October, 2014

MALDI-TOF MS as a supportive tool for the evaluation of bacterial diversity in soils from Africa and the Americas

Adrian Douglas Allen, *Howard University*

Maria Velez-Quinones

Broderick Eribo, *Howard University*

Vernon Morris, *Howard University*

MALDI-TOF MS as a supportive tool for the evaluation of bacterial diversity in soils from Africa and the Americas

Adrian Douglas Allen · Maria Velez-Quinones ·
Broderick E. Eribo · Vernon Morris

Received: 3 April 2014 / Accepted: 3 October 2014 / Published online: 18 October 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Identification and characterization of viable culturable bacteria (VCB) associated with soils from Africa and the Americas are significant for environmental and battlefield security. Such analyses are scarce, and their evaluation using traditional microbiological methods does not fully elucidate the structure and chemotaxonomic characteristics of the microbial community. In this study, matrix-assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS), spectrometry in addition to 16S rRNA sequencing, and diversity indices were employed to characterize VCB and their associated biomarkers. Nineteen genera were identified across all sample locations, but only four

(*Bacillus*, *Brevibacillus*, *Paenibacillus*, and *Terribacillus*) confirmed by ClustalW2 as being 98–99 % similar among locations. Further evaluation of soils showed bacterial diversity (H) of (1.0–8.4), evenness (E_H), (0.14–0.72), similarity (S_j), (0.0–0.38), and cfu/g soil (2.5×10^1 – 2.2×10^7). Analysis of representative bacteria using MALDI-TOF MS identified biomarkers for the genera *Bacillus* at m/z 6,778 (75 %), 9,437 (100 %); *Brevibacillus*, m/z 7,381 (86 %); *Paenibacillus*, 5,473 (63 %); and *Terribacillus*, 4,517, 6,532, 7,574 (67 %). Peptide mass fingerprinting of biomarkers identified partial peptide maps for several potential virulence factors such as hemagglutinin from a *Brevibacillus* spp. The data indicate an east (Sudan) to west (Jamaica; Mexico; Washington, DC; Baltimore) trend of potentially pathogenic endospore-forming bacteria.

A. D. Allen (✉)
Department of Comprehensive Sciences, Howard
University, 2441 Sixth Street, NW Locke Hall, Room 260,
Washington, DC 20059, USA
e-mail: adriandallen@yahoo.com

A. D. Allen · V. Morris
NOAA Center for Atmospheric Sciences, Howard
University, 1840 Seventh Street, NW Room 305,
Washington, DC 20001, USA
e-mail: vmorris@howard.edu

M. Velez-Quinones · B. E. Eribo
Department of Biology, Howard University Graduate
School, 415 College Street, NW, Washington, DC 20059,
USA

V. Morris
Department of Chemistry, Howard University,
525 College Street, NW, Washington, DC 20059, USA

Keywords Mass spectrometry · Aerobiology · Protein biomarkers · Saharan/sub-Saharan soil and dust · Caribbean · Virulence factor · Endospore-forming bacteria

1 Introduction

The translocation of Saharan soil westward into the Americas has been occurring prior to 1832, the point of earliest collection for ‘Darwin’s dust’ (Gorbushina et al. 2007). According to McTainsh et al. (1997), this translocation may be associated with the Bilma-Faya Largeau area of Lake Chad (October to April), the

Niger Delta in Mali (May to September), and other natural phenomena such as North Atlantic Oscillation (NAO) and El Nino Southern Oscillation (ENSO). A conservative estimate for the annual quantity of desert soil transported is approximately two billion metric tons, moving an estimated two quadrillion microorganisms ($\approx 10^9$ /gram topsoil), which include bacteria (diversity, $\approx 10,000$), viruses, fungi, and their biologically natural constituents/metabolites such as mycotoxins, endotoxins, and macrolides (Griffin et al. 2002, 2003; Garrison et al. 2003; Prospero et al. 2005; Kellogg and Griffin 2006; González-Toril et al. 2009). Previous research, to include that from the National Institute of Allergy and Infectious Diseases (NIAID), indicates that desert soil exposure may be associated with increased risk of respiratory diseases such as asthma, as well as various plant and animal diseases (Prospero 1999; Howitt 2000; Monteil et al. 2000; Kanatani et al. 2010; Cook et al. 2005; Sing and Sing 2010). For example, bacteria such as *Staphylococcus xylosum* and *Neisseria meningitidis* have been implicated as the cause of septicemia in a loggerhead turtle (*Caretta caretta*) off the Canary islands and meningitis in Nigeria respectively (Torrent et al. 2004). Previous analysis of air samples, using conventional methods, from the Caribbean identified the dominant genera *Microbacterium* (14.4 %), *Sphingomonas* (7.2 %), *Bacillus* (6.5 %), and *Streptomyces*, 4.0 % (Griffin et al. 2002). Similarly, dust samples collected in Mali, West Africa, identified twenty genera which include *Bacillus* (38 %), *Kocuria* (12.8 %), *Planococcus* (8.5 %), *Saccharococcus* (7.4 %), and 6.4 % for *Micrococcus* (Kellogg et al. 2004).

Given the potential impact of these and other pathogens on public health, efforts are required for the evaluation of not only air, but more importantly soil samples from sub-Saharan Africa and downwind locations such as the Americas (De Longueville et al. 2010; Leski et al. 2011). The analyses of bacteria in aerobiological samples from these locations have historically been ascertained using 16S rRNA sequencing, microscopy, and phenotypic assays. Although these methods provide valuable culture information, they also possess several limitations, such as, incomplete chemotaxonomic profiles for isolates, primarily designed for clinical isolates, increased time-to-results (48–72 h), and cost/sample at $\approx$ US\$18.0 (Seng et al. 2009). Therefore, the need for rapid, cost-effective, and supportive methods for

screening and/or identifying bacteria cannot be over-emphasized. One method now gaining attention for its robustness is matrix-assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS). The advantages offered by MALDI-TOF MS include soft ionization (nondestructive sampling), measurement of >200 kDa masses, cost-effectiveness (US\$2/sample), the generation of monoisotopic peaks, and minimum time (6–8.5 min) for sample processing and analysis (Cordwell et al. 1995; Haag et al. 1998; Arnold and Reilly 1998; Ruelle et al. 2004; Liu et al. 2007). Furthermore, MALDI-TOF MS has an overall greater accuracy (95–100 %) compared to the other methods (Nagy et al. 2009; Seng et al. 2009; Bessède et al. 2011; Bengali et al. 2011).

MALDI-TOF MS of whole bacterial cells can provide characteristic fingerprints (protein biomarkers) that can be used, in conjunction with a fingerprint-statistical-based or a proteomics approach to identify bacteria and fungi (Moura et al. 2003). This approach requires an extensive database of known organisms, which is used to match the unknown isolate, provide a matching score based on identified mass to charge ratios (m/z) and their intensity correlation. However, in the absence of such a database, the m/z of unknowns can be compared with type cultures by visual inspection, since spectra from identical isolates are similar but rarely identical (Lay 2001). Nevertheless, a single analysis of global biomarkers could provide intact genetic information which include antibiotic resistance, microbial community structures, environmental conditions, and elucidate close phylogenetic associations among closely related bacteria.

In this study, 16S rRNA sequencing, diversity indices, and matrix-assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS) were employed to evaluate viable culturable bacteria, their associated protein biomarkers, and their potential for transfer among nine locations.

2 Materials and methods

2.1 Sampling/diversity and evenness determination

Samples were collected from the Americas: [western Jamaica (rural)]; Baltimore (suburban); northern

Fig. 1 World map showing approximate locations (filled circle) of sample collection. The approximate coordinates for the locations are as follows: Americas (JM, western Jamaica, 18°18'44.478''; −78°7'29.4492''; BM, Baltimore, 39°20'51.5904''; −76°41'5.481''; NMX, northern Mexico; DC, Washington,

DC, 38°55'17.6916''; −77°1'8.2812''; and Africa (NML, northern Mali, 16°46'43.1646'' −3°0'36.9144''; SL, Morocco 31°47'30.1272''N; 7°5'33.4320''W; NNA, northern Nigeria; SN, Sudan, 15°38'N 032°32'E; NR, Mali-Niger River. Blank map was downloaded from <http://www.smartdraw.com>

Mexico (remote); Washington, DC (urban); Africa [northern Mali (remote)]; Morocco (urban); [northern Nigeria (remote)]; Sudan (urban), and Mali-Niger River (Fig. 1). About 25-g subsurface soil (6 inches from surface) was collected using a pre-sterilized stainless steel spade. Samples were immediately removed from direct sunlight and refrigerated (−20 °C) within 24 h.

Standard microbiological protocols were observed in storing, isolating, and evaluating the phenotypic characteristics of viable culturable bacteria (VCB). Briefly, standard plate count was performed by suspending about 1-g soil sample into synthetic broth, serially diluted (10 fold), and aliquots (100 µl) plated in duplicate onto R2A (Reasoner's 2 agar) and trypticase soy agar (TSA). R2A was prepared by mixing (g/L) the following: proteose peptone, caseino acids, yeast extract, dextrose, soluble starch, 0.6 g dipotassium phosphate, 0.1 g magnesium sulfate heptahydrate, 0.6 g sodium pyruvate, and 15 g agar, pH 7.2 at RTP. Plates were incubated at 25–30 °C for 24–48 h then evaluated. All media were sterilized at 121 °C for 15 min and stored at refrigeration

temperature until used. Chemicals were ascertained from Fisher Scientific (New Jersey, USA). In addition, the genetic diversity (H , Shannon–Wiener indices) and evenness (E_H) of all soil samples were determined according to the following equations.

$$H = - \sum P_i \ln P_i = H = - \sum (N_i/N) \ln(N_i/N) \quad (1)$$

$$E_H = H/\ln S \quad (2)$$

where P_i is the ratio between the number in a specific group (N_i) and the total number (N), S is the total number of morphotypes in a specific soil/medium. Representative isolates were randomly collected from each plate, subcultured on respective media (R2A, TSA), incubated at 25–30 °C, and further evaluated using a Voyager DE-STR MALDI-TOF MS (Applied Biosystems, Japan). The Shannon–Wiener diversity index (SWDI) is quantitative and reflects the different types (such as species) there are in a data set. Furthermore, the index concurrently evaluates how evenly the basic entities are distributed among those

types and increases both when the number of types increases and when evenness increases. In addition, the similarity between sample communities was evaluated using Jaccard's coefficient (S_j) (Eq. 3):

$$S_j = \frac{a}{a + b + c} \quad (3)$$

where 'a' represents a joint occurrence of species in two samples 'a' and 'b'; and 'b,' the number of species in sample 'b,' but not sample 'a'; and 'c' represents the number of species in sample 'a,' but not sample 'b.' Phylogenetic relationships were ascertained as outline elsewhere (Dereeper et al. 2008).

2.2 Molecular analysis

The 16S rRNA sequencing was performed with the primers (forward) AGAGTTTGATCCTGGCTCAG, (reverse) ACGGCTACCTTGTTACGACTT and the following parameters: denaturation at 95 °C for 5 min followed by 35 cycles of denaturation at 95 °C for 45 s, annealing at 51 °C for 45 s, extension at 72 °C for 2 min, and a final extension at 72 °C for 10 min (Genewiz® Inc., South Plainfield, NJ, USA). 16S rRNA gene sequences were then compared with those in Genbank (<http://www.ncbi.nlm.nih.gov>). Unique sequences were assigned accession numbers by NCBI-BankIt (<http://www.ncbi.nlm.nih.gov>). CLUSTAL/W2 multiple sequence alignment software (<http://www.ebi.ac.uk/Tools/msa/clustalw2>) and phylogeny.fr (<http://www.phylogeny.fr>, Dereeper et al. 2008) were used to construct phylogenetic relationships among sequenced isolates. Phylogenetic relationships were presented as neighbor-joining bootstrap tree using *Methanocaldococcus* spp. as the out-group.

2.3 Maldi-TOF MS

About eleven type cultures were obtained from either the ATCC (American Type Culture Collection, Bethesda, MD, USA) or previously characterized isolates (sequences deposited into Genbank) and used for the standardization of the protocol for MALDI-TOF MS. These include *Enterobacter aerogenes* (ATCC 35028), *Escherichia coli* 0157:H7, *Salmonella typhimurium* (ATCC 14028), *Pseudomonas fluorescens* (ATCC 13525), *Enterococcus faecium* (ATCC 51559), *Streptomyces* sp. JM3 (JN166713), and members from the genus *Bacillus*. Post-culturing, respective isolates were

transferred directly to MALDI-TOF target and mixed with 1.0 µl matrix: α-cyano-4-hydroxycinnamic acid (CHCA) or 3,5-dimethoxy-4-hydroxycinnamic acid (Sinapinic acid), in trifluoroacetic acid: acetonitrile. The matrices and TFA were acquired from Aldrich Chemical Company (Milwaukee, WI, USA). Data were acquired with optimized protocol, which include reflector mode at 20 kV, grid voltage 94 % with 150 transients. Data analysis was performed with the Applied Biosystems Voyager DE System 4098 software. The reproducibility of the MALDI spectra was tested by running duplicate samples. Bovine serum albumin was used as a control.

Data ascertained by MALDI-TOF MS were evaluated by principal component analysis (PCA). The method is multivariate, recognizes patterns in large data sets, and determines the projection with the largest gross variance. Therefore, for this analysis, the matrix was defined as bacteria (samples) by row and m/z (variable) by column. Subsequently, a covariance matrix was generated from the data matrix, and an eigen-analysis performed on the covariance matrix. This resulted in a set of eigenvalues and eigenvectors. These variables are orthogonal and linearly independent to each other. The eigenvector associated with each eigenvalue represents the amount of variance for the data set. PCA was performed with IBM SPSS statistics 20 software (IBM Corp. NY, USA).

2.4 Characterization of biomarkers

Global proteins, insoluble and soluble protein digests, were extracted from lysed 18 to 48 h cultures based on a modified protocol found elsewhere (Liu et al. 2007). Briefly, bacterial pellets were collected, and a solution containing 10 % trifluoroacetic acid (TFA, Aldrich Milwaukee, WI, USA) in acetone and 2 % β-mercaptoethanol (Fisher Scientific, Pittsburg, PA, USA) added. Proteins were precipitated for 18 h at −20 °C and precipitate collected by centrifuging at 5,000 g for 30 min, and washed thrice with 10 ml ice-cold acetone. Proteins were then concentrated (Savant, Global Medical Instrument, MN, USA) and frozen at −80 °C until utilized for SDS-PAGE and peptide mass fingerprinting (PMF). A standard recipe was used for 12.5 % denaturing gel SDS-PAGE. Samples were added in a 5:1 ratio (sample: sample buffer), and the gel electrophoresized at 100 V for 2.5 h. High molecular protein standard (Sigma Chemical

Company, St. Louis, Mo, USA) was used as the marker. Gels were stained with Coomassie Brilliant Blue R-250 (Biorad, Hercules, CA, USA), and images obtained (Kodak Gel Logic Imaging System, Rochester, NY, USA).

A standard method was used for in-gel tryptic digest of SDS-PAGE-isolated proteins. Protein bands were dissected (1–2 mm) from gel and destained with a solution containing 20 mM ammonium hydrogen carbonate (NH_4HCO_3)/50 % acetonitrile (ACN) at RTP. Gel pieces were then shrunk by adding 50 μl acetonitrile for 15 min at RTP and residual ACN removed by speed vacuuming. Gel pieces were washed, vortexed for 30 min (10 min interval) with 200 μl solution containing 50 % ACN and 50 mM NH_4HCO_3 . About 200 μl ACN was then added, and pieces incubated for 5 min at RTP, and ACN removed as before. Dried gel pieces were reduced and alkylated using 100 μl , 10 mM DTT (Sigma, St. Louis, MO, USA) in 50 mM NH_4HCO_3 (1 h, 56 °C), and 55 mM iodoacetamide (Sigma-Aldrich, St. Louis, MO, USA) in 50 mM NH_4HCO_3 (RTP, 45 min, dark), respectively. Protein was digested using 10 μl of 100 ng/ μl trypsin (Promega, Madison, WI, USA) for 18–24 h at 37 °C and peptides concentrated using C18 Zip TipTM (Millipore, Billerica, MA, USA) according to manufacturer and analyzed using MALDI-TOF MS. Briefly, purified peptides were mixed with α -cyano-4-hydroxycinnamic acid (10 mg/ml 50 % acetonitrile/1 % TFA) in a 2:5 μl ratio and 1 μl applied to target, allowed to dry freely, and data acquired as mentioned before. Peptide maps were generated using MASCOT/Swissprot peptide mass fingerprint software (<http://www.matrixscience.com>). The parameters include database, Swissprot; peptide tolerance ± 1.2 Da; mass values, MH^+ , and mono-isotopic. Furthermore, peptides were compared against NCBI Blastp with Delta-Blast (Domain Enhanced Lookup Time Accelerated BLAST) algorithm (<http://www.ncbi.nlm.nih.gov/blast>). Peptides with indicated conserved domains were further compared against NCBI's conserved domain database (CDD v3.11-45746 PSSMs) with expect value (E) threshold of 1.0×10^{-4} , (<http://www.ncbi.nlm.nih.gov/Structure/cdd>). E values in the range of ≥ 1 were considered putative false-positives (Marchler-Bauer et al. 2011).

3 Results

3.1 Phenotypic and molecular characterization

Bacterial load (cfu/g soil), as determined by standard plate count, ranged from 10^1 to 10^7 with highest proportions on medium R2A (Table 1) for N. Mali ($p < 0.05$). Diversity (H , Shannon–Wiener), calculated based on phenotypes and 16S rRNA sequencing, identified N. Mali and Niger River (Mali) as containing the most diverse/even bacterial populations respectively (Table 1; Fig. 2). Furthermore, bacterial similarity among all locations, evaluated using Jaccard's coefficient, identified Jamaica and N. Nigeria as the two most similar locations (S_j , 0.38), whereas only very low to no similarity was observed for all other locations (S_j , 0.0 to <0.38). In addition, the acidic soil ($\text{pH } 6 \pm 0.02$) sample from Morocco had the lowest cfu/g soil and diversity. However, the most basic soil ($\text{pH } 8 \pm 0.02$) from Niger River had the highest cfu/g soil, diversity, and evenness. Nevertheless, a plot of the percentage change in cfu/g versus geographical location indicates a distinct east to west trend of decreasing distribution (Sudan, Nigeria, Mali) and a reverse trend, that is, an increasing percentage change from the west to North America. A significant difference between the average cfu/g soil on media R2A ($5.6 \times 10^6/3.4 \times 10^5$) and TSA ($3.8 \times 10^6/1.9 \times 10^5$) was observed for desert/non-desert locations, respectively. However, there was no significant difference between the average diversity (2.5/3.4) and evenness (0.35/0.5) for desert/non-desert locations, respectively.

Molecular analysis, performed using 16S rRNA sequencing, identified nineteen bacterial genera across all soil samples (Fig. 1a, b). The partial unique 16S rRNA sequences were submitted to the Genbank BankIt database under accession numbers: JX434166, JX434094, JX434114, JX434104, JX434130, JX434149, JX434148, JX434126, JX434129, JX434120, JX434143, JX434140, JX434160, JX434137, JX434116, JX434092, JX434139, and JX434164. Phylogenetic analysis of all viable culturable bacteria indicated the following groups (Fig. 3): Gram-negative bacilli which include *Pseudomonas* (3 %), *Pontibacter* (2 %), *Microvirga* (1 %); Gram-negative cocci, *Paracoccus* (1 %); Gram-positive endospore-forming bacilli, *Paenibacillus* (8 %), *Brevibacillus* (7 %), *Bacillus* (58 %), *Aneurinibacillus* (2 %), *Terribacillus* (3 %); Gram-positive non-endospore-

Table 1 Comparison of CFU/g, diversity, and evenness of bacteria from various soils on the media R2A (Reasoner's 2 agar) and TSA (trypticase soy agar)

Sample location ^a	CFU/g ^b		Diversity (H) ^c $H = - \sum (N_i/N) \ln(N_i/N)(EH)$	Evenness ^d $EH = H/\ln(S)$
	R2A ^e	TSA ^f		
Sudan	1.8×10^5	8.3×10^4	1.95 ⁱ	0.3 ⁱ
N. Nigeria	2.3×10^2	2.5×10^1	1.38	0.14
Mali (Niger River)	$1.1 \times 10^{6*}$	6.6×10^5	8.4*	0.66
N. Mali ^g	2.2×10^7	1.5×10^7	3.16	0.44
Morocco	5.0×10^1	7.0×10^2	1.0	0.72
W. Jamaica ^h	$2.1 \times 10^{5*}$	1.5×10^5	1.45	0.18
N. Mexico	1.0×10^3	5.0×10^2	3.97	0.62
Baltimore	1.8×10^5	1.3×10^5	2.32	0.24
Washington, DC	$2.4 \times 10^{5*}$	9.0×10^4	3.3	0.45

^a Sample location, arrangement on table based on easternmost to westernmost direction^b CFU, colony-forming units/gram soil^c Diversity (Shannon–Wiener), microbial composition of sample^d Evenness, relative abundance of rear, and common species^e R2A, Reasoner's 2 Agar, incomplete medium used for stressed bacteria^f TSA, trypticase soy broth agar, ^g northern, ^h westernⁱ Shannon–Weiner calculations based on 16S rRNA sequencing

* $p < 0.05$ for SWDI between Niger and (N. Nigeria, N. Mali, W. Jamaica, N. Mexico, Baltimore and Washington, DC, USA); $p < 0.05$, for SWDI of N. Mali versus (W. Jamaica, N. Mexico, Baltimore, Washington, DC); $p < 0.05$ for SWDI for N. Mexico versus (Baltimore and Washington, DC, USA); $p < 0.05$ for SWDI between Baltimore and Washington, DC

forming cocci, *Janibacter* (1 %), *Staphylococcus* (1 %), *Rhodococcus* (2 %), *Microbacterium* (3 %), *Micrococcus* (1 %), *Kocuria* (2 %), *Arthrobacter* (1 %), *Rothia* (1 %), *Janibacter* (1 %), *Gordonia* (1 %), and *Streptomyces* (1 %). *Staphylococcus* sp. NRT5 was removed from the phylogenetic tree since it did not fit into an appropriate group. *Methanocaldococcus* spp. was used as an out-group, and the data were presented as neighbor-joining bootstrap tree (1,000 bootstraps). Isolates were cultured at 25–30 °C for 18–24 h prior to sequencing. Identities of isolates are based on sequence similarity (≥ 98 %) using NCBI Blast.

In addition, a similarity of 60–100 % was established among isolates using ClustalW2 multiple sequence analysis software. However, only some members from the genera *Bacillus*, *Terribacillus*, *Paenibacillus*, and *Brevibacillus* had ≥ 99 –100 % similarity and were common among several regions. For example, *Bacillus* sp. BMT3 (>98 % similar to *B. subtilis* in NCBI BLAST) and BMT8 (*B. aerophilus*) isolated from Baltimore were between 99 and 100 % similar to isolates from Sudan, N. Mali, Morocco, N. Nigeria, Jamaica. Likewise, *Bacillus*

sp. NMXR2 (*Bacillus litoralis*) and *Bacillus* sp. NMXR5 (*B. luciferensis*) from N. Mexico were 99 % similar to isolates from Washington, DC and W. Jamaica. Furthermore, *Brevibacillus* sp. HDR1 (*B. formosus*) from Washington, DC was 99 % similar to isolates from Baltimore and N. Mexico; *Paenibacillus* spp. 99 % similar to an isolate from N. Nigeria; *Terribacillus* sp. NMLT17 (*T. saccharophilus*) from N. Mali, 99 % similar to isolate from N. Mexico.

3.2 MALDI-TOF MS analysis

The mass spectra generated from the analysis of viable culturable isolates exhibited a complex mixture of discrete ions between 2,000 and 10,000 m/z (Fig. 3; Table 2). Most spectra were easily distinguishable by visual inspection and illustrate evidence of common and unique protein biomarkers. For example, the environmental isolate *B. subtilis* had unique biomarkers at m/z 6,949 (73 %) and 9,171 (60 %). Similarly, unique biomarkers were shown for *B. thuringiensis* at m/z 4,809 (100 %), 5,047, 5,919 (83 %); *B. aerophilus*, 6,778 (75 %); *B. acidecelator*, 9,437 (100 %); *B.*

Fig. 2 Characterization of nineteen genera of bacteria across nine soil samples based on 16S rRNA sequencing. **a** Diversity (H') and evenness (E_H) of bacteria, ascertained using the Shannon–Wiener diversity index under varying pH; arrangement of sample location is based on easternmost to westernmost direction and **b** percentage of bacteria genera/g soil based on location. Isolates were cultured at 25–30 °C for 18–24 h prior to sequencing. Identities of isolates are based on sequence similarity (≥ 98 %) using NCBI Blast

megaterium, 4,634 (91 %); 5,156, 7,600 (63 %); *Brevibacillus* spp. 7,381 (86 %); *Paenibacillus* spp. 5,473 (63 %), and *Terribacillus* spp. 4,517, 6,532, 7,574–67 % (Fig. 4; Table 2). Similar ion profiles were identified from replicate experiments. In addition, some biomarkers of the isolate *B. thuringiensis* were consistently expressed over 330 days (m/z 4,816.890, 5,023.33), while others only at 18 h, m/z 3,706.74, 5,722.79, 7,380.84.

PCA, utilized to verify the presence of trends within the data set, indicated a Kaiser–Myer–Olkin measure of sampling adequacy and Bartlett’s test of Sphericity of 0.689 and 0.8, respectively. Therefore, three eigenvalues could be extracted with extraction sum of squared loadings (% variance) of 35.33, 24.21, and 15.01 for components 1–3, respectively (total variance 74.55 %). Three dimensions in the component space accounted for 74.55 % of the variance. Observation of the correlation

matrix indicates a weak association among variables tested. The communalities, which represent the proportion of each variable’s variance, indicated an initial value of 1.0. Nevertheless, the extraction, defined as the proportion of each variable’s variance that can be explained by the principal components, indicated variables with high values (0.5–0.9). These values suggest that variables are well represented in the common factor space. Twelve components were identified from the analysis, albeit only three (components 1–3) accounted for the greatest variance. This is confirmed by the analysis of the scree plot—plot of eigenvalue against component number (figure not shown).

The component plot in rotated space (Fig. 5a) clearly shows three groupings of variables associated with each major component: 1–5, 6–8, and 9–12. Variables 1–5 (0.71, 0.765, 0.715, 0.545, 0.376, respectively) were strongly associated with component

Fig. 3 Phylogenetic tree illustrating the relationship among eighteen genera of soil isolates using 16S rRNA sequences (>600 bp) and with *methanocaldococcus* spp. as the out-group. Data are presented as neighbor-joining bootstrap tree. The Genbank accession number is shown for most isolates after the species name. The scale (0.3) indicates the number of substitutions per position; bootstrap (1,000). The closest phylogenetic affiliation/genbank accession numbers for isolates are as follows, NRTX, *Pontibacter korlensis*/NR044004.1; NRT2, *Rhodococcus ruber*/NR026185.1; NRT7, *Kocuria marina*/NR025723.1;

NRTY, *Janibacter anopheles*/NR 043218.1; BMT3, *Bacillus subtilis*/NR 027552.1; NRT1, *Microbacterium paraoxydans*/NR 025548.1; NRR13, *Paracoccus denitrificans*/NR026456.1; NRR11, *Gordonia westfalica*/NR 025468.1/NMLR4, *Microvirga guangxiensis*/NR 044563.1/BMT11, *Brevibacillus brevis*/NR 041524.1; NNAR8, *Paenibacillus kribbensis*/NR025169.1; NSNTu, *Aneurinibacillus migulanus*/NR 036799.1; BMT15, *Terribacillus saccharophilus*/NR 041356.1; BMR1, *Rothia dentocariosa*/NR044712.1; JMT1, *Arthrobacter ramosus*/NR 026193.1; NRT79, *Pseudomonas stutzeri*/NR 041715.1

2; 9–12 (0.718, 0.779, 0.737, .652) with component 1; and 6–8 with component 3 (−0.581, −0.717, −0.467). Two major clusters were identifiable from the PCA for the evaluation of biomarkers (m/z) desorbed from whole cells during MALDI-TOF MS (Fig. 5b); these include *Brevibacillus* (represented by ‘n’) and *Bacillus* (*subtilis* (‘c’); *B. megaterium* (‘f’); *Bacillus thuringiensis* (‘g’), *Terribacillus* (‘x’), and *Paenibacillus* (‘u’). There appears to be an association between some bacteria and multiple sample locations (Fig. 5c). For example, the clustering to the upper right consists of *Paenibacillus* spp. from N. Mali, Sudan and Baltimore; similarly, *B. subtilis* was distributed among Sudan, Niger River, Baltimore, Washington, DC, and Nigeria.

3.3 Characterization of biomarkers

3.3.1 In-gel tryptic digest/peptide mass fingerprinting

The representative SDS-PAGE (15 %) profile for whole cell proteins precipitated from various *Bacillus*

spp. is shown in Fig. 6a. The apparent molecular weights (M_w), occurred between 15 and 115 kDa for representative isolates. Partial peptide maps (Fig. 6b, c) were produced using Expasy, MASCOT peptide mass fingerprint software (<http://www.matrixscience.com>). Therefore, a protein from isolate *Brevibacillus* sp. BMT11 was similar to hemagglutinin (HA), and another from *Bacillus* sp. BMT3 to filamentous hemagglutinin (FHA) with protein scores of 50/70 (22 % coverage) and 52/70 (8 % coverage), respectively (Table 3a). Similarly, the spectrum for the control, myosin (not shown), ranged from m/z 714 to 5,600 and illustrates well-resolved peptides with varying masses. Only one protein had a significant score ($p < 0.05$) of 70/70 (33 % coverage) and was identified as ATP synthase subunit beta. Protein scores were automatically determined by MASCOT software using $-10 \cdot \log(P)$, where p is the probability that the observed match is a random event. Protein scores greater than or equal to maximum score are considered significant ($p < 0.05$).

Table 2 Protein markers identified by MALDI-TOF MS analysis of whole cells for members of type and viable culturable environmental isolates from the genera *Bacillus*, *Brevibacillus*, *Paenibacillus*, and *Terribacillus*

Isolate	Common markers ^a (m/z)	Unique markers ^b (m/z)
Type ^c		
<i>Bacillus subtilis</i>	2,047, 3,364, 3,530	3,014, 3,118
<i>Bacillus thuringiensis</i>	3,350, 7,370, 7,338	6,423, 5,900, 5,041, 6,721, 8,356, 10,090
Environment ^d		
<i>Bacillus subtilis</i>	3,152 [2/15]; 3,590 [5/15]; 3,352 [2/15]; 4,320 [4/15]; 5,006 [3/15]; 6,949 [11/15]; 9,171 [9/15]	3,406, 4,378, 7,794, 7,134, 7,284, 8,690, 7,305, 8,209, 8,425, 8,797, 7,494, 9,087
<i>Bacillus thuringiensis</i>	3,052 [2/6]; 3,002 [4/6]; 3,627 [2/6]; 3,807 [3/6]; 4,809 [6/6]; 5,047 [5/6]; 5,460 [4/6]; 5,919 [5/6]; 7,398 [3/6]; 7,372 [3/6]	3,533, 3,194, 4,106, 6,256, 7,564, 8,008, 9,884, 10,386, 6,220, 6,844
<i>Bacillus aerophilus</i>	3,588 [2/4]; 6,778 [3/4]	3,746, 4,050, 6,072, 6,103, 7,248
<i>Bacillus acidceles</i>	3,191 [2/3]; 4,672 [2/3]; 6,712 [2/3]; 9,437 [3/3]	6,806, 6,949, 7,885, 9,171
<i>Bacillus megaterium</i>	3,912 [3/11]; 4,634 [10/11]; 4,632 [6/11]; 5,156 [7/11]; 6,265 [4/11]; 7,600 [7/11]	3,164, 6,906, 7,318, 7,829
<i>Bacillus safensis</i>	3,049 [2/2]; 3,624 [2/2]; 5,236 [2/2]; 5,861 [2/2]; 7,250 [2/2]; 9,600 [2/2]	4,346, 6,223, 7,826
<i>Brevibacillus</i> spp.	3,532 [3/7]; 7,381 [6/7]; 7,407 [5/7]	8,156, 9,513, 7,242
<i>Paenibacillus</i> spp.	3,824 [3/8]; 3,004 [4/8]; 4,222 [3/8]; 4,753 [3/8]; 5,046 [4/8]; 5,473 [5/8]; 5,861 [3/8]; 5,918 [3/8]; 6,035 [2/8]; 6,213 [2/8]; 6,440 [2/8]; 7,252 [3/8]; 7,370 [2/8]; 7,733 [2/8]; 8,021 [3/8]; 8,990 [4/8]; 7,130 [3/8]; 7,150 [4/8]; 7,312 [3/8]	2,688, 3,734, 5,252, 3,359
<i>Terribacillus</i> spp.	4,517 [2/3]; 6,532 [2/3]; 7,574 [2/3]	5,776, 3,903, 6,303, 3,419, 6,719

All markers were identified by MALDI-TOF MS in reflector mode, 20 kV with CHCA matrix. Whole

^a Protein markers shared by 80–85 % of all members of a specified genus

^b Protein markers not shared by members of the same genus

^c American Type Culture Collection (ATCC) bacteria used as standard

^d Environmental isolates

Fig. 4 Representative MALDI-TOF mass spectrum of *Bacillus thuringiensis* post-18-h incubation (initial sampling) at 30 °C. Data were ascertained at 20 kV, 150 transients with 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), and

10 mg/ml 0.1 % TFA/ACN as the matrix. Comparison of common peaks observed for environmental and type isolates at both 18 h and 330 days: peak identified at only 18 h (*); peak identified at both 18 h and 330 days (**)

The peptides identified for *Brevibacillus* sp. BMT11 (B2), searched against NCBI Delta-Blast, indicated a 57 % similarity with hemagglutinin with possibly conserved domains. Furthermore, NCBI's conserved domain database (CDD) indicated a similarity between the queried and the haemagglutinin superfamily member with accession number c110565, *e* value of 6.80e^{-44} (Table 3b). Similarly, a NCBI Delta-Blast and CDD search of the partial peptides for *Brevibacillus* sp. BMT11 (B1) identified the peptides as belonging to the ATP synthase F1 beta subunit (*e* value, 4.01e^{-54} , threshold 1×10^{-4}) with conserved sequences for the ATP binding site, Walker B motif, and regions belonging to the RecA-like NTPases superfamily with *e* value 4.01e^{-54} .

4 Discussion

In the current study, nineteen genera of bacteria were identified and characterized across nine soil samples, from several countries in northern Africa and the Americas, using MALDI-TOF MS and 16S rRNA sequencing. The observed difference ($p < 0.05$) between the mean bacterial load (cfu/g soil) for the media utilized (R2A and TSA) arose due to R2A's nutrient limitation, which supports the recovery of stressed bacteria and minimizes the growth of isolates

with shorter generation times. Although higher bacterial loads were shown in samples from N. Mali (10^7) and the Niger River (Mali) (10^5), these concentrations may represent <1 % of the total viable culturable microbiota in any soil (González-Toril et al. 2009; Prospero et al. 2005). Thus, culturing, while significant, inadvertently restricts the composition and number of microbes that could be detected. The evaluation of bacterial diversity (*H*) and evenness (*E_H*) across all samples using Shannon–Wiener diversity index illustrates that either the usage of morphotypes or 16S rRNA sequencing could derive an estimation of these indices. Thus, Niger and Morocco contained the most and least diverse bacterial populations, respectively. Furthermore, (*H*, $p < 0.05$) value was shown for Mali (Niger River), N. Mali, and N. Mexico when compared to other regions. All other communities had similar evenness and were dominated by the genus *Bacillus*. Further evaluation of the similarity between communities (sample locations) using Jaccard's coefficient identified the two most similar communities as Jamaica and N. Nigeria (*Sj*, 0.38), even though these locations have very dissimilar climates. Thus, W. Jamaica is tropical, loamy soil ($\text{pH } 7.6 \pm 0.02$) and N. Nigeria, arid, sandy-clayey soil ($\text{pH } 7$). The computed values (*Sj*) for other communities were $\leq 0.38 \geq 0.0$, where a value of 1.0 and 0.0 reflects maximum and no similarities, respectively, between two communities.

Fig. 5 Principal component analysis (PCA) of isolates involved in possible transfers from N. Sudan to Baltimore/DC. **a** Component plot in rotated space for variables (m/z), **b** distribution of isolates/genera (c, *B. subtilis*; f, *B. megaterium*; g, *B. thuringiensis*; n, *Brevibacillus* spp.; u, *Paenibacillus* spp.; x, *Terribacillus* spp.) most common to locations; **c** distribution of bacteria across nine locations (N, Niger River; S, Sudan; ML,

N, Mali; B, Baltimore; W, Washington, DC; X, N. Mexico; NA, N. Nigeria; J, Jamaica; SL, Morocco). Kaiser–Myer–Olkin measure of sampling adequacy, 0.689 and Bartlett’s test of Sphericity, 0.8. Three components were extracted (from scree plot, not shown), with extraction sum of squared loadings (% variance) of 35.33, 24.21, and 15.01 for components 1–3, respectively (total variance 74.55 %)

Since it was unlikely to molecularly tag bacteria in soil sources and follow their distribution from Sudan through Mali, W. Jamaica, N. Mexico, Baltimore, and Washington, DC, pair-wise sequence alignment of 16S rRNA gene sequences identified the genera *Bacillus*, *Brevibacillus*, *Paenibacillus*, and *Terribacillus* as those most similar among locations evaluated. This observation is significant since, to our knowledge, no previous study has established such among the locations investigated. Likewise, this confirms that similarity of these genera

among locations did not result from background contamination and might be similar to, but not governed by, postulates outlined elsewhere since soil and air (dust) samples are two different environments (Virginia et al. 2003). Analysis of total ozone mapping spectrometer (TOMS) satellite images suggest that Mali is a source area for dust traveling westward (Prospero 1999; Prospero et al. 2005; Gillies and Nickling 1996), and our data suggest that regions as far east as Sudan (or mineral dust source regions possessing similar microbial loads) may

Fig. 6 **a** 15 %-SDS-PAGE profile for whole bacteria cell proteins, Lane 1, ladder (Benchmark Pre-stained protein ladder, Invitrogen); 2, *Bacillus mojavensis* ATCC; 3, *Bacillus* sp. BMT4; 4, *Bacillus* sp. BMT3; 5, *Bacillus* sp. NML19; 6, *Bacillus* sp. NML12; 7, *Bacillus* sp. JMT4; **b** Representative MALDI-TOF mass spectrum of in-gel-tryptic digest peptides of soil isolate *Bacillus* sp. BMT3 (bright band on gel). Data were

acquired in reflector mode at 25 kV with 150 transients using sinapinic acid as the matrix, **c** partial peptide map illustrating matched peptides from MALDI-TOF MS trypsinized protein for isolate. Colored (red) regions illustrate peptides matched. Protein was presumptively identified as filamentous hemagglutinin. See Table 3a for summary

also be potential source regions even if they are not associated with the heaviest dust. Many studies have indicated northern transfers of Saharan dust into Europe, and Asian dust (Gobi desert) into Washington State (Ke-Yi 2010). The migratory genera identified in our study are similar to previous studies, wherein air samples collected in Barbados from dust events occurring in 1812 identified the genera *Brevibacillus*, *Bacillus*, and *Paenibacillus* (Gorbushina et al. 2007; Kellogg et al. 2004). Other reports have identified between 5–27 % pathogens (animal, plant, opportunistic human) in air samples collected from Mali, air samples, which identified 96 % Gram-positive, spore formers, *Bacillus* spp., 38 % and include *Kocuria*, *Staphylococcus aureus*, *Gordonia*, *Microbacterium*, *Micrococcus*, *Paenibacillus*, *Paracoccus*, *Rhodococcus* (Kellogg et al. 2004). Similarly, our samples from Mali (Niger River and N. Mali) identified the genera *Bacillus*, *Staphylococcus*, *Microbacterium*, *Streptomyces*, *Micrococcus*, *Pontibacter*, *Pseudomonas*, *Paracoccus*, *Kocuria*; *Paenibacillus*; *Brevibacillus*, *Gordonia*; Nigeria—*Kocuria*, *Staphylococcus*, and *Stenotrophomonas maltophilia* with the prevalent bacteria

species of novel phylotypes was *Bacillus fusiformis* (Pasteur et al. 2002). We identified *Bacillus* and *Paenibacillus* in the samples from N. Nigeria. Similar locations which were previously evaluated by others include St. Thomas, US Virgin Island. *Bacillus* spp., *Arthrobacter* spp., *Kocuria* spp. *Microbacterium* spp., *Paracoccus* spp., *Pseudomonas* spp., *Streptomyces* spp. among other bacteria not identified in our current study (Griffin et al. 2003); Chad, the following genera among others from sand and dust samples: *Streptomyces*, *Nocardia*, *Micrococcus*, *Microbacterium*, *Pseudomonas*, *Bacillus*, *Staphylococcus*, *Stereptococcus*, *Lactobacillus*, *Paenibacillus*, and *Alcaligenes* (Favet et al. 2013).

Therefore, the higher proportions of these endospore formers in air reflect conditions at the source, soil (pH, low nutrients) and are consistent with our findings. For example, the soil sample from Morocco was shown to have pH at 6 ± 0.02 , the lowest diversity, evenness, and 100 % endospore-forming bacteria. Generally, higher cfu/g, diversity, and evenness were associated with near neutral to slightly alkaline pH ($7-7.3 \pm 0.02$) for other locations.

Table 3 (a) Summary of protein scores, matched peptides and sequence coverage from Mascot/SwissProt database, and NCBI's Delta-Blast data for representative soil isolates, (b) Graphical summary of conserved domain(s) on partial peptide map for hemagglutinin isolated from *Brevibacillus* sp. BMT11

Isolate	Presumptive protein	^a SCORE ($p < 0.05$)	^b PM	^c Cov(%)	^d NCBI Delta blast (Accession number)	^e Ident (%)
(a)						
<i>Bacillus</i> sp. BMT3	(B1) Filamentous hemagglutinin	52/70	15	8	filamentous hemagglutinin/ (gil566087614 WP_023999820.1)	38
<i>Brevibacillus</i> sp. BMT11	(B1) ATP synthase subunit beta	*70/70	9	33	ATP synthase F1, beta subunit/ (gil495990527 WP_008715106.1)	67
	(B2) Hemagglutinin	50/70	5	22	hemagglutinin/(gil123864011 Q38SQ8.1)	57
(b)						
<div> <div>10 20 30 40 50 60 70 80</div> <div> <div>lcl local_KTITNDQIEV 33 RSQQTVIPNIGSRPWVRG-----RSDAPIGICSSECTPN 67</div> <div>Cdd:pfam00509 205 TYNRSFVPEIGARPLVNGsggridfywtikpgdtitfesngnliaprygykisgrshssgiiKSDLPIGNCETKQTPK 284</div> <div>90 100 110 120 130 140 150 160</div> <div> <div>lcl local_KTITNDQIEV 68 GSIPNDKPFQNVNKITYGACPRY-----KQTRGIFGAIAGFIENGWEGMVDgwygfrhrhgifgaiagf 130</div> <div>Cdd:pfam00509 285 GGINTSKPFQNVSPITAGCEPKYvkskslkatglrnvpeIQSRGLFGAIAGFIEGWEGMID----- 347</div> <div>170 180 190</div> <div> <div>lcl local_KTITNDQIEV 131 iengwegmvdGWYGFRRHQNSEGTQAADLKS 161</div> <div>Cdd:pfam00509 348 -----GWYGFRRHQNAQGTGYAADLKS 368</div> </div> </div> </div> </div>						

^a Protein score is $-10 \times \log(P)$, where P is the probability that the observed match is a random event. Protein scores greater than or equal to maximum score are considered significant ($p < 0.05$). *Bacillus* sp. BMT3 similar to *B. subtilis*; *Brevibacillus* sp. BMT11 similar to *Brevibacillus brevis*

^b PM, peptides matched by database

^c Percent coverage of peptides matched to presumptive protein

^d Presumptive protein and accession number identified by NCBI Delta blast

^e Ident (%), percent similarity between presumptive and queried protein

(b), graphical summary of closest protein match, accession number cl10565, in NCBI conserved domain database (CDD), with Cd Length: 550 Bit Score: 152.50 e value: 6.80×10^{-44} and e value threshold 1.00×10^{-04} ; red color indicates matched amino acids and blue and black no matches; lcl (local_KTITNDQIEV) represents sample queried, and Cdd:pfam00509, the closest matched protein in CD database

The identification of bacteria previously implicated in various diseases is significant, for example, *Staphylococcus* sp. NRT5 (similar to *S. xylosus* KL 162, from NCBI blast), an isolate from Niger, was previously implicated as the causative agent for septicemia in a loggerhead turtle (*C. caretta*) off the Canary Islands (Torrent et al. 2004), other opportunistic infections in animals (Fthenakisi et al. 1994) and humans (Tselenis-Kotsowilis et al. 1982). Nevertheless, the identification of these genera and their associated chemotaxonomic signatures (genera specific protein markers) are valuable tools for rapid identification and could be used for the development of specific assays and database development. Therefore, environmental isolates such as *B. subtilis* had unique m/z at 6,949 (73 %), 9,171 (60 %); *B. thuringiensis* m/z 4,809 (100 %), 5,047, 5,919 (83 %); *B. aerophilus* m/z 6,778 (75 %); *B. acideceler* at m/z 9,437 (100 %); *B. megaterium* m/z 4,634 (91 %), 5,156, 7,600 (63 %); *Brevibacillus* spp. m/z 7,381 (86 %), 7,407 (71 %); *Paenibacillus* spp. m/z 5,473

(62 %); *Terribacillus* spp. m/z 4,517, 6,532, 7,574 (67 %). The ions at m/z 7,370 and 7,400 were common to all members of the type genus *Bacillus* except *B. subtilis* and were similar to previous observations (Krishnamurthy et al. 1996).

Although a qualitative association among isolates can be ascertained by visually analyzing the fingerprints (m/z) from MALDI-TOF spectra, this may be time-consuming for large data sets; therefore, PCA was used to determine any association among isolates. Three eigenvalues were extracted from the scree plot with an extraction sum of squared loading (% variance) of 74.6 %, and with a Kaiser–Meyer–Olkin measure of sample adequacy value of 0.68 is the suggested minimum. The Bartlett's test of sphericity was utilized to test the null hypothesis that the correlation matrix is an identity matrix. Since the significance was >0.05 , the null hypothesis was rejected. Taken together, these tests provide a minimum standard, which should be passed before a principal component analysis is conducted. Several associations were obvious for the distribution

among genera and their distribution among locations and may indicate the transference of bacteria among locations. This further confirms the relevance of utilizing PCA for the evaluation of large data sets. However, the distribution suggests that clusters are based on some other factor, which is not shared among isolates from specific genera, but by other factor(s), which is/are common to isolates (same genus/species) from differing locations.

In addition, the presumptive identification of partial peptide maps for several potential virulent factors is significant. These include hemagglutinin (HA) from *Brevibacillus* sp. BMT11, which could be associated with respiratory infections. These peptides were suggested by NCBI's conserved domain database (CDD) as having conserved domains belonging to the haemagglutinin superfamily and were similar to the member with accession number c110565. Other potential virulence factors suggested by MASCOT include phosphoglucosamine mutase for the formation of peptidoglycan precursors and was previously shown to contribute to the resistance of *Streptococcus gordonii* DL1 to polymorphonuclear leukocyte killing (Yajima et al. 2009); *Mycobacterium* spp. have been shown to produce structurally related phenolic glycolipid (PGL) antigens and glycosylated p-hydroxybenzoic acid methyl esters (p-HBAD), which are important virulence factors (Pérez et al. 2004); the siderophore, achromobactin, an important virulence factor of *Erwinia chrysanthemi* strain 3,937 is associated with soft rot disease in a variety of plants (Franza et al. 2005). Potential pathogenic bacteria may have acquired pathogenic genes through horizontal gene transfer and may be significant for understanding prey–predator relationships in non-infective natural ecosystems (Martinez 2013). Furthermore, of the peptide set evaluated by MASCOT, only one had a significant score ($p < 0.05$) with 70/70, 33 % coverage, and was identified as ATP synthase subunit beta. The known conservation ATP synthase was confirmed by the CCD, which identified conserved sequences for the ATP binding site, Walker B motif, and similarity to the RECA-like NTPases superfamily and multidomains consisting of atpD. ATP synthase has been shown to influence biofilm resistance to host defenses, particularly during inflammation (Paino et al. 2011). The presumptive protein, filamentous hemagglutinin (FHA), isolated from a *Bacillus* sp. BMT3, had 15 matched peptides, but only 8 % coverage ($p > 0.05$)

with no significant matches in the CCD. The presence of these presumptive virulent factors could have significant implications for the health (upper respiratory tract infections) of individuals. The isolate was 99 % similar to *Bacillus* spp. from Jamaica and N. Nigeria, and 100 % similar to other *Bacillus* spp. from Sudan and N. Mali. However, there is need for us to confirm these and other virulent factors highlighted in this study. Significant protein scores (partial peptides) were assessed by the Mascot database. Matched peptides with sequence coverage of ≥ 15 % and minimum of five matches were considered good (unambiguous) presumptive matches (Mann et al. 2001).

The current study represents the first published report to evaluate the diversity, establish the possible transfer of several genera of viable culturable bacteria (VCB) among the aforementioned regions, and elucidate some of their chemotaxonomic signatures. Although the evaluation of soil microbial diversity has limitations arising from taxonomy and methodology, the data suggest that the combination of MALDI-TOF MS and 16SrRNA sequencing may constitute a powerful tool to distinguish among VCBs from environmental sources. Furthermore, since the time-to-result is minimal and unique and/or common biomarkers can be rapidly identified, MALDI-TOF MS is an ideal support tool for characterizing microorganisms in the field, monitoring of environmental contaminations, and other potential hazards.

Acknowledgments This material is based upon (work) supported by the National Oceanic and Atmospheric Administration, Educational Partnership Program, US Department of Commerce, under Agreement No. NA11SEC4810003. This publication was made possible by the National Oceanic and Atmospheric Administration, Office of Education Educational Partnership Program award. Its contents are solely the responsibility of the award recipient and do not necessarily represent the official views of the US Department of Commerce, National Oceanic and Atmospheric Administration. We would also like to express sincere thanks to the Department of Biology and Chemistry, Howard University for providing the facilities to enable this research.

References

- Arnold, R. J., & Reilly, J. P. (1998). Fingerprint matching of *E. coli* strains with MALDI-TOF mass spectrometry of whole cells using a modified correlation approach. *Rapid Communications in Mass Spectrometry*, 12, 630–636.

- Bengali, C., Rossi, V., Dolina, M., Tonolla, M., & Petrini, O. (2011). Matrix assisted laser desorption ionization-time of flight mass spectrometry for the identification of clinically relevant bacteria. *PLoS ONE*, 6, 1–7.
- Bessède, E., Angla-Gre, M., Delagarde, Y., Sep Hieng, S., Ménard, A., et al. (2011). Matrix-assisted laser-desorption/ionization biotyper: experience in the routine of a University hospital. *Clinical Microbiology & Infection*, 17, 533–538.
- Cook, A. G., Weinstein, P., & Centeno, J. A. (2005). Health effects of natural dust—role of trace elements and compounds. *Biological Trace Element Research*, 103, 1–15.
- Cordwell, S. J., Wilkins, M. R., Cerpa-Poljak, A., Gooley, A. A., Duncan, M., et al. (1995). Cross-species identification of proteins separated by two dimensional gel electrophoresis using matrix-assisted laser desorption time of flight spectrometry and amino acid composition. *Electrophoresis*, 16, 438–443.
- De Longueville, F., Hountondji, Y.-C., Henry, S., & Ozer, P. (2010). What do we know about effects of desert dust on air quality and human health in West Africa compared to other regions? *Science of the Total Environment*, 409, 1–8.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet S., et al. (2008). *Phylogeny.fr: robust phylogenetic analysis for the non-specialist nucleic acids research 36 (web server issue):W465-9*. Electronic Publication, 2008 Apr 19. (PubMed).
- Favet, J., Lapanje, A., Giongo, A., Kennedy, S., Aung, Y.-Y., et al. (2013). Microbial hitchhikers on intercontinental dust: catching a lift in Chad. *The ISME Journal*, 7, 850–867.
- Franza, T., Mahe, B., & Expert, D. (2005). *Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection. *Molecular Microbiology*, 55, 261–275.
- Fthenakisi, G. C., Marples, R. R., Richardson, J. F., & Jones, J. E. T. (1994). Some properties of coagulase-negative staphylococci isolated from cases of bovine mastitis. *Epidemiology and Infection*, 112, 171–176.
- Garrison, V. H., Shinn, E. A., Foreman, W. T., Griffin, D. W., Holmes, C. W., et al. (2003). African and Asian dust: from desert soils to coral reefs. *BioScience*, 53, 469–480.
- Gillies, J. A., & Nickling, W. G. (1996). Dust concentrations and particle-size characteristics of an intense dust haze event: inland Delta region, Mali, West Africa. *Atmospheric Environment*, 30, 1081–1090.
- González-Toril, E., Amils, R., Delmas, R. J., Petit, J.-R., Komárek, J., & Elster, J. (2009). Bacterial diversity of autotrophic enriched cultures from remote, glacial Antarctic, Alpine and Andean aerosol, snow and soil samples. *Bio-geosciences*, 6, 33–44.
- Gorbushina, A. A., Kort, R., Schulte, A., Lazarus, D., Schnetger, B., et al. (2007). Life in Darwin's dust: intercontinental transport and survival of microbes in the nineteenth century. *Environmental Microbiology*, 9, 2911–2922.
- Griffin, D. W., Kellogg, C. A., Garrison, V. H., & Shinn, E. A. (2002). The global transport of dust: an atmospheric river of dust, micro-organisms and toxic chemicals crosses the oceans. *American Scientist*, 90, 228–235.
- Griffin, D. W., Kellogg, C. A., Garrison, V. H., Lisle, J. T., Borden, T. C., & Shinn, E. A. (2003). Atmospheric microbiology in the northern Caribbean during African dust events. *Aerobiologia*, 19, 143–157.
- Haag, A. M., Taylor, S. N., Johnston, K. H., & Cole, R. B. (1998). Rapid identification and speciation of *Haemophilus* bacteria by MALDI-TOF mass spectrometry. *Journal of Mass Spectrometry*, 33, 750–756.
- Howitt, M. E. (2000). Asthma management in the Caribbean—an update. *Postgrad Doctor Caribbean*, 16, 86–104.
- Kanatani, K. T., Ito, I., Al-Delaimy, W. K., Adachi, Y., et al. (2010). Desert dust exposure is associated with increased risk of asthma hospitalization in children. *American Journal of Respiratory and Critical Care Medicine*, 182, 1475–1481.
- Kellogg, C. A., & Griffin, D. W. (2006). Aerobiology and the global transport of desert dust. *Trends in Ecology & Evolution*, 21, 638–644.
- Kellogg, C. A., Griffin, D. W., Garrison, V. H., Peak, K. K., Royall, N., et al. (2004). Characterization of aerosolized bacteria and fungi from desert dust events in Mali, West Africa. *Aerobiologia*, 20, 99–110.
- Ke-Yi, C. (2010). The Northern path of Asian dust transport from the Gobi desert to North America. *Atmospheric and Oceanic Science Letters*, 3, 155–159.
- Krishnamurthy, T., Ross, P. L., & Rajamani, U. (1996). Detection of pathogenic and nonpathogenic bacteria by matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 10, 883–888.
- Lai, C. C., Wang, J. Y., Lin, S. H., Tan, C. K., Wang, C. Y., et al. (2011). Catheter-related bacteraemia and infective endocarditis caused by *Kocuria* species. *Clinical Microbiology & Infection*, 17, 190–192.
- Lay, J. O. (2001). MALDI-TOF mass spectrometry of bacteria. *Mass Spectrometry Reviews*, 20, 172–194.
- Leski, T. A., Malanoski, A. P., Gregory, M. J., Lin, B., & Stenger, D. A. (2011). Application of a broad-range resequencing array for detection of pathogens in desert dust samples from Kuwait and Iraq. *Applied and Environmental Microbiology*, 77, 4285–4292.
- Liu, H., Du, Z., Wang, J., & Yang, R. (2007). Universal sample preparation method for characterization of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Applied and Environmental Microbiology*, 73, 1899–1907.
- Mann, M., Hendrickson, C., & Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry*, 70, 437–473.
- Marchler-Bauer, A., et al. (2011). CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Research*, 39(D), 225–229.
- Martinez, J. (2013). Bacterial pathogens: from natural ecosystems to human hosts. *Environmental Microbiology*, 15, 325–333.
- McTainsh, G. H., Nickling, W. G., & Lynch, A. W. (1997). Dust deposition and particle size in Mali, West Africa. *Catena*, 29, 307–322.
- Monteil, M. A., Juman, S., Hassanally, R., et al. (2000). Descriptive epidemiology of asthma in Trinidad, West Indies. *Journal of Asthma*, 37, 677–684.
- Moura, H., Ospina, M., Woolfitt, A. R., Barr, J. R., & Visvesvera, G. S. (2003). Analysis of four human microsporidian

- isolates by MALDI-TOF mass spectrometry. *Journal of Eukaryotic Microbiology*, 50, 156–163.
- Nagy, E., Maier, T., Urban, E., Terhes, G., & Kostrzewa, M. (2009). ESCMID study group of antimicrobial resistance in anaerobic bacteria. Species identification of clinical isolates of *Bacteroides* by matrix-assisted laser-desorption/ionization time of flight mass spectrometry. *Clinical Microbiology & Infection*, 15, 796–802.
- Paino, A., Tuominen, H., Jääskeläinen, M., Alanko, J., Nuutila, J., et al. (2011). Trimeric form of intracellular ATP synthase subunit β of *Aggregatibacter actinomycetemcomitans* binds human Interleukin-1 β . *PLoS ONE*, 6, 1–11.
- Pasteur, B. J., Falkler, W. A., Jr, Enwonwu, C. O., Idigbe, E. O., Savage, K. O., et al. (2002). Prevalent bacterial species and novel phylotypes in advanced noma lesions. *Journal of Clinical Microbiology*, 40, 2187–2191.
- Pérez, E., Constant, P., Lemassu, A., Laval, F., & Daffé, M. (2004). Characterization of three glycosyltransferases involved in the biosynthesis of the phenolic glycolipid antigens from the *Mycobacterium tuberculosis* complex. *The Journal of biological chemistry*, 279, 42574–42583.
- Prospero, J. M. (1999). Long-term measurements of the transport of African mineral dust to the southeastern United States: implications for regional air quality. *Journal of Geophysical Research*, 104, 15917–15927.
- Prospero, J. M., Blades, E., Mathison, G., & Naidu, R. (2005). Interhemispheric transport of viable fungi and bacteria from Africa to the Caribbean with soil dust. *Aerobiologia*, 21, 1–19.
- Ruelle, V., El Moualij, B., Zorzi, W., Ledent, P., & De Pauw, E. (2004). Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 18, 2013–2019.
- Seng, P., Drancourt, M., Gouriet, F., Scola, B. L., Fournier, P. E., Rolain, J. M., et al. (2009). Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time of flight mass spectrometry. *Clinical Infectious Diseases*, 49, 543–551.
- Sing, D., & Sing, C. F. (2010). Impact of direct soil exposures from airborne dust and geophagy on human health. *International Journal of Environmental Research and Public Health*, 7, 1205–1223.
- Torrent, A., Deniz, S., Calabuig, P., & Oros, J. (2004). Septicemia due to *Staphylococcus xylosus* in a loggerhead sea turtle (*Caretta caretta*). <http://www.vet.uga.edu/esp/IVCVM/Oros1/>
- Tselenis-Kotsowilis, A. D., Koliomichalis, M. P., & Papavasiliou, J. T. (1982). Acute pyelonephritis caused by *Staphylococcus xylosus*. *Journal of Clinical Microbiology*, 16, 593–594.
- Virginia, H. G., Shinn, E. U., Foreman, W. T., Griffin, D. W., Holmes, C. W., Kellogg, C. A., et al. (2003). African and Asian dust: from desert soils to coral reefs. *BioScience*, 53, 469–480.
- Yajima, A., Takahashi, Y., Shimazu, K., Urano-Tashiro, Y., Uchikawa, Y., et al. (2009). Contribution of phosphoglucomutase to the resistance of *Streptococcus gordonii* DL1 to polymorphonuclear leukocyte killing. *FEMS Microbiology Letters*, 297, 196–202.