Systematic and Applied Microbiology 48 (2025) 126622

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.com/locate/syapm

Classification of MAGs associated with trace gas metabolism in volcanic soils named following SeqCode rules☆

Shamik Roy [a,b], Gary M. King [c], Marcela Hernández [a,*]

[a] School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK
[b] Chair for Forest Zoology, Technische Universität Dresden, Tharandt 01737, Germany
[c] Deptartment of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

ARTICLE INFO

Keywords:
Volcanic soils
CO-dehydrogenase
Trace gases
SeqCode
Chloroflexota
Archaea

ABSTRACT

Trace gas metabolism is important for nutrient flow in all ecosystems, particularly volcanic ecosystems. Microbes in volcanic ecosystems are among the early colonisers and can play key roles in ecological succession. Here, we describe the taxonomic and functional characteristics of two new metagenome-assembled genomes (MAGs), one belonging to *Bacteria* (MAG_1957-2.1) and one to *Archaea* (MAG_C2-3), retrieved from soils in volcanoes located in Chile (Llaima) and the USA (Kilauea), respectively. MAG_1957-2.1 has a genome size of 6.36 Mb with 96.21 % completeness. MAG_C2-3 has a genome size of 3.02 Mb with 97.57 % completeness. Phylogenetic analyses of the bacterial MAGs placed MAG_1957-2.1 in the class *Ktedonobacteria*, while the archaeal MAG_C2-3 was placed in the class *Nitrososphaeria*. Functional characterisation for potential trace gas metabolism showed that MAG_1957-2.1 contains a *coxL* gene encoding the large subunit of form I carbon monoxide dehydrogenase (CoxL), which is associated with the oxidation of carbon monoxide (CO). It also contains the form I *cox* gene cluster with a *coxMSL* arrangement. On the other hand, MAG_C2-3 contains gene subunit A (*amoA*) as well as subunit B (*amoB*), which encode for ammonia monooxygenase, the enzyme that catalyses the oxidation of ammonia. Based on the sequence characteristics and phylogenomic analyses we propose the names *Paraktedonobacter carboxidivorans sp.* nov for MAG_1957-2.1 and *Nitrososphaera maunauluensis sp.* nov for MAG_C2-3. The names are proposed following the rules of the published Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode).

Introduction

Atmospheric trace gases are essential determinants of climate change (Hansen et al., 2007). Therefore, any alteration in their relative abundance in the atmosphere can have disproportionate consequences for ecosystem processes such as nutrient cycling, primary productivity, and trophic interactions. Biogenic trace gas flux between soil and atmosphere is primarily regulated by microorganisms through their metabolism (Bay et al., 2021; Conrad, 1996; Fenchel et al., 2012). Microbial controls of the fluxes of trace gases can be direct or indirect (Bay et al., 2021; Conrad, 1996; Fenchel et al., 2012). Direct controls are mediated by the utilisation of trace gases for growth and maintenance, where the gases serve as electron donors or acceptors in the redox reactions and acquire energy. Indirect controls are mediated by the production of trace gases as end products of dissimilatory pathways. Although we know much about the biochemistry of trace gas metabolism, we know relatively little about the taxonomy of trace gas oxidisers, particularly in newly formed soils such as volcanic ecosystems (Fenchel et al., 2012; King, 2015; King and Weber, 2008). Identifying and classifying trace gas-oxidising microbes will enhance our understanding of their distribution, ecology, interactions with organisms, and their impact on biogeochemistry. Therefore, it is crucial to classify and investigate the taxonomic distribution of these microorganisms.

Volcanic ecosystems are distributed on all continents and provide model systems to study pedogenesis and ecological succession. Microorganisms are the first inhabitants of volcanic deposits (*e.g.*, lava, cinders, tephra, and ash) following a drop in temperature, and changes in moisture to tolerable levels (Gomez-Alvarez et al., 2007; Kim et al., 2018; King, 2003; Ohta et al., 2003). They are hypothesised to derive energy either from photosynthesis or through oxidation of trace gases

☆ This article is part of a Special issue entitled: 'Application of the SeqCode' published in Systematic and Applied Microbiology.

* Corresponding author.
E-mail address: marcela.hernandez@uea.ac.uk (M. Hernández).

https://doi.org/10.1016/j.syapm.2025.126622
Received 6 June 2024; Received in revised form 29 January 2025; Accepted 27 May 2025
Available online 28 May 2025

such as methane (CH_4), carbon monoxide (CO), hydrogen (H_2), and hydrogen sulfide (H_2S). Microbial oxidation of trace gases in the newly exposed or newly formed terrestrial surfaces following volcanic eruptions can enrich local biodiversity and stimulate primary productivity, which ultimately forms the functional basis of primary succession (Hernández et al., 2020a; Hernández et al., 2020b; Ni et al., 2023).

In addition to *Bacteria*, the *Archaea* domain represents a significant fraction of the microbial biomass on Earth. *Archaea* also perform crucial functions in the biogeochemical cycles and influence the dynamics of trace gas metabolism (Offre et al., 2013). For instance, methanogenesis by *Archaea* is estimated to result in about 1 Gt of methane production per year (Thauer et al., 2008). Still, compared to *Bacteria*, our knowledge and understanding of the diversity and ecology of *Archaea* is limited. This is because studies have heavily relied on culturing and single-gene diversity assessments to classify and characterise *Archaea*. However, recent advancements in genomic sequencing and computational approaches have resulted in the acceleration of culture-independent genome recovery following metagenomic assembly and subsequent clustering of DNA fragments into metagenome-assembled genomes (MAGs) (Nayfach et al., 2021; Rinke et al., 2021; Tyson et al., 2004). This has led to the description and naming of several new archaeal lineages along with their functional characterisation (Meng et al., 2014; Rinke et al., 2021; Vigneron et al., 2022).

In our previous study, we assembled MAGs of different carbon monoxide oxidisers from the soils in Llaima representing different successional stages (Hernández et al., 2020b). Two MAGs belonging to the order *Ktedonobacterales* (phylum *Chloroflexota*) from the youngest site (Lava deposition in 1957) were of particular interest as *Ktedonobacterales* are hypothesised to be important for early soil formation. However, our knowledge of the diversity and ecological significance of *Ktedonobacterales* is limited because very few cultured strains have been characterised (Hernández et al., 2020b). Most of the known *Ktedonobacterales* genomes are either available as draft genomes in the NCBI database or have been retrieved from metagenomic assemblies.

Analyses of metagenomes and MAGs have a vital role in our growing understanding of the ecological importance of microbial trace gas metabolism. The large number of genomes derived from metagenomic analyses of yet-to-be-cultured organisms underscores the significance of classifying and naming these new taxa (Konstantinidis et al., 2017). This includes those trace-gas metabolisers, which play a crucial role in global biogeochemical cycle. Recently, the Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode), a nomenclature system designed for naming prokaryotes based on genome sequences, including isolate genomes, metagenome-assembled genomes, or single-amplified genome sequences, has been made available (Hedlund et al., 2022; Whitman et al., 2022). SeqCode allows researchers to name both cultivated and uncultivated microorganisms based on genome sequence data, the common denominator for all microorganisms. Additionally, SeqCode does not violate the priority of names validly published under the rules of the International Code of Nomenclature of Prokaryotes (ICNP).

In this study, we present the classification of one yet-uncultivated bacterial species and one yet-uncultivated archaeal species represented by the recovered metagenome assembled genomes (MAGs) from early-stage volcanic soils following the rules of the newly published SeqCode (Hedlund et al., 2022; Whitman et al., 2022). In addition, we have characterised these MAGs with respect to their potential for metabolising trace gas. In a previous study, we characterised the bacterial MAG described here (Hernández et al., 2020b), but a detailed classification was lacking. Therefore, in this study, we follow the SeqCode rules to identify these new MAGs from volcanic deposits.

Methods

Sampling and retrieval of MAGs

The bacterial MAG characterised in this study was retrieved from volcanic deposits collected from sites established after the eruption in 1957 of Llaima (38.57953°S, 71.78722°W). The sampling campaign and the retrieval and initial characterisation of the MAG have been previously described in Hernández et al. (2020b). This MAG is part of the project that has been submitted to NCBI-SRA with code PRJNA602601 (Hernández et al., 2020b). The genome accession number is GCA_041156125.1 for MAG_1957-2.1.

The archaeal MAG was retrieved from a sampling campaign to Kilauea, Hawaii, in April 2022. Three soil samples were collected randomly from Mauna Ulu site (approximately 60 years old soil; 19.3388382°N, 155.2033492°W, permit number: HAVO-2022-SCI-0012). All soil samples were stored in polyethylene bags until transported to the UK. Total DNA was extracted from each soil sample (0.5 g of soil) by using the Qiagen DNeasy PowerSoil Pro Kit according to the manufacturer's recommendation with minor modifications (samples were lysed in a FastPrep bead beater for 30 s at 6 m s-1). DNA was quantified using Qubit, and the DNA from the three sites were combined before being sent for DNA sequencing to Novogene UK, using Illumina Sequencing PE150 (6 GB raw data per sample). After quality checking, reads were merged into scaffolds using *de novo* assembler SPAdes version 3.14.0 (Bankevich et al., 2012). Metagenomic binning of the assembled scaffolds larger than 1000 bp was carried out with the metaWRAP version 1.2.1 pipeline using three available binners (metabat2, maxbin2, and concoct) (Uritskiy et al., 2018) The resulting bins were collectively processed to produce consolidated MAGs using the bin_refinement module in metaWrap (criterion: completeness >70 %; contamination <5 %). The genome accession number is GCA_041156085.1 for MAG_C2–3 (NCBI-SRA).

For both MAGs, completion and contamination metrics was performed using CheckM program (Parks et al., 2015).

Genome annotation

Genome sequence annotation for the two MAGs were performed using MicroScope, an online platform by GenoScope (France) providing a collection of bioinformatic tools (Vallenet et al., 2020). Médigue et al. (2019) has described all the software and databases integrated into the MicroScope pipelines used to perform the genome annotation.

Phylogeny

Phylogenetic trees for both MAGs were constructed using single-copy marker proteins and the 16S rRNA gene sequence.

For single-copy marker proteins phylogeny, the bacterial MAG was phylogenetically placed using a concatenated alignment of 120 single-copy marker proteins, while the archaeal MAG was phylogenetically placed using a concatenated set of 53 marker proteins. This was done using the 'classify' workflow in GTDB-Tk (version v2.3.0+) (Chaumeil et al., 2022). The output of the 'classify' workflow was then used to construct maximum-likelihood trees in RAxML (Kozlov et al., 2019) using default parameters with *Nocardia higoensis* (RS GCF 000308595.1) as the bacterial outgroup and *Candidatus* Caldarchaeum subterraneum (informal name: uncultured crenarchaeote 10-H-08, RS GCF 000270325.1) as the archaeal outgroup. For 16S rRNA gene phylogeny, the 16S rRNA gene for the MAGs were retrieved from the metagenome using Barrnap v. 0.9 software (https://github.com/tseemann/barrnap). 16S rRNA gene sequences for the members of the class *Ktedonobacteria* for *Bacteria* and the order *Nitrososphaerales* for *Archaea* were collected from the sequences either available in NCBI or retrieved using Barrnap v. 0.9 software (https://github.com/tseemann/barrnap). The sequences were aligned using MEGA11, and trees were constructed. Similar to

single-copy marker proteins phylogeny, trees were constructed using maximum-likelihood method in RAxML (Kozlov et al., 2019) using default parameters with *Nocardia higoensis* (AB108778) as the bacterial outgroup and *Ca.* Caldarchaeum subterraneum (fosmid clone: JFF052_F07, AP011786) as the archaeal outgroup. Relative evolutionary distances (RED) were calculated (Parks et al., 2018) using the 'get_reds' function in the 'castor' R-package (Louca and Doebeli, 2018).

Genomic relatedness indices

The average nucleotide identity (ANI) was calculated using the CJ Bioscience's online Average Nucleotide Identity (ANI) calculator (https://www.ezbiocloud.net/tools/ani, Yoon et al., 2017). The average amino acid identity (AAI) between all genomes was calculated using the BLAST tool with aai.rb scripts (Rodriguez-R and Konstantinidis, 2016). *d*DDH (digital DNA-DNA hybridisation) was computed with GGDC (Genome-to-Genome Distance Calculator) (Meier-Kolthoff et al., 2013). The genomes of organisms with which our bacterial and archaeal MAG were compared belong to the members closely related in the phylogenetic tree. For the *Bacteria* domain, we selected *Ktedonosporobacter rubrisoli* SCAWS-G2^T, *Ktedonobacter racemifer* SOSP1-21^T, and *Thermogemmatispora carboxidivorans* PM5^T. For the Archaea domain, we selected *Ca.* Nitrososphaera evergladensis SR1, *Ca.* Nitrososphaera gargensis Ga9.2, *Nitrososphaera viennensis* EN76^T, and *Ca.* Nitrosocosmicus franklandus C13.

Functional characterisation for trace gas metabolism

The MAGs were functionally characterised based on their ability to metabolise trace gases, particularly carbon monoxide, methane, and ammonia. This was done using BLASTp for the translated sequences of genes involved in metabolic pathways of interest, such as the form I *coxL* gene (for carbon monoxide), *pmoA* gene (methane), and *amoA* gene (ammonia). Phylogenetic trees for AmoA and CoxL were constructed separately for bacterial and archaeal MAGs against reference sequences. The sequences were aligned using ClustalW, and the phylogenetic tree was constructed using the neighbour-joining method with a JJT matrix-based model and 100 bootstrap replicates.

Results and discussion

Genome assemblies

The MAG_1957-2.1 (Llaima) is of high quality, with 96.21 % completeness and 3.96 % contamination (Table 1). This MAG has 56 tRNA that correspond to 19 different amino acids. The archaeal MAG_C2-3 retrieved from Kilauea is of high quality, with 97.57 % completeness and 0 % contamination (Table 1). This MAG has 26 tRNA that correspond to 14 different amino acids.

Table 1
Genomic features of the two MAGs.

	MAG_1957-2.1 (Llaima Volcano)	MAG_C2-3 (Kilauea Volcano)
Proposed name	*Paraktedonobacter carboxidivorans*	*Nitrososphaera maunauluensis*
Binned size, Mb	6.36	3.02
% completeness	96.21	97.57
% contamination	3.96	0
% coding bases	83.74	72.04
% GC content	53.43	35.54
Average gene length, bp	211.45	251.66
Total number of genes	5470	3892
Total number of tRNA	56	26
Total number of protein-coding genes	5386	3851
Number of contigs	233	434
Size of the largest contig (bp)	95,722	53,170
N50	24,071	10,533
Accession number	SRX9386060 (PRJNA602601)	SRX24172872 (PRJNA1096838)

Phylogenetic analyses

The phylogenetic tree constructed for bacterial MAG using 120 concatenated single-copy protein markers placed MAG_1957–2.1 on a distinct branch compared to the type-strains available for the genera *Ktedonobacter, Ktedonospora, Ktedonosporobacter, Thermosporothrix, Tengunoibacter,* and *Dictyobacter*. The RED value for the nearest common node was 0.7160 (Fig. 1A). A similar pattern emerges from the phylogenetic tree build using 16S rRNA gene of the MAG_1957-2.1, with a length of 1480 bp. In this tree, MAG_1957-2.1 forms a separate branch compared to the genera *Ktedonobacter, Ktedonospora, Tengunoibacter,* and *Dictyobacter* with a RED value of 0.2473 for the nearest common node (Fig. 1B). The RED values for both trees with MAG_1957-2.1 were below 0.8, the lower range of RED values for bacterial genus (Parks et al., 2018).

The phylogenetic tree constructed for the archaeal MAG using concatenated 53 single-copy protein markers placed the MAG_C2-3 close to the genus *Nitrososphaera* (Fig. 2A), with the RED value of 0.7669 for the nearest common node. The phylogenetic tree built using the 16S rRNA gene placed MAG_C2-3 close to the genera *Nitrososphaera* and *Ca.* Nitrosocosmicus (Fig. 2B), with a RED value of 0.7736 for the common node. The RED values for both trees with MAG_C2-3 were below 0.8, which is at the lower end of the range of RED values typically observed for archaeal species (Rinke et al., 2021).

Genomic relatedness

The ANI values for MAG_1957-2.1 ranged from 68.74 % to 69.22 % when compared to *Ktedonosporobacter rubrisoli* SCAWS-G2^T, *Ktedonobacter racemifer* SOSP1-21^T, and *Thermogemmatispora carboxidivorans* PM5^T. The ANI values for MAG_C2-3 ranged from 67.02 % to 67.61 % when compared to *Ca.* Nitrososphaera evergladensis SR1, *Ca.* Nitrososphaera gargensis Ga9.2, *Nitrososphaera viennensis* EN76^T, and *Ca.* Nitrosocosmicus franklandus C13 (Table 2). The ANI values for both MAGs are below the 95 % threshold for species delineation (Konstantinidis et al., 2006; Richter and Rosselló-Móra, 2009). Similarly, the AAI values (Table 2) for MAG_1957–2.1 ranged from 58.36 % to 58.96 % when compared to *Ktedonosporobacter rubrisoli* SCAWS-G2^T, *Ktedonobacter racemifer* SOSP1-21^T, and *Thermogemmatispora carboxidivorans* PM5^T. For MAG_C2-3, the AAI values ranged from 53.71 % to 61.26 % when compared to *Ca.* Nitrososphaera evergladensis SR1, *Ca.* Nitrososphaera gargensis Ga9.2, *Nitrososphaera viennensis* EN76^T, and *Ca.* Nitrosocosmicus franklandus C13. The AAI values for both MAGs were below the threshold for species delineation (65–95 %) proposed for new or novel genera in the family (Konstantinidis et al., 2017). Although the AAI value for the archaeal MAG is below this threshold, we were hesitant to propose this MAG as a new genus due to the lack of type strains and *Candidatus* representation for the genus *Nitrososphaera*. The *d*DDH values (Table 2) for MAG_1957–2.1 ranged from 18.9 % to 23.2 % when compared to *Ktedonosporobacter rubrisoli* SCAWS-G2^T, *Ktedonobacter racemifer* SOSP1-21^T, and *Thermogemmatispora carboxidivorans* PM5^T, and for MAG_C2–3 ranged from 21.2 % to 29.7 % when compared to *Ca.* Nitrososphaera evergladensis SR1, *Ca.* Nitrososphaera gargensis Ga9.2, *Nitrososphaera viennensis* EN76^T, and *Ca.* Nitrosocosmicus franklandus C13. The *d*DDH values for both MAGs were below the proposed threshold of 50 % for genus delineation (Konstantinidis et al., 2006; Richter and Rosselló-Móra, 2009).

General genomic features

MAG_1957-2.1, the bacterial MAG, possess a high G + C content of

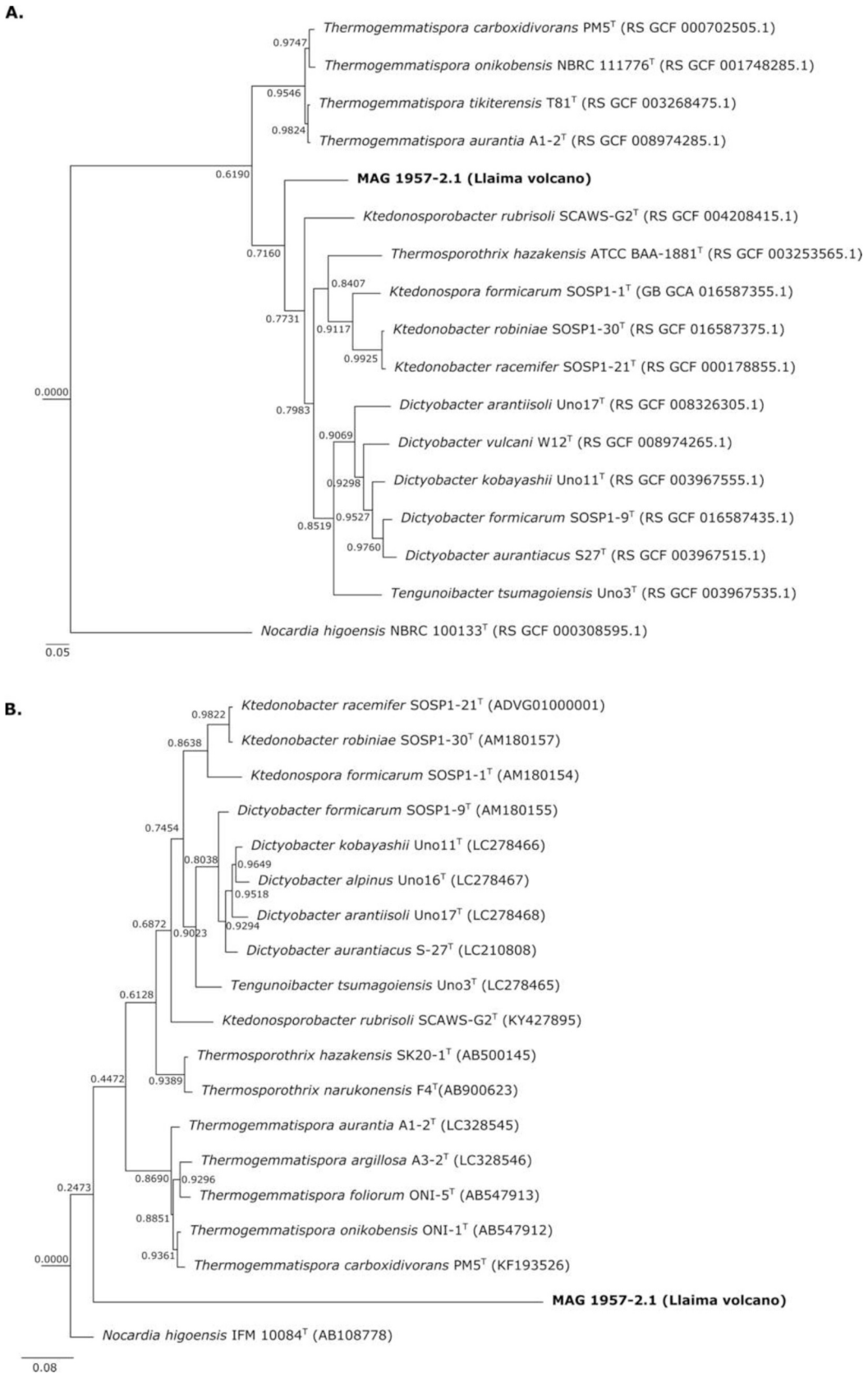

Fig. 1. Maximum-likelihood phylogenetic tree based on the concatenated protein alignment of 120 single-copy bacterial protein markers (A) and 16S rRNA gene (B) highlights the position of the bacterial MAG retrieved from volcanic deposits after the 1957 eruption of Llaima Volcano. The accession number of each sequence is given in brackets. The number on the node denotes the relative evolutionary distance (RED).

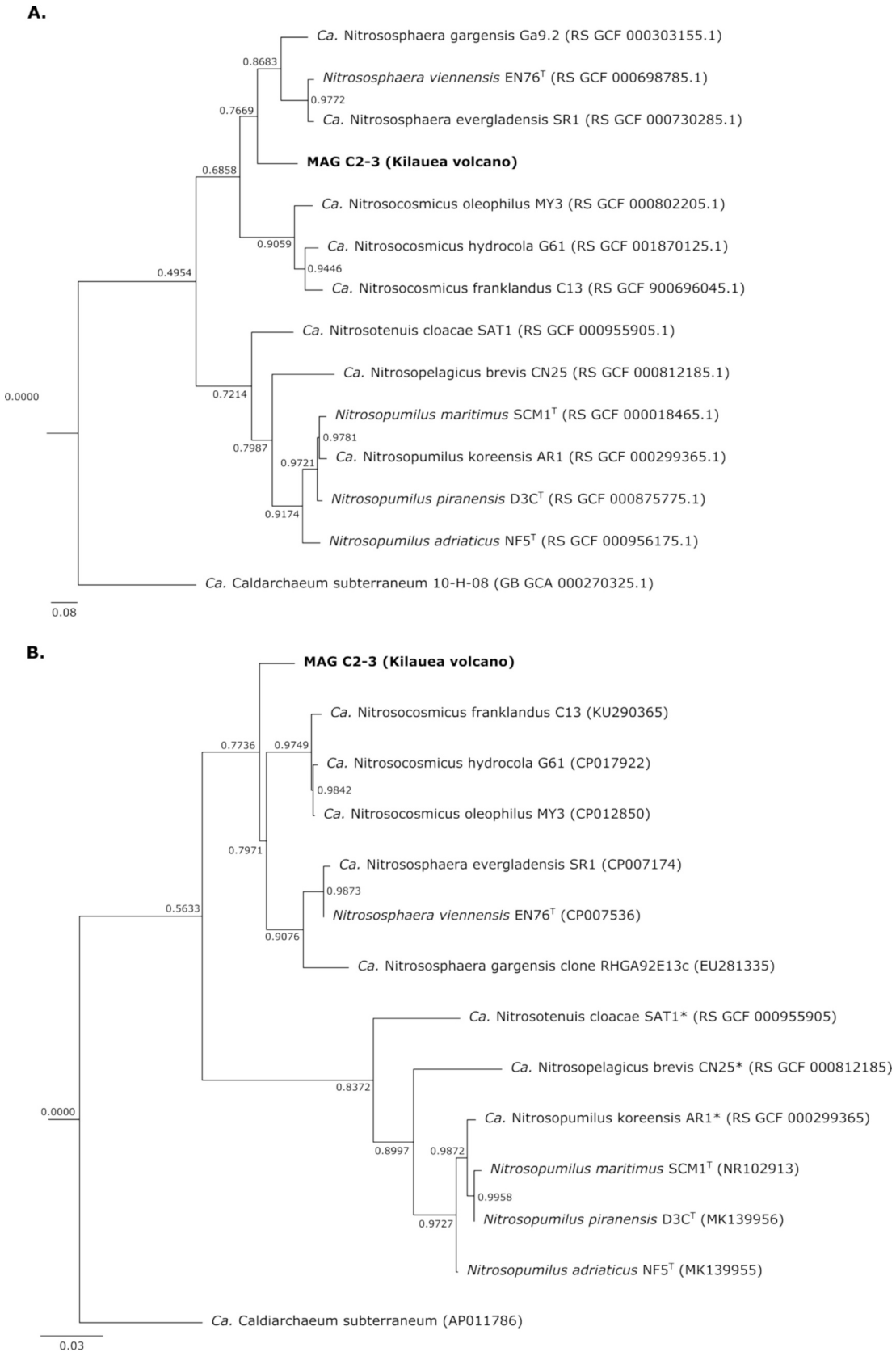

Fig. 2. Maximum-likelihood phylogenetic tree based on the concatenated protein alignment of 53 single-copy archaeal protein markers (A) and 16S rRNA gene (B) highlights the position of the archaeal MAG retrieved from volcanic deposits of Kilauea Volcano. The accession number of each sequence is given in brackets, with * marked sequences retrieved from the whole-genome sequence available in NCBI. The number on the node denotes the relative evolutionary distance (RED).

Table 2
Summary of the genomic relatedness of the MAGs and their closely related members.

MAG	Reference genomes (NCBI RefSeq assembly)	OrthoANI (%)*	AAI (%)**	dDDH (%)***
MAG_1957-2.1	*Ktedonosporobacter rubrisoli* SCAWS-G2^T (GCF_004208415.1)	68.74	58.96	19.7
	Ktedonobacter racemifer SOSP1-21^T (GCF_000178855.1)	69.22	58.36	23.2
	Thermogemmatispora carboxidivorans PM5^T (GCF_000702505.1)	68.93	58.41	18.9
MAG_C2-3	*Ca.* Nitrososphaera evergladensis SR1 (GCF_000730285.1)	67.33	59.66	23.9
	Ca. Nitrososphaera gargensis Ga9.2 (GCF_000303155.1)	67.61	61.26	29.7
	Nitrososphaera viennensis EN76^T (GCF_000698785.1)	67.02	59.88	23.7
	Ca. Nitrosocosmicus franklandus C13 (GCF_900696045.1)	67.90	53.71	28.3

* OrthoANI: orthologous Average Nucleotide Identity.
** AAI: average amino acid identity (reciprocal best hits (two-way AAI) between two genomic datasets of proteins).
*** dDDH: digital DNA-DNA hybridisation.

53.43 % with a relatively large genome size of 6.36 Mb. In comparison, *Ktedonobacter racemifer* SOSP1-21^T (RS GCF 000178855.1) has a G + C content of 53.79 % and genome size of 13.66 Mb, whereas *Thermogemmatispora carboxidivorans* PM5^T (RS GCF 000702505.1) has a G + C content of 60.88 % and genome size of 5.61 Mb. Genomic analyses indicate that the bacterial MAG is likely Gram-positive due to the absence of multiple genes involved in lipopolysaccharide synthesis and the presence of genes responsible for the formation of pentaglycine bridge. These are consistent with the cell structure of the members of the genus *Ktedonobacter*.

On the other hand, the archaeal MAG_C2–3 has a low G + C content of 35.54 % with a genome size of 3.02 Mb (Table 1). In comparison, *Nitrososphaera viennensis* EN76^T (RS GCF 000698785.1) has a G + C content of 52.72 % and a genome size of 2.52 Mb, *Ca. Nitrososphaera gargensis* Ga9.2 (RS GCF 000303155.1) has a G + C content of 48.35 % and a genome size of 2.83 Mb, and *Ca.* Nitrososphaera evergladensis SR1 (RS GCF 000730285.1) has a G + C content of 50.14 % and a genome size of 2.95 Mb.

Table 3
Functional characterisation of the MAGs based on their ability to utilize trace-gases. Numbers in parenthesis indicate NCBI accession number.

Gene (trace-gas)	MAG_1957-2.1 (Llaima)	MAG_C2–3 (Kilauea)
CoxL (carbon monoxide)	Yes 88.02 % *Ktedonobacter racemifer* SOSP1-21^T DSM_44,963 (WP_007909428.1)	No
PmoA (methane)	No	No
AmoA (ammonia)	No	Yes 100 % Ammonia monooxygenase subunit A, partial (ARD89532.1)

Trace gas metabolism

MAG_1957–2.1 contain the form I *coxL* gene, which encodes the large subunit of carbon monoxide dehydrogenase (CODH), with 'AYXCSFR' as the amino acid sequence motif at the active site (Table 3) (King and Weber, 2007). It also contains the canonical form I CODH encoding *cox* gene clusters, *coxMSL*. Phylogenetic analyses with the form I CoxL protein showed that it is related to the CoxL protein of *Ktedonobacter racimifer*^{T} (DSM 44963) (Fig. 3). Form I *coxL* is associated with the oxidation of carbon monoxide (CO) at both high concentrations and sub-atmospheric levels (King and Weber, 2007). These suggest that soil microbes from this group might oxidise CO, which is prevalent in recent volcanic environments (Hernández et al., 2020a). On the other hand, the archaeal MAG (MAG_C2-3) contains the *amoA* gene, which encodes for ammonia monooxygenase subunit A (AmoA) (Table 3), as well as ammonia monooxygenase subunit B. These subunits catalyse the oxidation of ammonia, the rate-limiting step in the nitrification process. The phylogenetic analyses showed that AmoA is related to the members of *Nitrosphaera* (Fig. 4). Ammonia oxidisers are widely distributed in *Bacteria* and *Archaea* across different ecosystems. Additionally, ammonia oxidising archaea (AOA) are also the pioneer organisms in volcanic soils (Hernández et al., 2014), indicating that they might also facilitate soil development. MAG_C2–3 does not encode for *coxL*, therefore it may not be involved in the oxidation of CO, suggesting a limited role in carbon cycle in this environment (Table 3).

The names of the MAGs were registered with SeqCode (Hedlund et al., 2022; Whitman et al., 2022). The SeqCode approach reflects advancements in genomics and microbiology, where the ability to sequence and analyse genomes has become a fundamental tool for studying and classifying microorganisms. Using genome sequences as nomenclatural types in the naming process provides a more accurate and detailed basis for classification compared to traditional methods (Protologue table).

Description of MAG_1957-2.1

Description of Paraktedonobacter gen. nov.

Paraktedonobacter (Pa.ra.kte.do.no.bac'ter. Gr. prep. *para*, alongside of, resembling; N.L. masc. n. *Ktedonobacter* a bacterial genus name; N.L. masc. n. *Paraktedonobacter*, besided *Ktedonobacter*.

Description of Paraktedonobacter carboxidivorans sp. nov.

Paraktedonobacter carboxidivorans (car.bo.xi.di.vo'rans. N.L. neut. adj. *carboxidum,* carbon monoxide; L. press. part. *vorans* devouring; N.L. part. adj. *carboxidivorans*, carbon monoxide-devouring).

This MAG was retrieved from Llaima Volcano, Chile. The genome size is 6.36 Mb, and the G + C content is 53.43 mol%. MAG_1957-2.1 contains the form I *cox* gene cluster, arranged in a *coxMSL* configuration, which is linked to the oxidation of carbon monoxide.

The type genome of the species has been deposited at NCBI under the accession number SRX9386060.

Classification: *Bacteria*; *Chloroflexota*; *Ktedonobacteria*; *Ktedonobacterales*; *Ktedonobacteraceae*; *Paraktedonobacter; Paraktedonobacter carboxidivorans*.

Description of MAG_C2-3

Nitrososphaera maunauluensis (mau.na.u.lu.en'sis N.L. masc./fem. adj. *maunauluensis* referring to Mauna Ulu, the place where the MAG was retrieved in Kilauea). This MAG was retrieved from Kilauea Volcano, USA. The genome size is 3.02 Mb, and the G + C content is 35.54 mol%. MAG_C2–3 encodes the ammonia monooxygenase gene subunit A (*amoA*) and subunit B (*amoB*), which are involved in catalysing the oxidation of ammonia.

The type genome of the species has been deposited at NCBI under the accession number SRX24172872.

Fig. 3. Phylogenetic tree of form I carbon monoxide dehydrogenase large subunit (CoxL) of the bacterial MAG retrieved from volcanic deposits after the 1957 eruption of Llaima against reference sequences. Bootstrap values (>50) are shown at the nodes. The scale bar represents 5 % sequence divergence.

Classification: *Archaea*; *Nitrososphaerota*; *Nitrososphaeria*; *Nitrososphaerales*; *Nitrososphaeraceae*; *Nitrososphaera; Nitrososphaera maunauluensis.*

Conclusion

In this work, we described one bacterial and one archaeal MAG, retrieved from two different volcanic ecosystems. The bacterial MAG contains genes encoding carbon monoxide dehydrogenase, responsible

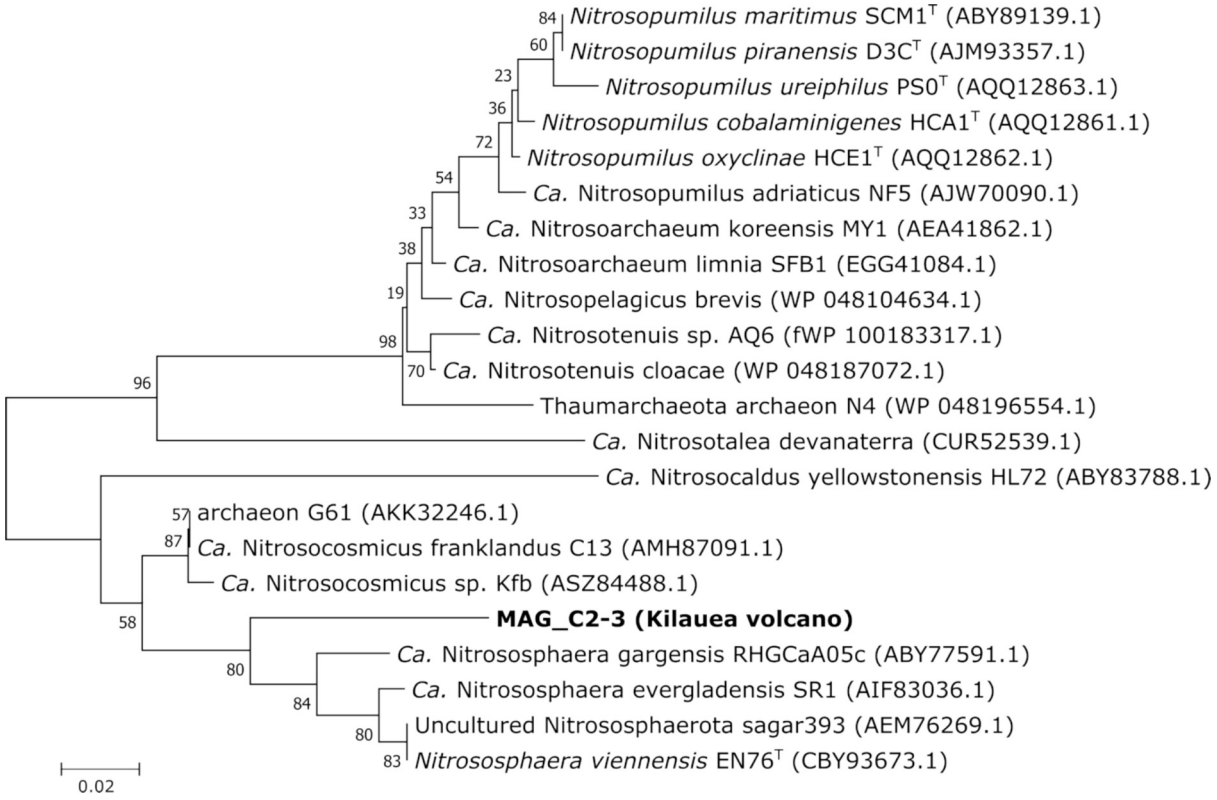

Fig. 4. Phylogenetic tree of ammonia monooxygenase subunit A (AmoA) of the archaeal MAG_C2–3 retrieved from Kilauea against reference sequences. Bootstrap values (>50) are shown at the nodes. The scale bar represents 2 % sequence divergence.

for CO oxidation, whereas the archaeal MAG contains genes encoding the ammonia monooxygenase, responsible for ammonia oxidation. These findings highlight the potential of microbes in volcanic ecosystems to metabolise trace gases and facilitate soil formation and development. Following SeqCode rules, we propose new names for these MAGs: the bacterial MAG is named *Paraktedonobacter carboxidivorans sp.* nov. and the archaeal MAG is named *Nitrososphaera maunauluensis sp.* nov.

CRediT authorship contribution statement

Shamik Roy: Writing – review & editing, Writing – original draft, Visualisation, Validation, Methodology, Investigation, Formal analysis, Data curation. **Gary M. King:** Writing – review & editing, Writing – original draft, Validation, Conceptualisation. **Marcela Hernández:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Acknowledgments

The research presented in this paper was carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service (RSCSS) at the University of East Anglia. A special thanks to Amjad Sayed in RSCSS for technical support. We thank Dr. Tomeu Viver for his help with Enveomics. We thank SeqCode curators, Dr. Maria Chuvochina and Dr. Luis-Miguel Rodríguez Rojas, for their assistance with the etymology. The authors are grateful to the valuable comments received by the editors and two anonymous reviewers. Shamik Roy was supported by a NERC Discipline Hopping (DH) for Discovery Science grant awarded to Marcela Hernández (NE/X018180/1) and a postdoctoral fellowship from European Union (ERC, BIOCOMP, GA 101075426). Marcela Hernández was supported by a Royal Society Dorothy Hodgkin Research Fellowship (DHF\R1\211076) and Royal Society Research Fellows Enhanced grants (RF\ERE\210050 and RF\ERE\231066).

Protologues of the new species description

	MAG_1957-2.1 (Llaima)	MAG_C2-3 (Kilauea)
Genus name	*Paraktedonobacter*	*Nitrososphaera*
Genus status	gen. nov.	
Genus etymology	Pa.ra.kte.do.no.bac'ter. Gr. prep. *para*, alongside of, resembling; N.L. masc. n. *Ktedonobacter* a bacterial genus name; N.L. masc. n. *Paraktedonobacter*, besided *Ktedonobacter*	
Specific epithet	*carboxidivorans*	*maunauluensis*
Species status	*sp.* nov.	*sp.* nov.
Species etymology	car.bo.xi.di.vo'rans. N.L. neut. n. carboxidum carbon monoxide; L. part. adj. vorans devouring; N.L. part. adj. carboxidivorans carbon monoxide-devouring	mau.na.u.lu.en'sis N.L. masc./fem. adj. *maunauluensis* referring to Mauna Ulu, the place where the metagenome-assembled genome was retrieved in Kilauea
Description of the new taxon and diagnostic traits	–	–
Country of origin	Chile	USA
Region of origin	Llaima Volcano, Araucanía region	Kilauea Volcano, Mauna Ulu site, Hawaii
Date of isolation	2019	2022
Source of isolation	Soil	Soil
Sampling date	15/02/2011	12/04/2022
Latitude	38.57953°S	19.3388382°N
Longitude	71.78722°W	155.2033492°W
Altitude (m)	741	828

(*continued on next page*)

(*continued*)

16S rRNA gene accession number	SRR32116977	SRR32116972
Genome accession number	GCA_041156125.1	GCA_041156085.1
Genome status	Draft – 96.21 % completeness	Draft – 97.57 % completeness
Information related to Nagoya protocol (NP)	Both Chile and the USA have not ratified the NP (Convention on Biological Diversity https://www.cbd.int/abs/nagoya-protocol/signatories, accessed on May 22nd 2024)	
Designation of the Type Strain or genomic assembly	–	–
Metagenomic raw data	PRJNA602601	PRJNA1096838
SeqCode registry URL	seqco.de/r:89dif5uc	seqco.de/r:89dif5uc

Data availability

Bacterial MAG_1957-2.1 has been deposited in the NCBI-SRA under the bioproject accession number PRJNA602601 (Hernández et al., 2020b). Archaeal MAG-C2-3 has been deposited in the NCBI-SRA under the bioproject accession number PRJNA1096838. The 16S rRNA gene sequences of these MAGs were deposited in the NCBI-SRA under SRA accession numbers SRR32116977 for MAG_1957-2.1 and SRR32116972 for MAG MAG-C2-3.

References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5), 455–477. https://doi.org/10.1089/cmb.2012.0021.

Bay, S.K., Dong, X., Bradley, J.A., Leung, P.M., Grinter, R., Jirapanjawat, T., Arndt, S.K., Cook, P.L.M., LaRowe, D.E., Nauer, P.A., Chiri, E., Greening, C., 2021. Trace gas oxidizers are widespread and active members of soil microbial communities. Nat. Microbiol. 6 (2), 246–256. https://doi.org/10.1038/s41564-020-00811-w.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics 38 (23), 5315–5316. https://doi.org/10.1093/bioinformatics/btac672.

Conrad, R., 1996. Soil microorganisms as controllers of atmospheric trace gases (H_2, CO, CH_4, OCS, N_2O, and NO). Microbiol. Rev. 60 (4), 609–640. https://doi.org/10.1128/MR.60.4.609-640.1996.

Fenchel, T., King, G.M., Blackburn, T.H., 2012. Bacterial biogeochemistry. In: The Ecophysiology of Mineral Cycling, 3rd ed. Elsevier.

Gomez-Alvarez, V., King, G.M., Nüsslein, K., 2007. Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. FEMS Microbiol. Ecol. 60 (1), 60–73. https://doi.org/10.1111/J.1574-6941.2006.00253.X.

Hansen, J., Sato, M., Kharecha, P., Russell, G., Lea, D.W., Siddall, M., 2007. Climate change and trace gases. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 365 (1856), 1925–1954. https://doi.org/10.1098/rsta.2007.2052.

Hedlund, B.P., Chuvochina, M., Hugenholtz, P., Konstantinidis, K.T., Murray, A.E., Palmer, M., Parks, D.H., Probst, A.J., Reysenbach, A.-L., Rodriguez-R, L.M., Rossello-Mora, R., Sutcliffe, I.C., Venter, S.N., Whitman, W.B., 2022. SeqCode: a nomenclatural code for prokaryotes described from sequence data. Nat. Microbiol. 7 (10), 1702–1708. https://doi.org/10.1038/s41564-022-01214-9.

Hernández, M., Dumont, M.G., Calabi, M., Basualto, D., Conrad, R., 2014. Ammonia oxidizers are pioneer microorganisms in the colonization of new acidic volcanic soils from South of Chile. Environ. Microbiol. Rep. 6 (1), 70–79. https://doi.org/10.1111/1758-2229.12109.

Hernández, M., Calabi, M., Conrad, R., Dumont, M.G., 2020a. Analysis of the microbial communities in soils of different ages following volcanic eruptions. Pedosphere 30 (1), 126–134. https://doi.org/10.1016/S1002-0160(19)60823-4.

Hernández, M., Vera-Gargallo, B., Calabi-Floody, M., King, G.M., Conrad, R., Tebbe, C.C., 2020b. Reconstructing genomes of carbon monoxide oxidisers in volcanic deposits including members of the class Ktedonobacteria. Microorganisms 8 (12), 1880. https://doi.org/10.3390/microorganisms8121880.

Kim, J.-S., Kim, D.-S., Lee, K.C., Lee, J.-S., King, G.M., Kang, S., 2018. Microbial community structure and functional potential of lava-formed Gotjawal soils in Jeju, Korea. PLoS ONE 13 (10), e0204761. https://doi.org/10.1371/journal.pone.0204761.

King, G.M., 2003. Contributions of atmospheric CO and hydrogen uptake to microbial dynamics on recent Hawaiian volcanic deposits. Appl. Environ. Microbiol. 69 (7), 4067–4075. https://doi.org/10.1128/AEM.69.7.4067-4075.2003.

King, G.M., 2015. Carbon monoxide as a metabolic energy source for extremely halophilic microbes: implications for microbial activity in Mars regolith. Proc. Natl. Acad. Sci. U. S. A. 112 (14), 4465–4470. https://doi.org/10.1073/pnas.1424989112.

King, G.M., Weber, C.F., 2007. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. Nat. Rev. Microbiol. 5 (2), 107–118. https://doi.org/10.1038/nrmicro1595.

King, G.M., Weber, C.F., 2008. Interactions between bacterial carbon monoxide and hydrogen consumption and plant development on recent volcanic deposits. ISME J. 2 (2), 195–203. https://doi.org/10.1038/ismej.2007.101.

Konstantinidis, K.T., Ramette, A., Tiedje, J.M., 2006. The bacterial species definition in the genomic era. Philos. Trans. R. Soc., B 361 (1475), 1929–1940. https://doi.org/10.1098/rstb.2006.1920.

Konstantinidis, K.T., Rosselló-Móra, R., Amann, R., 2017. Uncultivated microbes in need of their own taxonomy. ISME J. 11 (11), 2399–2406. https://doi.org/10.1038/ismej.2017.113.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35 (21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

Louca, S., Doebeli, M., 2018. Efficient comparative phylogenetics on large trees. Bioinformatics 34 (6), 1053–1055. https://doi.org/10.1093/BIOINFORMATICS/BTX701.

Médigue, C., Calteau, A., Cruveiller, S., Gachet, M., Gautreau, G., Josso, A., Lajus, A., Langlois, J., Pereira, H., Planel, R., Roche, D., Rollin, J., Rouy, Z., Vallenet, D., 2019. MicroScope—an integrated resource for community expertise of gene functions and comparative analysis of microbial genomic and metabolic data. Brief. Bioinform. 20 (4), 1071–1084. https://doi.org/10.1093/BIB/BBX113.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14 (1), 60. https://doi.org/10.1186/1471-2105-14-60.

Meng, J., Xu, J., Qin, D., He, Y., Xiao, X., Wang, F., 2014. Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. ISME J. 8 (3), 650–659. https://doi.org/10.1038/ismej.2013.174.

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.-M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T.B.K., Nielsen, T., Kirton, E., Faria, J.P., Edirisinghe, J.N., Henry, C.S., Eloe-Fadrosh, E.A., 2021. A genomic catalog of Earth's microbiomes. Nat. Biotechnol. 39 (4), 499–509. https://doi.org/10.1038/s41587-020-0718-6.

Ni, G., Lappan, R., Hernández, M., Santini, T., Tomkins, A.G., Greening, C., 2023. Functional basis of primary succession: traits of the pioneer microbes. Environ. Microbiol. 25 (1), 171–176. https://doi.org/10.1111/1462-2920.16266.

Offre, P., Spang, A., Schleper, C., 2013. Archaea in biogeochemical cycles. Annu. Rev. Microbiol. 67 (1), 437–457. https://doi.org/10.1146/annurev-micro-092412-155614.

Ohta, H., Ogiwara, K., Murakami, E., Takahashi, H., Sekiguchi, M., Koshida, K., Someya, T., Morishima, W., Rondal, J.D., Concepcion, R.N., Yoshida, M., Watanabe, M., 2003. Quinone profiling of bacterial populations developed in the surface layer of volcanic mudflow deposits from Mt. Pinatubo (the Philippines). Soil Biol. Biochem. 35 (8), 1155–1158. https://doi.org/10.1016/S0038-0717(03)00160-3.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25 (7), 1043–1055. https://doi.org/10.1101/GR.186072.114.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., Hugenholtz, P., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36 (10), 996–1004. https://doi.org/10.1038/nbt.4229.

Richter, M., Rosselló-Móra, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. 106 (45), 19126–19131. https://doi.org/10.1073/pnas.0906412106.

Rinke, C., Chuvochina, M., Mussig, A.J., Chaumeil, P.-A., Davín, A.A., Waite, D.W., Whitman, W.B., Parks, D.H., Hugenholtz, P., 2021. A standardized archaeal taxonomy for the Genome Taxonomy Database. Nat. Microbiol. 6 (7), 946–959. https://doi.org/10.1038/s41564-021-00918-8.

Rodriguez-R, L.M., Konstantinidis, K.T., 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.1900v1.

Thauer, R.K., Kaster, A.-K., Seedorf, H., Buckel, W., Hedderich, R., 2008. Methanogenic archaea: ecologically relevant differences in energy conservation. Nat. Rev. Microbiol. 6 (8), 579–591. https://doi.org/10.1038/nrmicro1931.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428 (6978), 37–43. https://doi.org/10.1038/nature02340.

Uritskiy, G.V., Diruggiero, J., Taylor, J., 2018. MetaWRAP - A flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6 (1), 1–13. https://doi.org/10.1186/s40168-018-0541-1.

Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S., Gautreau, G., Lajus, A., Langlois, J., Planel, R., Roche, D., Rollin, J., Rouy, Z., Sabatet, V., Médigue, C., 2020. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. Nucleic Acids Res. 48 (D1), D579–D589. https://doi.org/10.1093/NAR/GKZ926.

Vigneron, A., Cruaud, P., Lovejoy, C., Vincent, W.F., 2022. Genomic evidence of functional diversity in DPANN archaea, from oxic species to anoxic vampiristic consortia. ISME Commun. 2 (1). https://doi.org/10.1038/S43705-022-00088-6.

Whitman, W.B., Chuvochina, M., Hedlund, B.P., Hugenholtz, P., Konstantinidis, K.T., Murray, A.E., Palmer, M., Parks, D.H., Probst, A.J., Reysenbach, A.-L., Rodriguez-R, L.M., Rossello-Mora, R., Sutcliffe, I., Venter, S.N., 2022. Development of the SeqCode: A proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type. Syst. Appl. Microbiol. 45 (5), 126305. https://doi.org/10.1016/j.syapm.2022.126305.

Yoon, S.-H., Ha, S., Lim, J., Kwon, S., Chun, J., 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110 (10), 1281–1286. https://doi.org/10.1007/s10482-017-0844-4.