RSC Advances

PAPER

View Article Online
View Journal | View Issue

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Check for updates

Cite this: *RSC Adv.*, 2020, **10**, 14606

The use of principle component analysis and MALDI-TOF MS for the differentiation of mineral forming *Virgibacillus* and *Bacillus* species isolated from sabkhas

Rim Abdel Samad, Zulfa Al Disi, Mohammad Yousaf Mohammad Ashfaq, Sara Mohiddin Wahib and Nabil Zouari *

Occurrence of mineral forming and other bacteria in mats is well demonstrated. However, their high diversity shown by ribotyping has not been explained, although it could explain the diversity of formed minerals. Common biomarkers as well as phylogenic relationships are useful tools for clustering the isolates and the prediction of their potential role in the natural niche. In this study, a combination of MALDI-TOF MS with PCA was shown to be a powerful tool to categorize 35 mineral forming bacterial isolates isolated from Dohat Faishakh sabkha, northwest of Qatar (23 from decaying mats and 12 from living ones). The 23 strains from decaying mats belong to the *Virgibacillus* genus as identified by ribotyping and are shown to be highly involved in the formation of protodolomite and a diversity of minerals. They were used as internal references for the categorization of sabkha bacteria. Combination of the isolation of bacteria on selective mineral forming media, their MALDI TOF MS protein profiling and PCA analysis established their relationship in a phylloproteomic dendrogram based on protein biomarkers including *m*/*z* 4905, 3265, 5240, 6430, 7765, and 9815. PCA analysis clustered the studied isolates into 3 major clusters, showing strong correspondence to the 3 phylloproteomic groups that were established by the dendrogram. Both clustering analysis means have evidently demonstrated a relationship between known *Virgibacillus* strains and other related bacteria based on profiling of their synthesized proteins. Thus, larger populations of bacteria in mats can be easily screened for their potential to exhibit certain activities, which is of ecological, environmental and biotechnological significance.

Received 8th February 2020
Accepted 1st April 2020

DOI: 10.1039/d0ra01229g

rsc.li/rsc-advances

1 Introduction

Gulf countries, including Qatar, belong to the most arid coastal ecosystems of the world. The western coast of the Qatari peninsula known as sabkha is enriched with carbonate minerals formed during the Holocene, 4000–6000 years ago.[1] The Inland sabkhas and the coastal sabkhas in the Qatari peninsula are characterized by highly saline water.[2] Dohat Faishakh sabkha is a flat supratidal expanse. It passes laterally through algal mats near the intertidal region within the shallow marine area in the west of Qatar.[3] Since 1965, this sabkha is being recognized as one of the few places on Earth where dolomite forms at ambient temperature.[4] Dolomite is a common ancient carbonate rock. However, the rare occurrence of dolomite in recent environments and failure to synthesize it in laboratory at low temperature and pressure lead to an enigma frequently highlighted in literature as "dolomite problem".[5–8] Numerous recent studies have evidenced the role of microorganisms in the biomineralization process.[9–12] Moreover, it was shown that many microorganisms such as sulfur reducers, methanogens, phototrophs and heterotrophic aerobes have the ability to mediate dolomite formation.[13]

In Dohat Faishakh sabkha, it was shown that bacteria, especially those belonging to the *Virgibacillus* genus were able to mediate mineralization process.[14] Indeed, in the reported works, the identification of these isolates was routinely performed by ribotyping; the 16s rRNA gene sequencing technique.[15] These identified *Virgibacillus* strains demonstrated variable capabilities to mediate carbonate minerals with various magnesium content, including high magnesium calcites that are considered as potential precursors for dolomite formation.[14] Moreover, one of the most proposed hypotheses of biomineralization is attributed to exo-polymeric substances (EPS) which are synthesized by bacteria as a response to high temperature, high salinity and other stressors.[16–18] *Virgibacillus* strains from Dohat Faishakh were shown to respond differently to such stressors.[19] The

Department of Biological and Environmental Sciences, College of Arts and Sciences, Qatar University, P. O. B 2713, Doha, Qatar. E-mail: Nabil.Zouari@qu.edu.qa; Tel: +974-4403-4559

 This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

compared amplified 16s rRNA sequences was insufficient to explain the diversity of *Virgibacillus* in biomineralization potential. However, the relationship between the possible diversity of function and diverse capacities to form minerals in terms of magnesium incorporation into the carbonate minerals is not well elucidated. This correlation between the diversity in *Virgibacillus* strains and their EPS might be established through clustering these isolates based on their respective protein profiles, if appropriate cultural conditions are used for examining the biomineralization potentials. Indeed, a correlation between mineral formation, magnesium incorporation and composition of exopolymeric substances was recently reported.[19] It is now important to establish the relationship between the diversity of bacterial isolates, even at the level of species, their response to different stressors *via* EPS and minerals' diversity. Since the separate study of each of them in the huge bacterial population in the environment cannot be contemplated, a fast and easy prediction of such relationship through bacterial clustering is proposed in the current work. The matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF-MS) technique is suitable for the purpose. It is utilized for the identification and differentiation of microorganisms.[20–23] It mainly relies on protein profiles to identify the isolates of specific genera, species and subspecies.[24] Expressed genes at a cultural condition could also be profiled, leading to categorization of the corresponding strains.[13]

MALDI-TOF-MS technique is rapid and reliable.[25] However, the identification of microorganisms is restricted to available protein profiles in databases, which is not always the case.[26] If protein biomarkers among a bacterial genus or species are known, they can pave the way for the identification of bacterial isolates by proteins matching.

In the current study, a list of 23 *Virgibacillus* isolates, all previously identified by the 16s rRNA gene sequencing technique, were used in MALDI TOF MS, for the identification of new bacterial isolates from Qatari sabkhas using common biomarkers.[14] Bacteria which can be isolated from the same site may be identified by MALDI TOF MS, differentiated and categorized based on their protein profiles. An isolate belonging to a cluster would allow for the prediction of its function. The approach of integrating bacterial identification and protein profiling while focusing on protein biomarkers detected at conditions favorable to mineral formation would lead to their clustering as biomineral-forming bacteria. Differentiation of isolates belonging to the same species and principal component analysis (PCA) clussters are developed for further characterization of Qatari *Virgibacillus* and other bacterial isolates. On the other hand, this research would explain the diversity of minerals in the sabkhas, representing an advancement in studying the relationship between diversity of bacteria and minerals in sabkhas.

2 Methodology

2.1. Bacterial isolates from decaying mats, used as reference

Twenty-three bacterial strains were used as reference strains (Table 1). They are aerobic, halophilic and heterotrophic bacteria that were previously isolated from decaying mats sampled from Dohat Faishakh sabkha, northwest of Qatar. Their identification was performed by ribotyping and the access numbers of their DNA sequences were published.[14] 7 strains of *Virgibacillus marismortui*, 3 strains of *Virgibacillus salarius*, and 13 *Virgibacillus* sp. identified only at the genus level were tested in the current study. These strains were selected based on their significant role as aerobic microorganisms to mediate formation of Mg-rich carbonates.[14] Stock bacterial cultures were preserved at −80 °C in 60% glycerol (Microbiology and Biotechnology Lab, Qatar University, Doha) until use.

2.2. Sample preparation and protein extraction for MALDI-TOF

The sample preparation of the newly isolated strains was performed in two different techniques in order to generate the most reliable results. Extraction procedure using ethanol/formic acid (v/v) was used as reported by Wang *et al.* (2012).[27] Solid LB was used to grow the bacterial cultures. A colony was suspended in 300 μl of sterile water in an Eppendorf tube. Then the cells were re-suspended in 900 μl of absolute ethanol. The Eppendorf tubes were centrifuged at 13 000 rpm for 2 min. The supernatant obtained was discarded and the pellet was mixed with 1 ml 70% of formic acid and then 1 ml 100% acetonitrile. The mixture was re-centrifuged at 13 000 rpm for 2 min. From the supernatant, 1 μl was transferred onto a biotarget 48 sample spot. A total of 1 μl HCCA (α-cyano-4-hydroxycinnamic acid) matrix solution (50% acetonitrile and 2.5% trifluoroacetic acid in pure water) is then added onto the sample spots for protein extraction. Analyses were run in triplicates by spotting a colony into three different wells. Alternatively, since mass spectra obtained were not very clear, the whole cell method was performed. An amount of approximately 0.5 μl of material was taken from freshly grown colonies and transferred with a plastic loop into the well of the target plate and mass fingerprints were obtained allowing for better results and detection of biomarkers.

2.3. Identification of bacterial isolates

The similarities between individual mass spectra of the isolated strains and those of the database entries were expressed in the form of log(scores) obtained by default from the Biotyper software settings. The identification was carried out by the Bruker Biotyper software, where a log scale from 0.000 to 3.000 defines the identification matching level with the database. The score of 2.300–3.000 is highly probable species level identification with high confidence, which shows the identification result is highly accurate up to the species level. Score 2.000–2.299 is genus and probable species level identification, which shows the identification result is highly accurate up to the genus level, and probably correct at species level. Whereas scores between 1.700 and 1.999 indicates probable genus level identification, since the score is low, the result shows that the identification is probably correct at genus level.[23] Proteins having a m/z in the range of 2000 and 20 000 m/z are utilized for identification of bacterial strains based on individual mass peaks corresponding

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 1 Identification of the 35 bacterial strains isolated from decaying and living mats of Dohat Faishakh sabkha (Qatar), using MALDI-TOF MS. ND: not determined. NR: not reliable

Strain code	Identification by ribotyping	MALDI TOF score	Identification by MALDI TOF	Code for PCA analysis
DF112	*Virgibacillus marismortui*	<1.70	NR	31
DF221	*Virgibacillus* sp.	<1.70	NR	19
DF231	*Virgibacillus marismortui*	<1.70	NR	1
DF241	*Virgibacillus* sp.	<1.70	NR	18
DF252	*Virgibacillus* sp.	<1.70	NR	27
DF281	*Virgibacillus salarius*	<1.70	NR	21
DF282	*Virgibacillus* sp.	<1.70	NR	3
DF291	*Virgibacillus* sp.	<1.70	NR	4
DF322	*Virgibacillus marismortui*	<1.70	NR	7
DF341	*Virgibacillus marismortui*	<1.70	NR	14
DF351	*Virgibacillus salarius*	<1.70	NR	8
DF411	*Virgibacillus marismortui*	<1.70	NR	11
DF431	*Virgibacillus* sp.	<1.70	NR	30
DF451	*Virgibacillus* sp.	<1.70	NR	5
DF461	*Virgibacillus salaries*	<1.70	NR	2
DF472	*Virgibacillus marismortui*	<1.70	NR	15
DF491	*Virgibacillus marismortui*	<1.70	NR	20
DF2102	*Virgibacillus* sp.	<1.70	NR	22
DF2121	*Virgibacillus* sp.	<1.70	NR	12
DF2131	*Virgibacillus* sp.	<1.70	NR	13
DF2141	*Virgibacillus* sp.	<1.70	NR	28
DF2161	*Virgibacillus* sp.	<1.70	NR	29
DF2172	*Virgibacillus* sp.	<1.70	NR	10
K011	ND	1.81	*Bacillus licheniformis*	26
K1031A	ND	2.25	*Bacillus cereus*	33
K103B	ND	1.87	*Bacillus cereus*	6
K9-3-1	ND	1.76	*Bacillus circulans*	24
K012A	ND	<1.70	NR	35
K012B	ND	<1.70	NR	9
K9-1-1	ND	<1.70	NR	16
K9-1-2	ND	<1.70	NR	25
K9-1-4	ND	<1.70	NR	34
K915A	ND	<1.70	NR	32
K9-2-1	ND	<1.70	NR	17
K9-2-2	ND	<1.70	NR	23

to specific ribosomal proteins of distinct types of microorganisms.

2.4. Data processing

The protein profile of each bacterial isolate was obtained from the Bruker Flex Control software by using two sample spots. The Bruker Flex Control software was used to obtain mass spectra using linear and positive mode at 60 Hz laser frequency and intensity of 35%. The acceleration and source voltages were set as 20 and 18.7 kV, respectively. From different areas of the sample spot, 240 laser shots in 40-shot steps for each spectrum were obtained and analyzed using default settings. The protein profiles were processed and analyzed using the Flex Analysis and Biotyper RTC 3 software.

Since the mass spectra generated from MALDI are regarded as multivariate data, in which every mass signal represents a single molecular dimension, multivariate statistical methods are used for differentiation between bacterial species. Principle component analysis (PCA) was performed to decrease dimensionality of the data set and maintain the original information present. The PCA was based on the peaks acquired from MALDI-TOF. These peaks have the possibility to be proteins or peptides but cannot be identified as proteins or peptides. PCA allows the formation of clustered groups of spectra having similar variation characteristics and the visualization of the differences between them. The data can be represented in either a 2D or a 3D coordinate system; however, it is usually adequate to use the 2D which plots PC1 against PC2, since it generally offers more than 80% of the total variance between the samples. For a better visualization of the hierarchical relationship between the isolates and the reference strains used in this study, dendrogram clustering was also carried out using MALDI Biotyper Compass Explorer software adopting default settings according to the manufacturer's instructions.

All the analysis (PCA and dendrograms) was carried out as per the standard operating procedure for the instrument and built-in software. The raw spectra obtained for the three replicates was pre-processed which involved smoothing and

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

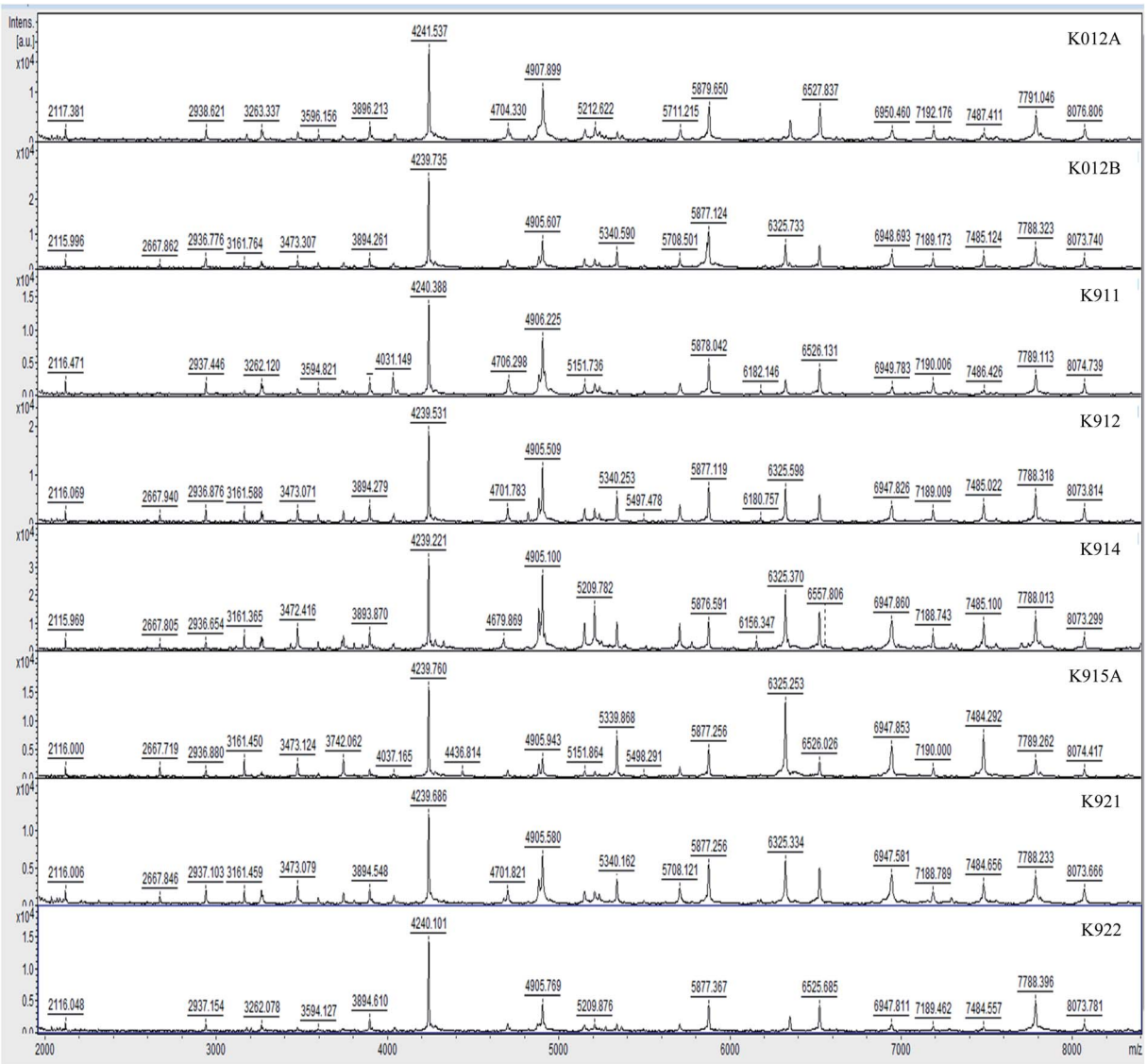

Fig. 1 MALDI-TOF spectra of the group of newly isolated strains used in this study.

baseline subtraction using Flex Analysis software followed by verification of quality of each spectrum measurement. Any individual spectrum with poor quality (having background noise or too high/low intensities) was excluded. The processed selected spectra were then used to create Main Spectra Projection (MSP) using the automated MSP creation functionality included in MALDI Biotyper 3.0 software. The MSP contains all the information about mean peak masses, mean peak intensities, and mean peak frequencies. These MSPs for each strain

Table 2 Intra- and interspecific percentage of similarity between species studied

Similarity percentages (%)					
	Virgibacillus marismortui	*Virgibacillus salarius*	*Bacillus cereus*	*Bacillus licheniformis*	*Bacillus circulans*
Virgibacillus marismortui	100	70.1–99.5	0.5–18.4	0.2–14.7	0.2
Virgibacillus salarius	70.1–99.5	100	0.9–16.2	3.8–5.2	0.2
Bacillus cereus	0.46–18.4	0.9–16.2	100	13.2–15.3	0.1–0.3
Bacillus licheniformis	0.2–14.7	3.8–5.2	13.2–15.3	100	0.2
Bacillus circulans	0.2	0.2	0.1–0.3	0.2	100

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 3 Characteristic peak masses of 17 *Virgibacillus* reference strains expressed as the arithmetic means of the *m*/*z* values of the three replicates of the corresponding strains of each species. Peak masses highlighted in bold are genus specific

PCA strain code	11	12	7	15	2	5	22	18	31	10	1	4	13	29	21	27	30
Strains	DF411	DF2121	DF322	DF472	DF461	DF451	DF2102	DF241	DF112	DF2172	DF231	DF291	DF2131	DF2161	DF281	DF252	DF431
Characteristic							2855				2855	2855					
peaks	3210		3210				3210					3210					
	3880		3880	3880		3880	3880	3880	3880	3880	3880	3880		3880	3880		3880
	4240		4240				4240		4240		4240	4240		4240	4240		4240
	4905	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**
						5240	5240				5240	5240					5240
	5910						5910		5910		5910	5910					5910
		6180	6180										6180	6180		6180	
	6430	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**	**6430**
		6750	6750										6750				
			6905		6905								6905	6905			
	7765	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**	**7765**
	9815	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**	**9815**
	10 485						10 485		10 485		10 485	10 485					

($n = 35$) were then fed to the functionality of PCA or dendrogram (MALDI Biotyper Compass Explorer), to carry out analysis and produce graphs.

3 Results and discussion

3.1. Identification of newly isolated strains by MALDI-TOF MS

Twelve bacterial isolates were isolated from the living mat sampled from Dohat Faishakh sabkha in Qatar by enrichment culture on the mineral inducing medium MD1. They were subjected to identification by MALDI-TOF MS. The twenty-three strains, previously isolated from decaying mat sampled from the same sabkha and identified by ribotyping as *Virgibacillus* bacteria, were used as reference for possible identification by MALDI-TOF mass spectrometric protein profiling, by matching their scores in the available database.

The results showed that the method based on ethanol/formic acid extraction of proteins before MALDI-TOF MAS analysis was not effective, since none of the isolates was reliably identified, although mass spectra were obtained. In fact, Wang *et al.* (2012) proceeded with extraction, because identification with whole cells of their bacteria was not effective.[27] Here, the last method was effective to identify *Bacillus* strains. *Virgibacillus* strains were not identified by both techniques. Table 1 shows the list of *Virgibacillus* strains used as reference and the 12 isolates from the living mat, with their corresponding code in the MALDI TOF MS protein profiling employed for the data analysis by PCA and dendrogram. All strains identified by ribotyping as *Virgibacillus* exhibited MALDI TOF MS scores less than 1.70 that cannot provide their reliable identification, according to the available database in the used machine. However, interestingly, a reproducible protein profile obtained by MALDI TOF MS profiling was obtained for all these *Virgibacillus* isolates.

From the 12 newly isolated isolates, four were identified at the level of *Bacillus* genus (two *Bacillus cereus,* one *Bacillus circulans* and one *Bacillus licheniformis*) with corresponding scores of 1.87, 2.25, 1.76, and 1.81 respectively. The others exhibited scores less than 1.70, which cannot identify them, reliably.

The significance of using this method for identifying bacterial isolates or closely related ones lies in the fact that masses and signals of proteins serving as markers, interlinked representatives of their corresponding expressed genes. Moreover,

Table 4 Characteristic peak masses of 6 *Virgibacillus* reference strains and 4 newly isolated strains of *Bacillus* genus expressed as the arithmetic means of the *m*/*z* values of the three replicates of the corresponding strains of each species. Peak masses found to exist in all strains of the group are highlighted in bold

PCA strain code	3	14	20	28	8	19	6	33	24	26
Species	DF282	DF341	DF491	DF2141	DF351	DF221	K1031B	K1031A	K931	K011
Characteristic peaks	3265	3265	3265	3265	3265	3265				
	3880		3880	3880		3880		3880		
	4240		4240			4240				
	4905	4905	4905	4905	4905	4905				
	5240	**5240**	**5240**	**5240**	**5240**	**5240**	**5240**	**5240**	**5240**	**5240**
	6430		6430	6430	6430	6430	6430	6430	6430	6430
	7765		**7765**	**7765**	**7765**	**7765**	**7765**			**7765**
	9815	9815	9815	9815	9815	9815				

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

mass spectra and protein profiles were established for each isolate.

3.2. Analysis of protein profiles of the bacterial strains

Hereby, MALDI-TOF-MS provides a platform for the comparison of mass spectra of the 35 studied isolates from Dohat Faishakh sabkha, analyzed using the data analysis software. The obtained peaks revealed admissible resolution using the whole cells, whereas proteins extraction method revealed no separate peaks. For each strain, an average spectrum was obtained from duplicated spectra of triplicated colonies (six in total). Examples of the spectra obtained are shown in Fig. 1. Growth media and conditions were similar to all isolates, to reduce culturing impact on the protein profiles, biomarkers and reproducibility of the profiles. The mass spectra demonstrated complex collections of discrete ions of *m*/*z* ratios ranging between 2000 and 10 000. Table 3 shows peak masses of 17 *Virgibacillus* reference strains and Table 4 shows those of 6 other *Virgibacillus* reference strains along with 4 newly identified *Bacillus* isolated strains (K011, K1031A, K1031B, and K931). Peak highlighted in bold are peaks obtained by most of strains.

The intensity of the generated peaks allows the detection of some unique proteins considered as biomarkers among highly related isolates. Indeed, peak of 4905 and 9815 *m*/*z* are only obtained with *Virgibacillus* strains, they can be considered as biomarkers of the genus *Virgibacillus*. However, peak masses of 4240, 6430 and 7765 *m*/*z* are found commonly in most *Bacillus* and *Virgibacillus* strains. This results is not surprising, since *Virgibacillus* genus is strongly similar to the *Bacillus* genus. It was reclassified fom *Bacillus* genus in 1998 based on analysis of the species *Virgibacillus pantothenticus*.[28] However, at the cultural conditions favorable to minerals formation, *Virgibacillus* strains produced much more proteins than *Bacillus* strains.

Therefore, these indicated peaks are regarded as genus specific biomarkers for the known genera included in the database. Thus, it can only be used for future identification of newly isolated bacterial strains belonging to these genera. Moreover, the mass spectra of *Virgibacillus* appear to be highly similar, at the species level. Indeed, the peak at 3880 *m*/*z* was detected in five *V. marismortui* strains in addition to the peak at 4240 *m*/*z* in four of them. This suggests that these two protein peaks could be possible species-specific biomarkers of *V. marismortui*. On the other hand, some common peaks were present in most of the isolates, but none was shown to be species specific in that sense. In a second group of *Virgibacillus* isolates (DF221, DF282, DF341, DF351, DF491 and DF2141) the peak at 3265 *m*/*z* is specific (Table 4). None of the other reference *Virgibacillus* strains revealed this peak, which suggests that the peak 3265 *m*/*z* is specific to some *Virgibacillus*, not to all. In addition, the biomarker 3265 *m*/*z* was not found in the mass spectra of any of the *Bacillus* strains grouped in this cluster. Six other peak masses were detected in most *Virgibacillus* strains (*m*/*z* 3880, 4240, 4905, 6430, 7765, and 9815), showing high similarity among all studied *Virgibacillus* sp.

In order to confirm the differences in term of biomarkers at the level of species, the intra- and interspecific percentage of

Table 5 Characteristic protein peaks of the unidentified bacterial isolates expressed as the arithmetic means of the *m*/*z* values of the three replicates of the corresponding strains of each species

	35	9	16	25	34	32	17	23
Isolates	K012A	K012B	K911	K912	K914	K915A	K921	K922
Characteristic peaks	2115	2115	2115	2115	2115	2115	2115	2115
		2667		2667	2667	2667	2667	
	2940	2940	2940	2940	2940	2940	2940	2940
		3160		3160	3160	3160	3160	
	3260		3260					3260
		3470		3470		3470	3470	
	3595		3595					3595
	3880	3880		3880	3880		3880	3880
			4035			4035		
	4240	4240	4240	4240	4240	4240	4240	4240
	4700		4700	4700	4670		4700	
	4905	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**	**4905**
	5880	5880	5880	5880	5880	5880	5880	5880
			6180	6180	6180			
	6525		6525		6525	6525		6525
	6950	6950	6950	6950	6950	6950	6950	6950
	7190	7190	7190	7190	7190	7190	7190	7190
	7490	7490	7490	7490	7490	7490	7490	7490
	7790	7790	7790	7790	7790	7790	7790	7790
	8070	8070	8070	8070	8070	8070	8070	8070
	9400	9400	9400	9400		9400	9400	9400
	9800	9800	9800	9800	9800	9800	9800	9800
	10 420	10 420	10 420	10 420	10 420	10 420	10 420	10 420
	11 415	11 415	10 415	11 415	11 415	11 415	11 415	11 415

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

similarity was established for strains identified at the species level using composite correlation matrix. Results are shown in Table 2. The differentiation showed existence of a wide range of similarity between strains belonging to *Virgibacillus* species which goes under "intraspecific similarity", as well as similarity between different *Virgibacillus* species known as "interspecific similarity". Thus, *V. marismortui* and *V. salaries* are shown highly related species with 70.1–99.5% interspecific similarity. Similar results were reported in 2015, but with phylogenic trees generated based on 16S rRNA sequences of *Virgibacillus* species.[29] Here, evidences were simply provided by the rapid and reliable MALDI-TOF MS proteins profiling. However, *B. cereus* and *B. lecheinformis* showed only 13–15% interspecific similarity. *B. circulans* was highly different from other *Bacillus* strains with extremely low similarity inter and intra-specifically.

In order to establish the relationship between the unidentified isolates, their principal peaks were determined and listed in Table 5. It is clear that all the isolates showed a characteristic peak at 4905 *m*/*z* which was also shown characteristic of *Virgibacillus* strains. The peak 3880 *m*/*z* is present in several *Virgibacillus* and *Bacillus* strains. 13 other peaks are common among the unidentified isolates, but not present in *Virgibacillus* nor *Bacillus* strains. Differences between these isolates are related to 11 other peaks. The comparison based on ions of *m*/*z* ratios ranging between 2000 and 10 000 was not effective to identify newly isolated strains. However, protein profiling is appropriate to differentiate newly isolated bacteria.

The isolates K012A and K922 exhibited the same characteristic peaks. They are highly similar or identical. The differences between others are mostly minor with one or two peaks only, showing the high sensitivity of the differentiation using proteins profiling. However, the differences could be more sensitive by considering peaks having *m*/*z* below 2000.

In order to establish the similarities at the genus and species level, the interspecific percentage of similarity of the unidentified isolates shows that the unidentified strains should be classified in a group distinct to *Virgibacillus* strains (Table 6). The results show existence of a wide range of similarity with strains of *Virgibacillus* identified at the genus level, as well as similarity with different *Virgibacillus* species. Generally, similarities fluctuate between 20% to 40%, with the exception of the isolate K012A which shows 12% and 7% only with *Virgibacillus* DF2131 (PCA code 13) and *Virgibacillus* DF2102 (PCA code 22) respectively. This study clearly shows that the unidentified strains should be classified in a group distinct to *Virgibacillus* strains.

3.3. Relationships and clustering of isolates using MALDI-TOF MS and PCA

Although the information about the relationship between bacterial isolates can be obtained by comparison of the protein *m*/*z* fingerprints in MALDI-TOF MS profiles, however, differentiation between the closely related isolates can be revealed by combination between MALDI-TOF MS analysis and PCA. PCA provides a rapid qualitative assessment tool for determining the association among the studied isolates and evaluating large data sets (protein profiles). MALDI-TOF MS instrument employed in this research has a built-in software for statistical analysis in which protein profiles can be directly analyzed (after

Table 6 Similarity percentages between unidentified isolates of group C and reference *Virgibacillus* sp. strains

		Unidentified strains of group C							
	Strain code	25	32	34	35	23	17	16	9
Virgibacillus sp.	3	29.6	21.6	21.2	24.9	28	16.4	22.06	33.2
	4	30.8	23.8	26.5	28.1	27.4	23.3	26.1	35.15
	5	36.7	29	30.1	28	33	24.2	0.29	37.9
	10	37.5	32.4	35.1	38.3	35.2	30.2	34.9	41.2
	12	33.8	28.8	28.4	30.7	31.5	21	28.2	38.03
	13	32.7	25.02	28.4	11.7	31.3	25.9	29.3	37.6
	18	37.2	34.8	36.4	36.4	37	32.1	37.2	43.24
	19	31.2	28.3	28.4	32.4	28.7	24.1	29.7	35.8
	22	37.8	32.6	34	7.3	34.5	28.5	33	41.5
	27	32.2	30	30.6	34	32.5	26	31.7	38.93
	28	35.5	26.1	25.8	26.2	32.3	20.8	26.7	33.8
	29	34.9	25	28.7	28.7	31.4	26.7	29.5	37.4
	30	30.7	24.5	26.3	26.3	29	23.8	27.6	38.8
Virgibacillus marismortui	31	40.8	39.6	41.5	35.9	41.7	35.7	41.0	39.9
Virgibacillus marismortui	1	36.2	29.2	33.8	34.1	33.5	30.0	33.2	39.6
Virgibacillus salarius	21	30.9	25.1	27.6	25.5	29.0	26.6	29.3	38.2
Virgibacillus marismortui	7	35.6	32.2	32.7	31.6	33.8	25.9	32.2	38.4
Virgibacillus marismortui	14	33.7	27.3	23.7	27.1	29.5	18.3	24.5	35.7
Virgibacillus salarius	8	28.6	23.2	23.7	28.5	28.5	20.3	25.3	35.5
Virgibacillus marismortui	11	30.5	24.4	24.3	26.3	29.8	17.9	23.8	36.5
Virgibacillus salarius	2	33.7	28.0	29.4	29.7	31.5	24.6	29.2	38.8
Virgibacillus marismortui	15	37.6	34.9	34.9	32.8	34.6	25.9	32.8	40.8
Virgibacillus marismortui	20	30.9	26.4	27.1	28.6	30.3	22.6	28.2	36.9

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

baseline subtraction and smoothing). The values used in the PCA are exact *m*/*z* values and their relative intensities. Principle component analysis (PCA) is one of the oldest methods used by statisticians in order to interpret large datasets.[30] This method analyzes matrices of variance–covariance and correlations of data. Its main objective is to use principal components as a way of reducing the dimensions of objects being studied. Furthermore, this reduction creates linear combinations of variables representing the objects being studied. The combinations are known as principal components.[31] The results of PCA are shown in Fig. 2. From Fig. 2a it is clear that the strains exhibit large biodiversity at proteins level. The total variance of the 10 PCA's is shown in Fig. 2b. It shows that the three principle components *i.e.* PC1 (32%), PC2 (21.5%) and PC3 (12.5%) combine to show 66% variability in the data. Using the first three components, three clusters were obtained. The distance between the clusters shows the variation at groups level, while the distance between the strains (within the cluster) shows the differences in protein profiles at strain level. Cluster 1, which has positive correlation to PC1 and negative correlation to PC2 and PC3, include unidentified strains (K012B, K911, K921, K922, K914, K912, K915A, and K012A). Whereas, cluster 2 has positive correlation to all three components and is mainly comprised of *B. cereus* and *B. licheniformis* (K1031B, K011 and K1031A) demonstrating variation in their protein profiles in comparison to other studied strains. These results are in complete agreement with previous studies showing that these two species give well detectable and easily distinguishable band pattern profiles.[32,33] Nevertheless, differentiation between *B. cereus* and *B. licheniformis* by BOX-PCR genomic fingerprinting has showed almost identical patterns of discrimination with distinct band patterns.[32] However, *B. circulans* (K931) although belonging to the genus *Bacillus* was grouped with the other 6 *Virgibacillus* strains belonging to cluster III, proving that their protein profiles are more closely related to those of other *Virgibacillus* strains than they are to *B. cereus* and *B. licheniformis* (cluster II). Remaining strains (DF231, DF291, DF2172, DF2131, DF2161, DF281, DF431, DF241, DF2102, DF461, DF112, DF252, DF322, DF451, DF472, DF411, DF2121) can be categorized separately, in which the variations between the strains are also higher since they are located relatively far from each other *e.g.* strains DF291 and DF2121 of respective PCA codes 4 and 12, as clearly shown in Fig. 2a.

3.4. Phylloproteomic tree

The PCA results lead to the establishment of a dendrogram as a phylloproteomic tree (Fig. 3). It shows that the studied strains may be categorized into three distinct groups based on their similarity matrix. The 23 *Virgibacillus* which are used as reference, are clearly divided into two separate groups (A and B) in the phylloproteomic tree. The isolates *V. marismortui* (DF112, DF231, DF322, DF411 and DF472), *Virgibacillus* sp. (DF 241, DF252, DF291, DF451, DF2121, DF2102, DF2131, DF2161 and DF2171), *V. salarius* (DF281, DF461 and DF431) were grouped together (group A). On the other hand, the other reference strains *Virgibacillus* sp. (DF221, DF282 and DF2141), *V. marismortui* (DF341 and DF491) and *V. salarius* 351 were categorized separately into another group (group B) along with four newly isolated strains that have been identified by MALDI-TOF, *Bacillus cereus* (K1031A and K1031B), *Bacillus licheniformis* (K011), and *Bacillus circulans* (K931). The eight remaining strains were categorized in one group (group C). Those strains are the ones isolated from living mats which could not be identified by MALDI-TOF (K012A, K012B, K911, K912, K914, K915A, K921, and K922). Attempts to show any similarities between these strains and other *Virgibacillus* or *Bacillus* species have failed. Clustering analysis by neither PCA nor dendrogram could show any significant relationship between isolates of group A and those of groups B and C; therefore, it can only be deduced that they have no resemblance with any *Bacillus* or *Virgibacillus* strains in this study. In comparison to each other, both approaches of clustering; PCA and dendrogram, have showed quite similar results. In fact, both approaches allow clustering of isolated and reference strains, which means that

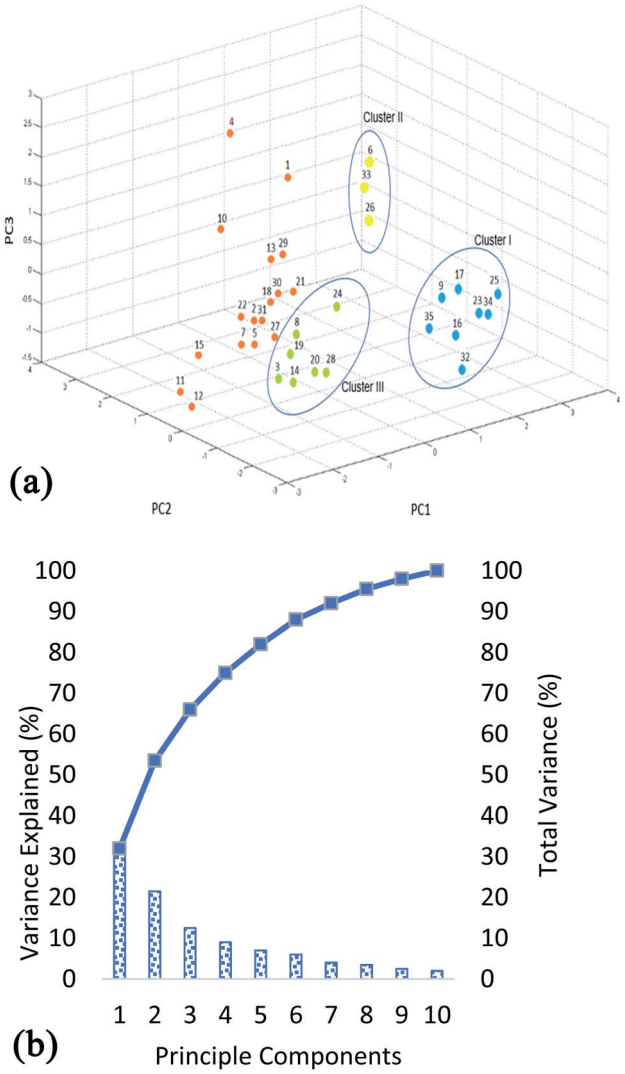

Fig. 2 Classification of strains using PCA (a) PCA plot (b) percentage of variance explained.

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

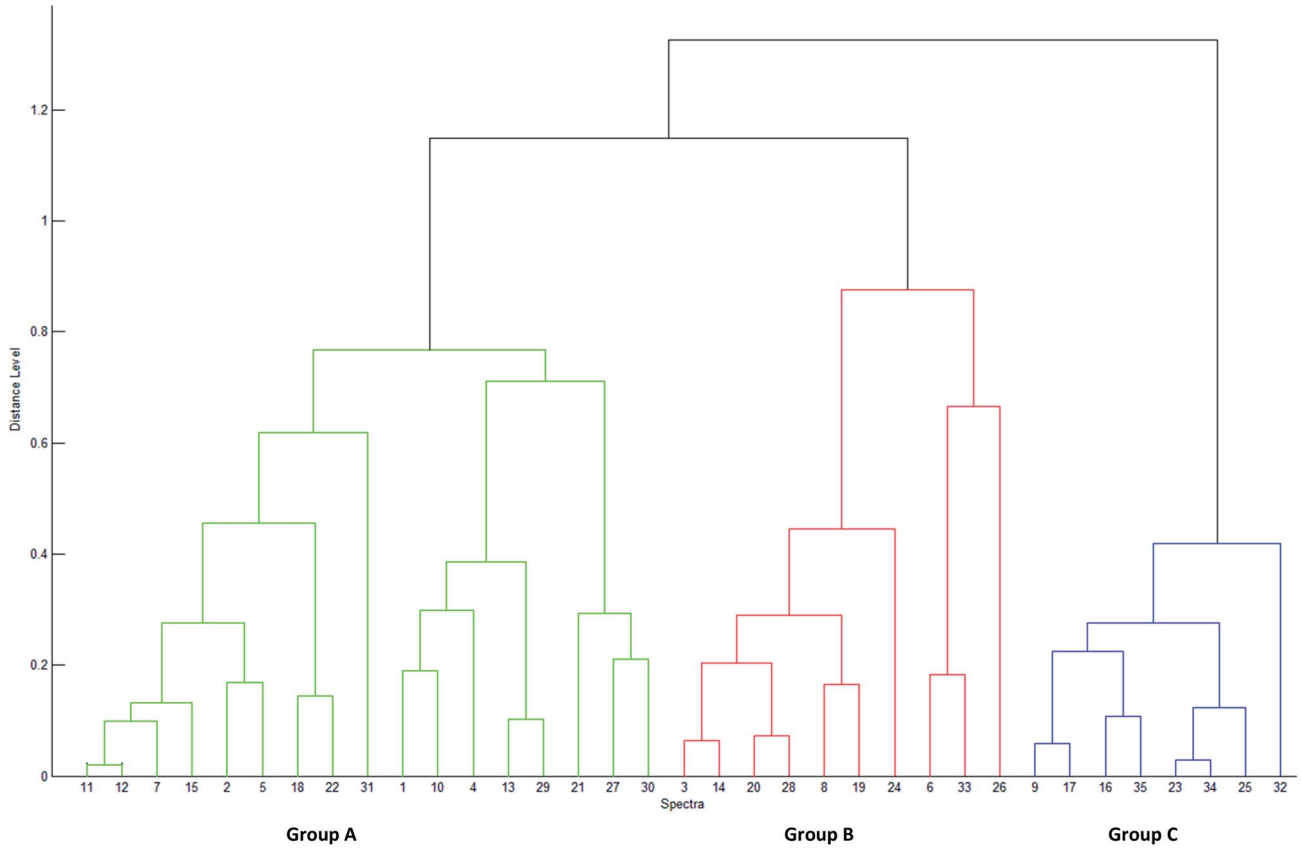

Fig. 3 Phylloproteomic tree illustrating the relationship among the strains used in this study using similarity coefficient and single linkage. The dendrogram's scale on the left represents the relative distance used in the clustering analysis.

the data obtained by MALDI-TOF MS could orient the prediction of the identity of unknown isolates belonging to any of the clusters. Nevertheless, strains are differentiated based on the presence/absence of one or more discriminating peaks without presenting any hierarchical relationship between them in the PCA approach, but it is only using the dendrogram that hierarchical clustering of samples is made possible, in which the relationship between the strains in the same group and those in different ones are presented.

4 Conclusions

Hence, combination of MALDI-TOF MS with PCA is shown to be a powerful tool for rapid identification and categorization of strains isolated from the same niche or comparable ones. The reliability of any identification method depends strictly on its database. With more updates of the database of MALDI-TOF, not only will it be a reliable tool for identification of bacteria in clinical fields of microbiology, but also in environmental ones. Here, combination of isolation of bacteria from sabkhas mats by enrichment cultures in a medium that allows formation of many types of minerals, to MALDI TOF MS protein profiling and PCA analysis guides on their rapid identification as a first step. Then, by establishing the strength of their relationship with known related bacteria based on profiling of their synthesized proteins at mineral forming conditions would help to prediction of their potential activity. Then, larger number of bacterial isolates can be easily screened for their potential to exhibit certain activities, which is of ecological, environmental and biotechnological significance. Moreover, these findings demonstrate the high occurrence and diversity of *Virgibacillus* strains in the same mat and their high similarity to others from other mats. They also explain the high diversity of their potential to form diverse minerals in sabkhas, related to their diversity of adaptation routes. Al Disi *et al.*, (2019)[19] showed high variability of EPS compositions as a tool employed by *Virgibacillus* to adapt differently to stressors mainly through modulating EPS composition. This result agrees with those reported by earlier in 2015 showing that MALDI-TOF MS and PCA were appropriate to elucidate the environmental significance of biodegradation potential of petroleum hydrocarbons by bacteria.[33] Similar conclusions were drawn regarding the classification into groups of these bacteria based on their enzymatic activities (ability to degrade hydrocarbons).

Author contributions

Nabil Zouari, and Zulfa Al Disi conceived and designed the experiments, contributed in analysis of the data and wrote the paper. Rim Abdelsamad, and Sara Wahib performed the experiments on microbiology and contributed in MALDI-TOF MS experiments and writing of the manuscript. Mohammad

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Yousaf Ashfaq, performed the MALDI-TOF MS experiments and contributed in writing of the manuscript.

Conflicts of interest

The authors declare no conflict of interest. All authors decided to publish this work. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Acknowledgements

This work was made possible by the QUST-1-CAS-2018-3 fund, a student grant from the Qatar University. Many thanks to Microbiology laboratory, Hamad Medical Corporation for providing facility for MALDI-TOF MS technique. The statements made herein are solely the responsibility of the authors.

References

1 M. Brauchli, M. Brauchli, J. A. McKenzie, C. J. Strohmenger, J. A. McKenzie and C. J. Strohmenger, The importance of microbial mats for dolomite formation in the Dohat Faishakh sabkha, Qatar, *Carbonates Evaporites*, 2016, **31**, 339–345.

2 M. M. Ashour, *Sabkhas In Qatar Peninsula. Landscape and Geodiversity*, vol. 1, 2013, pp. 10–35.

3 L. Illing and J. C. Taylor, Penecontemporaneous dolomitization in Sabkha Faishakh, Qatar; evidence from changes in the chemistry of the interstitial brines, *J. Sediment. Res.*, 1993, **63**, 1042–1048.

4 L. V. Illing, A. J. Wells and J. C. Taylor, *Penecontemporary Dolomite In The Persian Gulf*. Society for Sedimentary Geology Special Publications (SPEM), 1965.

5 J. McKenzie, The dolomite problem: an outstanding controversy, in *Evolution of Geological Theories in Sedimentology, Earth History and Tectonics*, ed. D. W. Müller, J. A. McKenzie and H. J. Weissert, Academic Press Ltd, London, pp. 37–54, 1991.

6 R. S. Arvidson and F. T. Mackenzie, The dolomite problem; control of precipitation kinetics by temperature and saturation state, *Am. J. Sci.*, 1999, **1**, 257–288.

7 D. A. Petrash, O. M. Bialik, T. R. Bontognali, C. Vasconcelos, J. A. Roberts, J. A. McKenzie and K. O. Konhauser, Microbially catalyzed dolomite formation: from near-surface to burial, *Earth-Sci. Rev.*, 2017, **171**, 558–582, DOI: 10.1016/j.earscirev.2017.06.015.

8 T. R. Bontognali, Anoxygenic phototrophs and the forgotten art of making dolomite, *Geology*, 2019, **47**(6), 591–592.

9 N. K. Dhami, M. S. Reddy and A. Mukherjee, Biomineralization of calcium carbonates and their engineered applications: a review, *Front. Microbiol.*, 2013, **4**, 314.

10 T. R. Bontognali, J. A. McKenzie, R. J. Warthmann and C. Vasconcelos, Microbially influenced formation of Mg-calcite and Ca-dolomite in the presence of exopolymeric substances produced by sulphate-reducing bacteria, *Terra. Nova*, 2014, **26**, 72–77.

11 T. Zhu and M. Dittrich, Carbonate Precipitation through Microbial Activities in Natural Environment, and Their Potential in Biotechnology: A Review, *Front. Bioeng. Biotech.*, 2016, **4**, 4.

12 Y. Zhao, H. Yan, J. Zhou, M. E. Tucker, M. Han, H. Zhao and Z. Han, Bio-Precipitation of Calcium and Magnesium Ions through Extracellular and Intracellular Process Induced by Bacillus Licheniformis SRB2, *Minerals*, 2019, **9**(526), 23.

13 V. Pauker, R. T. Bryan, G. Gregor, B. Pauline, H. Matthias, Z. Lothar and Z. Sabine, Improved Discrimination of Bacillus anthracis from Closely Related Species in the Bacillus cereus Sensu Lato Group Based on Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry, *J. Clin. Microbiol.*, 2018, **56**(5), 1900–1917.

14 Z. A. Al Disi, S. Jaoua, T. R. Bontognali, E. S. Attia, H. A. Al-Kuwari and N. Zouari, Evidence of a Role for Aerobic Bacteria in High Magnesium Carbonate Formation in the Evaporitic Environment of Dohat Faishakh Sabkha in Qatar, *Frontiers in Environmental Science*, 2017, **5**, 1.

15 J. E. Clarridge, Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases, *Clin. Microbiol. Rev.*, 2004, **17**(4), 840–862.

16 O. Y. Costa, J. M. Raaijmakers and E. E. Kuramae, Microbial Extracellular Polymeric Substances: Ecological Function and Impact on Soil Aggregation, *Front. Microbiol.*, 2018, **9**, 14.

17 D. H. Limoli, C. J. Jones and D. J. Wozniak, Bacterial Extracellular Polysaccharides in Biofilm Formation and Function, *Microbiol. Spectr.*, 2015, **3**(3), 223–247.

18 C. C. Carvalho and P. Fernandes, Production of Metabolites as Bacterial Responses to the Marine Environment, *Mar. Drugs*, 2010, **8**(3), 705–727.

19 Z. A. Al Disi, N. Zouari, M. Dittrich, S. Jaoua, H. A. Al-Kuwari and T. R. Bontognali, Characterization of the extracellular polymeric substances (EPS) of Virgibacillus strains capable of mediating the formation of high Mg-calcite and protodolomite, *Mar. Chem.*, 2019, **216**, 103693.

20 A. Miñán, A. Bosch, P. Lasch, M. Stämmler, D. O. Serra, J. Degrossi and D. Naumann, Rapid identification of Burkholderia cepacia complex species including strains of the novel Taxon K, recovered from cystic fibrosis patients by intact cell MALDI-ToF mass spectrometry, *Analyst*, 2009, **134**(6), 1138–1148.

21 A. P. Desai, T. Stanley, M. Atuan, J. McKey, J. J. Lipuma, B. Rogers and R. Jerris, Use of matrix assisted laser desorption ionisation–time of flight mass spectrometry in a paediatric clinical laboratory for identification of bacteria commonly isolated from cystic fibrosis patients, *J. Clin. Pathol.*, 2012, **65**(9), 835–838.

22 N. Al-Kaabi, M. A. Al-Ghouti, M. Oualha, M. Y. Mohammad, A. Al-Naemi, T. I. Sølling and N. Zouari, A MALDI-TOF study of bio-remediation in highly weathered oil contaminated soils, *J. Petrol. Sci. Eng.*, 2018, **168**, 569–576.

23 S. Bibi, M. Oualha, M. Y. Ashfaq, M. T. Suleiman and N. Zouari, Isolation, differentiation and biodiversity of

Open Access Article. Published on 09 April 2020. Downloaded on 11/14/2025 6:21:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

ureolytic bacteria of Qatari soil and their potential in microbially induced calcite precipitation (MICP) for soil stabilization, *RSC Adv.*, 2018, **8**(11), 5854–5863.

24 N. Singhal, M. Kumar, P. K. Kanaujia and J. S. Virdi, MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis, *Front. Microbiol.*, 2015, **6**, 791.

25 F. Cobo, Application of maldi-tof mass spectrometry in clinical virology: a review, *Open Virol. J.*, 2013, **7**, 84.

26 O. Šedo, S. Pekár and Z. Zdráhal, MALDI-TOF Mass Spectrometric Profiling of Spider Venoms. in *Snake and Spider Toxins*, Humana, New York, NY, 2020, pp. 173–181.

27 J. Wang, W. F. Chen and Q. X. Li, Rapid identification and classification of Mycobacterium spp. using whole-cell protein barcodes with matrix assisted laser desorption ionization time of flight mass spectrometry in comparison with multigene phylogenetic analysis, *Anal. Chim. Acta*, 2012, **716**, 133–137.

28 M. Heyndrickx, L. Lebbe, K. Kersters, P. De Vos, G. Forsyth and N. A. Logan, Virgibacillus: a new genus to accommodate Bacillus pantothenticus (Proom and Knight 1950). Emended description of Virgibacillus pantothenticus, *Int. J. Syst. Bacteriol.*, 1998, **48**, 99–106.

29 S. Khelaifia, O. Croce, J. Lagier, C. Robert, C. Couderc and F. Di Pinto, Noncontiguous finished genome sequence and description of Virgibacillus massiliensis sp. nov., a moderately halophilic bacterium isolated from human gut, *New Microbes New Infect.*, 2015, **8**, 78–88.

30 I. T. Jolliffe and J. Cadima, Principal component analysis: a review and recent developments, *Philos. Trans. R. Soc., A*, 2016, **374**(2065), 20150202.

31 A. Maćkiewicz and W. Ratajczak, Principal components analysis (PCA), *Comput. Geosci.*, 1993, **19**(3), 303–342.

32 M. Vyletělová, P. Švec, Z. Páčová, I. Sedláček and P. Roubal, Occurence of Bacillus cereus and Bacillus licheniformis strains in the course of UHT milk production, *Anim. Sci.*, 2002, **47**, 200–205.

33 N. Hua, A. Hamza-Chaffai, R. Vreeland, H. Isoda and T. Naganuma, Virgibacillus salarius sp. nov., a halophilic bacterium isolated from a Saharan salt lake, *Int. J. Syst. Evol. Microbiol.*, 2008, **58**(10), 2409–2414.

 This journal is © The Royal Society of Chemistry 2020