Heliyon 10 (2024) e27820

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A pan-genomic assessment: Delving into the genome of the marine epiphyte *Bacillus altitudinis* strain 19_A and other very close *Bacillus* strains from multiple environments

Natalia Beatriz Comba-González [a], Diego Chaves-Moreno [b], Johanna Santamaría-Vanegas [a,*], Dolly Montoya-Castaño [c]

[a] *Biological and Environmental Sciences Area, Jorge Tadeo Lozano University, Colombia*
[b] *Microbial Interactions and Processes Research Group, Helmholtz Centre for Infection Research, Braunschweig, Germany*
[c] *Bioprocesses and Bioprospecting Group, Biotechnology Institute, Universidad Nacional de Colombia, Colombia*

ARTICLE INFO

Keywords:
Bacillus altitudinis
Pan-genome
Epiphytic bacteria
Marine macroalgae
Molecular ecology

ABSTRACT

Marine macroalgae are the habitat of epiphytic bacteria and provide several conditions for a beneficial biological interaction to thrive. Although *Bacillus* is one of the most abundant epiphytic genera, genomic information on marine macroalgae-associated *Bacillus* species remains scarce. In this study, we further investigated our previously published genome of the epiphytic strain *Bacillus altitudinis* 19_A to find features that could be translated to potential metabolites produced by this microorganism, as well as genes that play a role in its interaction with its macroalgal host. To achieve this goal, we performed a pan-genome analysis of *Bacillus* sp. and a codon bias assessment, including the genome of the strain *Bacillus altitudinis* 19_A and 29 complete genome sequences of closely related *Bacillus* strains isolated from soil, marine environments, plants, extreme environments, air, and food. This genomic analysis revealed that *Bacillus altitudinis* 19_A possessed unique genes encoding proteins involved in horizontal gene transfer, DNA repair, transcriptional regulation, and bacteriocin biosynthesis. In this comparative analysis, codon bias was not associated with the habitat of the strains studied. Some accessory genes were identified in the *Bacillus altitudinis* 19_A genome that could be related to its epiphytic lifestyle, as well as gene clusters for the biosynthesis of a sporulation-killing factor and a bacteriocin, showing their potential as a source of antimicrobial peptides. Our results provide a comprehensive view of the *Bacillus altitudinis* 19_A genome to understand its adaptation to the marine environment and its potential as a producer of bioactive compounds.

1. Introduction

Macroalgae harbor a diverse group of epiphytic microorganisms [1] that possess a broad repertoire of biochemical traits that enable their algal hosts to absorb nutrients and cope with pathogen attack and abiotic stress [2,3]. The interaction between marine macroalgae and their epiphytic bacteria also helps the former to survive in many different marine habitats characterized by a wide range of physical and chemical conditions not found in terrestrial environments [4,5].

* Corresponding author. Carrera 4 # 22-61, Bogotá, 110311, Colombia.
E-mail address: johanna.santamaria@utadeo.edu.co (J. Santamaría-Vanegas).

https://doi.org/10.1016/j.heliyon.2024.e27820
Received 27 March 2023; Received in revised form 14 February 2024; Accepted 7 March 2024
Available online 20 March 2024

The salient features driving such epiphyte-macroalgal interactions include the production of enzymes to degrade algal-produced sugars [6], the production of secondary metabolites with antibiotic properties [7–9], the synthesis of phytohormones [10] and metabolites that promote motility and chemotaxis and facilitate biofilm formation [11], and the production of siderophores [12]. Consequently, these metabolically diverse bacterial communities [13] are the source of bioactive molecules with unique properties that drive ecological interactions and have biotechnological potential [14].

Microorganisms commonly associated with macroalgae include Actinobacteria, Betaproteobacteria, Deltaproteobacteria, Planctomycetes, Cyanobacteria, and Firmicutes [13]. Within the Firmicutes, *Bacillus* has been reported as one of the most abundant epibionts in several macroalgal species [15], and bioprospecting efforts on them are increasing. Recently, the *Bacillus altitudinis* 19_A strain isolated from the surface of the marine macroalga *Ulva lactuca* from the Colombian Caribbean [16] has been shown to be a source of enzymes and bioactive compounds with potential biotechnological applications, such as lipases, cellulases, and siderophores [16, 17]. Macroalgae-associated *Bacillus* species are economically and ecologically valuable sources, due to their interaction with their host and the full range of potentially bioactive compounds they produce, including bacteriocins, lipopeptides such as surfactins, iturins, and fengycins, and polyketides such as macrolactin, bacillaene, and difficidin. In addition, the identified bioactive compounds have antimicrobial, anticancer, anti-inflammatory, antihypertensive, and antidiabetic properties, as well as applications in heavy metal detoxification and as biocontrol agents [18,19]. However, these microorganisms remain largely unexplored compared to marine macroorganisms such as tunicates, polychaetes, sponges, bryozoans, mollusks, and other organisms [18].

To advance in the understanding of the epiphyte-macroalgal interaction and in the search for bioactive compounds, product of this interaction, we can take advantage of the availability of the sequenced genome of *Bacillus altitudinis* 19_A [17]. Next-generation sequencing technologies are helping to identify the genomic basis of metabolic and ecological traits that promote bacterial survival in different environments [18,20]. For example, genomic data facilitate the prediction of protein-coding genes, which in turn can reveal how an organism is biochemically equipped to interact with its habitat [21]. In addition, data on bacterial specialized metabolites can be recognized by the presence of biosynthetic gene clusters (BGCs) in the genome. For example, in *Bacillus* spp, between 5 and 10% of the genome is involved in antibiotic biosynthesis [18].

Bacillus also attracts attention for its ubiquity and high persistence [22]. Its good colonization capacity allows it not only to inhabit marine environments, but also to be present in soil, water, and air environmental matrices, making it one of the most studied bacteria [19]. The diversity of this genus, its range of adaptations to different environments, and the availability of several sequenced *Bacillus* genomes allow comparative studies of *Bacillus* from different habitats to explore the metabolic pathways and secondary metabolites involved in the interaction mechanisms of *Bacillus* species with their hosts and their survival success in a particular habitat. Genomic sequences allow these studies through comparative genomic analysis, identifying unique genetic mechanisms relevant to ecological adaptation and unique sequences encoding substances with biotechnological potential in a given bacterial species. In this sense, *Bacillus altitudinis* 19_A genes related to its epiphytic lifestyle and unique genomic features (of biotechnological interest) can be identified by analyzing and comparing the genome sequences of closely related bacteria from contrasting ecological niches.

Therefore, the objectives of this work were (i) to identify common and unique genomic features of *Bacillus altitudinis* 19_A compared to related *Bacillus* strains isolated from very different habitats. (ii) we sought to identify features within the accessory and unique genomes of *Bacillus altitudinis* 19_A that may play a role in the epiphyte-algal host interaction and their survival in the marine environment, and (iii) we searched for the genomic features of *Bacillus altitudinis* 19_A that can be translated into potential metabolites produced by this strain.

2. Methodology

2.1. Genome reassembly and annotation

The whole genome sequence of the strain *Bacillus altitudinis* 19_A, previously sequenced and published (by our laboratory), was obtained from DDBJ/ENA/GenBank, accession number NZ_CABEHU000000000.1 [17]. Considering our previous work [17], the assembly was performed without a reference genome (*de novo*) and it was only possible to reach the contig level. Here, we wanted to improve this assembly in order to perform a whole genome analyses. Therefore, we decided to reassemble the genome using SPAdes v3.13.1 [23] and guide the assembly with a closely related reference genome. After comparing average nucleotide identities (ANIs), we selected the genome sequence of the strain *Bacillus altitudinis* SCU11 (NZ_CP038517) as the assembly reference. Subsequently, different k-mer sizes were evaluated to select the best assembly based on Quast v5.0.2 statistics [24], including number of contigs, mean contig length, N50, number of contigs≥1 Kb, and longest contig as the main criteria used. The contigs and the reference genome were evaluated using MeDuSa5 [25] and Ragout [26] to order and orient the assembled genome contigs into scaffolds. Finally, the reassembled genome was examined for synteny using the Mauve aligner [27], and a dot plot (http://dgenies.toulouse.inra.fr/) comparison with the reference genome was generated.

Genome annotation was performed using Prokka v1.12 [28], and scaffolds were submitted to the RAST server (http://rast.nmpdr.org/rast.cgi). Transfer RNA (tRNA) and ribosomal RNA (rRNA) genes were detected using tRNAscan-SE v1.4 [29] and RNAmmer v1.2 [30], respectively.

2.2. Selection of Bacillus genomes for comparative analysis

The refined and reassembled genome of *Bacillus altitudinis* 19_A was compared with other bacterial genomes most closely related to *B. altitudinis*. Evolutionary proximity was assessed using ANI values ranging from 85 to 100% [31]. We selected the genomes closest to

B. altitudinis with the aim of finding genes specific to *B. altitudinis* 19_A that are involved in the epiphytic lifestyle of this strain and in the production of potential metabolites.

ANI values of *Bacillus* genomes were calculated using the EzGenome ANI web tool (http://www.ezbiocloud.net/ezgenome/ani) [32]. Strains with ANIs ≥95% were considered to belong to the same species [33]. In addition, *in silico* predictions of digital DNA-DNA hybridization (dDDH) between the *Bacillus altitudinis* 19_A genome and the genomes of other *Bacillus* species were calculated using the Genome-Genome Distance Calculator (GGDC) 2.0 BLAST+ and recommended formula 2 (https://www.dsmz.de/services/online-tools/genome-to-genome-distance-calculator-ggdc). A 70% dDDH value represents the boundary between different species [31,34].

To find the specific and common functions of *Bacillus altitudinis* 19_A involved in its epiphytic lifestyle, we compared the 19_A genome sequence with those of other *Bacillus* strains of the *Bacillus pumilus* group. We considered the following inclusion criteria for the comparative genomic analysis: (i) the ANI values obtained (in the range of 85%–100%), (ii) the availability of complete genome sequences (either fully closed or high draft sequences), and (iii) the platform used for sequencing (only those sequenced with Illumina technology were retained). Twenty-nine complete genomes representing eight species of the genus *Bacillus* were retrieved from the GenBank database at the National Center for Biotechnology Information (NCBI) (accessed in August 2020, Table 1). Among them, six were from the species *B. altitudinis*, six from *Bacillus safensis*, six from *Bacillus pumilus*, five from *Bacillus stratosphericus*, two from *Bacillus aerophilus*, two from *Bacillus xiamenensis*, one from *Bacillus zhangzhouensis,* and one from *Bacillus australimaris*. These strains were isolated from soils (eight genomes), marine environments (six genomes), plants (five genomes), extreme environments (five genomes), air (three genomes), and food (two genomes).

2.3. Genome comparisons

We performed a comparative analysis including all retrieved *Bacillus* genomes and another set of analyses grouping these genomes according to their corresponding bacterial habitats, namely plants, soil, air, food, marine, and extreme environments.

All genome sequences were re-annotated using Prokka v1.12 [35]. The following tools were used for pan-genome analysis: (i) Roary v3.11.2 [36], (ii) bacterial pan-genome analysis (BPGA) pipeline (v1.3) [37], (iii) panX software [38], and OrthoMCL v2.0.9 pipeline [39]. For these analyses, a minimum identity of 80% was considered for BLASTP and USEARCH. Core genes (genomes = 30), accessory genes (2 ≤ genomes ≤29), and unique genes (genome = 1) were identified.

Core and accessory genes from the respective genomes were submitted to the web-based server BlastKOALA v2.2 (http://www.kegg.jp/blastkoala/) [40] and to the Anvi'o platform (http://merenlab.org/software/anvio/) [41] for functional annotation based on Clusters of Orthologous Groups (COGs). Furthermore, the relative percentages of COG categories were calculated and visualized.

Table 1
Genomic features of *Bacillus altitudinis* 19_A and closely related *Bacillus* species.

Species	Strain	Habitat	Genome size (Mb)	GC %	CDSs	ANI (%)	dDDH (%)	NCBI accession
Bacillus altitudinis	19_A	Marine	3.76	41.28	3994	–	–	NZ_CABEHU000000000
Bacillus altitudinis	SCU11	Extreme/solid waste	3.75	41.40	3781	98.49	86.6	NZ_CP038517
Bacillus altitudinis	W3	Food/raw gallnut honey	3.74	41.40	3780	98.43	86.2	NZ_CP011150
Bacillus altitudinis	P-10	Soil/rice rhizosphere	3.76	41.50	3874	98.42	86.4	NZ_CP024204
Bacillus altitudinis	FD48	Plant/rice phylloplane	3.75	41.20	3859	98.36	85.2	CP025643
Bacillus altitudinis	Cr2-1	Soil/sludge	3.74	41.40	3791	98.34	85.4	NZ_CP031774
Bacillus altitudinis	SGAir0031	Air/tropical air	3.84	41.37	3851	98.02	83.1	NZ_CP022319
Bacillus stratosphericus	M53	Extreme/hot spring	3.69	41.30	3741	98.51	86.8	NZ_LFMM00000000
Bacillus stratosphericus	LK18	Plant	3.73	41.20	3765	98.48	86.2	NZ_LDWI00000000
Bacillus stratosphericus	BT2.2	Plant/dragon fruit tree	3.70	41.20	3743	98.48	86.5	NZ_JABMIW000000000
Bacillus stratosphericus	5Co	Plant	3.62	41.3	3677	98.47	86.6	NZ_MWKO00000000
Bacillus stratosphericus	LAMA 585	Marine/deep sea	3.70	41.19	3808	98.4	85.6	NZ_APAS00000000
Bacillus aerophilus	232	Air/tropical air	3.82	41.40	3889	98.45	86.1	NZ_CP026008
Bacillus aerophilus	C772	Soil/rice field	3.75	41.20	3806	98.10	82.4	NZ_JXRO00000000
Bacillus xiamenensis	HYC-10	Marine/intestinal tract	3.61	41.30	3692	91.49	44.1	NZ_AMSH00000000
Bacillus xiamenensis	VV3	Food/fermented rice	3.63	41.50	3700	91.52	44.3	NZ_CP017786
Bacillus zhangzhouensis	DW5-4	Marine/shrimp farm	3.73	41.40	3808	89.94	39.3	NZ_JOTP00000000
Bacillus safensis	FO-36b	Air/spacecraft facility	3.76	41.7	3800	88.89	36.5	NZ_CP010405
Bacillus safensis	Sami	Plant/*Ficus religiosa*	3.67	41.60	3688	88.85	36.3	CP032830
Bacillus safensis	BRM1	Marine/mangrove soil	3.74	41.80	3773	88.85	36.5	NZ_CP018100
Bacillus safensis	U41	Extreme/lake	3.73	41.5	3780	88.80	36.2	NZ_CP015610
Bacillus safensis	KCTC	Marine/marine sponge	3.97	41.36	4011	88.79	36.5	NZ_CP018197
Bacillus safensis	U17-1	Extreme/lake	3.74	41.5	3780	88.79	36.3	NZ_CP015611
Bacillus pumilus	PDSLzg-1	Soil/oil contaminated	3.71	41.99	3719	89.16	37.2	NZ_CP016784
Bacillus pumilus	ZB201701	Soil/maize rhizosphere	3.64	41.90	3607	89.16	37.2	NZ_CP029464
Bacillus pumilus	SH-B9	Soil/beet rhizosphere	3.87	41.46	3914	88.75	36.2	NZ_CP011007
Bacillus pumilus	145	Extreme/shallow water	4.05	41.16	4125	88.59	35.7	NZ_CP027116
Bacillus pumilus	150a	Soil/sediment top	3.74	41.40	3726	88.44	35.3	NZ_CP027034
Bacillus pumilus	SF-4	Soil	3.77	41.20	3790	88.27	35	NZ_CP047089
Bacillus australimaris	NH71_1	Marine/sediment	3.64	41.30	3667	88.62	35.8	NZ_LGYN00000000

For phylogenetic analysis, multiple sequence alignment was performed using Clustal_W with the retrieved 16S rRNA gene sequence data of the selected strains. The alignment was used to construct a cladogram in MEGA7 [42]. The distances were calculated using a joint maximum likelihood method with 1000 bootstrap replicates.

2.4. Codon bias analysis

We assessed codon bias for all 30 *Bacillus* genomes and in the set of core and accessory genes from each *Bacillus* genome using the program CodonW v1.4.2 (http://www.molbiol.ox.ac.uk/cu). We estimated codon bias using several metrics, including Codon Adaptation Index-CAI, Effective Number of Codons-Nc, GC content at the first, second, and third codon positions (GC1, GC2, and GC3), G or C frequency at the third codon position of synonymous codons (GC3s), and total GC and AT content. CAI values range from 0 (no bias) to 1 (maximum bias), and Nc values range from 20 (maximum bias) to 61 (no bias) [43]. We analyzed the correlations between the following pairs of codon bias estimators: CAI and GC3s, CAI and Nc, and nucleotide composition and Nc. This analysis included Spearman's rank correlation using R statistical software [44].

2.5. Mining of genes related to secondary metabolites in the accessory genome

Secondary metabolite gene clusters and non-ribosomal peptides were identified across all genomes using the online tool antiSMASH [45]. The web-based database BAGEL 3 was used to predict the coding sequences of bacteriocins and post-translationally modified peptides (RiPPs) of interest for their biological functions [46]. In addition, we used the NRPSpredictor2 tool [47] to identify the non-ribosomal peptide synthetases (NRPS) responsible for the synthesis of secondary metabolites. Through the computational workflow integrating the software tools BiG-SCAPE and CORASON [48], we detected biosynthetic gene clusters (BGCs) and compared their genomic context between the *Bacillus altitudinis* 19_A genome and the rest of the genomes included in this study.

3. Results

3.1. Genome reassembly and annotation

The *Bacillus altitudinis* 19_A genome assembly comprises 59 contigs (N50 3,750,214, L50 1) with a length of 3,762,775 bp (~3.76 Mb) and a G + C content of 41.28%. After annotation, a total of 3994 protein-coding genes were predicted. The tools tRNAscan-SE and RNAmmer predicted 47 tRNA- and 11 rRNA-coding genes, respectively. Approximately 23% of the genes potentially encode hypothetical proteins.

The synteny plot in Fig. 1, which contrasts the annotated genome of *Bacillus altitudinis* 19_A with the genome of strain *B. altitudinis* SCU11, showed that most of the genes on the chromosome of these two bacterial strains share the same order, reflecting a high level of synteny between the genomes of the two strains, except for some points of possible deletions or inversions represented by discontinuities in the syntenic line. This finding indicates the close genetic relationship between *Bacillus altitudinis* 19_A and *B. altitudinis* SCU11, since the synteny plots between *Bacillus altitudinis* 19_A and the other *Bacillus* species showed a lower degree of conservation in

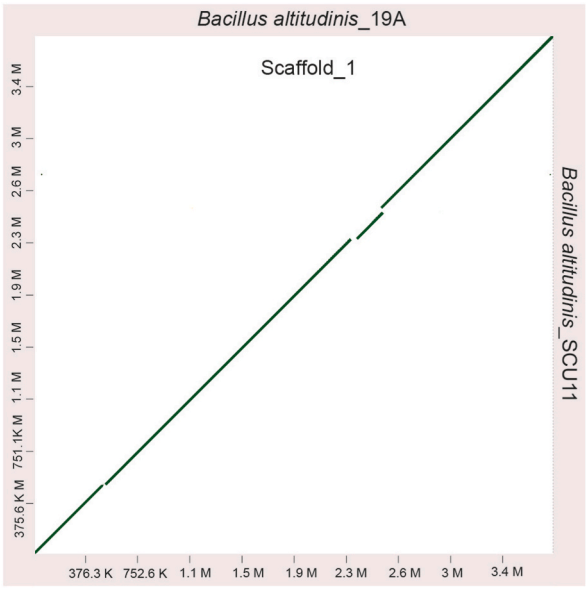

Fig. 1. Whole-genome dot-plot comparison between *Bacillus altitudinis* 19_A and *B. altitudinis* strain SCU11.

terms of gene order and content between these genomes. Rearrangements of large portions of the genomes were also observed.

3.2. Comparative genomics analyses

When comparing all the *Bacillus* genomes examined (Table 1), we observed heterogeneity in genome size, ranging from 3.61 to 4.05 Mbp. The GC content of these genomes varied from 41.10% (*B. altitudinis* strain Cr-2-1) to 41.99% (*B. pumilus* strain PDSLzg-1). The number of CDSs ranged from 3607 (*B. pumilus* strain ZB201701) to 4125 (*B. pumilus* strain 145), suggesting a high diversity in the genetic repertoire of *Bacillus* strains and their genome plasticity.

According to the average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) values, the *Bacillus altitudinis* 19_A genome showed the most similarities with the genomes of the species *B. altitudinis*, *B. stratosphericus,* and *B. aerophilus* (Table 1).

Our pan-genome analysis using Roary, BPGA, the panX software, and the OrthoMCL v2.0.9 pipeline yielded similar results. Comparison of the complete genome sequences of *Bacillus altitudinis* 19_A with the genomes of 29 *Bacillus* strains obtained from NCBI revealed a core genome containing 2708 protein-coding genes. In addition, we identified 3171 accessory genes and 2589 unique genes. Core, accessory, and unique genes accounted for 31.98%, 37.44% and 30.58% of the pan-genome (8468 genes), respectively. On average, 72% of the genes accounted for the core genome of each strain.

Our analysis also revealed a large share of unique genes in *B. pumilis* 145 (9.89%; extreme environment), *B. zhangzhouensis* (6.17%; marine), *B. pumilus* SHB9 (5.09%; soil), *B. safensis* KCTC (4.66%; marine), and *B. xiamenensis* HYC-10 (3.14%; marine), suggesting that these strains may have more genetic characteristics and ecological traits than strains with fewer unique genes (Fu et al., 2021). For *Bacillus altitudinis* 19_A, the analysis identified 47 unique genes (1.23%).

The cumulative curve of genes showed that, while the pan-genome size increases due to the emergence of new strain-specific gene families, its core genome remains stable (Fig. 2). The resulting curve equation is as follows:

$f(x) = 3138.22.x^{0.2842}$. In our study $\gamma = 0.2842$. According to Heap's Law, when $0 < \gamma < 1$, the pan-genome will continue to increase, which indicates that the pan-genome is open. This could mean that *Bacillus* is a genus that can quickly adapt to contrasting conditions as it can colonize any niche [49].

Using the Cluster of Orthologous Groups (COG) database, we classified core, accessory, and unique gene families into super-functional categories. Most of the core and accessory gene families belonged to the superfunctional categories: "Metabolism" (37% and 47%, respectively), "Cellular Processes and Signal Transduction" (23% and 18%, respectively), and "Information Storage and Processing" (22% and 17%, respectively) (Fig. 3).

Unique gene families were underrepresented in the "Metabolism" gene category (21%), and a large proportion of the unique genome fell into the super-functional COG categories "Information Storage and Processing" (36%) and "Cellular Processes and Signaling" (23%). In addition, genes in the "Poorly characterized" super-functional group accounted for 17%, 18%, and 19% of the core, accessory, and unique gene families, respectively (Fig. 3).

To analyze the evolutionary relationship among the 29 *Bacillus* isolates, we constructed a phylogenetic tree based on 16S rRNA gene sequences. In the tree shown in Fig. 4, *Bacillus altitudinis* 19_A belonged to the *Bacillus pumilus* (Bp) group and formed a separate clade, with strain SCU11 being the closest to *B. altitudinis*, *B. stratosphericus,* and *B. aerophilus*.

After grouping *Bacillus* genomes by habitat, we identified 3215 core genes in the genomes of strains isolated from air, 3204 core genes in isolates from food, 3133 core genes in isolates from plants, and 3075 core genes in isolates from extreme environments. We

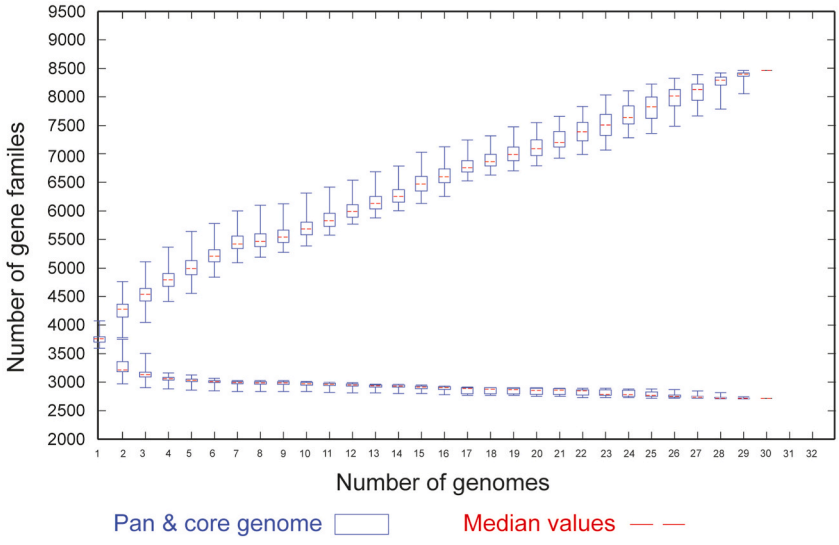

Fig. 2. Accumulation curves showing the number of gene families belonging to the pan and core genomes as calculated through the BPGA tool. The upper box plots represent the pangenome, and the lower box plots depict the core genome.

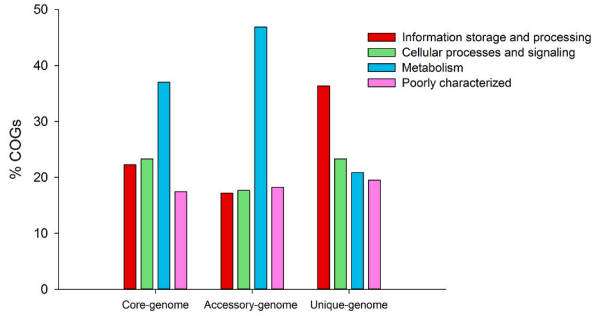

Fig. 3. Comparison of the abundance of COG super-functional categories in the core, accessory, and unique gene families in the 29 assessed *Bacillus* genomes.

also identified 3041 and 2884 core genes in strains isolated from soil and marine environments, respectively. Through comparative analysis, we observed a decrease in the number of accessory genes in the genome groups according to habitat and a 32% difference between the genomes of strains from marine environments and the genomes of strains isolated from air (Fig. 5).

When analyzing the accessory genome, we identified 141 genes shared among habitats and counted a total of 130, 93, and 71 unique accessory genes in genomes isolated from marine environments, soil, and extreme environments, respectively (Fig. 6).

As shown in Table 2, the unique accessory genes mainly correspond to the superfunctional categories: "Metabolism" in air, marine, and extreme environment *Bacillus* strains and "poorly characterized" categories in plant and soil strains. The least represented categories were "Cellular Processes and Signaling" in marine and soil strains; "Information Storage and Processing" in extreme environments and air; in plants, "Information Storage and Processing" and "Cellular Processes and Signaling" registered the same percentage.

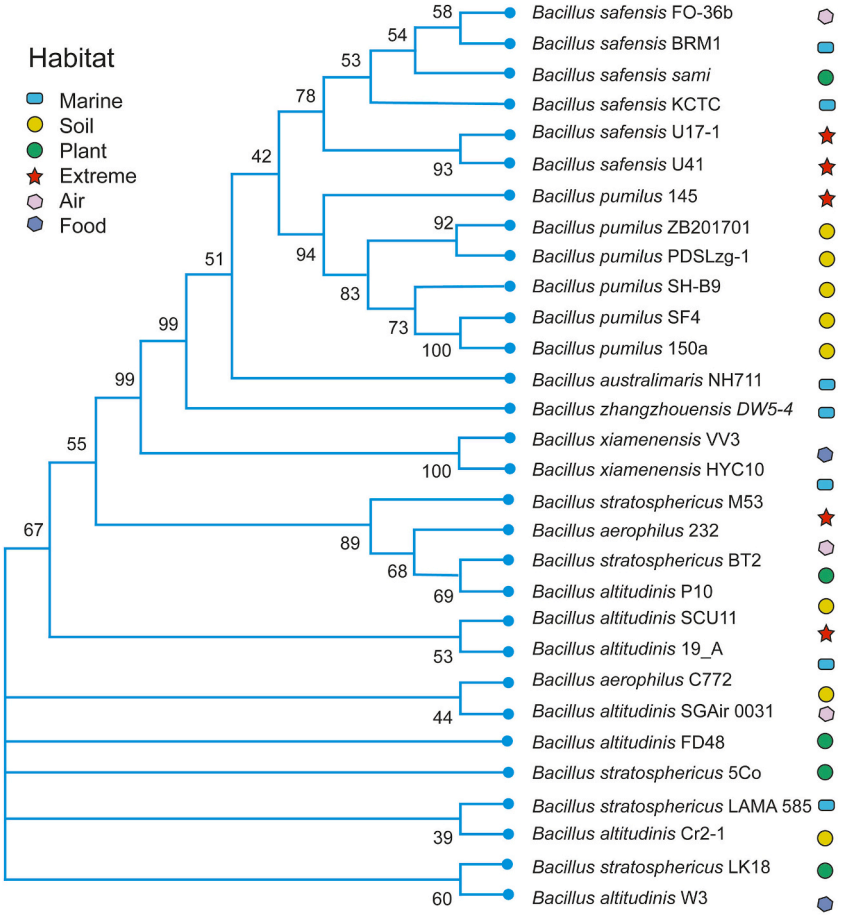

Fig. 4. Cladogram derived from 16S rRNA sequence data. Analyses were conducted in MEGA7, using the Maximum Likelihood method with 1000 bootstrap replicates. Bootstrap values are displayed as percentages on their respective branch nodes.

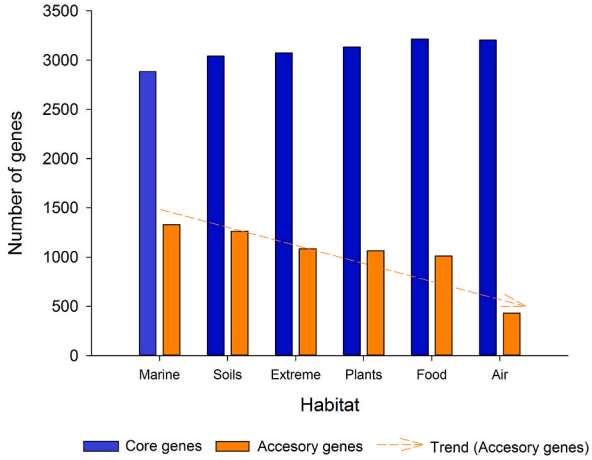

Fig. 5. Core and accessory genes in the genomes of the assessed *Bacillus* strains grouped according to habitat.

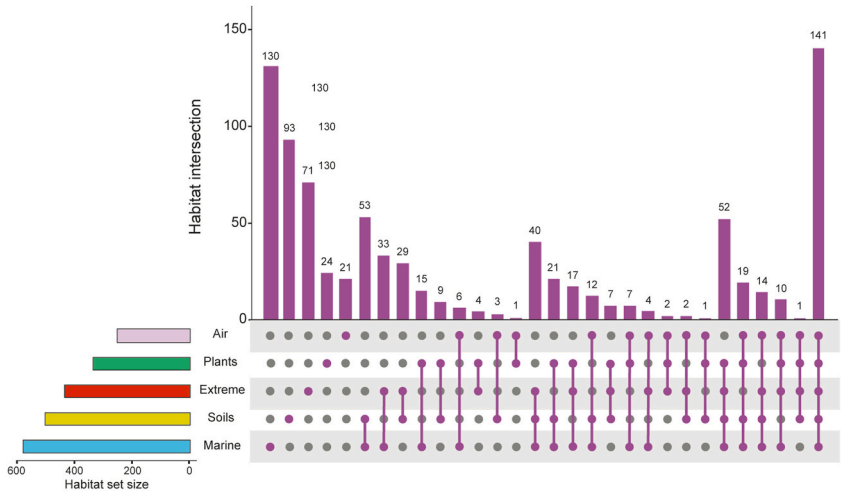

Fig. 6. Number of accessory genes shared among in the assessed *Bacillus* genomes grouped according to habitat.

In addition, eight of the unique accessory genes of strains from marine environments were also found in the *Bacillus altitudinis* 19_A genome. Genomic analysis revealed genes encoding proteins involved in bacterial osmoregulation (ABC-type proline-glycine-betaine transport systems), aromatic compound metabolism (phenylpyruvate tautomerase, phenolic acid decarboxylase), carbohydrate metabolism (ribulose-5-phosphate 4-epimerase, arabinofuranosidase, xylose isomerase, polysaccharide pyruvyltransferase), and sulfur cycle processes (formylglycine producing enzyme).

3.3. Codon bias analysis

We performed a codon usage analysis for all *Bacillus* strain genomes included in this study, in addition to the core and accessory genes belonging to each genome. The results of both approaches were comparable. The effective number of codons (Nc) values ranged

Table 2
Percentage of super-functional categories in unique accessory genes of the grouped genomes according to habitat.

Super-functional category	Habitat				
	Marine environment	Soil	Extreme environment	Plants	Air
Information storage and processing	12.97%	18.94%	19.17%	21.73%	9.52%
Cellular processes and signaling	12.21%	16.84%	21.91%	21.73%	14.28%
Metabolism	41.98%	26.31%	31.50%	26.08%	47.61%
Poorly characterized	32.82%	37.89%	27.39%	30.43%	28.57%

from 54.79 to 57.45, indicating a low codon usage bias (Supplementary Table S1).

Our analysis also calculated the GC3 index, which indicates the frequency of G or C nucleotides at the third position of synonymously variable codons. We used this index as a predictor variable for Nc variation. GC3s values in *Bacillus* strain genomes grouped by habitat ranged from 0.39 to 0.41 (Supplementary Table S1).

Our results also showed low codon adaptation index (CAI) values (Supplementary Table S1). We performed correlation analyses between CAI vs. GC3s and CAI vs. Nc. The obtained CAI and GC3 values were negatively correlated (R^2 -0.014), whereas CAI and Nc were significantly and positively correlated (R^2 0.397).

The obtained values of the codon bias parameters GC3s (Fig. 7a), CAI (Fig. 7b), and Nc (Fig. 7c) differed among the genomes of *Bacillus* strains from marine environments, plants, and soil, whereas among the genomes from extreme environments, air, and food, these parameter values were similar.

3.4. Unique genes in the Bacillus altitudinis 19_A genome

Our comparative genome analysis of *Bacillus altitudinis* 19_A revealed genes unique to this strain and confirmed their identity by BLAST searches. These genes mainly encode hypothetical proteins. In addition, we identified genes encoding proteins involved in horizontal gene transfer, DNA modification and repair processes, and transcriptional regulation. The strain 19_A genome also contains a gene encoding a peptidase involved in bacteriocin biosynthesis and some genes with a role in ribosome assembly and triple helical structure formation (Table 3).

3.5. Mining of genes related to secondary metabolites

Since the differences between the genomes of the *Bacillus* strains were mainly in genes within the metabolism category (according to the present comparative analysis), we screened the genome of *Bacillus altitudinis* 19_A for secondary metabolite genes.

Using the antiSMASH genome analysis tool, we identified six clusters of secondary metabolite genes in the genome of *Bacillus altitudinis* 19_A (Table 4). Three of these genes are located in the core genome: surfactin (lipopeptide), carotenoid (terpene), and aerobactin (type of siderophore). The other three clusters are in the accessory genome: bacteriocin, bacilysin (dipeptide), and sporulation-killing factor (sactipeptide). Our analysis revealed that at least 1.7% of the genome of *Bacillus altitudinis* 19_A contains genes involved in the biosynthesis, regulation, and transport of secondary metabolites.

Two relevant findings were the identification of the sporulation-killing factor operon in only four of the genomes included in our study (19_A, LAMA 585, KCTC, and SCU11) and the bacteriocin gene in the genomes of strains isolated from all the habitats studied (marine and extreme environments, plants, soil, air, and food); however, in the strains from the marine environment, the gene for this metabolite was present only in the *Bacillus altitudinis* 19_A and HYC-10 genomes (Table 4). Some of the characteristics of these two biosynthetic clusters are presented below.

3.5.1. Sporulation killing factor

The sactipeptide-encoding *skf* operon is located at positions 159,272 and 166,619 of the *Bacillus altitudinis* 19_A genome. This operon consists of eight genes (*skfA-H*, Fig. 8a) involved in immunity and the biosynthesis and release of the sporulation killing factor.

Thanks to the analysis performed with the AntiSMASH and BAGEL4 online tools, we identified the *skfB* gene (1235 nt), which encodes an enzyme called sactisynthase, involved in sactipeptide maturation. A blastp search of the amino acid sequence against the NCBI database showed 100% identity and 100% query coverage with the sporulation-killing factor system radical SAM maturase of *Bacillus* and 99.76% identity and 100% query coverage with that of *B. altitudinis*.

We compared the *skfB* gene structure between genomes isolated from marine (LAMA 585 and KCTC) and extreme environments (SCU11) using CORASON software. We identified differences in the *skfB* gene structure registered in strains from marine environments (Fig. 8b); however, when compared with strains from extreme environments, we observed that the gene organization was very similar (Fig. 8c).

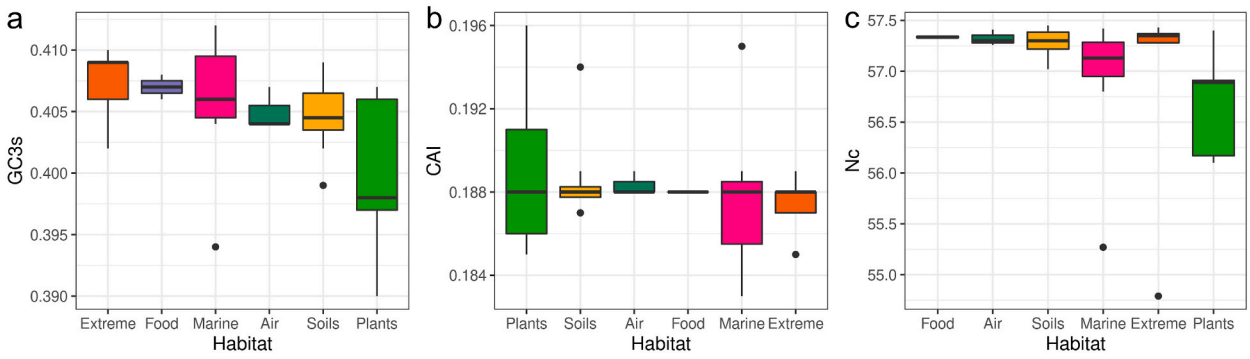

Fig. 7. Boxplots of codon bias parameters, GC3s (a), CAI (b), and Nc (c) among *Bacillus* strains genomes, grouped according to habitat.

Table 3
Unique protein-coding genes identified in the *Bacillus altitudinis* strain 19_A genome and the level of identity of their predicted amino sequence with other *Bacillus* proteins in the NCBI databank, as revealed by the BLASTp tool.

Category	Protein	Function	Percent of identity	Species
Horizontal gene transfer	IS3 family transposase	Insertion sequence	98%	*Bacillus* sp.
	*Sin*I family restriction endonuclease	Horizontal gene transfer modulation	100%	*B. altitudinis*
DNA modification	Exonuclease V	DNA repair	99.6%	*B. altitudinis*
	Restriction endonuclease (HNH family)	DNA modification and rearrangement	93.48%	*B. pumilus*
	DNA cytosine methyltransferase	DNA methylation	100%	*B. altitudinis*
Transcription regulation	AAA family ATPase	Transcription regulator	100%	*B. altitudinis*
	DNA-binding transcriptional regulator (XRE-family)	Gene regulation (DNA damage)	100%	*B. altitudinis*
Metabolite biosynthesis	C39 family peptidase	Bacteriocin biosynthesis	100%	*B. altitudinis*
Other	Collagen-like repeat preface domain-containing protein	Formation of the triple helical structure	95.24%	*B. altitudinis*
	50S ribosome-binding GTPase	50S ribosomal subunit assembly	100%	*B. altitudinis*

Table 4
Comparative analysis of genes involved in the secondary metabolites synthesis identified in the genome of *Bacillus altitudinis* 19_ A with those in the genomes of *Bacillus* strains isolated from marine, extreme, plant, soil, air, and food-associated environments.

Habitat	Strain	Compound					
		Sporulation killing factor	Surfactin	Carotenoid	Aerobactin	Bacteriocin	Bacilysin
Marine environment	19_A	+	+	+	+	+	+
	HYC-10	–	+	+	+	+	–
	BRM1	–	+	+	+	–	+
	DW5-4	–	+	+	+	–	+
	NH71_1	–	+	+	+	–	+
	LAMA 585	+	+	+	+	–	+
	KCTC	+	+	+	+	–	+
Extreme environment	M53	–	+	+	+	+	+
	SCU11	+	+	+	+	+	+
	145	–	+	+	+	+	+
	U17-1	–	+	+	+	+	+
	U41	–	+	+	+	+	+
Plant	5Co	–	–	+	+	–	+
	Sami	–	+	+	+	+	–
	BT2.2	–	+	+	+	+	+
	LK18	–	+	+	+	+	+
	FD48	–	+	+	+	+	+
Soil	ZB201701	–	+	+	+	+	+
	PDSLzg-1	–	+	+	+	–	+
	150[a]	–	+	+	+	–	+
	SF-4	–	+	+	+	+	+
	Cr2-1	–	+	+	+	+	–
	C772	–	+	+	+	+	+
	P-10	–	+	+	+	–	+
	SH-B9	–	+	+	+	+	+
Air	FO-36b	–	+	+	+	+	+
	SGAir0031	–	+	+	+	+	+
	232	–	+	+	+	+	+
Food	VV3	–	+	+	+	+	–
	W3	–	+	+	+	+	+

3.5.2. Bacteriocin

Bacteriocins are antimicrobial peptides of ribosomal synthesis. Analysis using the antiSMASH tool identified a gene encoding a circular bacteriocin in the *Bacillus altitudinis* 19_A genome between positions 2,342,586 and 2,342,912. A blastp search of the amino acid sequence against the NCBI database showed 98% identity and 100% query coverage with the circular bacteriocin, circularin A/uberolysin family, in *Bacillus*.

In addition, we observed differences in the organization of the circular bacteriocin gene between the genomes of *Bacillus altitudinis* 19_A and *B. xiamenensis* strain HYC10 isolated from marine environments (Fig. 9). When comparing the organization of this gene between *B. altitudinis* 19_A and the strains from other environments, no major differences were detected (data not shown).

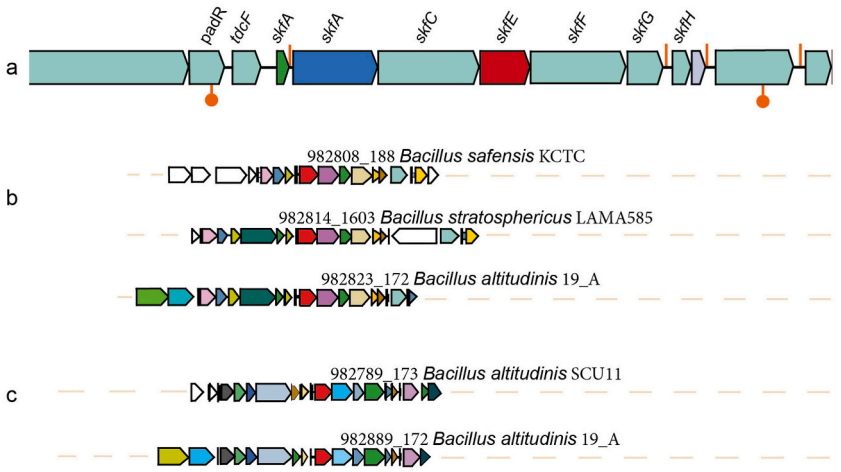

Fig. 8. Sporulation killing factor gene cluster in *Bacillus altitudinis* 19_A. a) Scheme of the *skf* operon. b) Comparison of *skfB* gene organization in genomes of strains isolated from marine environments. c) Comparison of *skfB* gene structure in the genome of a strain isolated from extreme environments.

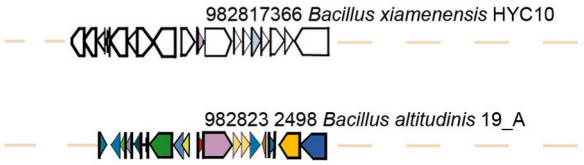

Fig. 9. Comparison between circular bacteriocin sequences of strains isolated from marine environments.

4. Discussion

4.1. Genome similarities of Bacillus altitudinis 19_A with other strains

In our analysis, we compared the genomes of a number of *Bacillus* isolates from different habitats. The ANIs suggested that our focal strain, *Bacillus altitudinis* 19_A, was closely related to strains within the *Bacillus pumilus* (Bp) group from different habitats. Members of the Bp group are often isolated from terrestrial ecosystems and marine environments, such as sediments, seawater, and marine animals [50]. Our phylogenetic reconstruction using 16S rRNA sequences is consistent with previous studies suggesting that some members of the Bp group tend to cluster, as lineage takes precedence over niche effects in defining evolutionary relationships within this group [50, 51].

Our *Bacillus* sp. pan-genome analysis is consistent with other studies using environmental strains [50,52]. The results showed a sequential addition of new genomes to the analysis, resulting in an accumulation curve showing a continuous increase in the number of new genes added to the pan-genome. The slope of the regression curve was 0.28. This value is comparable to the slopes of the regression curve of *B. pumilus*, *B. safensis* and *P. megaterium* (0.16, 0.17 and 0.23), indicating open pangenomes. However, these values show a tendency to saturate the number of new genes with the inclusion of a few new genomes [53].

The pan-genome of the 30 strains studied contains 8468 genes, and the proportions of core genes among the strains ranged from 66.48% to 75.43%, which is comparable to the proportions of core genes observed in the *B. pumilis group* (58.79%–68.78%) [50]. The proportions of accessory (37.44%) and strain-specific genes (30.58%) recorded in our study are also similar to those described for the *B. pumilis* group (35.5% and 39.3%, respectively) [50].

By grouping genomes according to habitat, we observed that the number of accessory genes is comparable across habitats, except for *Bacillus* strains from air. In this habitat, the low number of accessory genes in strains could be explained by barriers to horizontal gene transfer inherent to this environment, such as the physical distance separating microbes, lack of moisture sources for mobility and nutrient transport, or lack of compatibility between strains [52].

Through general genome comparisons between strains grouped by habitat, we found that marine strains had more strain-specific genes than strains from other habitats, consistent with a previous study suggesting that horizontal gene transfer events are more common among marine bacteria [50].

In terms of function, the most abundant accessory genes specific to air, marine, and extreme environments were associated with metabolism. This highlights the importance of this functional group in the response to different habitats [50]. In plant and soil bacterial genomes, most of the accessory genes have unknown functions.

Our analyses identified accessory genes related to the marine environment in the *Bacillus altitudinis* 19_A genome. COG functional classification revealed a gene encoding the protein ProV (ABC-type proline glycine betaine transport system), which is involved in bacterial osmoregulation and may play a role in osmotic stress conditions caused by salinity changes in marine environments [54]. Another aspect related to habitat adaptation in the epiphytic *Bacillus altitudinis* 19_A is nutrient availability; we identified genes involved in carbohydrate metabolism as well as a gene encoding a formylglycine producing enzyme required for the degradation of sulfated polysaccharides from macroalgal cell walls [55]. These genes may facilitate efficient uptake of polysaccharides from the macroalgal surface where this epiphytic strain establishes.

Our analysis also identified genes involved in the metabolism of aromatic compounds (phenylpyruvate tautomerase, phenolic acid decarboxylase), which some marine bacteria use as carbon and energy sources. Previous studies have described marine plankton and lignin derivatives from coastal marshes as precursors of aromatic compounds that are available to bacteria with the metabolic capacity to assimilate them [56].

4.2. Codon bias analysis

To gain insight into the adaptation of strains from different habitats, we analyzed codon bias in the genomes selected for this study. Codon bias refers to the preferential use of certain codons over their synonymous counterparts during the translation of genes into proteins [43]. Some factors such as GC content, lifestyle, generation time, and genome size can lead to codon bias [57].

We used the metrics CAI, Nc, and GC3 index to evaluate codon bias, and we found evidence of more codon bias in *Bacillus* strains from marine environments, plants, and soil than in strains from extreme environments, air, and food, in which the metric values were similar. However, the codon bias index values that we observed suggest low levels of codon bias in the genomes analyzed. Low levels of codon bias are common in species with horizontal gene transfer, which have all the tRNAs necessary to process each DNA fragment [50].

To explore the influence of different codon usage indices on each other, we performed pairwise correlations between index values: CAI vs. GC3s and CAI vs. Nc. CAI and GC3s values were negatively correlated, reflecting the independence of GC content at the third codon positions and gene expression. This is in agreement with a previous study in another bacterial group [58]. CAI and Nc values were positively correlated, probably indicating that other factors besides gene expression influence codon usage in these strains [59].

Studies have shown that codon bias may play an important role in the adaptation of microorganisms to different lifestyles and habitats; furthermore, the genomic plasticity of *Bacillus* indicates their ability to adapt to different niches [50]. However, our results do not show codon bias associated with the habitat of the strains analyzed in this study.

4.3. Specific features in the genome of Bacillus altitudinis 19_A

Comparative analysis of the reassembled genome of *Bacillus altitudinis* 19_A revealed the presence of unique genes encoding proteins involved in horizontal gene transfer, namely members of the IS3 family of transposases and the *Sin*I family of restriction endonucleases. Thanks to gene transfer, the potential for bacterial adaptation relies on the horizontal acquisition of genes transferred between neighboring bacteria [60–62].

Previous studies have shown that some horizontally transferred genes confer key functions that contribute to the adaptability of strains within their habitats [63,64]. Restriction endonucleases are components of prokaryotic DNA restriction modification mechanisms that protect microorganisms from invading foreign DNA, and insertion sequences are mobile genetic elements that are widely distributed among bacteria, involved in genetic information transfer, and influence genome content. These genes could play an important role in the organization, plasticity and evolution of the *Bacillus altitudinis* 19_A genome [65].

Our analysis also identified unique genes in the *Bacillus altitudinis* strain 19_A genome that encode proteins involved in DNA modification, including exonucleases and restriction endonucleases. These play a central role in site-specific recombination events that can result in insertion, deletion, mutation, or correction of DNA sequences [66]. In addition, we identified a gene encoding a DNA cytosine methyltransferase involved in defense against phage DNA [67]. A previous study showed several genes related to DNA recombination and repair in *B. altitudinis* genomes [50]. These features may be relevant for *B. altitudinis* strains such as 19_A to adapt to marine environments where exposure to UV radiation, desiccation, salinity, high temperature, and pollutants are common and can damage bacterial DNA.

The *Bacillus altitudinis* 19_A genome also contains unique genes encoding an AAA family ATPase and DNA-binding transcriptional regulators of the XRE family. ATPases are related to transcription factors, have chaperone roles, or serve as molecular matchmakers in the formation or activities of DNA-protein complexes [68]. Transcription factors that control the timing and level of gene expression, particularly those within the xenobiotic response element (XRE) family of proteins, are involved in metabolic functions such as toxin-antitoxin and nitrogen regulation systems [69].

Another important feature revealed by our analysis of *Bacillus altitudinis* 19_A is a gene encoding a member of the C39 peptidase family; proteins of this family consist mostly of bacteriocin-processing endopeptidases from bacteria. C39 peptidases are ABC transporters involved in bacteriocin maturation and translocation across the cytoplasmic membrane [70].

We also identified genes involved in structural functions, revealing a gene encoding a 50S ribosome-binding GTPase involved in ribosome biogenesis [71] and a gene encoding a collagen-like repeat preface domain-containing protein. These proteins form triple helical structures with protease resistance and melting temperatures similar to animal collagens. However, their function in environmental bacteria is unknown [72].

4.4. Mining of genes related to secondary metabolites

To survive in marine ecosystems, microorganisms have evolved various adaptations, including the synthesis of novel specific biomolecules. Some of these functions are encoded by accessory genes, which are responsible for intraspecific variation and environmental adaptation [50]. In the accessory genome of *Bacillus altitudinis* 19_A, we identified clusters of genes involved in the biosynthesis of sporulation-killing factor and bacteriocin.

The sporulation killing factor is a sactipeptide, a subclass of ribosomally synthesized and post-translationally modified peptides (RiPP) [73]. In *Bacillus* sp. this molecule is important for the sporulation process and nutrient limitation conditions. Its biosynthesis also involves other mechanisms that confer immunity to its toxicity. Previous studies have shown that *Bacillus subitilis* produces a sporulation killing factor to eliminate nonsporulating siblings susceptible to these toxins. This behavior is a form of cannibalism, as the dead cells release nutrients that can be assimilated by the surviving sporulating cells. This phenomenon delays sporulation and allows cells to grow and divide again [74].

The sporulation killing factor could be an adaptive advantage for *Bacillus altitudinis* strain 19_A in natural environments where other non-sporulating microorganisms may be present. Synthesis of this compound could sustain a small percentage of spores and most growing cells. In addition, the toxins produced could also kill other microorganisms and provide nutrients to delay the sporulation process. In the epiphytic lifestyle, expression of this factor could be an adaptive trait because sporulation becomes irreversible after its earliest stage, and delaying spore formation may be beneficial when nutrient deprivation is temporary. This is the case on macroalgal surfaces where high levels of nutrients are available for epiphytic microorganisms [13,74].

To compare the organization of genes in the cluster responsible for the biosynthesis of the sporulation killing factor, we identified the gene *SkfB*, which encodes the enzyme sactisynthase, involved in the post-translational modification of the precursor peptide SkfA [73]. In terms of organization, we observed differences in the *SkfB* gene organization between *Bacillus altitudinis* 19_A and other *Bacillus* strains isolated from marine environments. When we compared the *SkfB* gene structure between *Bacillus altitudinis* 19_A and *Bacillus altitudinis* SCU11, we observed that their gene organizations were very similar. This could indicate a conserved gene organization among strains of the same species regardless of their habitat.

Our analysis also identified a member of the circular bacteriocin gene family in the genome of the strain *Bacillus altitudinis* 19_A. Circular bacteriocins are ribosomally synthesized antimicrobial peptides characterized by an N-to-C-terminal covalent bond that forms a circular structure. The biosynthesis of these structures requires three steps: cleavage of a leader sequence, circularization, and export from the cell. However, many of the mechanisms involved in this process remain to be elucidated [75].

The arrangement of the circular bacteriocin gene in *Bacillus altitudinis* 19_A differed from that of the *B. xiamenensis* strain HYC10 isolated from a marine environment, consistent with studies of circular bacteriocins with low sequence identity [75].

Marine microorganisms have considerable potential for the production of bioactive compounds not found in terrestrial environments [4]. Our results showed that *Bacillus altitudinis* 19_A could be a potential source of bacterial antimicrobial peptides produced by ribosomal synthesis with pH and thermal stability, properties of interest for medical and environmental applications and use as natural preservatives and probiotics in the food industry [76].

5. Conclusion

Although our results do not show codon bias associated with the particular habitat of each studied strain, our comparative analysis highlighted the genomic plasticity of *Bacillus* sp. members and their potential for adaptation to multiple habitats. Our genomic analysis revealed that *Bacillus altitudinis* 19_A possesses specific genes encoding proteins involved in horizontal gene transfer, DNA repair, and transcriptional regulation. In the accessory genome of *Bacillus altitudinis* 19_A, we identified some genes that could be related to its epiphytic lifestyle, as well as clusters for the biosynthesis of secondary metabolites such as sporulation killing factor, surfactin, carotenoid, aerobactin, bacteriocin, and bacilysin. Our study provides a comprehensive genetic landscape of *Bacillus altitudinis* 19_A to thoroughly understand its adaptation to the marine environment and the potential of this epiphytic bacterium as a source of bacterial antimicrobial peptides produced by ribosomal synthesis. The information obtained will serve as a basis for designing future experiments to extract, characterize and evaluate the bioactive potential of the identified metabolites.

Data availability statement

The authors confirm that the GeneBank accession numbers of the data supporting the findings of this study are available within the article.

CRediT authorship contribution statement

Natalia Beatriz Comba-González: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Diego Chaves-Moreno:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Formal analysis. **Johanna Santamaría-Vanegas:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Dolly Montoya Castaño:** Validation.

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.

Acknowledgments

Authors gratefully thank the financial support provided by "Fondo Nacional de Financiamiento para la Ciencia, la Tecnología, y la Innovación -Fondo Francisco José de Caldas - Minciencias" and Universidad de Bogotá Jorge Tadeo Lozano through the grant No. 848-2019. Minciencias had no involvement in the manuscript authorship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27820.

References

[1] L. Van der Loos, B. Eriksson, J. Salles, The macroalgal holobiont in a changing sea, Trends Microbiol. 27 (2019) 635–650, https://doi.org/10.1016/j.tim.2019.03.002.
[2] Y. Xia, S. DeBolt, J. Dreyer, D. Scott, M.A. Williams, Characterization of culturable bacterial endophytes and their capacity to promote plant growth from plants grown using organic or conventional practices, Front. Plant Sci. 6 (2015) 490, https://doi.org/10.3389/fpls.2015.00490.
[3] H. Yao, X. Sun, C. He, P. Maitra, X.C. Li, L.D. Guo, Phyllosphere epiphytic and endophytic fungal community and network structures in a tropical mangrove ecosystem, Microbiome 7 (2019) 57, https://doi.org/10.1186/s40168-019-0671-0.
[4] F. Ameen, S. AlNadhari, A. Al-Homaidan, Marine microorganisms as an untapped source of bioactive compounds, Saudi J. Biol. Sci. 28 (2021) 1, https://doi.org/10.1016/j.sjbs.2020.09.052.
[5] A. Rotter, M. Barbier, F. Bertoni, A. Bones, M. Cancela, J. Carlsson, M. Carvalho, M. Cegłowska, J. Chirivella, M. Conk, M. Cueto, T. Dailianis, I. Deniz, A. Díaz, D. Drakulovic, A. Dubnika, C. Edwards, H. Einarsson, A. Erdoğan, .M. Vasquez, The essentials of marine Biotechnology, Front. Mar. Sci. 8 (2021) 629629, https://doi.org/10.3389/fmars.2021.629629.
[6] M. Barbato, V. Vacchini, A.H. Engelen, G. Patania, F. Mapelli, S. Borin, E. Crotti, What lies on the macroalgal surface: diversity of polysaccharide degraders in culturable epiphytic bacteria, Amb. Express 12 (2022) 1, https://doi.org/10.1186/s13568-022-01440-8.
[7] F.R. Goecke, A. Labes, J. Wiese, J.F. Imhoff, Chemical interactions between marine macroalgae and bacteria, Mar. Ecol. Prog. 409 (2010) 267–299, https://doi.org/10.3354/meps08607.
[8] B. Ortega-Morales, M. Chan, E. Miranda, M. Fardeau, J. Carrero, T. Stein, Antifouling activity of sessile bacilli derived from marine surfaces, J. Ind. Microbiol. 35 (2008) 9–15, https://doi.org/10.1007/s10295-007-0260-2.
[9] D. Velupillaimani, A. Muthaiyan, Potential of *Bacillus subtilis* from marine environment to degrade aromatic hydrocarbons, J. Environ. Sustain. 2 (2019) 381–389, https://doi.org/10.1007/s42398-019-00080-2.
[10] J.F. Ulrich, M.S. Gräfe, S. Dhiman, P. Wienecke, H. Arndt, T. Wichard, Thallusin quantification in marine bacteria and algae cultures, Mar. Drugs 20 (2022) 11, https://doi.org/10.3390/md20110690.
[11] R. Colin, B. Ni, L. Laganenka, V. Sourjik, Multiple functions of flagellar motility and chemotaxis in bacterial physiology, FEMS Microbiol. Rev. 45 (2021) 6, https://doi.org/10.1093/femsre/fuab038.
[12] L. Zoccarato, D. Sher, T. Miki, D. Segrè, H. Grossart, A comparative whole-genome approach identifies bacterial traits for marine microbial interactions, Commun. Biol. 5 (2022) 1, https://doi.org/10.1038/s42003-022-03184-4.
[13] S. Egan, T. Harder, C. Burke, P. Steinberg, S. Kjelleberg, T. Thomas, The seaweed holobiont: understanding seaweed-bacteria interactions, FEMS Microbiol. Rev. 37 (2012) 462–476, https://doi.org/10.1111/1574-6976.12011.
[14] A. Francis, K. Chakraborty, Marine macroalga-associated heterotroph *Bacillus velezensis* as prospective therapeutic agent, Arch. Microbiol. 203 (2021) 1671–1682, https://doi.org/10.1007/s00203-020-02169-3.
[15] P. Kumar, A. Verma, S. Sundharam, A. Ojha, S. Krishnamurthi, Exploring diversity and polymer degrading potential of epiphytic bacteria isolated from marine macroalgae, Microorganisms 10 (2022) 2513, https://doi.org/10.3390/microorganisms10122513.
[16] N. Comba, M.L. Ramírez, L. López, D. Montoya, Production of enzymes and siderophores by epiphytic bacteria isolated from the marine macroalga *Ulva lactuca*, Aquat. Biol. 27 (2018) 107–118, https://doi.org/10.3354/ab00700.
[17] N. Comba, D. Montoya, J.S. Montaña, Genome sequence of the epiphytic bacteria *Bacillus altitudinis* strain 19_A, isolated from the marine macroalga *Ulva lactuca*, Biotechnol. Rep. 30 (2021) e00634, https://doi.org/10.1016/j.btre.2021.e00634.
[18] K. Chakraborty, V. Kizhakkepatt, M. Joy, Polyketide-derived macrobrevins from marine macroalga-associated *Bacillus amyloliquefaciens* as promising antibacterial agents against pathogens causing nosocomial infections, Phytochemistry (Elsevier) 193 (2022) 112983, https://doi.org/10.1016/j.phytochem.2021.112983.
[19] C. Varghese, K. Chakraborty, S. Asharaf, Pharmacological potential of seaweed-associated heterotrophic bacterium *Bacillus atrophaeus*, Arch. Microbiol. 205 (2023) 6, https://doi.org/10.1007/s00203-022-03338-2.
[20] C. Li, P. Cheng, L. Zheng, Y. Li, Y. Chen, S. Wen, G. Yu, Comparative genomic analysis of two banana *Fusarium* wilt biocontrol endophytes *Bacillus subtilis* R31 and TR21 provides insights into their differences on phytobeneficial trait, Genomics 113 (2021) 900–909, https://doi.org/10.1016/j.ygeno.2021.02.006.
[21] V. Kizhakkepatt, K. Chakraborty, Marine macroalgae-associated heterotrophic Firmicutes and Gamma-proteobacteria: prospective anti-infective agents against multidrug resistant pathogens Arch, Microbiol. 202 (2020) 905–920, https://doi.org/10.1007/s00203-019-01800-2.
[22] A. Saxena, M. Kumar, H. Chakdar, N. Anuroopa, D. Bagyaraj, *Bacillus* species in soil as a natural resource for plant health and nutrition, J. Appl. Microbiol. 128 (2019) 1583–1594, https://doi.org/10.1007/s10295-007-0260-210.1111/jam.14506.
[23] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V. Lesin, S. Nikolenko, S. Pham, A. Prjibelski, A. Pyshkin, A. Sirotkin, N. Vyanni, G. Tesler, M. Alekseyev, P. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, https://doi.org/10.1089/cmb.2012.0021.
[24] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics 29 (2013) 1072–1075, https://doi.org/10.1093/bioinformatics/btt086.
[25] E. Bosi, B. Donati, M. Galardini, S. Brunetti, M. Sagot, P. Lió, P. Crescenzi, R. Fani, M. Fondi, MeDUSA: a multi-draft based scaffolder, Bioinformation 31 (2015) 2443–2451, https://doi.org/10.1093/bioinformatics/btv171.

[26] M. Kolmogorov, B. Raney, B. Paten, S. Pham, Ragout-a reference-assisted assembly tool for bacterial genomes, Bioinformatics 30 (2014) 12, https://doi.org/10.1093/bioinformatics/btu280.

[27] A.C. Darling, B. Mau, F.R. Blattner, N.T. Perna, Mauve: multiple alignment of conserved genomic sequence with rearrangements, Genome Res. 14 (2014) 1394–1403, https://doi.org/10.1101/gr.2289704.

[28] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics 30 (2014) 14, https://doi.org/10.1093/bioinformatics/btu153.

[29] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964, https://doi.org/10.1093/nar/25.5.955.

[30] K. Lagesen, P. Hallin, E.A. Rødland, H. Stærfeldt, T. Rognes, D. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res. 35 (2007) 3100–3108, https://doi.org/10.1093/nar/gkm160.

[31] J. Goris, K.T. Konstantinidis, J.A. Klappenbach, T. Coenye, P. Vandamme, J.M. Tiedje, DNA-DNA hybridization values and their relationship to whole-genome sequence similarities, Int. J. Syst. Evol. Microbiol. 57 (2007) 81–91, https://doi.org/10.1099/ijs.0.64483-0.

[32] I. Lee, Y.O. Kim, S.C. Park, J. Chun, OrthoANI: an improved algorithm and software for calculating average nucleotide identity, Int. J. Syst. Evol. Microbiol. 66 (2016) 2, https://doi.org/10.1099/ijsem.0.000760.

[33] C. Jain, L. Rodríguez, A. Phillippy, K. Konstantinidis, S. Aluru, High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries, Nat. Commun. 9 (2018) 5114, https://doi.org/10.1038/s41467-018-07641-9.

[34] J.P. Meier-Kolthoff, A.F. Auch, H.P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, BMC Bioinf. 14 (2013) 60, https://doi.org/10.1186/1471-2105-14-60.

[35] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics 30 (2014) 14, https://doi.org/10.1093/bioinformatics/btu153.

[36] A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T. Holden, M. Fookes, D. Falush, J.A. Keane, J. Parkhill, Roary: rapid large-scale prokaryote pan genome analysis, Bioinformatics 31 (2015) 22, https://doi.org/10.1093/bioinformatics/btv421.

[37] N.M. Chaudhari, V.K. Gupta, C. Dutta, BPGA-an ultra-fast pan-genome analysis pipeline, Sci. Rep. 6 (2016) 24373, https://doi.org/10.1038/srep24373.

[38] W. Ding, F. Baumdicker, R. Neher, panX: pan-genome analysis and exploration, Nucleic Acids Res. 46 (2019) 1–12, https://doi.org/10.1093/nar/gkx977.

[39] S. Fischer, B.P. Brunk, F. Chen, X. Gao, O.S. Harb, J.B. Iodice, D. Shanmugam, D. Roos, C. Stoeckert, Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups, Curr. Protoc, Bioinformatics 12 (2011) 11–19, https://doi.org/10.1002/0471250953.bi0612s35.

[40] M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences, J. Mol. Biol. 428 (2016) 4, https://doi.org/10.1016/j.jmb.2015.11.006.

[41] A, M. Eren, Ö.C. Esen, C. Quince, J.H. Vineis, H.G. Morrison, M.L. Sogin, T.O. Delmont, Anvi'o: an advanced analysis and visualization platform for omics data, PeerJ 3 (2015) e1319, https://doi.org/10.7717/peerj.1319.

[42] S. Kumar, G. Stecher, K. Tamur, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874, https://doi.org/10.1093/molbev/msw054.

[43] R. Prabha, D. Singh, S. Sinha, K. Ahmad, A. Rai, Genome-wide comparative analysis of codon usage bias and codon context patterns among cyanobacterial genomes, Mar. Genomics 32 (2017) 31–39, https://doi.org/10.1016/j.margen.2016.10.001.

[44] R Development Core Team, R: a Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, 2011. www.R-project.org/. (Accessed 20 June 2021).

[45] T. Weber, K. Blin, S. Duddela, D. Krug, H.U. Kim, R. Bruccoleri, S.Y. Lee, M.A. Fischbach, R. Müller, W. Wohlleben, R. Breitling, E. Takano, M.H. Medema, antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters, Nucleic Acids Res. 43 (2015) W13, https://doi.org/10.1093/nar/gkv437.

[46] A. Heel, A. Jong, M. Montalbán-López, J. Kok, O. Kuipers, BAGEL3: automated identification of genes encoding bacteriocins and (non-) bactericidal posttranslationally modified peptides, Nucleic Acids Res. 41 (2013) W448–W453, https://doi.org/10.1093/nar/gkt391.

[47] M. Röttig, M.H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher, NRPSpredictor2 - a web server for predicting NRPS adenylation domain specificity, Nucleic Acids Res. 39 (2011) W362–W367, https://doi.org/10.1093/nar/gkr323.

[48] J. Navarro, N. Selem, M. Mullowney, S. Kautsar, J. Tryon, E. Parkinson, E. De Los Santos, M. Yeong, P. Cruz, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez, L. Dias, A. Goering, R. Thomson, W. Metcalf, N. Kelleher, F. Barona, M. Medema, A computational framework to explore large-scale biosynthetic diversity, Nat. Chem. Biol. 16 (2020) 60–68, https://doi.org/10.1038/s41589-019-0400-9.

[49] H. Tettelin, D. Riley, C. Cattuto, D. Medini, Comparative genomics: the bacterial pan-genome, Curr. Opin. Microbiol. 11 (2008) 5, https://doi.org/10.1016/j.mib.2008.09.006.

[50] X. Fu, L. Gong, Y. Liu, Q. Lai, G. Li, Z. Shao, *Bacillus pumilus* group comparative genomics: toward pangenome features, diversity, and marine environmental adaptation, Front. Microbiol. 12 (2021) 571212, https://doi.org/10.3389/fmicb.2021.571212.

[51] Y. Liu, Q. Lai, C. Dong, F. Sun, L. Wang, G. Li, Z. Saho, Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis, PLoS One 8 (2013) 11, https://doi.org/10.1371/journal.pone.0080097.

[52] R. Blaustein, A. McFarland, S. Ben, A. Lopez, S. Castro, E. Hartmann, Pangenomic approach to understanding microbial adaptations within a model-built environment, the International Space Station, relative to human hosts and soil, mSystems 4 (2019) 1, https://doi.org/10.1128/mSystems.00281-18.

[53] E. Bach, C. Pinto, L. Alves, I. Pereira, Pangenome analyses of *Bacillus pumilus, Bacillus safensis*, and *Priestia megaterium* exploring the plant-associated features of bacilli strains isolated from canola, Mol. Genet. Genom. 297 (2022) 1063–1079, https://doi.org/10.1007/s00438-022-01907-0.

[54] A. D'ors, M. Bartolomé, S. Sánchez, Repercussions of salinity changes and osmotic stress in marine phytoplankton species, Estuar. Coast Shelf Sci. 175 (2016) 169–175, https://doi.org/10.1016/j.ecss.2016.04.004.

[55] W. Helbert, Marine polysaccharide sulfatases, Front. Mar. Sci. 4 (2017) 6, https://doi.org/10.3389/fmars.2017.00006.

[56] M. Moran, R. Belas, M. Schell, J. González, F. Sun, S. Sun, B. Binder, J. Edmonds, W. Ye, B. Orcutt, E. Howard, C. Meile, W. Palefsky, A. Goesmann, Q. Ren, I. Paulsen, L. Ulrich, L. Thompson, E. Saunders A. Buchan, Ecological genomics of marine Roseobacters, Appl. Environ. Microbiol. 73 (2007) 14, https://doi.org/10.1128/AEM.02580-06.

[57] D. Lal, M. Verma, S. Behura, R. Lal, Codon usage bias in phylum Actinobacteria: relevance to environmental adaptation and host pathogenicity, Res. Microbiol. 167 (2016) 669–677, https://doi.org/10.1016/j.resmic.2016.06.003.

[58] Y. Mondal, S. Sur, A. Bothra, A. Sen, Comparative analysis of codon usage patterns and identification of predicted highly expressed genes in five *Salmonella* genomes, Indian J. Med. Microbiol. 26 (2008) 4, https://doi.org/10.4103/0255-0857.43558.

[59] T. Yu, J. Li, Y. Yang, Q. Liu, B. Chen, F. Zhao, Q. Bao, J. Wu, Codon usage patterns and adaptive evolution of marine unicellular cyanobacteria *Synechococcus* and *Prochlorococcus*, Mol. Phylogenet. Evol. 62 (2012) 206–213, https://doi.org/10.1016/j.ympev.2011.09.013.

[60] A. Kouzuma, K. Watanabe, Exploring the potential of algae/bacteria interactions, Curr. Opin. Biotechnol. 33 (2015) 125–129, https://doi.org/10.1016/j.copbio.2015.02.007.

[61] J. Vandecraen, M. Chandler, A. Aertsen, R. Van Houdt, The impact of insertion sequences on bacterial genome plasticity and adaptability, Crit. Rev. Microbiol. 43 (2017) 6, https://doi.org/10.1080/1040841X.2017.130366.

[62] I. Mruk, T. Kaczorowski, A. Witczak, Natural tuning of restriction endonuclease synthesis by a cluster of rare arginine codons, Sci. Rep. 9 (2019) 5808, https://doi.org/10.1038/s41598-019-42311-w.

[63] A. Kouzuma, K. Watanabe, Exploring the potential of algae/bacteria interactions, Curr. Opin. Biotechnol. 33 (2015) 125–129, https://doi.org/10.1016/j.copbio.2015.02.007.

[64] E. Evanovich, P. Mattos, J. Guerreiro, Comparative genomic analysis of *Lactobacillus plantarum*: an overview, Int J Genomics. (2019) 4973214, https://doi.org/10.1155/2019/4973214.

[65] R. Tobes, E. Pareja, Bacterial repetitive extragenic palindromic sequences are DNA targets for Insertion Sequence elements, BMC Genom. 7 (2006) 62, https://doi.org/10.1186/1471-2164-7-62.

[66] B. Stoddard, Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification, Structure 9 (2011) 1–9, https://doi.org/10.1016/j.str.2010.12.003.
[67] S. Adhikari, P. Curtis, DNA methyltransferases and epigenetic regulation in bacteria, FEMS Microbiol. Rev. 40 (2016) 5, https://doi.org/10.1093/femsre/fuw023.
[68] A. Neuwald, L. Aravind, J. Spouge, E. Koonin, AAA+: a class of chaperone-like ATPases associated with assembly, operation, and disassembly of protein complexes, Genome Res. 9 (1999) 1, https://doi.org/10.1101/gr.9.1.27.
[69] H. Lu, L. Wang, S. Li, C. Pan, K. Cheng, Y. Luo, H. Xu, B. Tian, Y. Zhao, Y. Hua, Structure and DNA damage-dependent derepression mechanism for the XRE family member DG-DdrO, Nucleic Acids Res. 47 (2019) 18, https://doi.org/10.1093/nar/gkz720.
[70] K. Beis, S. Rebuffat, Multifaceted ABC transporters associated to microcin and bacteriocin export, Res. Microbiol. 170 (2019) 399–406, https://doi.org/10.1016/j.resmic.2019.07.002.
[71] L. Schaefer, W. Uicker, C. Wicker-Planquart, A. Foucher, J. Jault, R. Britton, Multiple GTPases participate in the assembly of the large ribosomal subunit in Bacillus subtilis, J. Bacteriol. 188 (2006) 23, https://doi.org/10.1128/JB.01213-06.
[72] Z. Yu, B. An, J. Ramshaw, B. Brodsky, Bacterial collagen-like proteins that form triple-helical structures, J. Struct. Biol. 186 (3) (2014) 451–461, https://doi.org/10.1016/j.jsb.2014.01.003.
[73] T. Grell, W. Kincannon, N. Bruender, E. Blaesi, C. Krebs, V. Bandarian, C. Drennan, Structural and spectroscopic analyses of the sporulation killing factor biosynthetic enzyme SkfB, a bacterial AdoMet radical sactisynthase, J. Biol. Chem. 293 (2018) 45, https://doi.org/10.1074/jbc.RA118.005369.
[74] J. González, Cannibalism: a social behavior in sporulating *Bacillus subtilis*, FEMS Microbiol. Rev. 35 (2011) 415–424, https://doi.org/10.1111/j.1574-6976.2010.00253.x.
[75] C. Gabrielsen, D. Brede, I. Nes, D. Diep, Circular bacteriocins: biosynthesis and mode of action, Appl. Environ. Microbiol. 80 (2014) 22, https://doi.org/10.1128/AEM.02284-14.
[76] H. Abriouel, C. Franz, N. Omar, A. Gálvez, Diversity and applications of *Bacillus* bacteriocins, FEMS Micriobiol Rev. 35 (2011) 201–232, https://doi.org/10.1111/j.1574-6976.2010.00244.x.