



HAL
open science

Metataxonomics of Tunisian phosphogypsum based on five bioinformatics pipelines: Insights for bioremediation

Houda Trifi, Afef Najjari, Wafa Achouak, mohamed Barakat, Kais Ghedira,
Faten Mrad, Mouldi Saidi, Haïtham Sghaier

► To cite this version:

Houda Trifi, Afef Najjari, Wafa Achouak, mohamed Barakat, Kais Ghedira, et al.. Metataxonomics of Tunisian phosphogypsum based on five bioinformatics pipelines: Insights for bioremediation. *Genomics*, 2019, 112 (1), pp.981-989. 10.1016/j.ygeno.2019.06.014 . cea-02327690

HAL Id: cea-02327690

<https://cea.hal.science/cea-02327690>

Submitted on 24 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Manuscript Details

Manuscript number	GENO_2019_16
Title	Metataxonomics of Tunisian phosphogypsum based on five bioinformatics pipelines: Insights for bioremediation
Short title	Metataxonomics of phosphogypsum: insights for its bioremediation
Article type	Research Paper

Abstract

Phosphogypsum (PG) is an acidic by-product from the phosphate fertilizer industry and it is characterized by a low nutrient availability and the presence of radionuclides and heavy metals which pose a serious problem in its management. Here, we have applied Illumina MiSeq sequencing technology and five bioinformatics pipelines to explore the phylogenetic communities in Tunisian PG. Taking One Codex as a reference method, we present the results of 16S-rDNA-gene-based metataxonomics abundances with four other alternative bioinformatics pipelines (MetaGenome Rapid Annotation using Subsystem Technology (MG-RAST), mothur, MICrobial Community Analysis (MICCA) and Quantitative Insights into Microbial Ecology (QIIME)), when analyzing the Tunisian PG. Importantly, based on 16S rDNA datasets, the functional capabilities of microbial communities of PG were deciphered. They suggested the presence of PG autochthonous bacteria recoverable into (1) removal of radioactive elements and toxic heavy metals (2) promotion of plant growth (3) oxidation and (4) reduction of sulfate.

Keywords	Bioremediation; Functional prediction; Illumina Miseq sequencing; Microbial community; Phosphogypsum
Taxonomy	Genetics, Bioinformatics, Ecology, Microbiology
Manuscript category	Methods Paper
Corresponding Author	houda TRIFI
Corresponding Author's Institution	National Center for Nuclear Sciences and Technology (CNSTN), Sidi Thabet Technopark
Order of Authors	houda TRIFI, afef najjari, Wafa ACHOUAK, Mohamed Barakat, Kais Ghedira, Faten MRAD, Mouldi SAIDI, Haïtham SGHAIER
Suggested reviewers	Maria Costa, Talitha Santini, Feth el Zahar Haichar, Noha Youssef

Submission Files Included in this PDF

File Name [File Type]

Cover letter.docx [Cover Letter]

Highlights.docx [Highlights]

Abstract.docx [Abstract]

Title page.docx [Title Page (with Author Details)]

Main document.docx [Manuscript File]

Figure 1.tiff [Figure]

Figure 2.tiff [Figure]

Table 1.docx [Table]

Table 2.docx [Table]

Table 3.docx [Table]

Conflict of interest.docx [Conflict of Interest]

Submission Files Not Included in this PDF

File Name [File Type]

Supplementary File 1.xlsx [Data in Brief]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Cover letter

Houda TRIFI

National Center for Nuclear Sciences

and Technology (CNSTN), Sidi Thabet, Tunisia

00216 25 867 494

trifi.houda@gmail.com

Dear Editor,

I am pleased to submit an original research article entitled "Metataxonomics of Tunisian phosphogypsum based on five Bioinformatics pipelines: Insights for bioremediation" by Trifi H, *et al*; to be considered for publication in *Genomics journal*. I hereby certify that this paper gives detailed picture and new information regarding the phylogenetic characterization of bacterial communities on phylum, class, order, family, genus and species levels in Tunisian phosphogypsum (PG) using Illumina MiSeq sequencing technology and five bioinformatics pipelines. We deciphered in this study indigenous bacteria which possess different traits such as multi-resistance to heavy metals, resistance to ionizing radiation, plant growth promoting capacities and the ability to oxidize or reduce sulfate compounds which provided a basis for exhaustive research leading to the development of an efficient tool for PG bioremediation. Our manuscript creates a paradigm for future studies of the eventual use of bacterial communities of PG in its bioremediation.

This study has not been duplicated publication or submission elsewhere, and if accepted, it will not be published elsewhere, including electronically in the same form, in English or in any other language, without the written consent of the copyright-holder. The authors received no financial support for the research and/or authorship of this article. The authors declare that they have no conflict of interest in the publication of this article, and they have participated in the study and concur with the submission and subsequent revisions submitted by the corresponding author.

Thank you for your consideration.

Best Regards,

Highlights

- The microbial communities of phosphogypsum (PG) were investigated using Illumina MiSeq.
- Five bioinformatics pipelines were used to analyze bacterial diversity of PG.
- The major phyla in Tunisian PG is Proteobacteria.
- The functional capabilities of microbial communities of PG were deciphered.
- Bacteria genera with bioremediation traits were detected in PG dumps.

Metataxonomics of Tunisian phosphogypsum based on five bioinformatics pipelines: Insights for bioremediation

Abstract

Phosphogypsum (PG) is an acidic by-product from the phosphate fertilizer industry and it is characterized by a low nutrient availability and the presence of radionuclides and heavy metals which pose a serious problem in its management. Here, we have applied Illumina MiSeq sequencing technology and five bioinformatics pipelines to explore the phylogenetic communities in Tunisian PG. Taking One Codex as a reference method, we present the results of 16S-rDNA-gene-based metataxonomics abundances with four other alternative bioinformatics pipelines (MetaGenome Rapid Annotation using Subsystem Technology (MG-RAST), mothur, MICRobial Community Analysis (MICCA) and Quantitative Insights into Microbial Ecology (QIIME)), when analyzing the Tunisian PG. Importantly, based on 16S rDNA datasets, the functional capabilities of microbial communities of PG were deciphered. They suggested the presence of PG autochthonous bacteria recoverable into (1) removal of radioactive elements and toxic heavy metals (2) promotion of plant growth (3) oxidation and (4) reduction of sulfate.

Metataxonomics of Tunisian phosphogypsum based on five bioinformatics pipelines: Insights for bioremediation

Houda TRIFI^{1,2§*}, Afef NAJJARI^{3,§}, Wafa ACHOUAK⁴, Mohamed BARAKAT⁴, Kais GHEDIRA⁵, Faten MRAD⁶, Mouldi SAIDI¹, Haïtham SGHAIER^{7,8*}

¹Laboratory of biotechnology and nuclear technology (LR16CNSTN01), National Center for Nuclear Sciences and Technology (CNSTN), Sidi Thabet Technopark, 2020, Tunisia

²Faculty of Sciences of Bizerte, Carthage University, 7021 Jarzouna, Tunisia

³Université de Tunis El Manar, Faculté des Sciences de Tunis, LR03ES03 Microorganismes et Biomolécules Actives, 2092, Tunis, Tunisia.

⁴Lemire, Laboratory of Microbial Ecology of the Rhizosphere and Extreme Environment, CEA Cadarache 13108 Saint Paul Lez Durance, France

⁵Laboratory of Bioinformatics, Biomathematics and Biostatistics - LR16IPT09, Institut Pasteur de Tunis, Université de Tunis el Manar, Tunis, 1002, Tunisia

⁶Tunisian Chemical Group, Route de Gabes Km 3.5, BP “3018”, Sfax, Tunisia

⁷Laboratory of Energy and Matter for Development of Nuclear Sciences (LR16CNSTN02), National Center for Nuclear Sciences and Technology (CNSTN), Sidi Thabet Technopark, 2020, Tunisia

⁸Associated with Laboratory of Biotechnology and Nuclear Technology (LR16CNSTN01) & Laboratory of Biotechnology and Bio-Geo Resources Valorization (LR11ES31), Sidi Thabet Technopark, 2020, Tunisia

§: Both authors contributed equally to this work. *Corresponding authors.

trifi.houda@ gmail.com; sghaier.haitham@gmail.com.

Number phone: 0021625867494

Metataxonomics of Tunisian phosphogypsum based on five bioinformatics pipelines: Insights for bioremediation

Houda TRIFI^{1,2§*}, Afef NAJJARI^{3,§}, Wafa ACHOUAK⁴, Mohamed BARAKAT⁴, Kais GHEDIRA⁵, Faten MRAD⁶, Mouldi SAIDI¹, Haïtham SGHAIER^{7,8*}

¹Laboratory of biotechnology and nuclear technology (LR16CNSTN01), National Center for Nuclear Sciences and Technology (CNSTN), Sidi Thabet Technopark, 2020, Tunisia

²Faculty of Sciences of Bizerte, Carthage University, 7021 Jarzouna, Tunisia

³Université de Tunis El Manar, Faculté des Sciences de Tunis, LR03ES03 Microorganismes et Biomolécules Actives, 2092, Tunis, Tunisia.

⁴Lemire, Laboratory of Microbial Ecology of the Rhizosphere and Extreme Environment, CEA Cadarache 13108 Saint Paul Lez Durance, France

⁵Laboratory of Bioinformatics, Biomathematics and Biostatistics - LR16IPT09, Institut Pasteur de Tunis, Université de Tunis el Manar, Tunis, 1002, Tunisia

⁶Tunisian Chemical Group, Route de Gabes Km 3.5, BP “3018”, Sfax, Tunisia

⁷Laboratory of Energy and Matter for Development of Nuclear Sciences (LR16CNSTN02), National Center for Nuclear Sciences and Technology (CNSTN), Sidi Thabet Technopark, 2020, Tunisia

⁸Associated with Laboratory of Biotechnology and Nuclear Technology (LR16CNSTN01) & Laboratory of Biotechnology and Bio-Geo Resources Valorization (LR11ES31), Sidi Thabet Technopark, 2020, Tunisia

§: Both authors contributed equally to this work. *Corresponding authors.

trifi.houda@gmail.com; sghaier.haitham@gmail.com.

Abstract

Phosphogypsum (PG) is an acidic by-product from the phosphate fertilizer industry and it is characterized by a low nutrient availability and the presence of radionuclides and heavy metals which pose a serious problem in its management. Here, we have applied Illumina MiSeq sequencing technology and five bioinformatics pipelines to explore the phylogenetic communities in Tunisian PG. Taking One Codex as a reference method, we present the results of 16S-rDNA-gene-based metataxonomics abundances with four other alternative

32 bioinformatics pipelines (MetaGenome Rapid Annotation using Subsystem Technology (MG-
33 RAST), mothur, MICRobial Community Analysis (MICCA) and Quantitative Insights into
34 Microbial Ecology (QIIME)), when analyzing the Tunisian PG. Importantly, based on 16S
35 rDNA datasets, the functional capabilities of microbial communities of PG were deciphered.
36 They suggested the presence of PG autochthonous bacteria recoverable into (1) removal of
37 radioactive elements and toxic heavy metals (2) promotion of plant growth (3) oxidation and
38 (4) reduction of sulfate.

39 **Keywords:** Bioremediation; Functional prediction; Illumina Miseq sequencing; Microbial
40 community; Phosphogypsum.

41 **1. Introduction**

42 During recent decades, the microbial communities in mine tailings have attracted a
43 considerable interest from many microbiologists [1; 2; 3]. Extreme tailing environments are
44 ideal targets for the study of microbial diversity and how microbial communities respond to
45 environmental changes [4]. Unlike the widespread idea about the negative effect of pollution
46 on biodiversity; microbial communities in mine site can be greater than thought. For instance,
47 an acidic and heavy metal contaminated mine can harbor rare and extremophile
48 microorganisms involved in the biogeochemical cycles in this site. In fact, these species may
49 be used in bioremediation processes of other contaminated soils [5].

50 Phosphogypsum (PG) is composed of over 90% of calcium sulfate and lower percentages of
51 silica, aluminum, phosphate, and fluoride [6]. Depending on the origin of the phosphate ores,
52 PG contains variable amounts of some impurities such as naturally occurring radionuclides
53 and heavy metals [7]. The stack of PG increases during the manufacture of phosphoric acid,
54 which makes its disposal an environmental concern.

55 Bioremediation could be used as an innovative PG management strategy. It is based on the
56 use of living organisms, especially microorganisms, to degrade or detoxify the environmental

57 contaminant as a safe and low-cost alternative [8]. However, the optimization of
58 bioremediation strategies is dependent on knowing and characterizing the *in situ* microbial
59 community. Indigenous bacteria able to remain active within the contaminant, can be
60 identified and potentially targeted for bioremediation [9].

61 Recently, the application of next generation sequencing technologies such as Illumina Miseq
62 has allowed to generate gigabases of high-quality raw data at low cost [10]. Metataxonomics
63 is a high-throughput process used to identify microorganisms within a complex environment
64 and create a metataxonomic tree in order to make taxonomic inferences [11]. Metagenomic
65 analysis of nucleic acids was a powerful approach to capture the entire spectrum of microbial
66 communities including both cultivable and uncultivable microorganisms as well as some rare
67 species [12].

68 The microbial diversity of Tunisian PG has been poorly understood. Recently, one paper has
69 studied the most important phyla of microbial communities in Tunisian PG using QIIME [13],
70 but there have been no additional reports of these communities. This study, therefore,
71 represents the first observation of the microbiology of PG using Illumina MiSeq approach and
72 five bioinformatics pipelines. It provides an evaluation of the potential applications of the
73 microbial communities of PG in bioremediation processes.

74 **2. Materials and methods**

75 **2.1 Site description and sample collection**

76 Sample collections were conducted in March; 2012 from the PG piles in accordance with
77 Smith et al. [14] sampling guidelines. PG dumps are located near to the fertilizer industry in
78 Sfax, Southeastern Tunisia. This region is characterized by a semi-arid to arid climate with an
79 annual precipitation of around 237.8 mm. The Sfax plant (SIAPE factory) stockpiles contain
80 \approx 30 million tons of PG covering more than 48 ha and measuring more than 50 m height [13].

81 The 10 cm of sediment layer was collected with sterile spatulas and spoons, and homogenized
82 in a pre-sterilized aluminum pan. Multiple aliquots of samples were placed into 50 mL
83 polypropylene tubes and were stored on ice until analysis prior to microbial analyses, while
84 others were dried for the determination of soil chemical properties and radionuclides content.
85 PG pH was determined using a glass combination electrode with a PG: water ratio of 1:2.5.
86 The physico-chemical characteristics of PG were determined using Analyzer and Inductive
87 Coupled Plasma Optical Emission Spectrometry (ICP-OES) in National Institute for Research
88 and Physicochemical Analysis (INRAP) in Tunisia. Gamma spectrometry was used for the
89 determination of radionuclides containing in the sample. Radioactivity measurements were
90 done in National Center for Nuclear Sciences and Technology (CNSTN) in Tunisia.

91 **2.2 DNA Extraction**

92 DNA samples were extracted according to the method of He et al. [15] with some
93 modification. 10 g of PG was mixed with 13.5 mL of DNA extraction buffer (100 mM Tris–
94 HCl [pH 8.0], 100 mM sodium EDTA [pH 8.0], 100 mM sodium phosphate [pH 8.0], 1.5 M
95 NaCl, 1% hexadecylmethylammonium bromide) and 10 mL of proteinase K (10 mg mL⁻¹) in
96 centrifuge tubes by horizontal shaking at 200 rpm for 30 min at 37°C. After the shaking
97 treatment, 1.5 mL of 20% dodecyl sulfate sodium (SDS) was added, and the samples were
98 incubated in a 65°C water bath for 2 h with gentle end-over end inversions every 15–20 min.
99 The supernatants were collected after centrifugation at 6,000 g for 10 min at room
100 temperature and transferred into 50 mL centrifuge tubes. The soil pellets were extracted two
101 more times by adding 4.5 mL of the extraction buffer and 0.5 mL of 20% SDS, vortexing for
102 10 s, incubating at 65°C for 10 min, and centrifuging as before. Supernatants from the three
103 cycles of extractions were combined and mixed with an equal volume of chloroform–
104 isopentanol (24:1, v/v). The aqueous phase was recovered by centrifugation and precipitated
105 with 0.6 volume of isopropanol at 4°C for 1 h. The pellet of crude nucleic acids was obtained

106 by centrifugation at 16,000 g for 20 min at 4°C, washed with a cold 70% ethanol, and
107 resuspended in sterile deionized water. DNA concentration and quality were assessed based
108 on spectral absorbance at 260 nm wavelength and absorbance ratios of 260/280 nm, using a
109 NanoDrop Spectrophotometer (NanoDrop Technology, USA). DNA samples were preserved
110 at –80°C until further processing.

111 **2.3 Amplification of 16S rRNA Genes**

112 DNA samples were subjected to high throughput sequencing (HTS) by Illumina MiSeq
113 technology, according to Illumina recommendations (16S Meta-genomic Sequencing Library
114 Preparation, Part #15044223_B) With the V3 chemistry (2 x 300 bp). V3-V4 domain of 16S
115 rRNA was amplified with tagged primers 16S rRNA Fwd primer [3'-
116 CCTACGGGNGGCWGCAG-5'] and 16S rRNA Rev primer [3'-
117 GACTACHVGGGTATCTAATCC-5']. The tagged primers structure was 5'-N₂₋₄X₆P_n-3',
118 with “N₂₋₄” were random bases, “X₆” were 6-bases tag and “P_n” was specific primer. The
119 polymerase chain reaction (PCR) mix contained 1 µg of template DNA, 10 µL of 5X Taq
120 reaction buffer, 1 µM of each primer, 200 µM dNTPs, and 1.25 units of Taq polymerase in a
121 50 µL reaction. The amplification procedure was as follows: an initial denaturation step at
122 95°C for 2 min, and 35 cycles of denaturing at 95°C for 1 min, annealing at 53°C for 1 min,
123 and extension at 72°C for 1 min, followed by a final extension step at 72°C for 5 min.

124 **2.4 Bioinformatics and statistical analysis**

125 The raw data generated from Illumina MiSeq technology run was processed and analyzed
126 with One Codex, using default parameters, and also, for comparison purposes, following the
127 pipelines of MG-RAST [16], mother [17], MICCA [18] and QIIME [19]. The merged file
128 contained 16,803 sequences totaling 7,639,017 base pairs with an average length of 455 bps.
129 The same generated merged FASTQ file was used as input for all five pipelines on or after

130 August 2017. Operational Taxonomic Units (OTUs) were calculated at phylum, order, family,
131 genus and species levels. Taxa with relative abundance higher than 1% were retained for
132 statistical analysis.

133 **2.4.1 One Codex**

134 In this work, first, we used the One Codex data platform [20] for the taxonomic assignment of
135 datasets because of the following advantages: (1) it exhaustively compares input samples
136 against a database of thousands of whole microbial genomes, numbering approximately
137 40,000 bacteria, viruses, protists, archaea and fungi (<https://onecodex.com/>) in seconds [21]
138 (2) it has best-in-class detection accuracy providing results with 0.00% false positives [22]
139 and (3) it has the highest degree of sensitivity and specificity compared to other methods, both
140 when detecting well-characterized organisms as well as "taxonomically novel" species [20].
141 For the metataxonomic analysis of the merged file, One Codex, used an assignment-first
142 approach. First, for the taxonomic assignment, all reads were compared to One Codex
143 database by checking for exact k-mer matches. Afterwards, based on those comparisons, the
144 lowest taxonomy possible to each read was assigned. Then, for taxonomic binning, reads were
145 gathered into various taxonomic units based on their annotations [20; 21] Metataxonomic
146 results were obtained as an overview, charts, phylogenies and tables.

147 **2.4.2 MG-RAST: MetaGenome Rapid Annotation using Subsystem Technology**

148 The merged file was uploaded to MG-RAST [16] for sequence analysis. The analysis
149 generated by the MG-RAST automated processing pipeline (<http://metagenomics.anl.gov>) is
150 as follows: Reads were given a taxonomic classification using the 'Best Hit Classification'
151 option in MG-RAST, a maximum E-value cut-off of $1e-05$, a % identity threshold of 60, a
152 minimum alignment length of 15 bp and a minimum abundance set to 1. Best Hit
153 Classification reports the highest scoring annotation(s) for each read. MG-RAST uses the

154 BLAT (BLAST-like alignment tool) [23] algorithm for the computation of similarities and to
155 identify rDNA sequences by searching a reduced rDNA database. The UCLUST algorithm
156 [24] is then used to cluster identified rDNA sequences. Representative sequences of each
157 cluster were then used to realize taxonomic assignments using the RDP database [25].

158 **2.4.3 Mothur**

159 Merged FASTQ file was quality-filtered and analyzed using mothur version 1.39.5, following
160 the MiSeq Standard Operating Process pipeline [26]. Briefly, contigs were screened to
161 remove sequences that did not align to Silva V4 database [27]. Duplicate sequences were
162 merged followed by the alignment to a reference database (SILVA 123). Chimeric sequences
163 were removed by using UCHIME algorithm [24]. Taxonomic classification was done based
164 on Ribosomal Database Project (RDP) trainset 14 and using the RDP Naïve Bayesian
165 Classifier [28]. Abundance was expressed as a percentage with respect to the total number of
166 sequences in the sample.

167 **2.4.5 MICCA**

168 Processing the analysis of amplicon metagenomic datasets with micca software pipeline
169 combines the following steps. First, the primers from both ends of merged FASTQ file were
170 trimmed using trim command in the MICCA software package. Second, the quality of
171 trimmed sequences was filtered using micca filter according to the maximum allowed
172 expected error. Third, filtered sequences were classified into OTUs based on the *de novo*
173 greedy clustering. Each representative sequence of generated OTUs was classified based on
174 the RDP classifier [28]. Abundance was expressed as a percentage with respect to the total
175 number of sequences in the sample.

176

177

178 **2.4.6 QIIME**

179 The Merged FASTQ file was quality-filtered and analyzed using QIIME v.1.9.0 pipeline [19].
180 Sequences with an average Phred quality score < 25 were discarded. Chimeric sequences
181 were identified using USEARCH [24] implemented in QIIME's pipeline using the reference
182 mode against the SILVA LTP database (release 119) [29]. OTUs were clustered with QIIME's
183 *de novo* OTU-picking pipeline using UCLUST [24] and rdp methods with Naive Bayes
184 classification [28]. Representative sequences for each OTU were classified taxonomically
185 using BLAST [30] against SILVA LTP database (release 119). Abundance was expressed as a
186 percentage with respect to the total number of sequences in the sample.

187 **2.5 Diversity comparisons and statistical analysis**

188 Alpha-diversity metrics, corresponding to variables describing diversity within the sample
189 (Good's coverage, Shannon diversity and Simpson indexes), was computed at a genus level.
190 Each pipeline (mothur, MICCA and QIIME) was used to calculate diversity measures.

191 **3. Results and Discussion**

192 **3.1 Assignment of metataxonomic datasets, comparison of bioinformatics pipelines and** 193 **taxonomic signature**

194 When analyzing 16S rDNA gene sequencing data from complex ecological niches, like PG,
195 careful consideration of sources of bias affecting each tool is required. The performance of
196 One Codex was confirmed by Lindgreen et al. [22] as previously indicated in this paper.
197 Taxonomic binning *via* One Codex allowed the assignment of the prevalent OTUs ($\geq 1\%$ of
198 the community) to four phyla which are Proteobacteria (46.76%), Actinobacteria (18.45%),
199 Firmicutes (15.11%) and Bacteroidetes (2.68%). The most frequent detected classes
200 comprising 36.4% of the proteobacterial phylotypes were, in order of abundance:

201 Gammaproteobacteria (15.74%), Alphaproteobacteria (14.07%) and Betaproteobacteria
202 (6.59%).

203 The results of One Codex were compared with four bioinformatics pipelines, MG-RAST,
204 mothur, MICCA and QIIME as well as with previously published data (Figure 1, Table 2,
205 Supplementary File 1). Among the 16,803 reads of the merged file, 98.19, 97.79, 99.45,
206 97.73 and 99.34% were classified using the One Codex, MG-RAST, mothur, MICCA and
207 QIIME tools, respectively (Supplementary File 1). Inspection of rarefaction curves
208 (Supplementary File 1) showed that they tended to asymptotes, thereby arguing for high
209 sequence coverage, a conclusion strongly supported by reference to Good's coverage metric
210 (Table 1). Indeed, estimation of the coverage provided by our metataxonomic data set showed
211 that One codex and mothur pipelines gave a good coverage with 99.19 and 99.00%,
212 respectively. Shannon's diversity and Simpson indices are presented in Table 1.

213 Taking One Codex as a reference method, we also showed that the four other bioinformatics
214 pipelines did not detect taxonomic units (*e.g.*, species) in similar abundances. Also,
215 differences were observed at different level (Table 2, Supplementary File 1). Differences at
216 the class and genus levels are shown in Figure 1. Moreover, in comparison with the reference
217 method, these four tools (MG-RAST, mothur, MICCA and QIIME) either misidentified many
218 phyla, classes, orders, families, genera and species that were not truly present (false positives)
219 or wrongly indicated that they are absent (false negatives) (Figure 1, Table 2, Supplementary
220 File). However, it was observed that One Codex and QIIME produced comparable results
221 (Table 2, Supplementary File 1).

222 In addition, many parameters can explain the difference between the study of Zouch et al. [13]
223 and our results (Table 2), including the DNA extraction method, 16S rDNA region targets and

224 the spatiotemporal sampling data. Moreover, these differences may be attributed to used
225 taxonomical classification algorithms and reference databases [31].

226 **3.2 Characteristics of Tunisian PG, functional predictions and insights for** 227 **bioremediation**

228 Fertilizer manufacturing industries release obvious sources of heavy metal and radionuclides
229 with severe impact on the environment. In Tunisia, one of the main phosphates producing
230 countries in the world, the phosphate fertilizer industries produce 10 Mt/year PG [32].
231 Tunisian PG is heavily polluted with heavy metals and radionuclides (Table 3). In order to
232 develop effective bioremediation strategies for these contaminants, the composition of
233 microbial communities in PG needs to be better understood. Therefore, a legitimate question
234 is to ask whether the composition of the core microbiome of Tunisian PG accords very closely
235 with its characteristics.

236 In this study, based on a reference method, One Codex, 193 species were detected
237 (Supplementary File 1). One Codex result indicated that Tunisian PG does not contain neither
238 high abundance (> 25% of classified reads) nor medium abundance (5–25% of classified
239 reads) microorganisms. However, it contains five low abundance (1–5% of classified reads)
240 bacteria, *Staphylococcus aureus*, *Empedobacter falsenii* (1.95%), *Geobacillus*
241 *thermoleovorans* (1.3%), *Polaromonas* sp. EUR3 1.2.1 (1.27%) and *Clostridium bifermentans*
242 (1.02%).

243 One codex analysis showed that PG was dominated at the phylum level by Proteobacteria
244 (46.76% of classified reads). In more detail, Proteobacteria-related sequences were composed
245 of five classes: Gammaproteobacteria, Alphaproteobacteria, Betaproteobacteria,
246 Deltaproteobacteria and Epsilonproteobacteria an order from the highest to the lowest
247 (Supplementary File 1). Proteobacteria are often shown to be predominant in various acidic

248 environments such as mine tailings, comprising up to 70% of 16S rRNA gene clone library
249 sequences [33; 34]. They are well-known to survive in oligotrophic environments, nitrate
250 reduction and metal resistance and they are able to reduce uranium. These characteristics were
251 frequently presented in Gammaproteobacteria [9; 35] which were detected as the dominant
252 class in this study.

253 Actinobacteria were the second most dominant phylum with a representative major order
254 (*Micrococcales*) and major family (*Micrococcaceae*) (Supplementary File 1). Members of this
255 phylum are ubiquitously distributed in different habitats, including metal-rich, acidic
256 ecosystems. Their metabolic versatility and resilience to harsh environments have been also
257 reported in mining-related areas [36]. Indeed, Actinobacteria have a range of mechanisms for
258 dealing with heavy metal resistance [37]. They have been widely used in bioremediation such
259 as biotransformation, biodegradation and other purposes [38]. El Baz et al.[39] showed that
260 Actinobacteria might be a good tool which can be used in a bioreactor configuration to treat
261 metal contaminated wastes prior to discharge into the environment.

262 Firmicutes was the third dominant phylum (Supplementary File 1). Members of this phylum
263 occupy a wide range of habitats and can be either beneficial or detrimental in diverse settings
264 [40] and they have an important role in industrial and clinical aspect [41].

265 Concordantly, Proteobacteria, Actinobacteria and Firmicutes were mainly detected in mine
266 tailing with low pH, intense weathering and iron precipitations [1; 42]. They are known to be
267 resistant to heavy metal toxicity and adapted for growth in low nutrient environments [35].

268 Deeper genus-level sequencing revealed that the majority of bacteria in PG can contribute to
269 the different biotechnological treatment of PG. The dominant bacteria genera were related to
270 *Arthrobacter* (13.25%), followed by the genera *Paracoccus* (12.32%), *Pseudomonas* (6.7%)
271 and *Staphylococcus* (3.63%) (Supplementary File 1). These genera have already been detected

272 in various environments contaminated with chemicals and heavy metal such as mine dumps,
273 sometimes as dominant populations. They have metabolically versatile and diverse
274 degradative capabilities and potential applications for bioremediation [43; 44; 45; 46].

275 Resolving metataxonomics data at the species level has great potential for deciphering
276 functional profiles within an ecological niche. These predictions suggested the presence of
277 PG-native strains valorisable into (1) removal of radioactive elements and toxic heavy metals
278 (2) promotion of plant growth (3) oxidation and (4) reduction of sulfate (Figure 2).

279 Many of detected strains in this study (Supplementary File 1) were isolated previously in
280 uranium (U) mine, and they are well-known for precipitation or reduction of U and other
281 radionuclides such as *Geobacter* sp. [3], *Idiomarina* sp. [47], *Clostridium* sp. [35]. The ability
282 of *Lactobacillus* sp. to remove U from aqueous solution was reported by Tsuruta [48]. The
283 removal of radioactive elements and toxic heavy metals from contaminated sources is a
284 worthwhile priority for environmental protection initiatives. Therefore, researchers have
285 concentrated on accumulation of U by microorganisms that can leach you from ore in mine
286 soil [49; 50]. These microorganisms from U deposits can be used as adsorbing agents for the
287 removal of nuclear elements which may be present in mine tailings [48]. The isolation of
288 these bacteria from a radioactive environment is very encouraging due to the possibility of
289 their being used in the radionuclide bioremediation process.

290 Importantly, although PG is a radioactive environment, we detected the presence of ionizing
291 radiation sensitive bacterial species such as *Corynebacterium* sp., *Acinetobacter baumannii*,
292 *Acinetobacter haemolyticus*, *Bacillus subtilis*, *Bacillus cereus* and *Bacillus megaterium*
293 (Figure 2). This can be explained based on previous findings that radioresistant microbial
294 cells may protect their radiosensitive neighboring species (even from a different phylum)
295 from high-level ionizing radiation [51]. Moreover, the Comparison of the concentration of
296 radionuclides found in this study with others carried out worldwide as reported by Ajam et al.

297 [32] showed that the radioactivity of ^{226}Ra in the Tunisian PG (200.58 becquerel / kilogramme
298 (Bq/Kg)) is lower than those in other PG (Morocco, 1420 Bq/kg; Florida, 1120 Bq/kg).

299 Concentrations of metal(loid)s in Tunisian PG (Table 3), particularly Sr, Zn and Cd, are
300 higher than those indicated in the environmental quality standard for agricultural applications
301 in Quebec [52]. Cd is one of the most phytotoxic metals especially in acidic soil with low
302 CEC and can cause agricultural yield reduction, contamination of the food crops [53].
303 Different bacterial species detected in Tunisian PG (Supplementary File 1) have been
304 characterized in extremely acidic areas characterized by low nutrient and high levels of heavy
305 metals like *Ralstonia* sp. [35; 54], *Corynebacterium* sp. [55], *Comamonas* sp. [56] and
306 *Pseudomonas* sp. [43] (Figure 2).

307 PG dumps are subjected to water erosion and wind dispersion representing a source of
308 contamination for nearby communities [57]. Numerous investigations have focused on the
309 search for a new way to restore mine areas by the creation of a vegetation cover which assure
310 long-term metal stabilization. Another method to valorize PG is its use as a fertilizer in
311 agricultural crops. In this study, different isolates were characterized by their plant growth
312 promoting traits as well as their resistances to various heavy metals (Figure 2) and may be
313 excellent candidates of bio-inoculants for enhancing the phytomanagement strategies of PG.
314 Among these isolates, *Bradyrhizobium* sp., a plant growth promoting rhizobacterium (PGPR)
315 used in the phytostabilization of heavy metals and to increase plant yield and nitrogen uptake
316 [58]. Other species in Tunisian PG (Supplementary File 1) can also resist to the toxicity of
317 heavy metals and promote plant growth (Figure 2) via various mechanisms such as secretion
318 of phytohormones (auxins), siderophores, ammonia (NH_3), 1-aminocyclopropane-1-
319 carboxylate (ACC) deaminase and solubilization of inorganic phosphate: *Delftia* sp. [59],
320 *Achromobacter* sp. [60], *Flavobacterium* sp., *Shingomonas* sp. [61], *Psychobacter* sp. [62],
321 *Rhizobium* sp. [63], *Planococcus* sp. [64], *Mesorhizobium* sp. [65], *Paenibacillus* sp. [66],

322 *Jeotgalicoccus* sp. [67], *Acinetobacter* sp. [68], *Pelomonas* sp. [69], *Burkholderia* sp. [70],
323 *Rhodococcus* sp. [71], *Chryseobacterium* sp. [72], *Bacillus* sp. [73], *Methylobacterium* sp.
324 [67], *Stenotrophomonas* sp. [74], *Massilia* sp. [75], *Kribbela* sp. [76].

325 Another strategy for PG bioremediation is the use of sulfate-oxidizing bacteria (SOB) and
326 sulfate-reducing bacteria (SRB) which well-known as good candidates for PG
327 biotransformation [77; 78]. SOB play an important role in detoxifying their environments,
328 and they can increase the availability of sulphate (SO_4^{2-}) for plant absorption which will
329 improve plant growth and agricultural production [79]. SOB not only oxidized sulfur, but they
330 may also oxidize iron (Fe) and glucose as an energy source and considered as mixotrophs.
331 This sulfur and Fe oxidation ability enables them to use for extraction of metals from low-
332 grade ores where the ferric Fe act as an oxidizing agent [80]. Furthermore, SRB may reduce
333 sulfate from PG into hydrogen sulfide. Thereafter, H_2S may be oxidized into sulfuric acid by
334 sulfide-oxidizing bacteria to be possibly reused in the fertilizer industry to produce
335 phosphoric acid. Moreover, H_2S produced by SRB may react with different with certain
336 dissolved metals, such as copper (Cu), Fe and Zn, forming insoluble precipitates [81]. For
337 instance, Martins et al. [78] showed the ability of SRB to remove Cu, Zn and Fe from the
338 sample. In this respect, biotechnological process based on SRB and SOB activity may be
339 considered as environmentally friendly for the treatment of metal-rich PG waste.

340 In accordance with the elevated sulfur concentration in PG, we detected SOB and SRB in this
341 study (Figure 2). The sulfur-oxidizing *Pseudomonas* sp., *Stenotrophomonas* sp.,
342 *Achromobacter* sp., *Thiokalvibrio* sp., *Methylophilus* sp., *Sedimenticola* sp., *Thiomicrospira*
343 sp., *Corynebacterium* sp., *Achrobacter* sp. and the sulfur-reducing *Geobacter* sp., *Clostridium*
344 sp. *Desulfotomaculum* sp. represented approximately 7.15 and 0.02%, respectively, of PG
345 community, suggesting that the oxidation-reduction of sulfurous compounds may occur in the
346 harsh conditions of PG.

347 The knowledge and detection of indigenous bacteria which possess different bioremediation
348 traits such as multi-resistant heavy metals/radionuclides PGPB with the capacity to oxidize or
349 reduce sulfate compounds provided a basis for exhaustive research leading to the
350 development of an efficient tool to detect, stabilize or remove heavy metals and radionuclides
351 with the ability to improve plant yield in extreme environments like the PG. For example,
352 *Pseudomonas* sp. and *Stenotrophomonas* sp. were SOB isolated from acid mine drainage [80],
353 they exhibited resistance to high concentrations of metalloids [82; 83]. Besides, they are able
354 to secrete phytohormones, siderophores and solubilize inorganic phosphate [74; 84].

355 **4. Conclusion**

356 In conclusion, using the high throughput Illumina sequencing technique and one reference
357 method (One Codex) with four alternative bioinformatics pipelines (MG-RAST, mothur,
358 MICCA and QIIME), provided a detailed picture and new information regarding the
359 phylogenetic characterization of bacterial communities on phylum, class, order, family, genus
360 and species levels in Tunisian PG. In light of *in silico* results of PG metataxonomics
361 generated using these five bioinformatics pipelines, the Tunisian PG is a mesmerizing
362 repository of metal-tolerant, radioresistant, PGP, SO and SR bacteria. However, clearly, when
363 analyzing 16S RDNA gene sequencing data from complex ecological niches, like PG, the use
364 of a reference method and careful consideration of sources of bias affecting each alternative
365 method are required. This work will improve our knowledge of bacteria suitable to be used in
366 PG bioremediation, as well as bioprospecting of enzymes and biotechnological compounds
367 hidden in PG.

368 **Acknowledgements**

369 This work was performed under the auspices of the Tunisian National Center for Nuclear
370 Sciences and Technologies (CNSTN) in collaboration with the CEA, France. It was supported

371 by grants from the Tunisian Ministry of Higher Education and Scientific Research and from
372 Laboratory of Microbial Ecology of the Rhizosphere and Extreme Environment (CEA,
373 France); without this generous and strong assistance, this study could not have been
374 undertaken.

375 Supplementary data of this work can be found in the online version of the paper.

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

- 400 [1] V.F. Bondici, J.R. Lawrence, N.H. Khan, J.E. Hill, E. Yergeau, G.M. Wolfaardt, J.
401 Warner, and D.R. Korber, Microbial communities in low permeability, high pH
402 uranium mine tailings: characterization and potential effects. *J Appl Microbiol* 114
403 (2013) 1671-86.
- 404 [2] L.-X. Chen, J.-T. Li, Y.-T. Chen, L.-N. Huang, Z.-S. Hua, M. Hu, and W.-S. Shu, Shifts in
405 microbial community composition and function in the acidification of a lead/zinc mine
406 tailings. *Environ Microbiol* 15 (2013) 2431–2444
- 407 [3] P.K. Dhal, and P. Sar, Microbial communities in uranium mine tailings and mine water
408 sediment from Jaduguda U mine, India: A culture independent analysis. *J Environ Sci*
409 *Health A Tox Hazard Subst Environ Eng* 49 (2014) 694–709.
- 410 [4] J. Liu, Z.-S. Hua, L.-X. Chen, J.-L. Kuang, S.-J. Li, W.-S. Shu, and L.-N. Huang,
411 Correlating microbial diversity patterns with geochemistry in an extreme and
412 heterogeneous environment of mine tailings. *Appl Environ Microbiol* 80 (2014) 3677–
413 3686.
- 414 [5] L.C. Batty, The potential importance of mine sites for biodiversity *Mine Water environ* 24
415 (2005) 101–103.
- 416 [6] H. Tayibi, M. Choura, F.A. Lopez, F.J. Alguacil, and A. Lopez-Delgado, Environmental
417 impact and management of phosphogypsum *J Environ Manage.* 90 (2009) 2377–
418 2386.
- 419 [7] M.B. Nisti, C.R. Saueia, L.H. Malheiro, G.H. Groppo, and B.P. Mazzilli, Lixiviation of
420 natural radionuclides and heavy metals in tropical soils amended with
421 phosphogypsum. *J Environ Radioactiv* 144 (2015) 120–126.
- 422 [8] J. Jalali, S. Magdich, R. Jarboui, M. Loungou, and E. Ammar, Phosphogypsum
423 biotransformation by aerobic bacterial flora and isolated *Trichoderma asperellum*
424 from Tunisian storage piles. *J Hazard Mater* 308 (2016) 362–73.
- 425 [9] D.M. Akob, H.J. Mills, and J.E. Kostka, Metabolically active microbial communities in
426 uranium contaminated subsurface sediments. *FEMS Microbiol Ecol* 59 (2007) 95–
427 107.
- 428 [10] C. Hong, Y. Si, Y. Xing, and Y. Li, Illumina MiSeq sequencing investigation on the
429 contrasting soil bacterial community structures in different iron mining areas. *Environ*
430 *Sci Pollut Res Int* 22 (2015) 10788–99.
- 431 [11] J.R. Marchesi, and J. Ravel, The vocabulary of microbiome research: a proposal.
432 *Microbiome* 3 (2015) 31.
- 433 [12] X. Zhang, J. Niu, i. Liang, X. Liu, and H. Yin, Metagenome-scale analysis yields insights
434 into the structure and function of microbial communities in a copper bioleaching heap.
435 *BMC Genetics* 17 (2016) 21.
- 436 [13] H. Zouch, F. Karray, F. Armougom, S. Chifflet, A. Hirschler-Rea, H. Kharrat, L.
437 Kamoun, W. Ben Hania, B. Ollivier, S. Sayadi, and M. Quemeneur, Microbial
438 Diversity in Sulfate-Reducing Marine Sediment Enrichment Cultures Associated with
439 Anaerobic Biotransformation of Coastal Stockpiled Phosphogypsum (Sfax, Tunisia).
440 *Front Microbiol* 8 (2017) 1583.
- 441 [14] K.S. Smith, C.A. Ramsey, and P.L. Hageman, Sampling strategy for the rapid screening
442 of mine-waste dumps on abandoned mine lands Open-File Report, USGS Numbered
443 Series, 2000.
- 444 [15] L.Y. He, Y.F. Zhang, H.Y. Ma, L.N. Su, Z.J. Chen, Q.Y. Wang, M. Qian, and X.F.
445 Sheng, Characterization of copper-resistant bacteria and assessment of bacterial

- 446 communities in rhizosphere soils of copper-tolerant plants. *Appl Soil Ecol* 44 (2010)
447 49–55.
- 448 [16] K.P. Keegan, E.M. Glass, and F. Meyer, MG-RAST, a metagenomics service for analysis
449 of microbial community structure and function. *Methods Mol Biol* 1399 (2016) 207–
450 33.
- 451 [17] P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A.
452 Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G.
453 Thallinger, D.J. Van Horn, and C.F. Weber, Introducing mothur: open-source,
454 platform-independent, community-supported software for describing and comparing
455 microbial communities. *Appl Environ Microbiol* 75 (2009) 7537–41.
- 456 [18] D. Albanese, P. Fontana, C. De Filippo, D. Cavalieri, and C. Donati, MICCA: a complete
457 and accurate software for taxonomic profiling of metagenomic data. *Sci Rep* 5 (2015)
458 9743.
- 459 [19] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello,
460 N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D.
461 Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M.
462 Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T.
463 Yatsunenko, J. Zaneveld, and R. Knight, QIIME allows analysis of high-throughput
464 community sequencing data. *Nat Methods* 7 (2010) 335–6.
- 465 [20] S.S. Minot, N. Krumm, and N.B. Greenfield, One Codex: A Sensitive and accurate data
466 platform for genomic microbial identification. *bioRxiv*. 23 (2015).
- 467 [21] L. Siegwald, H. Touzet, Y. Lemoine, D. Hot, C. Audebert, and S. Caboche, Assessment
468 of Common and Emerging Bioinformatics Pipelines for Targeted Metagenomics.
469 *PLoS One* 12 (2017) e0169563.
- 470 [22] S. Lindgreen, K.L. Adair, and P.P. Gardner, An evaluation of the accuracy and speed of
471 metagenome analysis tools. *Sci Rep* 6 (2016) 19233.
- 472 [23] W.J. Kent, BLAT—the BLAST-like alignment tool. *Genome Res* 12 (2002) 656–64.
- 473 [24] R.C. Edgar, Search and clustering orders of magnitude faster than BLAST.
474 *Bioinformatics* 26 (2010) 2460–2461.
- 475 [25] J.R. Cole, B. Chai, T.L. Marsh, R.J. Farris, Q. Wang, S.A. Kulam, S. Chandra, D.M.
476 McGarrell, T.M. Schmidt, G.M. Garrity, and J.M. Tiedje, The Ribosomal Database
477 Project (RDP-II): previewing a new autoaligner that allows regular up-dates and the
478 new prokaryotic taxonomy. *Nucleic Acids Res* . 31 (2003) 442–443.
- 479 [26] J.J. Kozich, S.L. Westcott, N.T. Baxter, S.K. Highlander, and P.D. Schloss, Development
480 of a dual-index sequencing strategy and curation pipeline for analyzing amplicon
481 sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79
482 (2013) 5112–20.
- 483 [27] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F.O.
484 Glockner, The SILVA ribosomal RNA gene database project: improved data
485 processing and web-based tools. *Nucleic Acids Res* 41 (2013) D590–6.
- 486 [28] Q. Wang, G.M. Garrity, J.M. Tiedje, and J.R. Cole, Naïve Bayesian Classifier for Rapid
487 Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ*
488 *Microbiol* 73 (2007) 5261–5267.
- 489 [29] P. Yarza, M. Richter, J. Peplies, J. Euzéby, R. Amann, K.-H. Schleifer, W. Ludwig, F.O.
490 Glöckner, and R. Rosselló-Móra, The all-species living tree project: a 16S rRNA-
491 based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* 31 (2008)
492 241–50.
- 493 [30] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, Basic local alignment
494 search tool. *J Mol Biol* 215 (1990) 403–10.

- 495 [31] J. Borthong, R. Omori, C. Sugimoto, O. Suthienkul, R. Nakao, and K. Ito, Comparison of
496 Database Search Methods for the Detection of *Legionella pneumophila* in Water
497 Samples Using Metagenomic Analysis. *Front Microbiol* 9 (2018) Article 1272.
- 498 [32] L. Ajam, M.B. Ouezdou, H.S. Felfoul, and R.E. Mensi, Characterization of the Tunisian
499 phosphogypsum and its valorization in clay bricks. *Constr Build Mater* 23 (2009)
500 3240–3247.
- 501 [33] Y.-Q. Chen, G.-J. Ren, S.-Q. An, Q.-Y. Sun, C.-H. Liu, and J.-L. Shuang, Changes of
502 Bacterial Community Structure in Copper Mine Tailings After Colonization of Reed
503 (*Phragmites communis*). *Pedosphere* 18 (2008) 731–740.
- 504 [34] C. Méndez-García, A.I. Peláez, V. Mesa, J. Sánchez, O.V. Golyshina, and M. Ferrer,
505 Microbial diversity and metabolic networks in acid mine drainage habitats. *Front.*
506 *Microbiol.* 6 (2015) 475.
- 507 [35] G. Rastogi, S. Osman, P.A. Vaishampayan, G.L. Andersen, L.D. Stetler, and R.K. Sani,
508 Microbial Diversity in Uranium Mining-Impacted Soils as Revealed by High-Density
509 16S Microarray and Clone Library. *Microb Ecol* 59 (2010) 94–108.
- 510 [36] E. Xiao, V. Krumins, Y. Dong, T. Xiao, Z. Ning, Q. Xiao, and W. Sun, Microbial
511 diversity and community structure in an antimony-rich tailings dump. *Appl Microbiol*
512 *Biotechnol* 100 (2016) 7751–63.
- 513 [37] S.A. Wakelin, R.R. Anand, F. Reith, A.L. Gregg, R.R.P. Noble, K.C. Goldfarb, G.L.
514 Andersen, T.Z. DeSantis, Y.M. Piceno, and E.L. Brodie, Bacterial communities
515 associated with a mineral weathering profile at a sulphidic mine tailings dump in arid
516 Western Australia. *FEMS Microbiol Ecol* 79 (2012) 298–311.
- 517 [38] T. Hema, K. Getha, G. Tan, H.L. Sahira, A.M. Syamil, and M.N. Fairuz, Actinobacteria
518 isolates from tin tailings and forest soil for bioremediation of heavy metals. *J Trop For*
519 *Sci* 26 (2014) 153–162.
- 520 [39] S. El Baz, M. Baz, M. Barakate, L. Hassani, A. El Gharmali, and B. Imzilin, Resistance to
521 and accumulation of heavy metals by actinobacteria isolated from abandoned mining
522 areas. *Sci World J* 2015 (2015) 761834.
- 523 [40] M. Haakensen, C.M. Dobson, H. Deneer, and B. Ziola, Real-time PCR detection of
524 bacteria belonging to the *Firmicutes* Phylum. *Int J Food Microbiol* 125 (2008) 236–
525 41.
- 526 [41] E. Kim, S. Choi, J.-W. Bae, C.-J. Cha, W.-T. Im, K.-Y. Jahng, K.-s. Joh, and H. Yi, A
527 report of 10 unrecorded bacterial species of Korea, belonging to the phylum
528 *Firmicutes*. *J Species Res* 5 (2016) 235–240.
- 529 [42] H. Korehi, M. Blothe, and A. Schippers, Microbial diversity at the moderate acidic stage
530 in three different sulfidic mine tailings dumps generating acid mine drainage. *Res*
531 *Microbiol* 165 (2014) 713–718.
- 532 [43] D.M. Kassab, and T.M. Roane, Differential responses of a mine tailings *Pseudomonas*
533 isolate to cadmium and lead exposures. *Biodegradation* 17 (2006) 379–87.
- 534 [44] D. Siddavattam, T.B. Karegoudar, S.K. Mudde, N. Kumar, R. Baddam, T.S. Avasthi, and
535 N. Ahmed, Genome of a novel isolate of *Paracoccus denitrificans* capable of
536 degrading N,N-dimethylformamide. *J Bacteriol* 193 (2011) 5598–9.
- 537 [45] K. Eddouaouda, S. Mnif, A. Badis, S.B. Younes, S. Cherif, S. Ferhat, N. Mhiri, M.
538 Chamkha, and S. Sayadi, Characterization of a novel biosurfactant produced by
539 *Staphylococcus* sp. strain 1E with potential application on hydrocarbon
540 bioremediation. *J Basic Microbiol* 52 (2012) 408–18.
- 541 [46] L. Xu, W. Shi, X.C. Zeng, Y. Yang, L. Zhou, Y. Mu, and Y. Liu, Draft genome sequence
542 of *Arthrobacter* sp. strain B6 isolated from the high-arsenic sediments in Datong
543 Basin, China. *Stand Genomic Sci* 12 (2017) 11.

- 544 [47] E. Islam, D. Paul, and P. Sar, Microbial diversity in uranium deposits from Jaduguda and
545 Bagjata uranium mines, India as revealed by clone library and denaturing gradient gel
546 electrophoresis Analyses. *Geomicrobiol J* 31 (2014) 862–874
- 547 [48] T. Tsuruta, Applications using bacteria isolated from the uranium deposits. in: I. Ahmed,
548 F. Ahmed, and J. Pichtel, (Eds.), *Microbes and Microbial Technology: Agricultural
549 and Environmental Applications*, Springer, New York, 2011, pp. 267–282.
- 550 [49] S. Choudhary, E. Islam, S.K. Kazy, and P. Sar, Uranium and other heavy metal
551 resistance and accumulation in bacteria isolated from uranium mine wastes. *J Environ
552 Sci Health A Tox Hazard Subst Environ Eng* 47 (2012) 622–37.
- 553 [50] A.P. Chung, T. Sousa, A. Pereira, and P. V. Morais, Microorganisms – Tools for
554 bioremediation of uranium contaminated environments. *Proced Earth Plan Sci* 8 (2014
555) 53–58
- 556 [51] I. Shuryak, V.Y. Matrosova, E.K. Gaidamakova, R. Tkavc, O. Grichenko, P.
557 Klimenkova, R.P. Volpe, and M.J. Daly, Microbial cells can cooperate to resist high-
558 level chronic ionizing radiation. *PLoS ONE* 12 (2017) e0189261.
- 559 [52] M. Beaulieu, Guide d'intervention-Protection des sols et réhabilitation des terrains
560 contaminés. Ministère du Développement durable, de l'Environnement et de la Lutte
561 contre les changements climatiques ISBN 978-2-550-76171-6 (2016) 192.
- 562 [53] Y. Xie, J. Fan, W. Zhu, E. Amombo, Y. Lou, L. Chen, and J. Fu, Effect of heavy metals
563 pollution on soil microbial diversity and bermudagrass genetic variation. *Front Plant
564 Sci* 7 (2016) 755.
- 565 [54] M.J. Kwon, J.S. Yang, S. Lee, G. Lee, B. Ham, M.I. Boyanov, K.M. Kemner, and E.J.
566 O'Loughlin, Geochemical characteristics and microbial community composition in
567 toxic metal-rich sediments contaminated with Au-Ag mine tailings. *J Hazard Mater*
568 296 (2015) 147–57.
- 569 [55] O. Bruneel, N. Mghazli, R. Hakkou, I. Dahmani, A. Filali Maltouf, and L. Sbabou, In-
570 depth characterization of bacterial and archaeal communities present in the abandoned
571 Kettara pyrrhotite mine tailings (Morocco). *Extremophiles* 21 (2017) 671–685.
- 572 [56] Y.E. Navarro-Noya, J. Jan-Roblero, C. Gonzalez-Chavez Mdel, R. Hernandez-Gama,
573 and C. Hernandez-Rodriguez, Bacterial communities associated with the rhizosphere
574 of pioneer plants (*Bahia xylopoda* and *Viguiera linearis*) growing on heavy metals-
575 contaminated soils. *A Van Leeuw* 97 (2010) 335–49.
- 576 [57] M.O. Mendez, and R.M. Maier, Phytostabilization of mine tailings in arid and semiarid
577 environments—An emerging remediation technology. *Environ Health Persp* 116
578 (2008) 278–283.
- 579 [58] M. Dary, M.A. Chamber-Pérez, A.J. Palomares, and E. Pajuelo, “*In situ*”
580 phytostabilisation of heavy metal polluted soils using *Lupinus luteus* inoculated with
581 metal resistant plant-growth promoting rhizobacteria. *J Hazard Mater* 177 (2010) 323–
582 330.
- 583 [59] X. Wang, D.V. Mavrodi, L. Ke, O.V. Mavrodi, M. Yang, L.S. Thomashow, N. Zheng,
584 D.M. Weller, and J. Zhang, Biocontrol and plant growth-promoting activity of
585 rhizobacteria from Chinese fields with contaminated soils. *Microb Biotechnol* 8
586 (2015) 404–18.
- 587 [60] Y. Ma, C. Zhang, R.S. Oliveira, H. Freitas, and Y. Luo, Bioaugmentation with
588 Endophytic Bacterium E6S Homologous to *Achromobacter piechaudii* Enhances
589 Metal Rhizoaccumulation in Host *Sedum plumbizincicola*. *Front Plant Sci* 7 (2016) 75.
- 590 [61] S.-L. Luo, L. Chen, J.-L. Chen, X. Xiao, T.-Y. Xu, Y. Wan, C. Rao, C.-B. Liu, Y.-T. Liu,
591 C. Lai, and G.-M. Zeng, Analysis and characterization of cultivable heavy metal-
592 resistant bacterial endophytes isolated from Cd-hyperaccumulator *Solanum nigrum* L.
593 and their potential use for phytoremediation. *Chemosphere* 85 (2011) 1130–1138.

- 594 [62] Y. Ma, M. Rajkumar, and H. Freitas, Improvement of plant growth and nickel uptake by
595 nickel resistant-plant-growth promoting bacteria. *J Hazard Mater* 166 (2009) 1154–61.
- 596 [63] X. Yu, Y. Li, C. Zhang, H. Liu, J. Liu, W. Zheng, X. Kang, X. Leng, K. Zhao, Y. Gu, X.
597 Zhang, Q. Xiang, and Q. Chen, Culturable heavy metal-resistant and plant growth
598 promoting bacteria in V-Ti magnetite mine tailing soil from Panzhihua, China. *PLoS*
599 *One* 9 (2014) e106618.
- 600 [64] C. Saha, G. Mukherjee, P. Agarwal-Banka, and A. Seal, A consortium of non-rhizobial
601 endophytic microbes from *Typha angustifolia* functions as probiotic in rice and
602 improves nitrogen metabolism. *Plant Biol* 18 (2016) 938–946.
- 603 [65] S. Mahieu, J. Escarré, B. Brunel, A. Méjamolle, S. Soussou, A. Galiana, and J.-C.
604 Cleyet-Marel, Soil nitrogen balance resulting from N fixation and rhizodeposition by
605 the symbiotic association *Anthyllis vulneraria/Mesorhizobium metallidurans* grown in
606 highly polluted Zn, Pb and Cd mine tailings. *Plant Soil* 375 (2014) 175–188.
- 607 [66] Y.E. Navarro-Noya, E. Hernández-Mendoza, J. Morales-Jiménez, J. Jan-Roblero, E.
608 Martínez-Romero, and C. Hernández-Rodríguez, Isolation and characterization of
609 nitrogen fixing heterotrophic bacteria from the rhizosphere of pioneer plants growing
610 on mine tailings. *Appl Soil Ecol* 62 (2012) 52–60.
- 611 [67] A.S. Sánchez-López, S. Thijs, B. Beckers, M.C. González-Chávez, N. Weyens, R.
612 Carrillo-González, and J. Vangronsveld, Community structure and diversity of
613 endophytic bacteria in seeds of three consecutive generations of *Crotalaria pumila*
614 growing on metal mine residues. *Plant Soil* (2017) 51–66.
- 615 [68] N. De La Torre-Ruiz, V.M. Ruiz-Valdiviezo, C.I. Rincon-Molina, M. Rodriguez-
616 Mendiola, C. Arias-Castro, F.A. Gutierrez-Miceli, H. Palomeque-Dominguez, and R.
617 Rincon-Rosales, Effect of plant growth-promoting bacteria on the growth and fructan
618 production of *Agave americana* L. *Braz J Microbiol* 47 (2016) 587–96.
- 619 [69] L. Chen, S. Luo, J. Chen, Y. Wan, X. Li, C. Liu, and F. Liu, A comparative analysis of
620 endophytic bacterial communities associated with hyperaccumulators growing in mine
621 soils. *Environ Sci Pollut Res Int* 21 (2014) 7538–47.
- 622 [70] G.H. Huang, H.H. Tian, H.Y. Liu, X.W. Fan, Y. Liang, and Y.Z. Li, Characterization of
623 plant-growth-promoting effects and concurrent promotion of heavy metal
624 accumulation in the tissues of the plants grown in the polluted soil by *Burkholderia*
625 strain LD-11. *Int J Phytoremediation* 15 (2013) 991–1009.
- 626 [71] M.P. McLeod, R.L. Warren, W.W.L. Hsiao, N. Araki, M. Myhre, C. Fernandes, D.
627 Miyazawa, W. Wong, A.L. Lillquist, D. Wang, M. Dosanjh, H. Hara, A. Petrescu,
628 R.D. Morin, G. Yang, J.M. Stott, J.E. Schein, H. Shin, D. Smailus, A.S. Siddiqui,
629 M.A. Marra, S.J.M. Jones, R. Holt, F.S.L. Brinkman, K. Miyauchi, M. Fukuda, J.E.
630 Davies, W.W. Mohn, and L.D. Eltis, The complete genome of *Rhodococcus* sp. RHA1
631 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A* 103 (2006)
632 15582–15587.
- 633 [72] H. Moreira, S.I.A. Pereira, A.P.G.C. Marques, A.O.S.S. Rangel, and P.M.L. Castro,
634 Mine land valorization through energy maize production enhanced by the application
635 of plant growth-promoting rhizobacteria and arbuscular mycorrhizal fungi. *Environ*
636 *Sci Pollut Res* (2015) 6940–6950.
- 637 [73] L.E. de-Bashan, J.-P. Hernandez, Y. Bashan, and R.M. Maier, *Bacillus pumilus* ES4:
638 Candidate plant growth-promoting bacterium to enhance establishment of plants in
639 mine tailings. *Environ Exp Bot* 69 (2010) 343–352.
- 640 [74] W. Liu, C. Yang, S. Shi, and W. Shu, Effects of plant growth-promoting bacteria isolated
641 from copper tailings on plants in sterilized and non-sterilized tailings. *Chemosphere*
642 97 (2014) 47–53.

- 643 [75] L. Benidire, S.I.A. Pereira, P.M.L. Castro, and A. Boularbah, Assessment of plant growth
644 promoting bacterial populations in the rhizosphere of metallophytes from the Kettara
645 mine, Marrakech. *Environ Sci Pollut Res* 23 (2016) 21751–21765.
- 646 [76] E. Tamburini, S. Sergi, L. Serreli, G. Bacchetta, S. Milia, G. Cappai, and A. Carucci,
647 Bioaugmentation-Assisted Phytostabilisation of Abandoned Mine Sites in South West
648 Sardinia. *Bull Environ Contam Toxicol* 98 (2017) 310–316.
- 649 [77] S. Azabou, T. Mechichi, and S. Sayadi, Sulfate reduction from phosphogypsum using a
650 mixed culture of sulfate-reducing bacteria. *Int Biodeterior Biodegrad* 56 (2005) 236–
651 242.
- 652 [78] M. Martins, M.L. Faleiro, R.J. Barros, A.R. Veríssimo, M.A. Barreiros, and M.C. Costaa,
653 Characterization and activity studies of highly heavy metal resistant sulphate-reducing
654 bacteria to be used in acid mine drainage decontamination. *J Hazard Mater* 166 (2009)
655 706–713.
- 656 [79] M.S.P. Marvi, A.A. Pourbabaee, H.A. Alikhani, A. Haidari, and Z. Manafi, The
657 diversity of sulfur-oxidizing bacterial populations at an Iranian copper mine and the
658 surrounding agricultural soils. *Appl ecol env res* 14 (2016) 509–533.
- 659 [80] W. Sajjad, T.M. Bhatti, F. Hasan, S. Khan, M. Badshah, A.A. Naseem, and A.A. Shah,
660 Characterization of sulfur-oxidizing bacteria isolated from acid mine drainage and
661 black shale samples. *Pak J Bot* 48 (2016) 1253–1262.
- 662 [81] J.S. Benedetto, S.K.d. Almeida, H.A. Gomes, R.F. Vazoller, and A.C.Q. Ladeira,
663 Monitoring of sulfate-reducing bacteria in acid water from uranium mines. *Minerals*
664 *Engineering* 18 (2005) 1341–1343.
- 665 [82] S. Choudhary, and P. Sar, Characterization of a metal resistant *Pseudomonas* sp. isolated
666 from uranium mine for its potential in heavy metal (Ni^{2+} , Co^{2+} , Cu^{2+} , and Cd^{2+})
667 sequestration. *Bioresour Technol* 100 (2009) 2482–92.
- 668 [83] N. Hamamura, K. Fukushima, and T. Itai, Identification of antimony- and arsenic-
669 oxidizing bacteria associated with antimony mine tailing. *Microbes Environ* 28 (2013)
670 257–63.
- 671 [84] P. Chang, K.E. Gerhardt, X.D. Huang, X.M. Yu, B.R. Glick, P.D. Gerwing, and B.M.
672 Greenberg, Plant growth-promoting bacteria facilitate the growth of barley and oats in
673 salt-impacted soil: implications for phytoremediation of saline soils. *Int J*
674 *Phytoremediation* 16 (2014) 1133–47.

675
676
677

678

Figure captions

679

680 **Figure 1:** (A) Taxonomic identities at the class level of PG as determined *via* Illumina MiSeq
681 16S rDNA gene amplicon sequencing and classification against the One Codex reference tool.
682 (B) Box-and-whisker plots comparing 16S-rDNA-gene-based metataxonomics abundances
683 for taxonomic units at the class level when analyzing PG using the One Codex, MG-RAST,
684 mothur, MICCA and QIIME tools. (C) Taxonomic identities at the genus level of PG as
685 determined *via* Illumina MiSeq 16S rDNA gene amplicon sequencing and classification
686 against the One Codex reference tool. (D) Box-and-whisker plots comparing 16S-rDNA-
687 gene-based metataxonomics abundances for taxonomic units at the genus level when
688 analyzing PG using the One Codex, MG-RAST, mothur, MICCA and QIIME tools. The
689 classes and genera shown are those that were >1% of the community.

690 **Figure 2:** Phylogenetic tree representing PG-native strains valorizable into (1) removal of
691 radioactive elements and toxic heavy metals (2) promotion of plant growth (3) oxidation and
692 (4) reduction of sulfate. Histograms indicate reads' percentage for each species as determined
693 by the One Codex reference tool. See text for details.

694

695 **Table 1:** Observed generic OTUs (0.03 cutoff) and Shannon's diversity indices at the genus
 696 level.
 697

Analysis method	Good's library coverage (%)	Observed OTUs	Shannon's diversity index (H)
One Codex	98.19	128	3.90
MG-RAST	98.04	83	ND
mothur	99	135	3.92
MICCA	33	66	4.24
QIIME	76	249	5.77
QIIME [13]	ND	224	4.51

698

699

700

701 **Table 2:** Major phyla, classes, orders, families, genera and species of Tunisian PG based on
 702 One Codex, as a reference method, and four other pipelines (MG-RAST, mothur, MICCA and
 703 QIIME).

	One Codex	MG-RAST	mothur	MICCA	QIIME	QIIME [13]
Phylum	Proteobacteria (46.76%)	Firmicutes (24.03%)	Proteobacteria (50.48%)	Proteobacteria (48.24%)	Proteobacteria (41.21%)	Proteobacteria (70%)
Class	Actinobacteria (18.27%)	Bacilli (23.76%)	Gammaproteobacteria (27.95%)	Gammaproteobacteria (29.75%)	Actinobacteria (37.34%)	Gammaproteobacteria (≈55%)
Order	Micrococcales (13.55%)	Bacillales (23.22%)	Pseudomonadales (15.85%)	Bacillales (23.8%)	Bacillales (17%)	ND
Family	<i>Micrococcaceae</i> (13.41%)	<i>Micrococcaceae</i> (11.82%)	<i>Pseudomonadaceae</i> (13.76%)	<i>Micrococcaceae</i> (16.24%)	<i>Micrococcaceae</i> (26.997%)	ND
Genus	<i>Arthrobacter</i> (13.25%)	<i>Bacillus</i> (15.61%)	<i>Pseudomonas</i> (4.58%)	<i>Arthrobacter</i> (12.88%)	<i>Arthrobacter</i> (20.5%)	<i>Desulfobacterium</i> (1.74%)
Species	<i>Staphylococcus aureus</i> (2.9%)	<i>Bacillus cereus</i> (10.91%)	<i>Pseudomonas fluorescens</i> (1.068%)	<i>Staphylococcus aureus</i> (2.71%)	<i>Acidithiobacillus albertensis</i> (13%)	ND

704 Abundance is expressed in parentheses.

705

706

707 **Table 3:** Chemical composition and radioactivity of Tunisian phosphogypsum (PG).

Composition	PG
pH	3.36
Calcium (Ca)	20.73%
Sulfur (S)	11.193%
Phosphate (P)	0.62%
Iron (Fe)	0.022%
Magnesium (Mg)	0.018%
Strontium (Sr)	646.33 ppm
Zinc (Zn)	113.65 ppm
Cadmium (Cd)	20.19 ppm
Nickel (Ni)	5.20 ppm
Copper (Cu)	4.56 ppm
²¹⁴ Pb	210.86±1.70 Bq/kg
²²⁶ Ra	200.58±2.46 Bq/kg
²¹⁴ Bi	190.31±1.78 Bq/kg
²³⁸ U	93.92±15.54 Bq/kg
²²⁸ Ac	28.49±4.42 Bq/kg
²³² Th	28.49±4.92 Bq/kg
²¹² Pb	27.50±2.44 Bq/kg
⁴⁰ K	12.94±10.40 Bq/kg

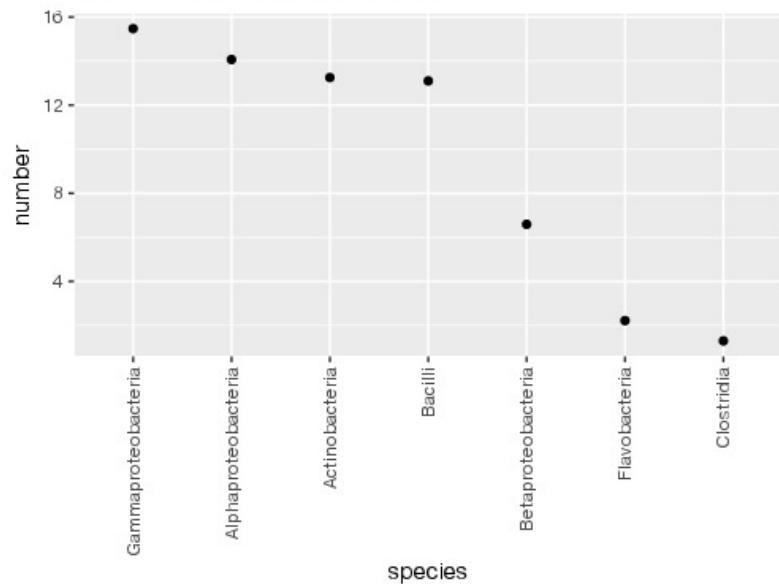
708

709

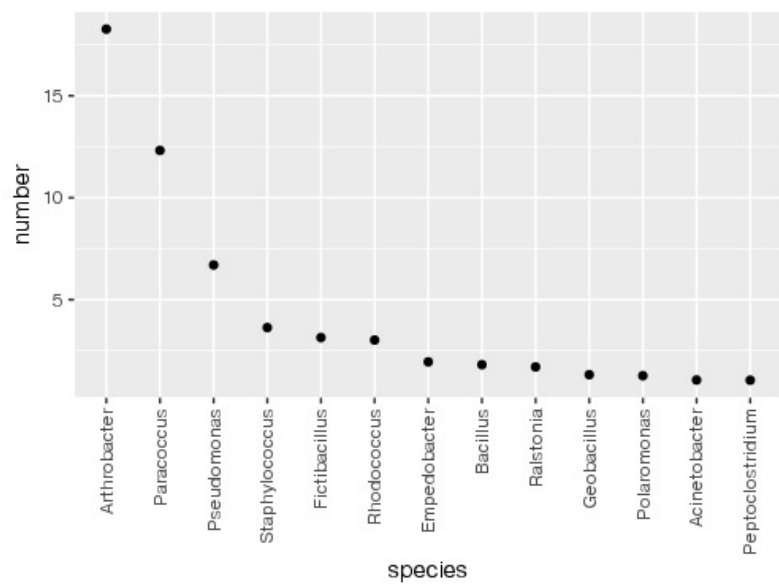
710

711

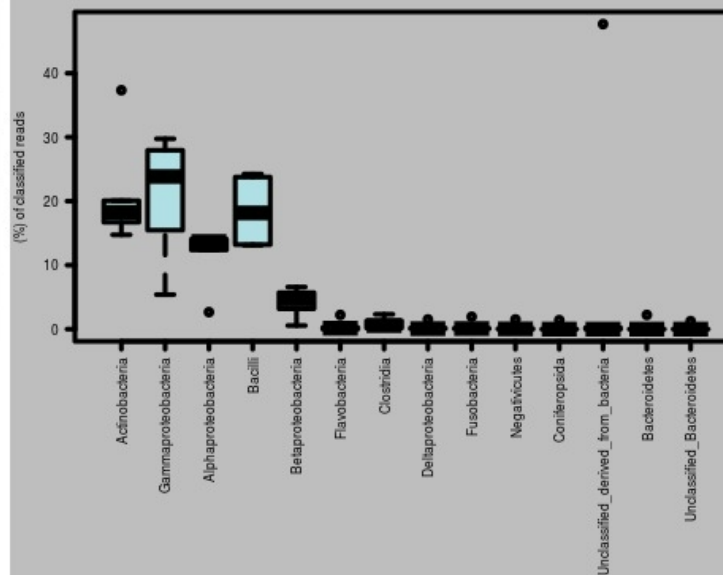
OneCodex Class distribution



OneCodex Genus distribution



Bacterial Class distribution using the 5 tools



Bacterial Genus distribution using the 5 tools

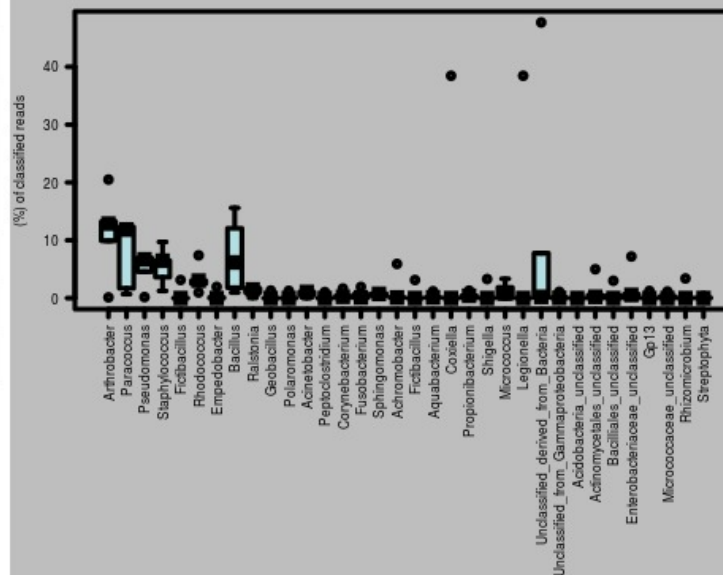


Table 1: Observed generic OTUs (0.03 cutoff) and Shannon's diversity indices at the genus level.

Analysis method	Good's library coverage (%)	Observed OTUs	Shannon's diversity index (H)
One Codex	98.19	128	3.90
MG-RAST	98.04	83	ND
mothur	99	135	3.92
MICCA	33	66	4.24
QIIME	76	249	5.77
QIIME (Zouch et al., 2017)	ND	224	4.51

Table 2: Major phyla, classes, orders, families, genera and species of Tunisian PG based on One Codex, as a reference method, and four other pipelines (MG-RAST, mothur, MICCA and QIIME).

	One Codex	MG-RAST	mothur	MICCA	QIIME	QIIME (Zouch et al. 2017)
Phylum	Proteobacteria (46.76%)	Firmicutes (24.03%)	Proteobacteria (50.48%)	Proteobacteria (48.24%)	Proteobacteria (41.21%)	Proteobacteria (70%)
Class	Actinobacteria (18.27%)	Bacilli (23.76%)	Gammaproteobacteria (27.95%)	Gammaproteobacteria (29.75%)	Actinobacteria (37.34%)	Gammaproteobacteria (≈55%)
Order	Micrococcales (13.55%)	Bacillales (23.22%)	Pseudomonadales (15.85%)	Bacillales (23.8%)	Bacillales (17%)	ND
Family	<i>Micrococcaceae</i> (13.41%)	<i>Micrococcaceae</i> (11.82%)	<i>Pseudomonadaceae</i> (13.76%)	<i>Micrococcaceae</i> (16.24%)	<i>Micrococcaceae</i> (26.997%)	ND
Genus	<i>Arthrobacter</i> (13.25%)	<i>Bacillus</i> (15.61%)	<i>Pseudomonas</i> (4.58%)	<i>Arthrobacter</i> (12.88%)	<i>Arthrobacter</i> (20.5%)	<i>Desulfobacterium</i> (1.74%)
Species	<i>Staphylococcus aureus</i> (2.9%)	<i>Bacillus cereus</i> (10.91%)	<i>Pseudomonas fluorescens</i> (1.068%)	<i>Staphylococcus aureus</i> (2.71%)	<i>Acidithiobacillus albertensis</i> (13%)	ND

Abundance is expressed in parentheses.

Table 3: Chemical composition and radioactivity of Tunisian phosphogypsum (PG).

Composition	PG
pH	3.36
Calcium (Ca)	20.73%
Sulfur (S)	11.193%
Phosphate (P)	0.62%
Iron (Fe)	0.022%
Magnesium (Mg)	0.018%
Strontium (Sr)	646.33 ppm
Zinc (Zn)	113.65 ppm
Cadmium (Cd)	20.19 ppm
Nickel (Ni)	5.20 ppm
Copper (Cu)	4.56 ppm
²¹⁴ Pb	210.86±1.70 Bq/kg
²²⁶ Ra	200.58±2.46 Bq/kg
²¹⁴ Bi	190.31±1.78 Bq/kg
²³⁸ U	93.92±15.54 Bq/kg
²²⁸ Ac	28.49±4.42 Bq/kg
²³² Th	28.49±4.92 Bq/kg
²¹² Pb	27.50±2.44 Bq/kg
⁴⁰ K	12.94±10.40 Bq/kg

Conflict of interest

The authors received no financial support for the research and/or authorship of this article. The authors declare that they have no conflict of interest in the publication of this article, and they have participated in the study and concur with the submission and subsequent revisions submitted by the corresponding author.