics of the natural habitat of AZUL, we sequenced the genome of this strain. Details on the quality of sequencing and assembly are shown in Additional File 1. Briefly, the genomic library of AZUL was composed of double-stranded DNA fragments of sizes ranging 300–500 bp. Cluster density was 600 K/mm², better than the average value expected for the technology used. The data had good quality parameters, with a 90.8% of all the reads with a Qscore > Q30

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 4 of 19

(error rate < 0.001, probability of incorrect base call < 1 in 1,000 bases, base call accuracy > 99.9%). Final total reads were 22,300.084, with a 99.14% of the reads passing internal quality filtering procedures (PF reads). The final assembly yielded 120 contigs for a total genome length of 6,050,441 bp, an N50 value of 272,881 and an L50 value of 6. The length of the shortest contig was 304 bp, while the longest was 1,180,838 bp. Quality of the assembly predicted that the genome of AZUL had a 100% completeness and a 2% contamination (Additional File 1).

Table 1 summarizes features of the genome of AZUL in comparison with those of other 38 *Rhodopseudomonas* strains with available genome sequences, as well as information on strain isolation and selected references. The genome of AZUL is larger than 6 Mb, and only *Rhodopseudomonas* sp. AAP120 and *Rhodopseudomonas* sp. BAL398 have genomes of similar size. As a consequence, the number of coding regions in these strains is also larger (Table 1). The GC content is, as for the rest of the strains, at the top end of the spectrum when compared to the genomes of other bacterial groups [38].

In order to establish the similarity of AZUL with other strains, we compared the ANI values of the 39 *Rhodopseudomonas* strains in Table 1. Figure 1 shows that the genome sequence of AZUL is most similar to that of *Rhodopseudomonas* sp. strain AAP120 and, by the high ANI value (96.2%), it could be presumed that both strains belong to the same species. The cluster formed by AZUL and AAP1120 lies within a larger group that contains the *Rhodopseudomonas* sp. strains ATH 2.1.18 (ANI 89.9%), HaA2 (ANI 87%), NC1818 (ANI 87.9%), B29 (ANI 84.9%), BisB5 (ANI 86.3%) and two *Rhodopseudomonas palustris* strains, *R. palustris* strain 4810 (ANI 86%) and the type strain *R. palustris* DSM 123 [T] (ANI 86%). As previously determined for AAP120 [60], these strains could be recognized as separate species compared to AZUL and AAP120 (ANI values lower than 90%).

Figure 1 also shows that the genome sequence of AZUL is quite different from those of the *R. pentothenatexigens* type strain JA575[T] and several *R. rutila* strains, including the type strain R1[T]. The ANI values between AZUL and these strains are approximately 85%, which indicates a clear separation at the species level. The ANI values between AZUL and several *R. faecalis* strains, including the type strain JCM 11668 [T], are even lower and range 80%.

Rhodopseudomonas pangenomic analysis

With the purpose of determining which genes are common to the genus and which are specific to AZUL or to subgroups that include this strain, we did pangenomic analyses using the *Rhodopseudomonas* genome sequences of the strains in Table 1 that passed quality filtering by PATRIC (31 strains, Table S1). We compared the results of the softwares Roary and GET HOMOLOGUES (GH), and adjusted the data to the model developed by Tettelin et al. [33] and to the mixed binomial model [34]. The core genome (genes shared by all strains) obtained with the GH package was calculated using the clustering algorithms BDBH, OMCL and COGt, either independently or as a consensus between the three [29]. This produced core genomes of 1,055, 1,073, 1,041 and 987 gene clusters, respectively (Table 2). Roary [35] produced a core genome of 1,217 gene clusters (Table 2). The core genome predicted by Roary was used to build a phylogenetic tree of the relationships of the strains used in the pangenomic analysis. Figure 2 shows that, in agreement with the ANI analysis, AZUL is most closely related to *Rhodopseudomonas* sp. strain AAP120, and these two strains lie within larger clusters comprised of the *Rhodopseudomonas* sp. strains 2.1.18, NC 1818, B29 and HaA2 and the *Rhodopseudomonas palustris* strains DSM 123 [T] and BisB5.

Figure 3 shows the adjustment of the core genome produced by OMCL-GH to the model used by Tettelin et al. [33], which follows an equation of exponential decay. As expected, the number of genes shared by all strains decreased as more genomes were added to the analysis and reached a value of 1,397 gene clusters with the addition of the 31st strain (Fig. 3A and Table 2).

The size of the pangenome (the whole gene collection of the genus) was also estimated. Table 2 shows that Roary, COGt-GH and OMCL-GH predicted pangenomes of 26,385, 20,757 and 10,507 gene clusters, respectively. Figure 3B shows that adjustment of the OMCL-GH data to the model in Tettelin et al. [33] predicts a pangenome of 27,595 gene clusters when the 31st strain is added, similar to the results obtained with Roary.

The models described above have been criticized on the basis of the restrictions imposed by the classification of genes into only two categories (core genome and novel or specific genes). To overcome these restrictions, mixed models have been developed, such as one developed by Snipen et al. [34]. Table 2 shows that the core genome of the genus becomes much larger when the number of genes present in at least 30 strains are considered (*n-1*, soft core). In addition, it is noteworthy that a great proportion of genes comprised within the *Rhodopseudomonas* pangenome are either strain-specific or rare, the so called cloud genome.

The results presented above show that the precise characteristics of each algorithm markedly affects the estimation of the pangenome size. Despite these differences, our results evidence that the core genome comprises only a small proportion of the gene repertoire of *Rhodopseudomonas*, roughly between 5 and 12%. This

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 5 of 19

Table 1 General features of the Rhodopseudomonas genomes analyzed in this study

Proposed name[a]	Genome assembly ID	Size	GC	CDS	Plasmid	Strain isolated from/Isolation information	Reference[b]
Rhodopseu-domonas faecalis JCM 11668T	GCA_003217325.1	4.07	64.20	3,681	—	Anaerobic digester treating chicken manure, China	[39]
Rhodopseu-domonas faecalis JSC-3b	GCA_000504425.1	4.07	64.20	NA[c]	—	Freshwater canal adjacent to a veg-etable field, China	[40]
Rhodopseu-domonas faecalis PSBS	GCA_002895035.1	3.95	64.10	NA[c]	—	Swine sewage wastewater, China	[41]
Rhodopseu-domonas palustris 4810	GCA_014145305.1	5.35	63.40	4,961	—	Contamination in cultures of Thio-spirillum jenense DSM 216[T]	[42]
Rhodopseu-domonas palustris DSM 123T	GCA_900110435.1	5.27	64.60	4,747	—	Surface water or mud	[43]
Rhodopseu-domonas rhenoba-censis DSM 12706	GCA_014203125.1	5.18	65.60	4,612	—	Sediment of eutrophic pond, Rheinbach, Ger-many	[44]
Rhodopseu-domonas pentoth-enatexigens JA575T	GCA_900218015.1	5.38	66.00	4,822	—	Paddy soils, India	[45]
Rhodopseu-domonas rutila CGA009	GCA_000195775.1	5.47	64.99	4,913	1	--	[3]
Rhodopseu-domonas rutila DSM 126	GCA_002937155.1	5.39	65.00	4,895	—	Freshwater pond, Germany: near Zeulenroda	[46]
Rhodopseu-domonas rutila ELI 1980	GCA_002026345.1	5.65	65.09	5,090	1	Freshwater pond, Suffolk County, NY, USA.	[47]
Rhodopseu-domonas rutila PS3	GCA_003031265.1	5.27	65.30	4,780	—	Paddy field, Taipei City, Taiwan	[48]
Rhodopseu-domonas rutila R1T	GCA_003547145.1	5.31	64.90	4,803	—	Rice fields, Japan	[49]
Rhodopseu-domonas rutila TIE-1	GCA_000020445.1	5.74	64.90	5,227	—	iron-rich mat, Woods Hole, Mas-sachusetts, USA	[50]
Rhodopseu-domonas rutila YSC3	GCA_003031245.1	5.37	65.20	4,871	—	Paddy field, Yilan County, Taiwan	[48]
Rhodopseu-domonas sp. CGMCC 1.2180	GCA_013415845.1	5.32	65.00	4,792	—	China	
Rhodopseu-domonas sp. 42OL	GCA_001020905.1	5.13	65.70	4,724	—	Sugar refinery waste treatment pond, Castiglion Fiorentino, Italy	[51]
Rhodopseu-domonas sp. AAP120	GCA_001295845.1	6.16	65.70	5,516	—	Lagoon, Inner Mongolia, China.	

Guardia *et al. BMC Microbiology* (2022) 22:297

Page 6 of 19

Table 1 (continued)

Proposed name[a]	Genome assembly ID	Size	GC	CDS	Plasmid	Strain isolated from/Isolation information	Reference[b]
Rhodopseu-domonas sp. ATH 2.1.18	GCA_003591005.1	5.63	65.40	5,084	—	Isolated and purified by C.B. van Niel in 1944 and later transferred to T.E. Meyer at the University of Arizona	
Rhodopseu-domonas sp. ATH 2.1.37	GCA_003591275.1	5.49	65.00	4,932	—	Isolated and purified by C.B. van Niel in 1944 and later transferred to T.E. Meyer at the University of Arizona.	[52]
Rhodopseu-domonas sp. B29	GCA_000333455.1	5.52	65.00	4,940	—	Rice fields, Japan	
Rhodopseu-domonas sp. BAL398	GCA_000935205.1	6.12	64.20	5,844	—	Sea surface water, Baltic Sea	[53]
Rhodopseu-domonas sp. BR0C11	GCA_010820945.1	5.29	65.10	4,989	—	Bromeliad Phytotelma, Carite Forest, Puerto Rico	
Rhodopseu-domonas sp. BR0G17	GCA_010907025.1	5.52	65.10	5,017	—	Bromeliad Phytotelma, Guajataca Forest, Puerto Rico	
Rhodopseu-domonas sp. BR0M22	GCA_010907035.1	5.29	65.20	4,761	—	Bromeliad Phytotelma, Maricao Forest, Puerto Rico	
Rhodopseu-domonas sp. Cfx3-05	GCA_013377015.1	4.83	64.40	4,511	—	Metagenomic assembly from the Ca. Chlorohelix allophototropha enrichment, Canada: near Kenora	
Rhodopseu-domonas sp. GJ-22	GCA_007005445.1	5.04	65.80	4,550	—	Wastewater from a pesticide factory, Changsha (Hunan, China)	[54]
Rhodopseu-domonas sp. HaA2	GCA_000013365.1	5.33	66.00	4,731	—	Uncontaminated freshwater marsh sediment location A, Haren, The Netherlands	
Rhodopseu-domonas sp. NC 1818	GCA_016215605.1	5.79	66.20	5,274	—	Metagenomic assembly obtained from the groundwater metagenome BioSample: SAMN15459608	[55]
Rhodopseu-domonas sp. RCB100	GCA_016584445.1	5.46	64.99	4,907	1	Creek soil, Cascadilla Creek, Ithaca, NY, USA	[56]
Rhodopseu-domonas sp. RI 341	GCA_016124795.1	5.14	63.00	4,784	—	Hot springs and caldera lake in Raoul Island, New Zealand.	
Rhodopseu-domonas sp. SK50-23	GCA_018279705.1	5.86	61.90	5,281	—	Nonpolluted garden soil, Japan	[57]

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 7 of 19

Table 1 (continued)

Proposed name[a]	Genome assembly ID	Size	GC	CDS	Plasmid	Strain isolated from/Isolation information	Reference[b]
Rhodopseudomonas sp. WA056	*GCA_010906995.1*	*5.06*	*65.80*	*4,562*	—	Water reservoir, Puerto Rico	
Rhodopseudomonas sp. XCP	*GCA_003226555.1*	*5.59*	*65.20*	*5,096*	—	Contaminant of a green bacterial culture, La Jolla, CA, USA	[58]
Rhodopseudomonas thermotolerans JA576	*GCA_003387125.1*	*5.38*	*66.00*	*4,825*	—	Paddy soils, India	[45]
Rhodopseudomonas palustris BisB5	*GCA_000013685.1*	*4.89*	*64.40*	*4,397*	—	Uncontaminated freshwater marsh sediment location B, Haren, The Netherlands	[1]
Rhodopseudomonas sp. BisB18	*PRJNA15750*[c]	*5.51*	*65.00*	*4,867*	—	Uncontaminated freshwater marsh sediment location B, Haren, The Netherlands	[1]
Rhodopseudomonas sp. BisA53	*PRJNA15751*[c]	*5.51*	*64.40*	*4,852*	—	Uncontaminated freshwater marsh sediment location A, Haren, The Netherlands	[1]
Rhodopseudomonas sp. DX-1	*GCA_000177255.1*	*5.4*	*65.40*	*4,849*	—	Microbial fuel cell	[59]
Rhodopseudomonas sp. AZUL	*GCA_024330085.1*	*6.05*	*65.60*	*5,596*	—	High altitude shallow lake, Andes Region, Catamarca, Argentina	[7], this work

[a] Taxonomy as proposed by Imhoff et al. [60].

[b] Main reference cited when available

[c] Not available

was further evidenced by adjustment to the equations proposed by Tettelin et al. [33], since the addition of strains to the analysis increased the size of the total gene repertoire with a linear trend and reduced the size of the core genome. These results are expected, given the great metabolic adaptability of this genus. In this scenario, our results reinforce the idea that *Rhodopseudomonas* strains evolve from the core metabolic flexibility of the genus into the generation of distinct ecotypes highly adapted to specific microenvironments [2].

Functional annotation of *Rhodopseudomonas* sp. AZUL
Functional annotation and clustering allowed determining which features of the AZUL genome are part of the basal capabilities of the *Rhodopseudomonas* genus (core, soft-core and shell genomes) and which are either strictly specific or shared with a few strains (cloud genome). For genome comparison, we selected

the output of Roary, since it resulted adequate at discriminating paralogs in selected protein families, as determined in HMMER clustering analyses (data not shown). The identity cut-off for clustering was 70%. The proteins clustered separately through this method are not necessarily absent in other strains, but simply less than 70% identical to orthologous sequences from other strains. This percentage has been reported elsewhere as a suitable value for the determination of orthologs in interspecific analyses at the genus level [61]. All proteins that resulted specific or rare in the AZUL genome had nhmmer inter-cluster E-values that were at least four-fold higher than the corresponding intra-cluster E-values (data not shown). Table S1 shows Roary pangenomic clustering and annotation, including *Rhodopseudomonas* sp. AZUL predicted gene inventory. In addition, the total number of genes from different protein families, classes and subsystems were determined for the 31 strains using PATRIC (Fig. 4).

Guardia *et al. BMC Microbiology* (2022) 22:297

Page 8 of 19
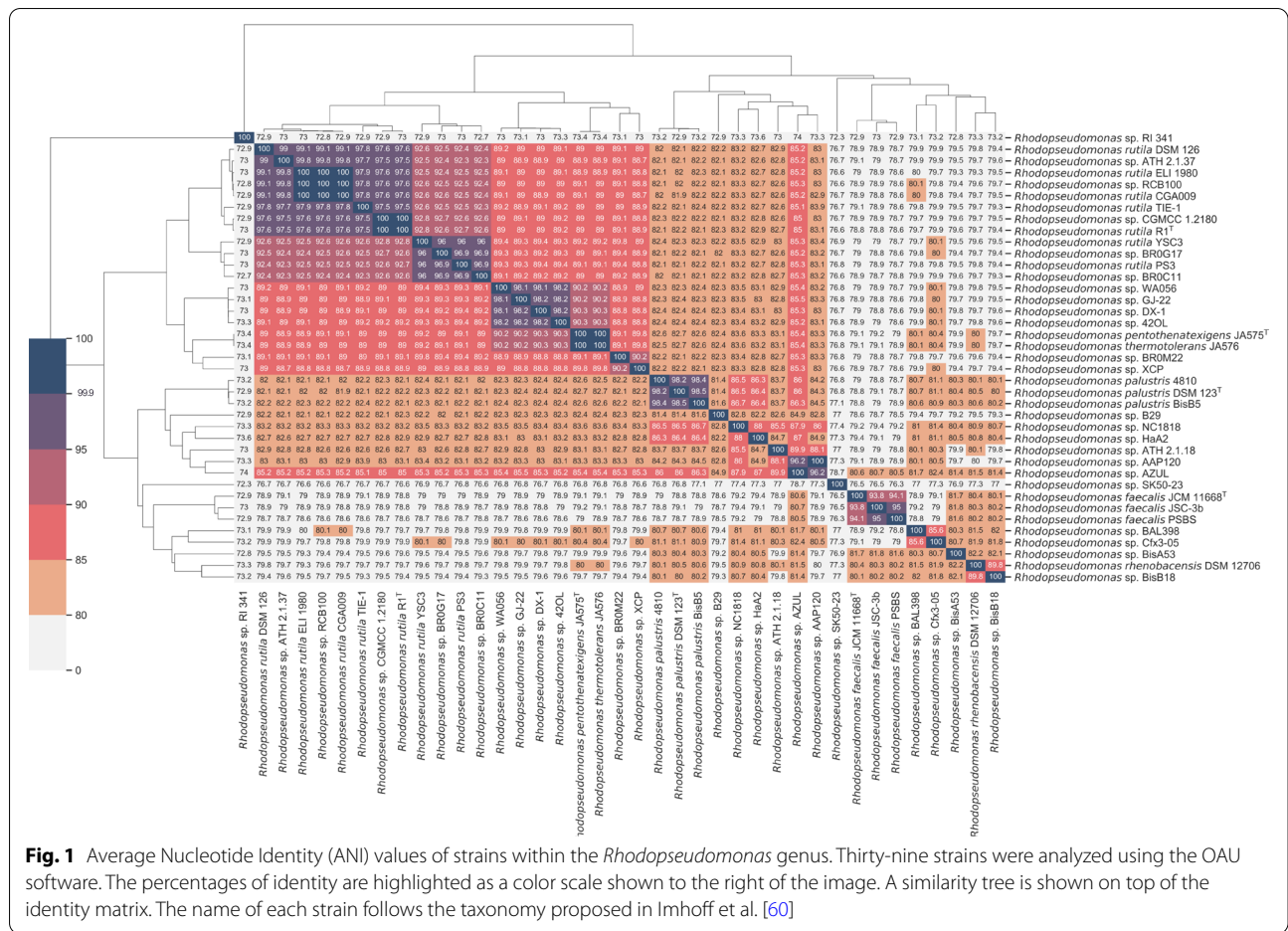
Fig. 1 Average Nucleotide Identity (ANI) values of strains within the *Rhodopseudomonas* genus. Thirty-nine strains were analyzed using the OAU software. The percentages of identity are highlighted as a color scale shown to the right of the image. A similarity tree is shown on top of the identity matrix. The name of each strain follows the taxonomy proposed in Imhoff et al. [60]

Table 2 Pangenome characteristics of the genus *Rhodopseudomonas* based on different clustering algorithms

Genome cluster	Def inition	Roary	GET Homologues			OMCL-exponential model[b]	OMCL-mixed model[c]
			BDBH	COGt	OMCL		
Core genome	Genes contained in all genomes/taxa	1,217	1,055 (987)[a]	1,073 (987)[a]	1,041 (987)[a]	1,397	1,014
Soft-core	Genes contained in n - 1 of the considered genomes/taxa	569	—	—	—	—	1,851
Shell	Moderately conserved genes, present in several genomes/taxa	4,553	—	—	—	—	3,069
Cloud	Rare genes present only in a few genomes/taxa	20,046	—	—	—	—	4,010
Pangenome	Complete gene collection	26,385	—	20,757 (8,944)	10,507 (8,944)	27,595	8,944

[a] The size of the core genome obtained from the consensus of the algorithms used by GH is shown between parentheses

[b] According to the model described in Tettelin et al.[33]

[c] According to the model of Snipen et al. [34]. In this model, the soft core includes the core, whereas it does not in Roary

Main predicted metabolic pathways of *Rhodopseudomonas* sp. AZUL (core, soft-core and shell genomes)

AZUL has the genes necessary for carrying out anoxygenic photosynthesis present in the other *Rhodopseudomonas* strains (Table S1). They include the light-harvesting complexes, electron carrier proteins, proteins involved in photo-phosphorylation and those linked to pigment biosynthesis.

Among the sequences associated with carbohydrate metabolism, we could identify genes related to glycolysis,

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 9 of 19

Fig. 2 Unrooted tree of *Rhodopseudomonas* strains based on the core genome. Core genes obtained with Roary were concatenated, aligned and used to infer phylogenetic relationships with the neighbour joining method using FastTree, all part of the Roary pipeline. The tree scale in number of nucleotide substitutions per site is shown

gluconeogenesis, TCA and pentose-phosphate cycles. AZUL carbohydrate metabolism also features genes that participate in oligosaccharide and organic acid metabolism, fermentation and carbon dioxide fixation. One distinguishing characteristic of the *Rhodopseudomas* genus is the ability to degrade aromatic compounds, which, in AZUL, is represented by 30 main predicted genes (Table S1).

The subsystem of nitrogen metabolism includes proteins involved in nitrogen fixation (nitrogenases), ammonium assimilation and ammonification, nitrification and denitrification, cyanide hydrolysis and nitrosylative stress. Similarly to a few other strains, the genome of AZUL has genes for the three types of nitrogenases (iron, molybdenum and vanadium) (Table S1). Since they vary in their affinities for atmospheric N_2 and in their catalytic efficacies [62], this would allow AZUL to fix nitrogen in the limited oligotrophic basin that constitutes its habitat [19]. *Rhodopseudomonas* strains have also shown the ability to form both monospecific biofilms [7] and
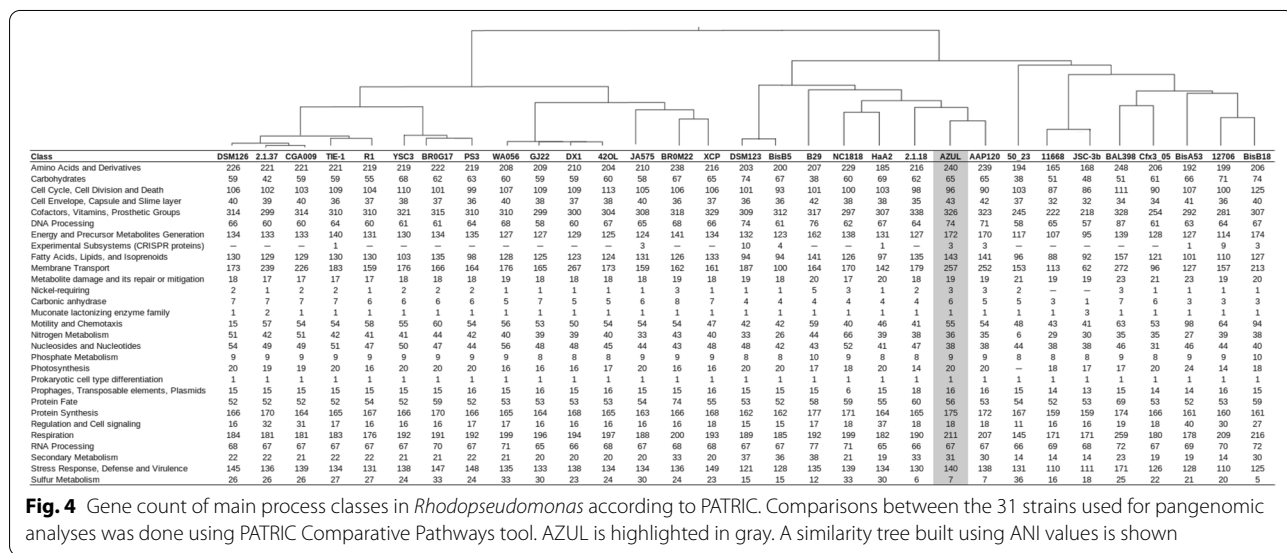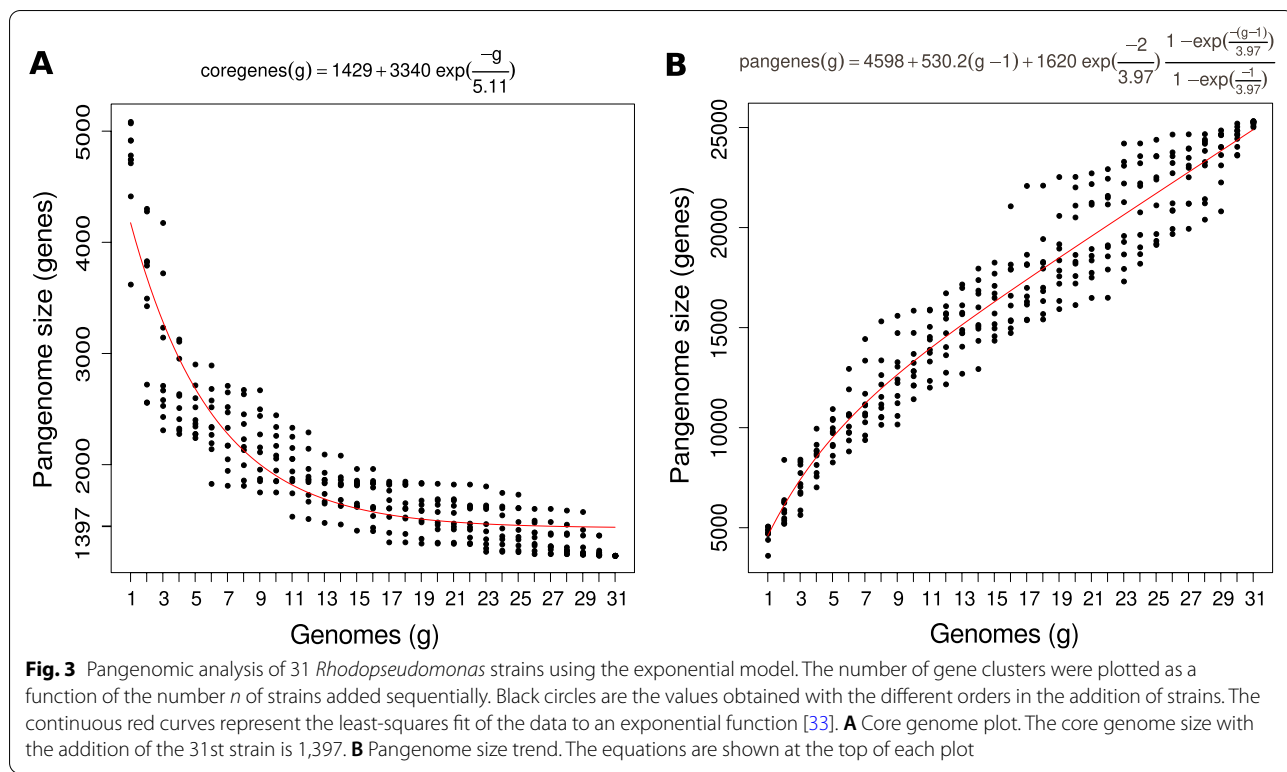
mixed-species microbial mats [63], which could further limit the accessibility to nitrogen and other nutrients [64].

Similarly to other purple bacteria, the genome of AZUL has the genes to perform the four types of carbon and energy metabolism: photoautotrophy, photoheterotrophy, chemoautotrophy and chemoheterotrophy. AZUL has the genes for both form I and II of Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBiscCO), CbbM and CbbS, and a Rubisco-like protein (Rlp2), which enable this strain to carry out autotrophic growth. During this type of metabolism, hydrogen and thiosulfate could be used as electron donors in the reverse electron flow [3, 65]. Related to this, *Rhodopseudomas* strains have *hoxL* and *hoxK* genes, which code for hydrogenases. In addition, genes that code for several proteins involved in sulfate/thiosulfate uptake are shared between AZUL and the rest of the *Rhodopseudomas* strains, such as the import ATP-binding protein CysA, the thiosulfate sulfurtransferase GlpE and the thiosulfate-binding protein CysP, although some of these homologs seem to have diverged

Fig. 3 Pangenomic analysis of 31 *Rhodopseudomonas* strains using the exponential model. The number of gene clusters were plotted as a function of the number n of strains added sequentially. Black circles are the values obtained with the different orders in the addition of strains. The continuous red curves represent the least-squares fit of the data to an exponential function [33]. **A** Core genome plot. The core genome size with the addition of the 31st strain is 1,397. **B** Pangenome size trend. The equations are shown at the top of each plot

Equation A:
$$\text{coregenes}(g) = 1429 + 3340\,\exp\!\left(\frac{-g}{5.11}\right)$$

Equation B:
$$\text{pangenes}(g) = 4598 + 530.2(g-1) + 1620\,\exp\!\left(\frac{-2}{3.97}\right)\frac{1-\exp\!\left(\frac{-(g-1)}{3.97}\right)}{1-\exp\!\left(\frac{-1}{3.97}\right)}$$

Class	DSM126	2.1.37	CGA009	TIE-1	R1	YSC3	BR0G17	PS3	WA056	GJ22	DX1	42OL	JA575	BR0M22	XCP	DSM123	BisB5	B29	NC1818	HaA2	2.1.18	AZUL	AAP120	50_23	11668	J5C-3b	BAL398	Cfx3_05	BisA53	12706	BisB18
Amino Acids and Derivatives	226	221	221	221	219	219	222	219	208	209	210	204	210	238	216	203	200	207	229	185	216	240	239	194	165	168	248	206	192	199	206
Carbohydrates	59	42	59	59	55	68	62	63	60	59	59	60	58	67	65	74	67	38	60	69	62	65	65	38	51	48	51	61	66	71	74
Cell Cycle, Cell Division and Death	106	102	103	109	104	110	101	99	107	109	109	113	105	106	106	101	93	101	100	103	98	96	90	103	87	86	111	90	107	100	125
Cell Envelope, Capsule and Slime layer	40	39	40	36	37	38	37	36	40	38	37	38	40	36	37	36	36	42	38	38	35	43	42	37	32	32	34	34	41	36	40
Cofactors, Vitamins, Prosthetic Groups	314	299	314	310	310	317	315	310	310	299	300	304	308	318	329	309	312	317	297	307	338	326	323	245	222	218	328	254	292	281	307
DNA Processing	66	60	60	64	60	61	61	64	68	58	60	67	65	68	66	74	61	76	62	67	64	74	71	58	65	57	87	61	63	64	67
Energy and Precursor Metabolites Generation	134	133	133	140	131	130	134	135	127	127	129	125	124	141	134	132	123	162	138	131	127	172	170	117	107	95	139	128	127	114	174
Experimental Subsystems (CRISPR proteins)	–	–	–	–	–	–	–	–	–	–	–	–	–	3	–	–	10	4	–	–	1	–	3	–	–	–	–	–	1	9	3
Fatty Acids, Lipids, and Isoprenoids	130	129	129	130	130	103	135	98	128	125	123	124	131	126	133	94	94	141	126	97	135	143	141	96	88	92	157	121	101	110	127
Membrane Transport	173	239	226	183	159	176	166	164	176	165	267	173	159	162	161	187	100	164	170	142	179	257	252	153	113	62	272	96	127	157	213
Metabolite damage and its repair or mitigation	18	17	17	17	17	18	18	18	19	18	18	18	18	19	18	19	18	20	17	20	18	19	19	21	19	19	23	21	23	19	20
Nickel-requiring	2	1	2	2	1	2	2	2	1	1	1	1	1	3	1	1	1	5	3	1	2	3	3	2	–	–	3	1	1	1	1
Carbonic anhydrase	7	7	7	7	6	6	6	6	5	7	5	5	6	8	7	4	4	4	4	4	1	6	5	5	3	1	7	6	3	3	3
Muconate lactonizing enzyme family	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	2	–	–	3	1	1	1	1	
Motility and Chemotaxis	15	57	54	54	58	55	60	54	56	53	50	54	54	54	47	42	42	59	40	46	41	55	54	48	41	63	53	98	64	94	
Nitrogen Metabolism	51	42	51	42	41	41	44	42	40	39	39	40	33	43	40	33	26	44	66	39	38	36	35	6	29	30	35	35	27	39	38
Nucleosides and Nucleotides	54	49	49	51	47	50	47	44	56	48	48	45	44	43	40	48	42	43	52	41	47	38	38	44	38	38	46	31	46	44	40
Phosphate Metabolism	9	9	9	9	9	9	9	9	9	8	8	8	9	9	9	8	8	10	9	8	8	9	9	8	8	9	9	9	10		
Photosynthesis	20	19	19	20	16	20	20	20	16	16	16	17	20	16	20	20	17	18	20	14	20	20	–	18	17	17	20	24	14	18	
Prokaryotic cell type differentiation	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Prophages, Transposable elements, Plasmids	15	15	15	15	15	15	15	16	15	16	15	16	15	15	16	1	15	15	6	15	18	16	16	15	14	13	15	14	14	15	
Protein Fate	52	52	52	52	54	52	59	52	53	53	53	53	54	74	55	53	52	58	55	60	56	53	54	52	53	69	53	52	53	59	
Protein Synthesis	166	170	164	165	167	166	170	166	165	164	168	165	163	166	168	162	162	177	171	164	165	175	172	167	159	159	174	166	161	160	161
Regulation and Cell signaling	16	32	31	17	16	16	16	17	17	16	16	16	16	18	15	15	17	18	37	18	18	18	11	16	16	19	18	40	30	27	
Respiration	184	181	181	183	176	192	191	192	199	194	197	188	200	193	189	185	192	199	182	190	211	207	145	171	171	259	180	178	209	216	
RNA Processing	68	67	67	67	67	67	70	67	71	65	66	68	67	68	68	67	67	77	71	65	66	67	67	66	69	68	72	67	69	70	72
Secondary Metabolism	22	22	21	22	22	21	21	22	21	20	20	20	20	33	20	17	36	38	21	19	31	30	14	14	23	19	19	14	30		
Stress Response, Defense and Virulence	145	136	139	134	131	138	147	148	135	133	138	134	134	136	149	121	128	135	139	134	130	140	138	131	110	111	171	126	128	110	125
Sulfur Metabolism	26	26	26	27	27	24	33	24	33	30	23	24	30	24	23	15	15	12	33	30	6	7	7	36	16	18	25	22	21	20	5

Fig. 4 Gene count of main process classes in *Rhodopseudomonas* according to PATRIC. Comparisons between the 31 strains used for pangenomic analyses was done using PATRIC Comparative Pathways tool. AZUL is highlighted in gray. A similarity tree built using ANI values is shown

considerably between strains. A few of the strains, including AZUL, have the additional thiosulfate sulfur-transferase RhdA (Table S1).

Genes of carbon monooxide and formate dehydrogenases are also part of the collection of the genus. These proteins could contribute to generating reducing power through their oxidation. Other common dehydrogenases are related to ethanol and methanol metabolism, which,

though not present in all strains, are part of the *Rhodopseudomas* repertoire (Table S1).

Within the subsystem of Respiration, 211 main sequences were identified in the genome sequence of AZUL, many within the core genome, including a great variety of oxidoreductases, cytochromes and proteins related to the biogenesis of these compounds. Oxidoreductase genes are either related to aerobic or anaerobic

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 11 of 19

respiration. An example refers to the presence of a *dmsA* gene, which encodes a dimethyl sulfoxide/trimethylamine N-oxide reductase, an enzyme that catalyzes the reduction of dimethyl sulfoxide (DMSO) and trimethylamine N-oxide (TMAO) to dimethyl sulfide (DMS) and trimethylamine, respectively [66, 67] and can also use other sulfoxides and N-oxide compounds as terminal electron acceptors [68]. Another example is the *dsbB* gene, encoding a thiol-disulphur oxidoreductase involved in the reduction of the metalloid tellurium that is present in all the *Rhodopseudomonas* strains (Table S1). In *Rhodobacter capsulatus,* tellurium reduction due to the action of DsbB was proposed to act as an electric wire between the metalloid and the quinone pool [69], which suggests that this enzyme can be used for both detoxification and electron transport purposes.

The *Rhodopseudomonas* genus evidences an ability to synthesize multiple siderophore receptors. As an example, eleven sequences corresponding to the outer membrane receptor of the siderophore ferrichrome [70, 71] (five *fcuA* and six *fhuA* genes), as well as three sequences corresponding to the ferripyoverdine receptor FpvA [72] were identified in AZUL (Table S1).

The *Rhodopseudomonas* genus has a core ability to handle zinc limitation, although the gene sequences have diverged enough among strains to be located in different clusters (Table S1) Sequences of *zntB*, *znuC* and *yciC* were detected in all the strains including AZUL. The proteins encoded by these genes participate, in different ways, in Zn^{2+} uptake in conditions of metal limitation [73–75].

Special predicted features of the *Rhodopseudomonas* sp. AZUL genome

Figure 4 shows that strains AZUL and AAP120 have remarkable similarities in the gene count of all the process classes analyzed by PATRIC. In general, these two strains are on the top end of gene count for most pathways, which could be, in part, due to the larger number of coding sequences present in their genomes. In general, this is not shared by other moderately related strains (those with ANI values above 85%).

Energy generation

Within the category of Energy and Precursor Metabolites Generation, it is remarkable in AZUL and AAP120 the presence of 4 genes that code for soluble methane monooxygenases (sMMO) (Table S1), present in only other 4 of the 31 strains analyzed (Table S1). This enzyme is a three-component non-heme iron oxygenase that catalyzes the initial step of the methane oxidation pathway, the conversion of methane to methanol. SMMOs can co-oxidize a very wide range of substrates together with methane, including alkanes, alkenes, alcohols, ethers, alicyclics, aromatics and chlorinated organic compounds such as the pollutant trichloroethene [76].

Urea and nitrogen metabolism

Within the category of Amino Acids and Derivatives, AZUL and AAP120 have a considerably larger number of urea amidolyase and urease genes (Table S1) than the rest of the *Rhodopseudomonas* strains (Table S1). Both enzymatic complexes participate in the utilization of urea as a nitrogen source by its sequential transformation to ammonium [77, 78]. A larger number of genes might allow more efficient nitrogen utilization under a wider array of conditions. As mentioned for nitrogenases, this would prove very useful in an oligotrophic environment with changing levels of nitrogen from various origins.

Membrane transport

The category of Membrane Transport is represented by 257 genes in AZUL and 252 genes in AAP120, only surpassed by *Rhodopseudomonas* sp. strain DX1 (267 genes) (Fig. 4). A closer inspection shows that, despite the precise gene count, different strains have different gene distribution within the subsystems of this class. Remarkable to AZUL and AAP120 are the number of genes involved in metal homeostasis, such as proteins that participate in copper and magnesium transport and multi-subunit cation antiporters (Table S1). In view of this, we did a semi-quantitative physicochemical analysis of the water sample where the strain was originally isolated from. Table 3 shows that metals and metalloids such as copper, zinc, chromium and arsenic are present in the water. Due to the fluctuating nature of the habitat of this strain, these results evidence, at the very least, that AZUL is periodically exposed to these metals.

Copper homeostasis

Within this category are Copper-translocating P-type ATPases CopA, CopB and ActP, as well as the copper resistance proteins CopD and C and the repressor CsoR of the *copZA* operon. While the genomes of most *Rhodopseudomonas* strains have at least 1 *copA* gene homolog, a *copB* ortholog was only detected in the genome of AZUL. Furthermore, *actP* homologs were identified in strains AZUL, BisB5, BAL398 and DSM123 (Table S1). Both *csoR* orthologs, as well as sequences of its homolog *ricR* (one *ricR1* and one *ricR2*) were also found in AZUL.

Table 4 summarizes the number of genes related to metal homeostasis detected in the AZUL genome, as well information regarding their protein families and general mechanisms. In addition to translocating P-type ATPases, copper (and silver) resistance in bacteria is

Table 3 Elements present in the Laguna Azul water sample as determined by Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

Element	Parts per billion (ppb)
B	10,640.76
Si	6,378.11
Li	2,948.40
P	1,240.92
Br	627.62
I	120.90
Fe	94.25
As	58.23
Rb	20.57
Cr	18.05
Xe	15.57
V	12.17
Ti	11.75
Se	6.20
Sr	6.10
Mo	3.56
U	2.29
Zn	2.04
Cu	2.00
Sc	1.74
Al	1.55
Ba	1.17

exerted via the resistance–nodulation–cell division (RND) metal efflux superfamily [79], which includes the partner proteins named membrane fusion protein (MFP) and outer membrane factors (OMF) [80]. These genes are part of the operon *cusCFBA* [81]. Although the *cus* operon is widespread in *Rhodopseudomonas*, AZUL has a larger number of genes compared to the rest of the strains, precisely 11 (Table 4).

Figure 5 shows the organization of metal homeostasis genes within the *Rhodopseudomonas* sp. strain AZUL genome. The *cus* system in AZUL is represented by three main operons (Fig. 4B), in two cases intercalated with *cop* genes, forming larger copper-resistance units that also contain cupredoxin genes (*cdx*). Cupredoxins function as electron shuttles between proteins, which opens the possibility that AZUL uses copper for bioenergetic processes, as reported for other organisms [82]. These copper resistance units are located in the proximity of other genes related to resistance, such as the gene for a multidrug outer membrane efflux protein, *tolC* [83].

Several sequences that code for extracellular proteins involved in copper sequestration were identified. For instance, two homologous genes of the mycobacterial multicopper oxidase, *mmcO,* were detected in the

genome of AZUL and three in the genome of BAL398, as opposed to single sequences in the rest of the strains (Table S1). MmcO possibly acts by oxidation of toxic Cu(I) in the periplasm [84]. In addition, single copies of the *pcoC* gene were detected in the genome of AZUL and several other strains (Table S1). PcoC is a periplasmic protein that binds excess copper ions and increases the level of resistance above that conferred by the *copA* operon alone [85]. The *pcoC* gene is part of one of the two *cop* operons present in AZUL, together with a single *copD* gene, which encodes a copper resistance protein of unclear function [86] (Fig. 4B).

CopA and CopB detoxify copper under aerobic conditions in many bacteria [87], whereas the *cus* operon is most frequently activated under anaerobic conditions [81]. In addition, actP ATPase exports copper at low pH [88]. Thus, the genome sequence of AZUL suggests that this strain is well appointed for copper export under varying environmental conditions, as well as for neutralizing toxic copper ions in the periplasm. The presence of the *ycnJ* and *ricR* genes, which code for transcriptional repressors [89, 90] evidences that AZUL can also respond to copper limitation. The location of an extra copy of the silver resistance gene *silP* adjacent to the *cus* operon suggests that, similar to other bacteria [91], the *cus* operon might be used for both silver and copper detoxification in AZUL.

Czc resistance operons

Another RND protein within the Cation Transporters subsystem is CzcA, a cation-proton antiporter that transports excess Cd^{2+}, Zn^{2+}, and Co^{2+} whose gene is located within an operon that contains genes for the MFP CzcB and the OMF protein CzcC [92]. Fourteen *czcA*, *czcB* and *czcC* genes were identified in the genome of AZUL (Table 4), most of them shared with other strains, although a few additional homologs were identified. The Czc system of AZUL is organized in at least six different operons, in some cases nearby other genes involved in efflux and resistance, such as those of the lead, cadmium, zinc and mercury-transporting ATPase ZntA, the cobalt–nickel resistance protein CnrA, the outer membrane efflux protein OprM and the multidrug efflux proteins MdtC, MdtE and BepC (Fig. 4A). One of these operons includes a gene for the cation diffusion facilitators (CDF) CzcD, which, in AZUL, is represented by single *czcD1* and *czcD2* gene sequences. CzcD was first described as a regulator of expression of the CzcCBA operon described above, but this protein is also able to mediate a small degree of $Zn^{2+}/Co^{2+}/Cd^{2+}$ resistance in the absence of the high resistance CzcCBA system [93], which implies that the strains expressing this protein have the ability to activate low-level metal efflux.

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 13 of 19

Table 4 Main genes involved in metal resistance in *Rhodopseudomonas sp.* AZUL

Metal resistance system	Resistance component	Genes
Cadmium, zinc and cobalt high resistance RND-MFP-OMF system (czcCBA)	Cobalt-zinc-cadmium resistance RND protein CzcA	czcA_1
		czcA_2
		czcA_3
		czcA_4
		czcA_5
		czcA_6
		czcA_7
	Cobalt-zinc-cadmium resistance MFP CzcB	czcB_1
		czcB_2
		czcB_3
		czcB_4
		czcB_4
	Cobalt-zinc-cadmium resistance OMF CzcC	czcC_1
		czcC_2
Additional czc operon-related proteins	Metal cation efflux system protein CzcD	czcD_2
		czcD
	Cobalt-nickel resistance protein A	cnrA
Copper/silver resistance RND-MFP-OMF system (cusCFBA)	Copper-silver resistance RND protein CusA	cusA_1a
		cusA_1b
		cusA_2
		cusA
	Copper-silver resistance MFP CusB	cusB_1
		cusB_2
		cusB_4
	Periplasmic copper-binding protein CusF	cusF_1
		cusF_2
	Transcriptional regulatory protein CusR	cusR
	Sensor kinase CusS	cusS
Export P-type ATPases	Zinc export P-type ATPases	zntA
	Copper export P-type ATPases	cop A
		copA_1
		copA_2
		copA_3
		copB
		actP
	Silver resistance P-type ATPases	silP_1
		silP
Additional cop operon-related proteins	Periplasmic proteins	pcoC
		copD
	Transcriptional repressors	csoR
		ricR_1
		ricR_2
Chromiun efflux system (chrBACF operon)	Chromate efflux transmembrane proteins	chrA1
		chrA_1
		chrA_2
		srpC_1
		srpC_2
	Chromate-sensing transcriptional regulator chrB	chrB
	Chromium-dependent Superoxide dismutase [Fe]	chrC
Chromium reduction system	Chromate reductases	chrR_2
		chrR_3
		chrR_1
Mercury resistance system (mer operon)	Mercuric reductase	merA
	Mercury resistance transport proteins MerF/T	merF
		merT
	Mercuric resistance periplasmic component	merP
	Mercuric resistance regulatory protein	merR1_1
		merR1_2
Arsenic resistance system (ars operon)	Arsenate reductases	arsC_1
		arsC_3
	Glutaredoxin arsenate reductases	arsC_3
		arsC_4
	NADPH-dependent FMN reductase ArsH	arsH_1
		arsH_2
	Arsenical-resistance permeases	acr3
		acr3_2
	Arsenic resistance transcriptional activators	hlyU
	Arsenic resistance transcriptional repressors	arsR_1

Fig. 5 Schematic view of the main operons involved in metal resistance in the AZUL genome. **A** Genes of the *czc* system for resistance to cadmium, cobalt and zinc (yellow boxes). **B** Copper resistance *cop* (blue boxes) and *cus* (purple boxes), cupredoxin genes (*cdx₁* and *cdx-2*, dark gray boxes), *silP* and *tolC* (green). **C** Chr chromate efflux genes *chrA, chrB, chrC, srpC* and *chrp* (pink boxes). **D** Mercury resistance genes *merP, merF, merT* and *merR1* genes, located in the vicinity of *tniQ, tniB* and *tnsB* genes, which participate in transposition (all orange boxes). **E** The main regulatory units of the AZUL arsenic resistance system (blue boxes) contain *arsC, acr3, arsH* and regulatory *arsR* and *hlyU* genes. In all the operons, box arrowheads show the orientation of each coding region. Light gray boxes represent unrelated or hypothetical coding regions and green boxes represent genes for resistance to other stressors. The numbered black lines indicate the size of each operon in base pairs. The scale is maintained for each resistance system, but different systems have different scales, in bp. For details about the function of each gene, please refer to Table 4

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 15 of 19

Chromium resistance

Additional genes that participate in metal homeostasis/resistance were identified within the Stress Response, Defense and Virulence gene class. Many of them are present in most strains, albeit in different numbers. For instance, the genomes of AZUL and AAP120 feature a larger number of genes involved in chromium resistance than the rest of the strains (7 genes each, Table S1). These genes belong to the CHR family, composed of transmembrane proteins that act as chromate efflux pumps driven by a chemiosmotic gradient [94]. Three copies of both *chrA1* and *chrA2* genes were identified in the genome of AZUL, as well as two copies of the *chrA* homolog *srpC* and one copy of the transcriptional regulator gene *chrB* [95]. Some of the genes involved in chromate efflux are organized in AZUL within two main operons (Fig. 5C), while the rest of the copies are located in other parts of the genome.

Chromate reductase genes *chr1* and *chr2* were identified in the genomes of all the *Rhodopseudomonas* strains, whereas *chr3* homologs were only found in AZUL and a few other strains. In addition, single gene copies of the chromium-activated superoxide dismutase SodM-like protein ChrC/F were identified in AZUL and AAP120. Chromate reduction confers resistance to this metal in conjunction with chromate transport [96].

Mercury resistance

A cluster of sequences homologous to genes of the *mer* operon were identified in AZUL, namely one *merA*, two *merR*, two *merT/F* and one *merP*. MerA is mercury reductase, which catalyzes the reduction of Hg(II) to volatile Hg(0), while the rest of the *mer* genes encode proteins involved in regulation (MerR), cytoplasmic transport (MerT/F) and periplasmic binding (MerP) [97]. *Mer* sequences were found in other 8 of the *Rhodopseudomonas* strains, although only one *merR* and no *merT* were identified in any of them (Table S1). Interestingly, close to the AZUL *mer* operon, we identified *tniQ*, *tniB* and *tnsB* genes, which participate in transposition of the *mer* operon (Fig. 4D) [98], suggesting that this could have been the mechanism for acquisition of these genes in AZUL.

Arsenic resistance

We detected sequences homologous to genes from the *ars1 or ars2* operons, which confer resistance to arsenite and arsenate in a number of species [99, 100]. Specifically, *arsC1* and *arsC2* (arsenate reductases)*, acr3* (bile/arsenite/riboflavin-type permease), and *arsR* (transcriptional repressor) were identified in all the *Rhodopseudomonas* strains (Table S1). Additional sequences for *arsH1* and *arsH2* were only found in AZUL, with only one ortholog present in a few of the *Rhodopseudomonas* strains. ArsH proteins are organoarsenical oxidases that detoxify trivalent methylated and aromatic arsenicals by oxidation to the pentavalent species [101]. Contrary to CGA009 [102], we did not find evidence of an arsenite S-adenosylmethionine methyltransferase gene (*arsM*) in the genome of AZUL, suggesting that the pathway for As (III) methylation and subsequent volatilization through this protein is not among the mechanisms of resistance to arsenic in this strain. The *ars* genes are organized in two main operons in AZUL, one of them upstream of a chromate transporter gene *chrA* (Fig. 4E).

Multidrug resistance

The genomes of *Rhodopseudomonas* sp. AZUL and AAP120 feature many genes that take part in efflux systems for multidrug resistance, represented in AZUL by 94 detected genes, among the largest number detected in the 31 *Rhodopeudomonas* strains (Table S1). Multidrug efflux pumps export a wide variety of compounds in bacteria [103]. Interestingly, several multidrug genes within the AZUL genome were identified contiguous to genes of metal resistance (Fig. 4).

The appearance of multidrug resistance in conjunction with resistance to metals within the same genome might not be casual. A study conducted in Laguna Azul showed a correlation between multi-resistance to UV radiation, arsenic and antibiotics, despite the fact that selective pressure for antibiotic resistance is not expected in that habitat [12]. The authors proposed that, under the extreme UV radiation that the HAALs are exposed to, bacteria have increased mutation rates in which spontaneous resistance to antibiotics can emerge. Indeed, our results show that, in AZUL, multidrug genes are often located adjacent to metal resistance operons, which could have important implications related to their origin and/or regulation.

Protein and nucleoprotein secretion system, Type IV

Within the Membrane Transport class, AZUL, AAP120 and BAL398 have several genes involved in Vir-like type 4 secretion (Table S1). The type IV (T4SS) is one of several types of secretion systems that microorganisms use for the transport of macromolecules such as proteins and DNA across the cell envelope [104]. It is the most versatile family of secretion, mediating transport of monomeric proteins as well as multi-subunit protein toxins and nucleoprotein complexes. In the case of DNA transfer, it is known to greatly increase genomic

plasticity, helping microorganisms to adapt to changes in their environment [104].

Metal uptake

Many genes that take part in metal uptake are common among all or most of the *Rhodopseudomonas* strains. One exception is the *mgtB* gene, which encodes a P-type ATPase involved in Mg^{2+} intake under limiting conditions [105] that was only detected in AZUL. In addition, two homologs of the *hoxN* gene, involved in high-affinity nickel uptake [106] are present in AZUL, while only one homolog, or none, were found in the rest of the strains (Table S1).

Response to UV

The sequences related to tolerance to UV are, in general, common between AZUL and the rest of the *Rhodopseudomonas* strains. Nevertheless, there are a few exceptions. For instance, single homologs encoding RuvA and RuvB, the subunits of an ATP-dependent DNA helicase that participates in recombinational repair of UV damage [107], were identified in all the *Rhodopseudomonas* strains, while an extra gene for a RuvB subunit was only found in the genome sequence of AZUL. As another example, homologs of the deoxyribodipyrimidine photolyase gene *phrA,* involved in photorepair [108], were identified in all the strains analyzed, whereas its counterpart, *phrB*, was only identified in AZUL and 2.1.18.

Conclusions

In this work, we have presented the draft genome sequence of *Rhodopseudomonas sp.* AZUL, a strain isolated from an endorheic basin in the Argentinean Andean region. In addition, we have done comparative genomic analyses that included most of the *Rhodopseudomona*s strains with available sequenced genomes. This has led us to conclude that the *Rhodopseudomonas* genus has an open pangenome, as the addition of strains to the analysis increased the size of the total gene repertoire with a linear trend. These results are expected given the great metabolic adaptability of this genus. In this scenario, our results reinforce the idea that *Rhodopseudomonas* strains evolve from the basal metabolic flexibility of the genus into the generation of ecotypes greatly adapted to their specific niches [2]. Our genomic analysis has shown that, in addition to the metabolic complexity of the genus, AZUL has numerous mechanisms for both the uptake and export of minerals and other chemicals, as well as for chemical detoxification. Being able to accomplish the four types of carbon and energy metabolism, it is likely that AZUL

has evolved to not only tolerate but also use some of these compounds as electron carriers.

The genome of AZUL has also several transport genes for DNA and nucleoprotein uptake, which could confer remarkable genomic plasticity that surely helps this strain adapt to changes in its surroundings. This is ideal considering that the natural habitat of AZUL experiences drastic changes in environmental parameters and seasonal contraction and expansion of the water levels that modify the amount of minerals and other compounds from geochemical and volcanic origin [17, 109]. In addition to a remarkable adaptation to a hostile environment, the genome of AZUL makes us envision promising industrial applications of this microorganism. Future experimentation will interrogate its capacity to bioremediate both metals and organic compounds.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-022-02685-w.

Additional file 1. Report for *R. palustris* AZUL genome sequencing and assembly. Project type: WGS.

Additional file 2: (Table S1):Table S1-1. Results of genome quality analysis of the 39 Rhodopseudomonas strains used in this work, obtained using the Genome Annotation service at PATRIC. **Table S1-2.** Full clustering analysis of the 31 strains of Rhodopseudomonas with Roary, using 70% identity cut-off and the 'do not split paralogs' option. **Table S1-3.** Gene annotation of the Rhodopseudomonas genome according to Prokka and Roary (split paralogs option). **Table S1-4.** Gene count of selected subsystems from the PATRIC comparative pathway output in the 31 Rhodopseudomonas strains.

Availability of data and materials
This whole-genome shotgun project is available at the RAST server (https://rast.nmpdr.org/) and also at NCBI GenBank, with the assembly accession GCA_024330085.1.

Declarations

Ethics approval and consent to participate
Not applicable.

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 17 of 19

Author details
[1]Ingeniería de Interfases y Bioprocesos, Instituto de Tecnología de Materiales (INTEMA-CONICET-UNMdP), Mar del Plata, Argentina. [2]Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Zavalla, Argentina. [3]Facultad de Ciencias Bioquímicas y Farmacéuticas, Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET-UNR), Universidad Nacional de Rosario, Rosario, Argentina. [4]Instituto de Investigaciones Biológicas (IIB-CONICET-UNMdP), Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina.

References

1. Oda Y, Wanders W, Huisman LA, Meijer WG, Gottschal JC, Forney LJ. Genotypic and phenotypic diversity within species of purple nonsulfur bacteria isolated from aquatic sediments. Appl Environ Microbiol. 2002;68:3467–77.
2. Oda Y, Larimer FW, Chain PSG, Malfatti S, Shin MV, Vergez LM, et al. Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. Proc Natl Acad Sci U S A. 2008;105:18543–8.
3. Larimer FW, Chain P, Hauser L, Lamerdin J, Malfatti S, Do L, et al. Complete genome sequence of the metabolically versatile *photosynthetic bacterium Rhodopseudomonas palustris*. Nat Biotechnol. 2004;22:55–61.
4. Braatsch S, Bernstein JR, Lessner F, Morgan J, Liao JC, Harwood CS, et al. *Rhodopseudomonas palustris* CGA009 has two functional ppsR genes, each of which encodes a repressor of photosynthesis gene expression. Biochemistry. 2006;45:14441–51.
5. Scheuring S, Gonçalves RP, Prima V, Sturgis JN. The photosynthetic apparatus of *Rhodopseudomonas palustris*: Structures and organization. J Mol Biol. 2006;358:83–96.
6. Simmons SS, Isokpehi RD, Brown SD, McAllister DL, Hall CC, McDuffy WM, et al. Functional annotation analytics of *Rhodopseudomonas palustris* genomes. Bioinform Biol Insights. 2011;5:115–29.
7. Guardia AE, Beligni MV, Cortéz N, Busalmen JP. Electrochemistry of R. palustris AZUL during phototrophic growth. Electrochim Acta. 2020;355:136757.
8. Farías ME, Revale S, Mancini E, Ordoñez O, Turjanski A, Cortez N, et al. Genome sequence of Sphingomonas sp. S17, isolated from an Alkaline, hyperarsenic, and hypersaline volcano-associated lake at high altitude in the Argentinean Puna. J Bacteriol. 2011;193:3686–7.
9. Albarracín VH, Kurth D, Ordoñez OF, Belfiore C, Luccini E, Salum GM, et al. High-up: a remote reservoir of microbial extremophiles in Central Andean wetlands. Front Microbiol. 2015;6:1404.
10. Escudero L, Chong G, Demergasso C, Farías ME, Cabrol NA, Grin E, et al. Investigating microbial diversity and UV radiation impact at the high-altitude Lake Aguas Calientes, Chile. In: Hoover RB, Levin G V, Rozanov AY, Davies PCW, editors., et al., Instruments, Methods, and Missions for Astrobiology X. SPIE. 2007. p. 66940Z.
11. Zannier F, Portero LR, Ordoñez OF, Martinez LJ, Farías ME, Albarracin VH. Polyextremophilic bacteria from High Altitude Andean Lakes: arsenic resistance profiles and biofilm production. Biomed Res Int. 2019;2019:1231975.
12. Dib J, Motok J, Zenoff VF, Ordoñez O, Farías ME. Occurrence of resistance to antibiotics, UV-B, and arsenic in bacteria isolated from extreme environments in high-altitude (above 4400 m) Andean wetlands. Curr Microbiol. 2008;56:510–7.
13. Dib J, Weiss A, Neumann A, Ordonez O, Estevez M, Farias M. Isolation of bacteria from remote High Altitude Andean Lakes able to grow in the presence of antibiotics. Recent Pat Antiinfect Drug Discov. 2009;4:66–76.
14. Flores MR, Ordoñez OF, Maldonado MJ, Farías ME. Isolation of UV-B resistant bacteria from two high altitude Andean lakes (4,400 m) with saline and non saline conditions. J Gen Appl Microbiol. 2009;55:447–58.
15. Di Capua C, Bortolotti A, Farías ME, Cortez N. UV-resistant Acinetobacter sp. isolates from Andean wetlands display high catalase activity. FEMS Microbiol Lett. 2011;317:181–9.
16. Albarracín VH, Pathak GP, Douki T, Cadet J, Borsarelli CD, Gärtner W, et al. Extremophilic *Acinetobacter* strains from High-Altitude Lakes in Argentinean Puna: remarkable UV-B resistance and efficient DNA damage repair. Orig Life Evol Biosph. 2012;42:201–21.
17. Lara J, González LE, Ferrero M, Díaz GC, Pedrós-Alió C, Demergasso C. Enrichment of arsenic transforming and resistant heterotrophic bacteria from sediments of two salt lakes in Northern Chile. Extremophiles. 2012;16:523–38.
18. Bequer Urbano S, Albarracín VH, Ordoñez OF, Farías ME, Alvarez HM. Lipid storage in High-Altitude Andean Lakes extremophiles and its mobilization under stress conditions in Rhodococcus sp. A5, a UV-resistant actinobacterium. Extremophiles. 2013;17:217–27.
19. Fernández-Zenoff MV, Estévez MC, Farías ME. Diurnal variation in bacterioplankton composition and DNA damage in the microbial community from an Andean oligotrophic lake. Rev Argentina Micrología. 2014;46:358–62.
20. Kurth D, Amadio A, Ordoñez OF, Albarracín VH, Gärtner W, Farías ME. Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. Sci Rep. 2017;7:1–16.
21. Alonso-Reyes D, Farias ME, Albarracín VH. Uncovering cryptochrome/photolyase gene diversity in aquatic microbiomes exposed to diverse UV-B regimes. Acquatic Microb Ecol. 2020;85:141–54.
22. Ordoñez OF, Lanzarotti E, Kurth D, Gorriti MF, Revale S, Cortez N, et al. Draft genome sequence of the polyextremophilic Exiguobacterium sp. strain S17, isolated from hyperarsenic lakes in the Argentinian Puna. Genome Announc. 2013;1:480–93.
23. Gutiérrez-Preciado A, Vargas-Chávez C, Reyes-Prieto M, Ordoñez OF, Santos-García D, Rosas-Pérez T, et al. The genomic sequence of Exiguobacterium chiriqhucha str. N139 reveals a species that thrives in cold waters and extreme environmental conditions. PeerJ. 2017;5:e3162.
24. Pérez V, Dorador C, Molina V, Yáñez C, Hengst M. Rhodobacter sp. Rb3, an aerobic anoxygenic phototroph which thrives in the polyextreme ecosystem of the Salar de Huasco, in the Chilean Altiplano. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol. 2018;111:1449–65.
25. Mandakovic D, Cintolesi Á, Maldonado J, Mendoza S, Aite M, Gaete A, et al. Genome-scale metabolic models of *Microbacterium* species isolated from a high altitude desert environment. Sci Rep. 2020;10:5560.
26. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043.
27. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
28. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110:1281–6.
29. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 2013;79:7696–701.
30. Overbeek R, Fonstein M, D'Souza M, Push GD, Maltsev N. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci U S A. 1999;96:2896–901.
31. Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin EV, et al. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics. 2010;26:1481–7.
32. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.
33. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." Proc Natl Acad Sci U S A. 2005;102:13950–5.
34. Snipen L, Almøy T, Ussery DW. Microbial comparative pan-genomics using binomial mixture models. BMC Genomics. 2009;10:385.

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 18 of 19

35. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3.

36. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.

37. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

38. Lassalle F, Périan S, Bataillon T, Nesme X, Duret L, Daubin V. GC-Content evolution in bacterial genomes: the biased gene conversion hypothesis expands. PLoS Genet. 2015;11:1–20.

39. Zhang D, Yang H, Huang Z, Zhang W, Liu S-J. Rhodopseudomonas faecalis sp. nov., a phototrophic bacterium isolated from an anaerobic reactor that digests chicken faeces. Int J Syst Evol Microbiol. 2002;52:2055–60.

40. Zhang S, Luo X, Cheng J, Peng J, Zhang D, Liu Y. Genome Sequence of Pyrethroid-Degrading Bacterium *Rhodopseudomonas palustris* Strain JSC-3b. Genome Announc. 2014;2:e01228-13.

41. Wei H, Okunishi S, Yoshikawa T, Kamei Y, Maeda H. Isolation and Characterization of a Purple Non-Sulfur Photosynthetic Bacterium *Rhodopseudomonas faecalis* Strain A from Swine Sewage Wastewater. Biocontrol Sci. 2016;21:29–36.

42. Imhoff JF, Meyer TE, Kyndt JA. The genome sequence of the giant phototrophic gammaproteobacterium *Thiospirillum jenense* gives insight into its physiological properties and phylogenetic relationships. Arch Microbiol. 2006;203:97–105.

43. Venkata Ramana V, Kalyana Chakravarthy S, Shalem Raj P, Vinay Kumar B, Shobha E, Ramaprasad EVV, et al. Descriptions of Rhodopseudomonas parapalustris sp. nov., Rhodopseudomonas harwoodiae sp. nov. and Rhodopseudomonas pseudopalustris sp. nov., and emended description of Rhodopseudomonas palustris. Int J Syst Evol Microbiol. 2012;62(Pt 8):1790–8.

44. Hougardy A, Tindall BJ, Klemme JH. Rhodopseudomonas rhenobacensis sp. nov., a new nitrate-reducing purple non-sulfur bacterium. Int J Syst Evol Microbiol. 2000;50(Pt 3):985–92.

45. Vinay Kumar B, Ramprasad EVV, Sasikala C, Ramana CV. Rhodopseudomonas pentothenatexigens sp. nov. and Rhodopseudomonas thermotolerans sp. nov., isolated from paddy soils. Int J Syst Evol Microbiol. 2013;63(Pt 1):200–7.

46. Imhoff JF, Rahn T, Künzel S, Keller A, Neulinger SC. Osmotic Adaptation and Compatible Solute Biosynthesis of Phototrophic Bacteria as Revealed from Genome Analyses. Microorganisms. 2021;9:1–32.

47. Crovadore J, Xu S, Chablais R, Cochard B, Lukito D, Calmin G, et al. Metagenome-Assembled Genome Sequence of *Rhodopseudomonas palustris* Strain ELI 1980, Commercialized as a Biostimulant. Genome Announc. 1980;2017:5.

48. Hsu SH, Shen MW, Chen JC, Lur HS, Te LC. The Photosynthetic Bacterium *Rhodopseudomonas palustris* Strain PS3 Exerts Plant Growth-Promoting Effects by Stimulating Nitrogen Uptake and Elevating Auxin Levels in Expanding Leaves. Front Plant Sci. 2021;12:93.

49. Akiba T, Usami R, Horikoshi K. *Rhodopseudomonas* rutila, a New Species of Nonsulfur Purple Photosynthetic Bacteria. Int J Syst Evol Microbiol. 1983;33:551–6.

50. Jiao Y, Kappler A, Croal LR, Newman DK. Isolation and Characterization of a Genetically Tractable Photoautotrophic Fe(II)-Oxidizing Bacterium, *Rhodopseudomonas palustris* Strain TIE-1. Appl Environ Microbiol. 2005;71:4487–96.

51. Adessi A, Spini G, Presta L, Mengoni A, Viti C, Giovannetti L, et al. Draft genome sequence and overview of the purple non sulfur bacterium *Rhodopseudomonas palustris* 42OL. Stand Genomic Sci. 2016;11:24.

52. du Toit JP, Pott RWM. Heat-acclimatised strains of *Rhodopseudomonas palustris* reveal higher temperature optima with concomitantly enhanced biohydrogen production rates. Int J Hydrogen Energy. 2021;46:11564–72.

53. Elbeltagy A, Nishioka K, Suzuki H, Sato T, Sato YI, Morisaki H, et al. Isolation and characterization of endophytic bacteria from wild and traditionally cultivated rice varieties. Soil Sci Plant Nutr. 2000;46:617–29.

54. Yin L, Li X, Liu Y, Zhang D, Zhang S, Luo X. Biodegradation of cypermethrin by *Rhodopseudomonas palustris* GJ-22 isolated from activated sludge. Fresenius Environ Bull. 2012;21:397–405.

55. He C, Keren R, Whittaker ML, Farag IF, Doudna JA, Cate JHD, Banfield JF. Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. Nat Microbiol. 2021;6:354–65.

56. Harwood CS, Gibson J. Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium *Rhodopseudomonas palustris*. Appl Environ Microbiol. 1988;54:712.

57. Bao Z, Sato Y, Kubota M, Ohta H. Isolation and Characterization of Thallium-tolerant Bacteria from Heavy Metal-polluted River Sediment and Non-polluted Soils. Microbes Environ. 2006;21:251–60.

58. Ambler RP, Daniel M, Hermoso J, Meyer TE, Bartsch RG, Kamen MD. Cytochromes C2 sequence variation among the recognised species of purple nonsulphur photosynthetic bacteria. Nat. 1979;278:659–60 1979 2785705.

59. Xing D, Zuo Y, Cheng S, Regan JM, Logan BE. Electricity Generation by *Rhodopseudomonas palustris* DX-1. Environ Sci Technol. 2008;42:4146–51.

60. Imhoff J, Meyer T, Kyndt J. Genomic and genetic sequence information of strains assigned to the genus *Rhodopseudomonas* reveal the great heterogeneity of the group and identify strain *Rhodopseudomonas palustris* DSM 123 T as the authentic type strain of this species. Int J Syst Evol Microbiol. 2020;70:3932–8.

61. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of Escherichia coli: Comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol. 2008;190:6881–93.

62. Harris DF, Yang ZY, Dean DR, Seefeldt LC, Hoffman BM. Kinetic understanding of N_2 reduction versus H_2 evolution at the E_4(4H) Janus state in the three nitrogenases. Biochemistry. 2018;57:5706–14.

63. Mehrabi S, Ekanemesang UM, Aikhionbare FO, Kimbro KS, Bender J. Identification and characterization of Rhodopseudomonas spp., a purple, non-sulfur bacterium from microbial mats. Biomol Eng. 2001;18:49–56.

64. Roberts ME, Stewart PS. Modeling antibiotic tolerance in biofilms by accounting for nutrient limitation. Antimicrob Agents Chemother. 2004;48:48–52.

65. McEwan AG. Photosynthetic electron transport and anaerobic metabolism in purple non-sulfur phototrophic bacteria. Antonie Van Leeuwenhoek. 1994;66:151–64.

66. Bilous PT, Weiner JH. Molecular cloning and expression of the *Escherichia coli* dimethyl sulfoxide reductase operon. J Bacteriol. 1988;170:1511–8.

67. McCrindle SL, Kappler U, McEwan AG. Microbial dimethylsulfoxide and trimethylamine-N-oxide respiration. Adv Microb Physiol. 2005;50:147–98.

68. Valentine-Serano A, Hudspeth MES, Meganathan R. Reduction of N-oxides and sulfoxide by the same terminal reductase in *Proteus mirabilis*. Curr Microbiol. 1991;23:271–6.

69. Borsetti F, Francia F, Turner RJ, Zannoni D. The thiol:disulfide oxidoreductase DsbB mediates the oxidizing effects of the toxic metalloid tellurite (TeO_3^{2-}) on the plasma membrane redox system of the facultative phototroph *Rhodobacter capsulatus*. J Bacteriol. 2007;189:851–9.

70. Koebnik R, Hantke K, Braun V. The TonB-dependent ferrichrome receptor FcuA of *Yersinia enterocolitica*: evidence against a strict co-evolution of receptor structure and substrate specificity. Mol Microbiol. 1993;7:383–93.

71. Killmann H, Benz R, Braun V. Properties of the FhuA channel in the *Escherichia coli* outer membrane after deletion of FhuA portions within and outside the predicted gating loop. J Bacteriol. 1996;178:6913–20.

72. Poole K, Neshat S, Krebes K, Heinrichs DE. Cloning and nucleotide sequence analysis of the ferripyoverdine receptor gene fpvA of *Pseudomonas aeruginosa*. J Bacteriol. 1993;175:4597–604.

73. Patzer SI, Hantke K. The ZnuABC high-affinity zinc uptake system and its regulator Zur in *Escherichia coli*. Mol Microbiol. 1998;28:1199–210.

74. Gaballa A, Wang T, Ye RW, Helmann JD. Functional analysis of the *Bacillus subtilis* Zur regulon. J Bacteriol. 2002;184:6508–14.

75. Gati C, Stetsenko A, Slotboom DJ, Scheres SHW, Guskov A. The structural basis of proton driven zinc transport by ZntB. Nat Commun. 2017;8:1–8.

76. Smith TJ, Murrell JC. Mutagenesis of Soluble Methane Monooxygenase. Methods Enzymol. 2011;495:135–47.

77. Collins CM, D'Orazio SEF. Bacterial ureases: structure, regulation of expression and role in pathogenesis. Mol Microbiol. 1993;9:907–13.

Guardia *et al. BMC Microbiology*   (2022) 22:297

Page 19 of 19

78. Zhao J, Zhu L, Fan C, Wu Y, Xiang S. Structure and function of urea amidolyase. Biosci Rep. 2018;38:BSR20171617.

79. Routh MD, Zalucki Y, Su CC, Long F, Zhang Q, Shafer WM, et al. Efflux pumps of the Resistance-nodulation-division family: A perspective of their structure, function, and regulation in gram-negative bacteria. Adv Enzymol Relat Areas Mol Biol. 2010;77(1):109–46.

80. Paulsen IT, Park JH, Choi PS, Saier MH. A family of Gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from Gram-negative bacteria. FEMS Microbiol Lett. 1997;156:1–8.

81. Mealman TD, Blackburn NJ, McEvoy MM. Metal export by CusCFBA, the periplasmic Cu(I)/Ag(I) transport system of *Escherichia coli*. In: Current Topics in Membranes. Academic Press Inc. 2012. p. 163–96.

82. Choi M, Davidson VL. Cupredoxins - A study of how proteins may evolve to use metals for bioenergetic processes. Metallomics. 2011;3:140–51.

83. Zgurskaya HI, Krishnamoorthy G, Ntreh A, Lu S. Mechanism and function of the outer membrane channel TolC in multidrug resistance and physiology of enterobacteria. Front Microbiol. 2011;2:189.

84. Rowland JL, Niederweis M. A multicopper oxidase is required for copper resistance in *Mycobacterium tuberculosis*. J Bacteriol. 2013;195:3724–33.

85. Lee SM, Grass G, Rensing C, Barrett SR, Yates CJD, Stoyanov JV, et al. The Pco proteins are involved in periplasmic copper handling in *Escherichia coli*. Biochem Biophys Res Commun. 2002;295:616–20.

86. Mills SD, Jasalavich CA, Cooksey DA. A two-component regulatory system required for copper-inducible expression of the copper resistance operon of *Pseudomonas syringae*. J Bacteriol. 1993;175:1656–64.

87. Völlmecke C, Drees SL, Reimann J, Albers SV, Lübben M. The ATPases CopA and CopB both contribute to copper resistance of the thermoacidophilic archaeon *Sulfolobus solfataricus*. Microbiol (United Kingdom). 2012;158:1622–33.

88. Reeve WG, Tiwari RP, Kale NB, Dilworth MJ, Glenn AR. ActP controls copper homeostasis in Rhizobium leguminosarum bv. viciae and Sinorhizobium meliloti preventing low pH-induced copper toxicity. Mol Microbiol. 2002;43:981–91.

89. Chillappagari S, Miethke M, Trip H, Kuipers OP, Marahiel MA. Copper acquisition is mediated by YcnJ and regulated by YcnK and CsoR in *Bacillus subtilis*. J Bacteriol. 2009;191:2362–70.

90. Festa RA, Jones MB, Butler-Wu S, Sinsimer D, Gerads R, Bishai WR, et al. A novel copper-responsive regulon in *Mycobacterium tuberculosis*. Mol Microbiol. 2011;79:133–48.

91. Randall CP, Gupta A, Jackson N, Busse D, O'Neill AJ. Silver resistance in Gram-negative bacteria: A dissection of endogenous and exogenous mechanisms. J Antimicrob Chemother. 2014;70:1037–46.

92. Rensing C, Pribyl T, Nies DH. New functions for the three subunits of the CzcCBA cation-proton antiporter. J Bacteriol. 1997;179:6871–9.

93. Anton A, Große C, Reißmann J, Pribyl T, Nies DH. CzcD is a heavy metal ion transporter involved in regulation of heavy metal resistance in Ralstonia sp. strain CH34. J Bacteriol. 1999;181:6876–81.

94. Pimentel BE, Moreno-Sánchez R, Cervantes C. Efflux of chromate by *Pseudomonas aeruginosa* cells expressing the ChrA protein. FEMS Microbiol Lett. 2002;212:249–54.

95. Aguilar-Barajas E, Jerónimo-Rodríguez P, Ramírez-Díaz MI, Rensing C, Cervantes C. The ChrA homologue from a sulfur-regulated gene cluster in cyanobacterial plasmid pANL confers chromate resistance. World J Microbiol Biotechnol. 2012;28:865–9.

96. Ramírez-Díaz MI, Díaz-Pérez C, Vargas E, Riveros-Rosas H, Campos-García J, Cervantes C. Mechanisms of bacterial resistance to chromium compounds. Biometals. 2008;21:321–32.

97. Boyd ES, Barkay T. The mercury resistance operon: from an origin in a geothermal environment to an efficient detoxification machine. Front Microbiol. 2012;3:349.

98. Kholodii GY, Mindlin SZ, Bass IA, Yurieva OV, Minakhina SV, Nikiforov VG. Four genes, two ends, and a res region are involved in transposition of Tn5053: a paradigm for a novel family of transposons carrying either a mer operon or an integron. Mol Microbiol. 1995;17:1189–200.

99. Yang HC, Rosen BP. New mechanisms of bacterial arsenic resistance. Biomedical Journal. 2016;39:5–13.

100. Fekih IB, Zhang C, Li YP, Zhao Y, Alwathnani HA, Saquib Q, et al. Distribution of arsenic resistance genes in prokaryotes. Front Microbiol. 2018;9:2473.

101. Chen J, Bhattacharjee H, Rosen BP. ArsH is an organoarsenical oxidase that confers resistance to trivalent forms of the herbicide monosodium methylarsenate and the poultry growth promoter roxarsone. Mol Microbiol. 2015;96:1042–52.

102. Qin J, Rosen BP, Zhang Y, Wang G, Franke S, Rensing C. Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. Proc Natl Acad Sci U S A. 2006;103:2075–80.

103. Nikaido H. Multidrug Resistance in Bacteria. Annu Rev Biochem. 2009;78:119.

104. Wallden K, Rivera-Calzada A, Waksman G. Type IV secretion systems: versatility and diversity in function. Cell Microbiol. 2010;12:1203.

105. Moncrief MBC, Maguire ME. Magnesium and the role of mgtC in growth of *Salmonella typhimurium*. Infect Immun. 1998;66:3802–9.

106. Wolfram L, Friedrich B, Eitinger T. The *Alcaligenes eutrophus* protein HoxN mediates nickel transport in *Escherichia coli*. J Bacteriol. 1995;177:1840–3.

107. Baharoglu Z, Bradley AS, Le Masson M, Tsaneva I, Michel B. RuvA mutants that resolve holliday junctions but do not reverse replication forks. PLoS Genet. 2008;4:1000012.

108. Dorrell N, Ahmed AH, Moss SH. Photoreactivation in a phrB mutant of *Escherichia coli* K-12: evidence for the role of a second protein in photorepair. Photochem Photobiol. 1993;58:831–5.

109. Flynn HC, Mc Mahon V, Diaz GC, Demergasso CS, Corbisier P, Meharg AA, et al. Assessment of bioavailable arsenic and copper in soils and sediments from the Antofagasta region of northern Chile. Sci Total Environ. 2002;286:51–9.

Publisher's Note