

Comparison of Actinobacteria communities from human-impacted and pristine karst caves

Buresova Andrea^{1,2,3}, Kopecky Jan³, Sagova-Mareckova Marketa³, Alonso Lise¹, Vautrin Florian¹, Moëgne-Loccoz Yvan¹, Rodriguez-Nava Veronica¹

¹Univ Lyon, Université Claude Bernard Lyon 1, CNRS, INRAE, VetAgro Sup, UMR5557 Ecologie Microbienne, F-69622 Villeurbanne, France,

²Department of Ecology, Faculty of Science, Charles University in Prague, Vinicna 7, CZ-12800, Prague 2, Czech Republic,

³Laboratory for Epidemiology and Ecology of Microorganisms, Crop Research Institute, Drnovska 507, CZ-16106 Prague 6, Ruzyně, Czech Republic

Corresponding author. V. Rodriguez-Nava, UMR Ecologie Microbienne, CNRS5557, INRAE1418, UCBL, Université de Lyon, VetAgro Sup, Institut des Sciences Pharmaceutiques et Biologiques de Lyon, France.

E-mail address: veronica.rodriguez-nava@univ-lyon1.fr

Abstract

Actinobacteria are important cave inhabitants, but knowledge of how anthropization and anthropization-related visual marks affect this community on cave walls is lacking. We compared *Actinobacteria* communities among four French limestone caves (Mouflon, Reille, Rouffignac, Lascaux) ranging from pristine to anthropized, and within Lascaux Cave between marked (wall visual marks) and unmarked areas in different rooms (Sas-1, Passage, Apse, Diaclase). In addition to the 16S rRNA gene marker, 441 bp fragments of the *hsp65* gene were used and an *hsp65*-related taxonomic database was constructed for identification of *Actinobacteria* to the species level by Illumina-MiSeq analysis. The *hsp65* marker revealed higher resolution for species and higher richness (99% OTU cutoff) versus 16S rRNA gene; however, more taxa were identified at higher taxonomic ranks. *Actinobacteria* communities varied between Mouflon and Reille caves (both pristine), and Rouffignac and Lascaux (both anthropized). Rouffignac displayed high diversity of *Nocardia*, suggesting human inputs, and Lascaux exhibited high *Mycobacterium* relative abundance, whereas *Gaiellales* were typical in pristine caves and the Diaclase (least affected area of Lascaux Cave). Within Lascaux, *Pseudonocardiaceae* dominated on unmarked walls and *Streptomyetaceae* (especially *Streptomyces mirabilis*) on marked walls, indicating a possible role in mark formation. A new taxonomic database (<https://zenodo.org/record/5576074>) was developed. Although not all *Actinobacteria* species were represented, the use of the *hsp65* marker enabled species-level variations of the *Actinobacteria* community to be documented based on the extent of anthropogenic pressure. This approach proved effective when comparing different limestone caves or specific conditions within one cave.

KEYWORDS:

16S rRNA gene sequencing, *hsp65* sequencing, metabarcoding, *Actinobacteria*, cave anthropization, Paleolithic cave

1. INTRODUCTION

Limestone caves are isolated habitats that are extremely limited in organic carbon (Barton et al., 2007). They are associated with the development of microbial strategies that enable adaptation to oligotrophic and high calcium conditions. In addition, caves are considered to be extreme environments, since they are oligotrophic, due to the complete darkness, low and constant temperatures, and high humidity (Barton & Jurado, 2007; De Mandal et al., 2017; Ortiz et al., 2014). Some of these caves contain Paleolithic artwork and, consequently, have been frequently visited by tourists for a long time (Alonso et al., 2019; Bontemps et al., 2021; Schabereiter-Gurtner et al., 2002). Thus, human-mediated dispersion of endogenous as well as exogenous nutrients and microorganisms inside the caves, together with human metabolism may induce changes in internal microclimatic and microbiological conditions (Hoyos et al., 1998; Mulec, 2014). Microorganisms can be somewhat resilient to the changes (Johnston et al., 2012; Tomczyk-Żak & Zielenkiewicz, 2016). However, community shifts may also occur, which deserves attention if those shifts favor microorganisms with biodeterioration ability (Bastian et al., 2010; Bontemps et al., 2021; Fernandez-Cortes et al., 2011; Sánchez-Moral et al., 1999) or increase the proportion of species harboring strains with pathogenic potential (Bercea et al., 2019; Neral et al., 2015; Rajput et al., 2012). In such cases, these shifts could represent a threat to cave conservation and the health of visitors.

Actinobacteria represent important cave dwellers, and many potentially pathogenic (Jurado et al., 2010), biotechnologically relevant (Bhullar et al., 2012; Hamedi et al., 2019; Syiemiong & Jha, 2019), and novel (Fang et al., 2017; Jurado et al., 2008; Jurado, Kroppenstedt, et al., 2009; Rangseekaew & Pathom-Aree, 2019) species have been identified in these environments. *Actinobacteria* can dominate in both anthropized (Gonzalez-Pimentel et al., 2018) and pristine caves (De Mandal et al., 2017); however, some actinobacterial taxa are specific to one cave type. For example, the autochthonous presence of the *Pseudonocardiaceae* family has been reported as part of the core microbiome of pristine caves (Riquelme et al., 2015, (Lavoie et al., 2017; Porca et al., 2012). However, some species, such as *Pseudonocardia*

hispaniensis, have been detected in the moonmilk of a cave associated with urban contamination (Miller et al., 2018), and its presence may be due to its ability to degrade exogenous organic compounds (Lavoie et al., 2017; Porca et al., 2012).

This is also true for the *Streptomycetaceae* family, which has also been identified among pristine cave inhabitants (Maciejewska et al., 2017). Furthermore, *Pseudonocardiaceae* and *Streptomycetaceae* seem to be prominent in pigment-forming communities on the walls of certain show caves (Cuezva et al., 2012; Porca et al., 2012). In contrast, *Euzebyales* (Gonzalez-Pimentel et al., 2018), *Nocardiaceae* (De Mandal et al., 2017), and *Gaiellales* (Zhu et al., 2019) represent oligotrophic taxa that dominate in pristine caves, most likely because of their ability to degrade complex compounds present in the cave, or to fix CO₂.

Identification of *Actinobacteria* taxa in these environments has been based on culture-dependent approaches and 16S rRNA gene sequencing (De Mandal et al., 2017; Yasir, 2018). Culture-dependent approaches limit community coverage and should, therefore, only be used as a supportive determinant to study their specialized metabolisms (Hamedi et al., 2019; Long et al., 2019). The 16S rRNA gene is the most frequently used marker for bacterial systematics; however, it is unable to resolve individual sequences to the species level (Fox et al., 1992; Hirsch et al., 2010). In addition, identical 16S rRNA gene sequences can be found in different taxa or different 16S rRNA gene sequences within a single species (Větrovský & Baldrian, 2013). However, other (protein-coding) genes with high sequence polymorphisms used for multilocus sequence analysis (MLSA) (*i.e.*, *rpoB*, *gyrB*, *hsp65*) may help to distinguish closely related species from environmental DNA. Thus, they may present greater discriminative power at the species level compared with the 16S rRNA gene (Gao & Gupta, 2012; Takeda et al., 2010; Vos et al., 2012). Thus, using a metabarcoding approach, Aigle et al (2021) utilized another gene (*tpm*) to evaluate the biodiversity of the bacterial community associated with the degradation of pollutants in anthropized environments at a species level.

In the case of *Actinobacteria*, *hsp65* encodes a molecular chaperone (a 65-kDa heat-shock protein), which, when amplified by specific primers (Telenti et al., 1993), has been used as a molecular marker to identify a large number of isolates (Rodríguez-Nava et al., 2007; Rodríguez-Nava et al., 2006). A single

copy of this gene is present in *Actinobacteria* genomes, unlike the 16S rRNA gene marker (Colaco & MacDougall, 2014). Additionally, the length of the amplified *hsp65* fragment (441 bp) is sufficient for amplicon sequencing. While this marker has been used to assess some *Actinobacteria* groups from environmental samples, it has mostly been used with regards to particular groups, such as *Nocardia* (Vautrin et al., 2021) or *Mycobacterium* (van der Wielen et al., 2013). Therefore, considering these features, the *hsp65* gene may be a good candidate for detailed assessments of the whole *Actinobacteria* community in extreme environments, such as caves. To make this possible, a reference database encompassing the *Actinobacteria* taxonomic group is required, which should be publicly available for taxonomic identification.

Since the 1960s, the Lascaux Cave has undergone several microbial outgrowth events until recently and has been subjected to (often unsuccessful) chemical treatments, which have resulted in the formation of various visual marks (including stains on walls) that have threatened the Paleolithic artwork (Bastian et al., Bontemps et al., 2021; 2010; Martin-Sanchez et al., 2015). Lascaux was closed to the public in 1963 to ensure cave conservation; despite this, microbial imbalance persisted and surface alterations kept forming, leading to visual marks on wall surfaces (Bontemps et al., 2021). Although the few microbiological studies conducted in this cave have suggested that *Actinobacteria* is a prevalent bacterial group (Alonso et al., 2018, 2019; Martin-Sanchez et al., 2012), a more complete investigation of the *Actinobacteria* community is lacking, especially at the species level. *Actinobacteria* possess broad metabolic variability, and key functional traits may be related to precise species within the *Actinobacteria*; therefore, we hypothesized that such an assessment would better discriminate *Actinobacteria* communities according to ecological conditions.

The objective of this study was to assess the usefulness of *hsp65* to complement 16S rRNA gene sequence data using a metabarcoding approach for a deeper, species-level appraisal of the *Actinobacteria* community in caves, and to assess variation according to cave condition and human disturbance (Mammola, 2019). We compared communities belonging to the phylum *Actinobacteria* between pristine limestone caves (Reille and Mouflon Caves) and anthropized caves (Lascaux and Rouffignac Caves), and within

Lascaux Cave between marked (wall visual marks) and unmarked areas in different rooms (*i.e.*, Sas-1, Passage, Apse, and Diaclase). The second objective was to create an *hsp65*-based *Actinobacteria* database that is accessible to the scientific community.

2. MATERIALS AND METHODS

2.1. Study site description and sample collection

Samples were collected in May and June 2016 from the walls of four limestone caves located in the Dordogne department of southwestern France: Lascaux (45°03'13" N 1°10'12"E, 235 m length); Rouffignac (45°00'31"N 0°59'16"E, 8,000 m length); Reille (45°07'09.5"N 1°07'11.2"E, 2,000 m length); and Mouflon (44°55'N 1°5'E, 280 m length) (Figure 1A).

Lascaux and Rouffignac caves contain Paleolithic wall paintings and engravings and are listed as UNESCO World Heritage Sites. Both caves are anthropized; Rouffignac has been open to tourist visits since 1959 (with up to 500 visitors per day) (Alonso et al., 2019), whereas Lascaux closed after 15 years of tourism (with up to 2000 visitors per day). Lascaux Cave was closed because various visual marks of microbial origin had developed on cave walls. Extensive chemical treatments have been applied to protect the wall paintings from microbial deterioration, which probably triggered further alterations (Bastian et al., 2010; De La Rosa et al., 2017; Martin-Sanchez et al., 2012). Mouflon and Reille Caves remain in a pristine state (Alonso et al., 2019). The entrance of Mouflon Cave was locked shortly after discovery, whereas Reille is occasionally explored by seasoned speleologists.

In Lascaux Cave, samples were taken from the walls of four distinct rooms representing different cave environments that are located progressively farther from the entrance in the following order: Sas-1, Passage, Apse, and Diaclase (Figure 1B). Sas-1 is a man-made, calcareous airlock entrance, and is isolated from the cave interior and exterior by doors. The Passage is a central alley connecting different rooms of Lascaux Cave. In the Passage, samples were taken from two mineral substrates: clay deposits (Passage banks, later referred to as Passage B) and near-vertical limestone walls (Passage inclined planes, later referred to as Passage IP). In the Apse, samples were collected from limestone inclined planes. The Diaclase

is a distinct compartment located below the main part of the cave, which has been least affected by humans. It is separated from the Apse by a trap door, has received very limited biocide treatment, and was never open to tourists. Samples originated from limestone walls located immediately below the Apse, in an area of the Diaclase termed the Shaft of the Dead Man.

Within each room, samples were collected from visual marks (when occurring) and unstained surfaces, as previously described (Alonso et al., 2018). Approximately 50 mg of wall material samples (three to six replicates) were collected using sterile scalpels and placed in liquid nitrogen for transportation to the laboratory, where they were stored at -80°C until DNA extraction. Six samples from Rouffignac, six from Mouflon, five from Reille, and forty-one from Lascaux (five from Sas-1, six from Passage B, six from Passage IP, eighteen from Apse, and six from Diaclase) were used for subsequent metabarcoding analyses (Table A1). Sampling was performed in accordance with the cave rules and regulations.

2.2. DNA extraction and amplicon sequencing

DNA was extracted using the FastDNATM SPIN Kit for Soil (MP BiomedicalsTM, Illkirch, France) following the manufacturer's instructions, and adapted to the low sample size, as previously described (Alonso et al., 2018). The elution step was achieved using two volumes of 50 μL elution buffer per sample.

The primers 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACTACHVGGGTATCTAATCC-3'), which target the V3-V4 region (428 bp fragments), were used for bacterial 16S rRNA gene amplification (Herlemann et al., 2011). The PCR cycling protocol consisted of initial denaturation at 95°C for 5 min; followed by 25 cycles of denaturation (95°C for 40 s), annealing (55°C for 30 s), and extension (72°C for 1 min); and a final extension step at 72°C for 7 min. Amplification and sequencing were performed by the Fasteris Company (Geneva, Switzerland) using the Illumina MiSeq Reagent Kit with V3 chemistry (600 cycles) in the paired-end mode, resulting in 2×300 bp sequence reads. We used a subset of the sequences obtained by Alonso et al. (Alonso et al., 2018).

The 441 bp fragments of the 65 kDa heat shock protein gene (*hsp65*) were amplified by PCR using the specific primers TB11 (5'-ACCAACGATGGTGTGTCCAT-3') and TB12 (5'-

CTTGTCGAACCGCATACCCT-3') from Telenti et al. (Telenti et al., 1993). Amplifications were performed in a reaction mixture of 50 μ L in packaged PCR tubes (PuReTaq Ready-To-Go™ PCR Beads; GE Healthcare; 2.5 U of Taq polymerase puRe Taq, 10 mM Tris-HCl [pH 9], 50 mM KCl, 1.5 mM MgCl₂, and 200 μ M of each deoxynucleoside triphosphate) with primers (0.2–0.4 μ M, depending on the template concentration), and DNA (1–40 ng). The PCR cycling protocol consisted of initial denaturation at 94°C for 5 min, then 35 cycles of denaturation (94°C for 60 s), annealing (55°C for 60 s), and elongation (72°C for 60 s) (Rodríguez-Nava et al., 2006). Sequencing, purification, quality control were performed by the Biofidal Company (Biofidal, Vaulx-en-Velin, France; <http://www.biofidal.com>) using the Illumina MiSeq Flow Cell V3 in paired-end mode, resulting in 2 \times 300 bp sequence reads. Overall, 50 and 45 samples were successfully sequenced for 16S rRNA gene and *hsp65*, respectively, while sequences for both markers were obtained for only 37 samples (Table A1).

2.3. Processing and analysis of sequence data

Paired sequence reads obtained from sequencing were merged using FLASH (Fast Length Adjustment of Short reads) (Magoč & Salzberg, 2011) with a maximum of 25% mismatches in the overlapping region. The sequences were then filtered and aligned using reference alignments of *hsp65* gene sequences (this work) or 16S rRNA gene sequences from the Silva database V132 (Quast et al., 2012). Chimeric sequences were removed using the integrated Vsearch tool (Rognes et al., 2016) according to the MiSeq standard operating procedure (MiSeq SOP, February 2018) (Kozich et al., 2013) in Mothur v1.39.5 software (Schloss et al., 2009). Sequence libraries were taxonomically assigned in Mothur using the *hsp65* reference database for *hsp65* sequences and the recreated SEED database subset of the Silva Small Subunit rRNA Database, release 132 (Yilmaz et al., 2014), adapted for use in Mothur (https://mothur.org/w/images/a/a4/Sylva.seed_v132.tgz), as the reference database for 16S rRNA gene sequences.

The sequences of plastids and mitochondria, and those not classified in the domain Bacteria (from the 16S rRNA gene sequence library), as well as sequences identified as homologs for *hsp65* (*groEL*, in the

database renamed as GROESL; see the *hsp65* reference database design section; from the *hsp65* sequence library), were discarded. The sequence library was clustered into operational taxonomic units (OTUs) using the Uparse pipeline in Usearch v10.0.240 software (OTUs at a 97% cutoff) (Edgar, 2013) and Mothur (OTUs at a 99% cutoff). Two cutoffs were applied, as the 97% cutoff is used as a standard threshold for estimating bacterial diversity for the 16S rRNA gene. Therefore, it might underestimate total diversity, especially that of the *hsp65* marker, which has higher variability than the 16S rRNA gene marker (see Figure 2 in the Results section). The tighter 99% cutoff was used for a more reliable estimation of *Actinobacteria* diversity at the lowest taxonomic ranks, particularly for the *hsp65* marker (Chen et al., 2013). The OTU table was further processed using tools implemented in Mothur. For 16S rRNA genes, only sequences corresponding to *Actinobacteria* were used.

To compare *Actinobacteria* among the caves, 25 samples for *hsp65* and 24 samples for 16S rRNA genes were used (Table A1). Apse and Diaclase (using unmarked areas) were selected as representative samples for Lascaux Cave, except for the analysis of the core microbiome (i.e., common OTUs present in all caves), for which all samples from Lascaux Cave samples were included. To examine differences among locations in Lascaux Cave, 33 samples for *hsp65* (16 from marked areas and 17 from unmarked areas) and 38 samples for the 16S rRNA gene (20 from marked areas and 18 samples from unmarked areas) were used (Table A1). For the 16S rRNA gene marker, samples from the Sas-1 unmarked area were missing for the analysis, because only a low number of replicates were successfully sequenced. To obtain the highest number of replicates per sample, all samples were included in the analysis, even when they were not common to both markers. A total of 37 common samples for both markers were used to compare these markers, i.e., rarefaction analysis, comparison of diversity indices, and several taxa recovered at different taxonomy ranks.

Rarefaction analysis was performed to analyze the richness of the *Actinobacteria* OTUs based on two cutoffs (i.e., 97 and 99%) as a function of sample number. The alpha diversity indices (richness: Chao 1, evenness: Simpson Evenness, and diversity: Inverse Simpson) were calculated in Mothur (97 and 99% OTU cutoff). Significant differences among samples were calculated by ANOVA with Tukey's post hoc

tests ($P < 0.05$; for caves or locations in Lascaux Cave), *Student t*-tests, and F tests ($P < 0.05$; for marked/unmarked areas) in Past 4.02 software (Hammer et al., 2001).

Bray-Curtis distance matrices were calculated to describe differences in bacterial community composition among individual samples. *Actinobacteria* communities in the caves, in different locations of Lascaux Cave, and between marked and unmarked areas within each location were compared by nonmetric multidimensional scaling (NMDS) according to both gene markers (97% OTU cutoff). Analysis of molecular variance (AMOVA) and homogeneity of molecular variance (HOMOVA) were calculated in Mothur with 100,000 iterations for the same factors (97% OTU cutoff).

Pairwise comparisons were made with the Metastats method using Fisher's exact tests (White et al., 2009), and the linear discriminant analysis (LDA) effect size (LEfSe) (Segata et al., 2011) was calculated to identify the *Actinobacteria* OTUs that differed significantly among the respective caves, locations, and marked/unmarked areas in Lascaux Cave (99% OTU cutoffs). Metastats was also used to compare relative abundances of *Streptomyces* sequences among the marked/unmarked areas (*hsp65* marker). Venn diagrams were created in Mothur to test the number of OTUs that encompassed the core microbiome for the different caves and rooms (*i.e.*, Sas-1, Passage, Apse, and Diaclase) in Lascaux Cave (99% OTU cutoff).

Taxonomic classification of OTUs was performed with Mothur using the *hsp65* and 16S rRNA gene SILVA V132 reference databases, and species-level identification with the *hsp65* marker was verified by BLASTN in the online databases of the National Center for Biotechnology Information (NCBI; U.S. National Library of Medicine, MD, USA). In parallel, a phylogenetic analysis based on *hsp65* was performed using retrieved sequences from caves belonging to most representative OTUs and those from the Genbank database after using the nucleotide BLAST (BLASTN) tool. Retrieved sequences from Genbank were those with the highest similarity with selected OTUs. Multiple alignments were generated using ClustalX (Thompson et al, 2003). MEGA6 software was used to compute the maximum likelihood of Neighbor-Joining phylogenetic trees from the observed amino acid sequence divergence (%) (Tamura et al 2013). The taxonomical composition of the core microbiome was constructed using the Krona tool

(Ondov et al., 2011). The number of taxa recovered with the common samples by *hsp65* and 16S rRNA gene analyses at different taxonomic ranks was calculated and displayed using Venn diagrams.

Co-occurrence networks were used to predict how marked/unmarked areas and different locations influenced the relative abundance of *Actinobacteria* in Lascaux Cave. Spearman correlation coefficients were computed in Mothur for Lascaux Cave OTUs and were categorized based on their significantly different proportions in the marked/unmarked areas, regardless of the location in Lascaux Cave (Metastats) or for the Lascaux Cave locations (Lefse) (99% OTU cutoff). Only significant ($P < 0.03$) correlations higher than 0.8/0.8 and lower than -0.35/-0.5 for the *hsp65*/16S rRNA gene markers were visualized in Gephi 0.9.2. (M. Bastian et al., 2009) by Fruchterman-Reingold spatialization (Fruchterman & Reingold, 1991). The average degree, eigenvector centrality, and modularity were computed, and the minimum number of links was filtered to five (using a degree range filter).

Figures were plotted using the vegan package (Oksanen et al., 2018) in the R computing environment (R Core Team, 2018) and Inkscape (v0.92; <http://www.inkscape.org>).

2.4. *In silico* analysis

Differences in the variability of the *hsp65* and 16S rRNA gene partial sequences between pairs of *Actinobacteria* strains at different taxonomical levels (*e.g.*, intraclass, intraorder, intrafamily, and intragenus) were determined using genomes retrieved from the Integrated Microbial Genomes and Microbiomes database (IMG; University of California, CA, USA, <https://img.jgi.doe.gov/>). For the analysis, sequences from the available genomes of 47 species from the order *Streptomycetales* (*Streptomyces*, *Kitakatospora*, Table A2), and 58 species from the order *Corynebacteriales* (family *Corynebacteriaceae*: *Corynebacterium*; family *Gordoniaceae*: *Gordonia*; family *Mycobacteriaceae*: *Mycobacterium*; family *Nocardiaceae*: *Nocardia*, *Rhodococcus*; Table A2) were used. Sequences were aligned and trimmed to the amplified length of both gene markers in BioEdit 7.2.5 software (Hall et al., 2011). Pairwise distances between sequences were calculated in Mothur v. 1.39.5 software (Schloss et al.,

2009). Plots were created and the coefficient of determination (R^2) was calculated in Microsoft Excel (2016).

2.5. *hsp65* reference database design

For taxonomic assignment using the *hsp65* marker, the reference database of the *Actinobacteria hsp65* gene was constructed similarly to that of the *rpoB* marker by (Ogier et al., 2019). Here, we built a new reference library based on the *hsp65* gene encompassing both newly obtained and public sequences from the most important *Actinobacteria* taxa. Part of this new reference library was constructed with *Nocardia* and *Gordonia* sequences from all type strains, from strains from clinical collections (supplied by the French Observatory for Nocardiosis, OFN <http://ofn.univ-lyon1.fr>), and from environmental strains (supplied by CRB-EML <http://eml-brc.org/>) sequenced by our means for which sequence quality was verified by double sense sequencing. For the *Mycobacterium* genus, *hsp65* sequences were available via the BIBI database (<https://umr5558-bibiserv.univ-lyon1.fr/lebibi/lebibi.cgi>), which is fairly robust as it contains all described *Mycobacterium* species. The rest of the sequences were collected from the online IMG and NCBI databases using sequences homologous to the 65 kDa heat shock protein (*hsp65*) of the reference strain *Nocardia asteroides* ATCC 14759 (Rodríguez-Nava et al., 2006) amplified by the TB11 and TB12 primers (Telenti et al., 1993). Paralogs of heat shock proteins, such as *groEL* genes from *Escherichia coli* (Colaco & MacDougall, 2014; Duchêne et al., 1994; Kumar et al., 2015), were identified using maximum-likelihood, FastTree 2.1 (Price et al., 2010), and were retained in the reference database (renamed GROESL sequences) as an outgroup that enabled amplified sequences belonging to this variant to be discarded. This database was named “ACTIhsp65_V1.0.0.fas” and is available at <https://doi.org/10.5281/zenodo.5576073>

The *hsp65* database contained 198 genera and 1066 unique taxa, the whole represented by 5,165 sequences, of which 1,782 were denoted as GROESL at the time of the analysis. TB11 and TB12 primer complementarity with sequences from the reference database was evaluated using the OligoAnalyzer 3.1 tool (<http://www.idtdna.com/calc/analyzer>). Percent coverage of primers for *Actinobacteria* genomes from different classes and orders was determined with primer BLAST using the RefSeq reference genomes

(NCBI). Only targets with no mismatches for the last three nucleotides at the 3' end and no more than four overall mismatches were included.

3. RESULTS

3.1. Sequence polymorphism of *hsp65* versus 16S rRNA genes for cave *Actinobacteria*

The rarefaction curves of 16S rRNA genes for both 97% and 99% cutoffs were more tilted, indicating an accumulation of identical OTUs due to their repeated sampling, which did not increase with a stricter cutoff, in contrast to *hsp65* (Figure 2A). These results are consistent with the *in silico* analysis of *hsp65* and 16S rRNA gene partial sequences obtained from 105 genomes (47 species from the order *Streptomyetales* and 58 species from the order *Corynebacteriales*) available on IMG. Pairwise distances between *hsp65* sequences, even for closely related species, were higher than those between 16S rRNA genes (Figure 2B).

The *hsp65* primers targeted *hsp65* sequences from the reference database with lower numbers of mismatches compared to the outgroup (*i.e.*, distant homologs renamed in the database as GROESL, *GroEl* of *Escherichia coli*, and Chloroflexi members) (Table A3). However, the primers only targeted select groups from the class *Actinobacteria*, as demonstrated in the analysis with representative genomes from online databases of the National Center for Biotechnology Information (NCBI; U.S. National Library of Medicine, MD, USA) (Table A4). The number of taxa identified at class or lower taxonomic levels is shown in Figure A1. Data indicated that 20% of the classes, 32% of the orders, 36% of the families, and 25% of the genus-level taxa detected with *hsp65* were also recovered through 16S rRNA gene analysis. Although there was a trend for a higher number of taxa uniquely detected by the *hsp65* marker at lower taxonomic ranks, contrary to the 16S rRNA gene marker, *hsp65* sequences detected a lower number of taxa overall. However, at the species level, 168 species were recovered with the *hsp65* marker only, and the highest species numbers were found within the *Streptomyces*, *Mycobacterium*, and *Nocardia* genera (Table A5).

In summary, the 16S rRNA gene marker was more suited for broad taxonomic profiling of the *Actinobacteria* community. In comparison, the *hsp65* marker could help to distinguish between selected

Actinobacteria species, which the 16S rRNA gene marker failed to achieve; therefore, it complemented the results obtained with the 16S rRNA gene marker (Table A5).

3.2. *Actinobacteria* diversity in caves based on *hsp65* versus 16S rRNA genes

In the diversity analysis based on 37 common samples of both markers, the *hsp65* marker recorded higher richness (Chao-1 index) and evenness (Simpson evenness index) than the 16S rRNA gene marker when considering the 99% OTU cutoff (Table A6). The diversity analysis based on all samples indicated that Rouffignac Cave exhibited the highest *Actinobacteria* richness (Chao 1 index, for the *hsp65* marker; Figure A2) but lower evenness (Simpson evenness index; for both markers; Figure A2) and diversity (Inverse Simpson index; for the 16S rRNA gene marker; Figure A2) among other caves. In contrast, the Lascaux and Rouffignac Caves had lower evenness than the pristine Reille (for the 16S rRNA gene marker) and Mouflon (for the *hsp65* marker; Figure A2) Caves) When comparing locations within Lascaux Cave, significantly higher *Actinobacteria* richness was detected in Sas-1 compared with Passage B using only the 16S rRNA gene marker (Figure A2). The average diversity of the unmarked areas was higher than that of the marked areas for both markers (Figure A2).

3.3. *Actinobacteria* community structure in caves based on *hsp65* versus 16S rRNA genes

NMDS analysis was conducted to compare the *Actinobacteria* communities in caves (97% OTU cutoff). These communities did not differ significantly between pristine caves, Reille and Mouflon (AMOVA), based on both *hsp65* and 16S rRNA genes. NMDS distinguished three groups of communities corresponding to Lascaux, Rouffignac, and the two pristine caves, Reille and Mouflon, which was further supported by AMOVA (Table A7 A). The AMOVA comparison of anthropized (Lascaux and Rouffignac) versus pristine (Reille and Mouflon) caves revealed that they significantly differed according to both markers (*hsp65*: $F = 3.986$, $P < 0.001$; 16S rRNA genes: $F = 8.610$, $P < 0.001$). Within Lascaux Cave, the *Actinobacteria* communities differed largely according to location (*i.e.*, Sas-1, Passage B, Passage IP, Apse, and Diaclase) with both markers, except that differences between Passage IP, Sas-1 (both markers), and

Diaclase (*hsp65*) were not significant due to the low number of samples (Table A7 A). Overall, the effect of surface alterations (i.e., marked versus unmarked areas) in Lascaux Cave was significant for both markers (Table A7 A), with higher *Actinobacteria* homogeneity in unmarked versus marked areas (Table A7 B). Passage B presented the highest community stability compared to Passage IP, Apse, and Diaclase (Table A7 B), which was reflected in the NMDS plot, where the marked and unmarked areas in Passage IP were the least separated (Figure 3).

3.4. Core and cave-specific *Actinobacteria* microbiomes based on *hsp65* versus 16S rRNA genes

Venn diagrams showed that the number of OTUs shared between different caves (Rouffignac, Lascaux, Mouflon, and Reille) was two (0.04%) for the *hsp65* marker and 72 (11%) for the 16S rRNA gene marker (Figure A3). Similarly, the number of OTUs shared between different Lascaux rooms (Sas-1, Passage, Apse, and Diaclase) was 23 (0.89%) for the *hsp65* marker and 88 (15.9%) for the 16S rRNA gene marker. The core microbiome was mainly constituted by *Pseudonocardiales* and *Streptomycetales*, as determined by *hsp65*, and *Gaiellales*, *Pseudonocardiales*, and strain IMCC26256_ge, as determined by 16S rRNA genes (Figure A4).

Pairwise comparison of cave conditions (Rouffignac, Lascaux's Apse, and Diaclase, Mouflon, Reille) using significantly different OTUs (Metastats; White et al., 2009) showed that Rouffignac (based on *hsp65*) and Lascaux's Apse (based on 16S rRNA genes) were most enriched in OTUs, separating them from the other caves (Figure 4). Among them, *Pseudonocardiaceae*, *Crosiella*, and *Nocardia* were the most proportionally abundant taxa that distinguished Rouffignac (for both markers), while *Nocardioides*, *Rhodococcus*, and *Pseudonocardia* distinguished Lascaux's Apse from the other caves and Diaclase. *Mycobacterium* was uniquely found in Lascaux's Apse based on both markers. Pristine caves were typified by *Streptomyces* (*hsp65* marker), *Gaiellales*, and *Pseudonocardiales*, and the environmental clones MB-A2-108_ge and IMCC26256_ge (16S rRNA genes) (Figure 4).

3.5. Proportions of cave *Actinobacteria* in sequence libraries based on *hsp65* versus 16S rRNA genes

A total of 1,799,680 *hsp65* and 1,919,236 16S rRNA gene sequences were obtained, out of which 696,716 *hsp65* gene sequences were mapped to 968 (97% cutoff) and 4,269 (99% cutoff) OTUs, while 136,183 16S rRNA gene sequences were mapped to 299 (97% cutoff) and 718 (99% cutoff) OTUs.

Within that, 696,007 *hsp65* and 99,980 16S rRNA gene sequences were allocated to the *Actinobacteria* class, whereas 293,923 and 3803, respectively, could not be allocated to a defined order. Concerning the *hsp65* gene, the *Streptomyces* genus was the most proportionally abundant, with 69,245 sequences representing about 9.9% of the identified *Actinobacteria*; however, 9,283 sequences from this genus could not be affiliated to a specific species within the *Streptomyces* genus. The *Mycobacterium* and *Nocardia* genera were also important groups representing about 7.8 and 1.3% of the identified *Actinobacteria*, respectively.

For the 16S rRNA gene data, the *Streptomyces* genus was the second most represented after *Pseudonocardia*, with 13,263 sequences representing about 9.7% of the identified *Actinobacteria*. The *Mycobacterium* and *Nocardia* genera were other important groups representing about 0.79 and 0.85% of the identified *Actinobacteria*, respectively.

Taxonomic analysis showed that 27 families were identified by both gene markers, but 5 families were uniquely identified by the *hsp65* marker but not by the 16S rRNA gene marker, while 40 families were identified by the 16S rRNA gene marker but not by the *hsp65* marker. Families uniquely found by the *hsp65* marker were *Jiangellaceae*, *Actinobacteria incertae sedis*, *Brevibacteriaceae*, *Nocardiopsaceae*, and *Gordoniaceae* (Figure 5).

Nocardioidaceae, *Streptomycetaceae*, and *Pseudonocardiaceae* dominated the sequence libraries of both gene markers when considering all cave conditions together. Based on both gene markers, there was a higher proportion of *Streptomycetaceae* in marked areas compared with unmarked areas of Lascaux Cave (Figure 5). With a closer focus on *Streptomyces* using only the *hsp65* marker, which can identify species or species-like groups, a *S. mirabilis*-like grouping (no more accurate association could be made and this was supported by a bootstrap of just over 60%, Figure A5), *S. niveus* and *S. fulvissimus* (bootstrap > 80%)

dominated in marked areas within Lascaux Cave (Metastats $P < 0.005$; Figure A6). This marker could also identify other species, such as *S. exfoliatus* and *S. albus*, with good discriminatory power (supported by a high phylogenetic resolution $n = 99\%$ bootstrap) in Rouffignac. The opposite result was found for *Pseudonocardiaceae*, which were present in higher proportions in unmarked versus marked areas in Lascaux Cave, except for Passage IP (by *hsp65*) and Passage B (by 16S rRNA gene marker) (Figure 5). Among the most proportionally abundant obtained OTU sequences belonging to *Pseudonocardiaceae*, none could be associated with a precise species. *Mycobacteriaceae* were found to be highly represented in Lascaux Cave, and the *hsp65* marker also identified them in Mouflon. Three of the most proportionally abundant *Mycobacteria* species in Lascaux Cave uncovered with the *hsp65* marker were *M. sp.*, *M. algericum*, mainly in Passage B) and *M. lentiflavum* (Figure A7). *Jiangellaceae* was a typical group in Rouffignac and Passage IP (*hsp65* marker). Closer analysis of *Nocardiaceae* with the *hsp65* marker uncovered the greatest richness of *Nocardia* species in Rouffignac (including *N. carnea*, *N. paucivorans*, and *N. abscessus*). Moreover, high proportions of *N. jejuensis* were found in Lascaux and Mouflon compared with *N. cummidelens* in Rouffignac, Apse, and Reille (Figure A8). *N. cyriacigeorgica* was only found in Mouflon Cave and *N. farcinica* only in SAS-1. Both pristine caves presented similar compositions of *Actinobacteria* communities, especially *Gaiellaceae*, taxa related to the environmental clones MB-A2-108_fa and IMCC26256_fa (16S rRNA gene marker), and high proportions of unclassified *Actinobacteria* (*hsp65*). High proportions of unclassified *Actinobacteria* were found in Lascaux's Diaclose, as found for the pristine caves (Figure 5).

The phylogenetic tree encompassing sequences of the most representative OTUs (99% cutoff) obtained in this study, together with their corresponding closest sequences obtained from GenBank is shown in Figure A5. This analysis revealed that the most proportionally abundant *Mycobacterium* (*M. algericum* and *M. lentiflavum*), *Streptomyces* (*S. mirabilis*-like grouping, *S. niveus* and *S. fulvissimus*), *Nocardia* (*N. jejuensis*, *N. salmonicida/cummidelens* clade, *N. cyriacigeorgica*, and *N. farcinica*), and *Pseudonocardia* species are supported by a bootstrap $> 80\%$. Less represented species were also identified; for example, *M. fluoranthenvivorans*, *M. mucogenicum*, and *M. stephanolepidis/salmoniphilum* clade, *M. gallinarum*

(bootstrap > 80%), *S. pratensis*, and *S. albus/sampsonii* clade (bootstrap > 90%); and *N. ninae* (bootstrap = 98%). Identification of these OTUs according to the *hsp65* reference database and corresponding closest sequences obtained from GenBank are listed in the Supplementary Table at <https://doi.org/10.5281/zenodo.6312516>.

3.6. Co-occurrence of *Actinobacteria* in Lascaux Cave based on *hsp65* versus 16S rRNA genes

Co-occurrence networks revealed Spearman correlations between the relative abundances of *Actinobacteria* OTUs in Lascaux Cave (Figure 6), for both Lascaux locations (calculated by LefSe; Segata et al., 2011) or marked/unmarked areas (calculated by Metastats; White et al., 2009). More co-occurrence connections were identified for *hsp65* than for the 16S rRNA gene. For the *hsp65* marker, no more correlated OTUs were indicated by Lefse or Metastats, as typical for the Lascaux locations or marked/unmarked areas, respectively. Negative correlations between the OTUs were only identified with the 16S rRNA gene marker.

Clusters of correlated OTUs dominated the unmarked areas, which also differed significantly in the Apse and Diaclase based on both gene markers (Figure 6). Additionally, OTU clusters that dominated the marked areas were typical for the Diaclase and Sas-1 areas based on the 16S rRNA gene marker. There was a correlation between OTUs from such distinct areas (Diaclase and Sas-1), suggesting that the marked-area factors affected the *Actinobacteria* communities in both locations (Figure 6). However, the *hsp65* gene marker revealed that the OTU clusters typical of marked areas were not typical of any Lascaux Cave locations. Finally, based on the 16S rRNA gene marker, the highest number of correlations between OTUs that did not differ between the marked and unmarked areas was found for those that dominated in Passage B, Sas-1, and Diaclase (Figure 6). For those OTUs, the location factor was more influential than the marked/unmarked areas. Networks based on both markers included variable actors (except for the OTUs in unmarked areas of Apse and Diaclase, regardless of marked/unmarked areas), which indicated that each marker covered different fractions of the community and their interactions.

4. DISCUSSION

This study represents an attempt to exploit the potential of *hsp65* as a taxonomic marker to assess the whole *Actinobacteria* community in an extreme environment, such as caves, highlighting its advantages over the 16S rRNA gene marker. When comparing both genes, Venn results revealed higher variability for *hsp65* sequences compared with *Actinobacteria* 16S rRNA gene sequences, and higher OTU richness (at the 99% cutoff) for *Actinobacteria*. Moreover, using the *hsp65* marker, we found 168 species, which could not be identified in the same samples using the 16S rRNA gene marker. Therefore, the low interspecific polymorphism of 16S rRNA genes leads to an underestimation of diversity, which might be compensated for by using the *hsp65* marker, as found for other protein-coding genes, such as *rpoB* (Vos et al., 2012). However, a higher number of higher-rank taxa were identified with the 16S rRNA gene compared with the *hsp65* marker, which highlights gaps in the *hsp65* reference database, preventing the complete taxonomic assignment of sequences. Another reason might be related to primer mismatches. Indeed, *in silico* analysis revealed that the primers amplified the gene from diverse *Actinobacteria* genomes; however, targets without perfect homology might be amplified with lower efficiency, leading to the underestimation of some taxa (Deagle et al., 2014). Overall, the data indicated that *hsp65* is a suitable complement for the 16S rRNA gene marker in high-throughput sequencing methods for *Actinobacteria*, especially for taxonomic analysis at the species level; however, primer bias may need to be considered for community analyses at higher taxonomic ranks. Moreover, the *hsp65* reference database is biased against representation in nature, due to uneven selection of restricted taxa, especially compared to those available in online databases. The 16S rRNA gene is available to the scientific community for all species because it must be mandatorily supplied as part of the submission. However, *hsp65* gene submission is not necessary when describing a new species, explaining the lack of databases. The availability of bacterial genomes is increasing and in the future, many more sequences will be added that will increase the coverage of our database, allowing for more complete detection of taxa within the *Actinobacteria*.

In this study, the combination of *hsp65* and 16S rRNA gene markers revealed that *Actinobacteria* diversity could vary according to anthropization in caves, on both higher (i.e., between caves) and lower

(i.e., between locations within one cave) spatial scales. Indeed, the actinobacterial communities of anthropized (Lascaux and Rouffignac) and pristine caves (Reille and Mouflon) differed. The occurrence of these caves in the same region and limestone vein probably facilitated this observation, since geographic distance (Barraclough et al., 2012) and geological type (Zhu et al., 2019) are significant factors influencing cave communities. The number of *Actinobacteria* that could be identified as part of the core microbiome was relatively low, which suggested that different cave features can lead to different microbial communities (Alonso et al. 2019). Nevertheless, taxa belonging to this core microbiome (*Pseudonocardiales*, *Streptomycetales*, and *Gaiellales*) can resist anthropogenic disturbance (Shade et al., 2012).

Mouflon and Reille are pristine caves with evenly distributed taxa in the *Actinobacteria* community, possibly as a result of their stable environments without man-made disturbances (Mammola, 2019). Typical pristine cave taxa that were identified included *Gaiellales* (Zhu et al., 2019) and *Actinobacteria* related to the environmental clone MB-A2-108_ge (De, 2019; Zhang et al., 2019), according to the 16S rRNA gene marker. High proportions of *Streptomyces* and unclassified *Actinobacteria* based on *hsp65* data suggest that each marker identifies particular *Actinobacteria* community members. Less anthropized areas, such as the pristine caves of Reille and Mouflon, and the Diaclase area of Lascaux, yielded a higher proportion of 'Actinobacteria unclassified' when using the *hsp65* gene compared with the 16S rRNA gene in the same zones. Thus, these areas host species of families whose *hsp65* gene sequences are missing in our database. Species may already be described for which the *hsp65* sequence is not available, or they may simply represent new and as yet undescribed species. Therefore, our database, which was conceived to reach the species level, seems to be more suitable to identify sequences derived from anthropized environments.

Environmental selection in oligotrophic habitats may promote rapid molecular evolution (Kuo & Ochman, 2009; Sagova-Mareckova et al., 2015) and result in a high number of novel species isolated from caves (De, 2019). A similar result was also found for the *Streptomyces* genus (Hamm et al., 2017), whose genomes may undergo a high rate of evolution (Cheng et al., 2015). For pristine caves, high proportions of *Streptomyces* were obtained from the Diaclase, which is an isolated compartment that has been exposed to

little human impact within the anthropized Lascaux Cave. This observation seemed to confirm the hypothesis of Rangseekaew et al., who reported that, not only individual caves but also less-exposed locations within anthropized caves maintained typical *Actinobacteria* cave community, which probably include novel taxa (Rangseekaew & Pathom-Aree, 2019).

In contrary to a previous study based on the whole bacterial community (Alonso et al., 2019), high richness but low evenness was observed in the visited Rouffignac Cave compared to the pristine caves, based on *hsp65* analysis. The high richness of *Actinobacteria* in this cave might reflect a disturbance effect, which promoted the cohabitation of ecologically different microorganisms (Galand et al., 2016).

In all four caves, the most abundant *Nocardia* species were non-pathogenic, typically associated with caves. In Lascaux and Mouflon *N. jejuensis* dominated, a species that was first isolated from a natural cave, while in Reille and Rouffignac *N. cummidelens* dominated, a species previously isolated from rocks of visited Altamira cave (Jurado, Fernandez-Cortes, et al., 2009). Conversely, certain opportunistic pathogenic *Nocardia* species were also identified. In Rouffignac, *N. carnea* (Boiron et al., 1992), *N. paucivorans* (Watanabe et al., 2006), *N. testacea* (Taj-Aldeen et al., 2013), and *N. abscessus* complex (Kageyama et al., 2004) were identified, this latter already found in anthropized environments (Vautrin et al., 2021). Other opportunistic pathogenic species such as *N. farcinica* were found in SAS-1, which is one of the most anthropized areas of Lascaux Cave. Its presence may be associated with human disturbance. The OTUs associated with this species were very close to the GenBank sequences obtained from a patient with pulmonary co-infection with *M. tuberculosis* (KF432743.1) and Madura foot (CP031418.1) (Zhang et al., 2013). Other opportunistic pathogenic species of *Streptomyces* genera such as *S. albus* could also be found, at a low proportion, in Rouffignac cave, which may also be a consequence of human presence (Martín et al., 2004).

However, pathogenic species were also found in pristine caves, including *N. cyriacigeorgica*, which is one of the most frequent *Nocardia* opportunistic pathogenic species in France. The finding in Mouflon represents the first detection of this species in caves. Thus, this species may belong to the autochthonous core microbiome of the caves. The OTUs associated with this species were close to the GenBank sequences

obtained from patients (EF127505.1; EF127507.1; EF127504.1) and urban sediments (VBUT00000000.1) positioned in phylogroups I and III of the *N. cyriacigeorgica* complex (Vautrin et al., 2021; Schlaberg, Huard, & Della-Latta, 2008). Our study has succeeded in positioning retrieved sequences in different phylogroups inside this complex.

As shown by Zoropogui et al (2013), the whole genome of *N. cyriacigeorgica* presents the acquisition of new genetic elements allowing it to thrive in multiple environments. This is confirmed by its isolation in a cave (this study), in the urban sediments of an infiltration basin (Vautrin et al., 2021), or even in humans. Its presence under different extreme environments could explain the evolution of this species towards several different phylogroups, which may possess different levels of virulence. The sequences found in this cave may belong to the less virulent and most preserved phylogroups, unlike phylogroup II (the most virulent of this species), which was not identified here.

Other opportunistic pathogens belonging to the *Mycobacteriaceae* family were also identified in Lascaux Cave. The two most proportionally abundant species, *M. lentiflavum*, and *M. algericum* are causal agents of pulmonary infections in humans (Chida et al., 2021) and animals (Sahraoui et al., 2011), respectively. In addition, the presence of these species has been reported in aquatic medium (Moradi et al., 2019; Makovcova et al., 2015; Mrlik et al. 2012). OTUs obtained in our study that belonged to these two species were close to the GenBank sequences; one of them was obtained from *Pinna nobilis* (GenBank submission code MN854410.1). Thus, the presence of these species could be due to human input or the presence of water, possibly due to infiltration.

This study reports the occurrence of the most frequent opportunistic pathogenic *Actinobacteria* in France from caves. It is unknown whether environmental strains from these species present any pathogenic potential; however, the occurrence of these species in caves suggests a possible allochthonous input by tourists. In turn, this raises potential health concerns, as previously proposed (Jurado et al., 2010), although *Actinobacteria* disease related to cave visitation has never been reported. In contrast, *Jiangellaceae* are a family typical of pristine caves (Rangseekaew & Pathom-Aree, 2019) and are another important group in Rouffignac Cave. According to the intermediate disturbance hypothesis (Roxburgh et al., 2004), the

coexistence of taxa typical of pristine or anthropized conditions in the same habitat suggests that Rouffignac Cave is under intermediate environmental pressure.

The actinobacterial communities in caves were highly specific even within very short distances (Zhu et al., 2019) since each Lascaux room hosted its own *Actinobacteria* populations. Lascaux Cave was typified by high relative abundances of *Mycobacterium*. Similar to *Nocardia* in Rouffignac, this group can signify external contamination (Jurado et al., 2010; Modra et al., 2017), although it is commonly found in natural environments (Kopecky et al., 2011). Confirming the previous results from Alonso et al. (2018), which were based on whole bacterial community analysis, the different geological substrates of Lascaux's Passage had a stronger role than the occurrence of visual marks; however, the marks had formed in the past and were stable. The effect of the geological substrate could also be observed when comparing Diaclase and Apse, because these two locations displayed different *Actinobacteria* communities, especially in unmarked areas, as shown by the network analysis. Relative isolation and distance from the cave entrance may represent efficient filters for alien microorganisms and enable the maintenance of the cave oligotrophic community, typical for unmarked areas (Cuezva et al., 2009; Mammola, 2019).

Our results showed that the *Actinobacteria* community, mainly *Streptomyces* taxa, varied according to the presence of visual marks. Marked areas, where *Streptomycetaceae* dominated, exhibited much lower diversity compared to unmarked zones, where *Pseudonocardiaceae* were prevalent. Unfortunately, comparisons of retrieved sequences with those in public databases yield low similarity scores that do not permit species associations to be made; therefore, we cannot hypothesize whether the presence of these OTUs is due to human disturbance. Similar to our study, the group *Pseudonocardiaceae* has been confirmed as a true cave rock dweller (Zhu et al., 2019). Moreover, this group has been identified in pigmented zones of pristine caves (Lavoie et al., 2017). Therefore, retrieved sequences may correspond to autochthonous species of the caves. In addition, members of the group *Pseudonocardiaceae* degrade complex molecules (Lavoie et al., 2017; Porca et al., 2012), which suggests that the unmarked areas were not completely oligotrophic (Tomczyk-Żak & Zielenkiewicz, 2016), perhaps as a result of past treatment with biocides. However, the higher community diversity in unmarked areas compared to marked areas suggests that

cooperative relationships prevail, which is typical for oligotrophic cave environments (Tomczyk-Żak & Zielenkiewicz, 2016). In this study, specific features of visual marks were not considered according to room location and geological substrate, as the number of replicates was not sufficient; therefore, we focused on findings obtained at the scale of all visual marks considered together.

The lower diversity of marked areas might result from the competitive advantage of invading species (Hamm et al., 2017; Van Elsas et al., 2012). In isolated cave systems, invasions are mainly expected from the cave entrance, similar to Altamira Cave, where *Streptomyces* is the most dominant group, especially in stains (Groth et al., 1999). Our study reports the detection of *S. mirabilis*-like grouping, *S. fulvissimus*, and *S. niveus*, in caves. These species, which were found in stained areas of Lascaux Cave, have not been previously reported as pathogenic, but rather as producers of secondary metabolites with antimicrobial activity, some of which are used for bioremediation (Saber-Riseh et al., 2021; Flinspach et al., 2014; Schütze et al., 2014; Myronovskyi et al., 2013; El-Sayed, 2012; Kominek, 1972). Regarding *S. fulvissimus* and *S. niveus*, the OTUs belonging to these species were close to the GenBank sequences obtained from torrent muddy soil (CP071044.1), rhizosphere (CP048397.1), and deep-sea sediment (CP018047.1). This seemed to confirm the ability of these species to survive in extreme environments away from human activity. Regarding the *S. mirabilis*-like grouping, the most proportionally abundant species in Lascaux stains found in this study, OTUs were close to the GenBank sequences obtained from fresh creek bank soil (CP074102.1). The set of OTUs associated with this group is positioned in a phylogenetic clade with only 60% bootstrap, which suggests that they could represent a new taxon within *Actinobacteria* communities, producers of stains in Lascaux Cave. Considering the reported antimicrobial activity of *S. mirabilis*, we believe that the ability of this probable new taxon to produce secondary metabolites may assist, through interactions with other microorganisms, in its survival in this stained environment.

Hypothetically, *Streptomyces* might also contribute to pigment production (Abdel-Haliem et al., 2013; Cuezva et al., 2012), or affect the pigment-forming fungi in Lascaux Cave (De La Rosa et al., 2017; Frey-Klett et al., 2011) based on their ability to cooperate with fungi and promote mycelial extension and secondary metabolite production (Frey-Klett et al., 2011). A previous study revealed the co-occurrence of

Streptomyces and pigment-forming fungi, *Acremonium* and *Exophiala*, in Lascaux Cave, although only in unmarked areas (Alonso et al., 2018). Moreover, *Streptomyces mirabilis* is resistant to heavy metals (Schütze et al., 2014); this might play a role in the formation of pigmented marks since protection against toxicity of heavy metals and other chemicals has also been linked to melanin production (Nosanchuk & Casadevall, 2003). Therefore, the role of *Streptomyces* in the development of visual marks should be studied in more detail to evaluate whether this group may represent a missing link in our understanding of stain development in Lascaux Cave.

5. CONCLUSION

Despite the limitations of our database, which yielded a high number of unclassified sequences, the *hsp65* gene demonstrated its utility to complement the 16S rRNA gene which cannot resolve species but encompasses a larger part of the actinobacterial community. Use of the *hsp65* gene allowed us to detect, with a robust bootstrap, several species not previously reported in caves, demonstrating that it can be used reliably to differentiate *Actinobacteria* species in such environments. We expect these limitations to be addressed with the enrolment of new *Actinobacteria* sequences. This can be performed periodically by requesting other public databases such as BLAST, or by enrolling sequences of previously identified strains isolated from environmental studies. The *Actinobacteria* community at different spatial levels reflected the natural quality of the caves and different locations within Lascaux Cave. Anthropization was shown to shape *Actinobacteria* communities, which altered typical cave communities and was exhibited by the *Actinobacteria* indicator taxa. Even if rare opportunistic pathogens from the *Actinobacteria* phylum were detected in anthropized caves, it is unknown whether these detected microorganisms represent an important health risk for visitors. We also found that marked areas on the surfaces of Lascaux Cave are important for shaping *Actinobacteria* communities, with particular reference to *Streptomyces* species, whose role in the formation of visual marks requires further research. Further studies are needed to isolate the *Actinobacteria* taxa detected in these caves to determine the true infectious risk associated with these pathogens and their

role in the pigmentation of the marked areas. A better understanding of *Actinobacteria* functioning in caves will be important to guide efforts in the conservation of Paleolithic cave heritage.

ACKNOWLEDGEMENTS

Funding was provided by DRAC *Nouvelle Aquitaine* (Bordeaux, France; grant ‘Ecologie microbienne de la grotte de Lascaux’), Grant for Cotutelle Studies from the French Government (Campus France), and the Josef Hlavka Foundation and Charles University in Prague. We are grateful for being permitted to sample Rouffignac Cave (F. Plassard) and La Reille (Gaillard family) and to G. Delorme† and *Compagnie des Beunes* potholers for help with pristine caves. We thank S. Géraud, J.C. Portais, A. Rieu, and M. Mauriac (DRAC *Nouvelle Aquitaine*) for providing key information and assistance, and D. Henry-Lormelle and her restoration team for providing technical help with Lascaux sampling, and the Lascaux Scientific Board for their helpful discussions. We are grateful to Emmanuelle Bergeron and Delphine Mouniée for providing technical assistance with building the *hsp65*-based reference taxonomic database used in this study, especially with its enlargement by obtaining sequences from strains of different origins.

We thank Observatoire Français des Nocardiosis and CRB-EML (UMR5557) for supplying clinical and environmental strains, which largely contributed to the construction of the database.

AUTHOR CONTRIBUTIONS

Andrea BURESOVA

Data curation-Lead, Formal analysis-Lead, Investigation-Equal, Methodology-Lead, Writing – original draft-Lead

Jan Kopecky

Formal analysis-Equal, Software-Lead

Marketa SAGOVA-MARECKOVA

Investigation-Equal, Validation-Equal, Writing – review & editing-Equal

Lise Alonso

Investigation-Supporting, Resources-Equal

Florian VAUTRIN

Data curation-Equal, Investigation-Supporting, Methodology-Supporting, Software-Supporting

Yvan Moënne-Loccoz

Conceptualization-Lead, Funding acquisition-Lead, Investigation-Equal, Resources-Lead, Writing – review & editing-Equal

Veronica RODRIGUEZ NAVA

Conceptualization-Lead, Funding acquisition-Lead, Investigation-Lead, Methodology-Lead, Supervision-Lead, Validation-Lead, Writing – review & editing-Lead

DATA AVAILABILITY STATEMENT

The Illumina MiSeq amplicon sequences have been deposited in the NCBI Sequence Read Archive under BioProject PRJNA695576 (*hsp65*; <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA695576>) and PRJNA694921 (16S rRNA gene; <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA694921>). The Supplementary Table is available in Zenodo at <https://doi.org/10.5281/zenodo.6312516> (*hsp65* OTUs, cutoff 99% - their inferred taxonomic allocations according to the *hsp65* database and, for selected OTUs, closest species obtained from GenBank BLAST with percent identity). The *hsp65* metabarcoding DNA sequence database for taxonomic allocations using Mothur is available in Zenodo at <https://doi.org/10.5281/zenodo.5576073>

CONFLICT OF INTERESTS

None declared

ETHICS STATEMENT

None required

ORCID

Andrea Buresova <https://orcid.org/0000-0001-8779-8702>

Jan Kopecky <https://orcid.org/0000-0002-3437-257X>

Marketa Sagova-Mareckova <https://orcid.org/0000-0002-0331-8570>

Lise Alonso <https://orcid.org/0000-0002-4137-9876>

Florian Vautrin <https://orcid.org/0000-0003-0521-6057>

Yvan Moëgne-Loccoz <https://orcid.org/0000-0002-9817-1953>

Veronica Rodriguez-Nava <https://orcid.org/0000-0001-7083-2441>

REFERENCES

Abdel-Haliem, M. E. F., Sakr, A. A., Ali, M. F., Ghaly, M. F., & Sohlenkamp, C. (2013).

Characterization of *Streptomyces* isolates causing colour changes of mural paintings in ancient Egyptian tombs. *Microbiological Research*, *168*(7), 428–437.

<https://doi.org/10.1016/j.micres.2013.02.004>

Aigle, A., Colin, Y., Bouchali, R., Bourgeois, E., Marti, R., Ribun, S., Marjolet, L., Pozzi, A. C. M., Misery, B., Colinon, C., Bernardin-Souibgui, C., Wiest, L., Blaha, D., Galia, W., & Cournoyer, B. (2021). Spatio-temporal variations in chemical pollutants found among urban deposits match changes in thiopurine S-methyltransferase-harboring bacteria tracked by the *tpm* metabarcoding approach. *Science of the Total Environment*, *767*, 145425. doi: 10.1016/j.scitotenv.2021.

- Alonso, L., Creuzé-des-Châtelliers, C., Trabac, T., Dubost, A., Moëgne-Loccoz, Y., & Pommier, T. (2018). Rock substrate rather than black stain alterations drives microbial community structure in the passage of Lascaux Cave. *Microbiome*, *6*(1), 216. <https://doi.org/10.1186/s40168-018-0599-9>
- Alonso, L., Pommier, T., Kaufmann, B., Dubost, A., Chapulliot, D., Doré, J., Douady, C. J., & Moëgne-Loccoz, Y. (2019). Anthropization level of Lascaux Cave microbiome shown by regional-scale comparisons of pristine and anthropized caves. *Molecular Ecology*, *28*(14), 3383–3394. <https://doi.org/10.1111/mec.15144>
- Barracough, T. G., Balbi, K. J., & Ellis, R. J. (2012). Evolving concepts of bacterial species. *Evolution Biology*, *39*, 148–157. <https://doi.org/10.1007/s11692-012-9181-8>
- Barton, H. A., & Jurado, V. (2007). What's up down there? Microbial diversity in caves microorganisms in caves survive under nutrient-poor conditions and are metabolically versatile and unexpectedly diverse. *Microbe*, *2*(3), 132–138.
- Barton, H. A., Taylor, N. M., Kreate, M. P., Springer, A. C., Oehrle, S. A., & Bertog, J. L. (2007). The impact of host rock geochemistry on bacterial community structure in oligotrophic cave environments. *International Journal of Speleology*, *36* (2), 93–104.
- Bastian, F., Jurado, V., Novakova, A., Alabouvette, C., & Saiz-Jimenez, C. (2010). The microbiology of Lascaux Cave. *Microbiology*, *156*(3), 644–652. <https://doi.org/10.1099/mic.0.036160-0>
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the Third International ICWSM Conference (2009)*, 361–362. <https://doi.org/10.1136/qshc.2004.010033>
- Bercea, S., Năstase-Bucur, R., Moldovan, O. T., Kenesz, M., & Constantin, S. (2019). Yearly microbial cycle of human exposed surfaces in show caves. *Subterranean Biology*, *31*, 1–14. <https://doi.org/10.3897/subtbiol.31.34490>
- Bhullar, K., Waglechner, N., Pawlowski, A., Koteva, K., Banks, E. D., Johnston, M. D., Barton, H. A., & Wright, G. D. (2012). Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE*, *7*(4), e34953. <https://doi.org/10.1371/journal.pone.0034953>

Boiron, P., Provost, F., Chevrier, G., & Dupont, B. (1992). Review of nocardial infections in France 1987 to 1990. *European Journal of Clinical Microbiology & Infectious Diseases*, *11*(8), 709–714.

<https://doi.org/10.1007/BF01989975>Bontemps, Z., Alonso, L., Pommier, T., Hugoni, M., & Moënnelocoz, Y. (2021). Microbial

ecology of tourist Paleolithic caves. *Science of The Total Environment*, 151492.

Chen, W., Zhang, C. K., Cheng, Y., Zhang, S., & Zhao, H. (2013). A Comparison of Methods for Clustering 16S rRNA Sequences into OTUs. *PLoS ONE*, *8*(8), e70837.

<https://doi.org/10.1371/journal.pone.0070837>

Cheng, K., Rong, X., Pinto-Tomás, A. A., Fernández-Villalobos, M., Murillo-Cruz, C., & Huang, Y.

(2015). Population genetic analysis of *Streptomyces albidoflavus* reveals habitat barriers to homologous recombination in the diversification of streptomycetes. *Applied and Environmental Microbiology*, *81*(3), 966–975. <https://doi.org/10.1128/AEM.02925-14>

Chida, K., Yamanaka, Y., Sato, A., Ito, S., Takasaka, N., Ishikawa, T., & Kuwano, K. (2021). Solitary pulmonary nodule caused by pulmonary *Mycobacterium lentiflavum* infection. *Respiratory Medicine Case Reports*, *34*, 101510. doi: 10.1016/j.rmcr.2021.101510

Colaco, C. A., & MacDougall, A. (2014). Mycobacterial chaperonins: The tail wags the dog. *FEMS Microbiology Letters*, *350*(1), 20–24. <https://doi.org/10.1111/1574-6968.12276>

Cuezva, S., Fernandez-Cortes, A., Porca, E., Pašić, L., Jurado, V., Hernandez-Marine, M., Serrano-Ortiz, P., Hermosin, B., Cañaveras, J. C., Sanchez-Moral, S., & Saiz-Jimenez, C. (2012). The biogeochemical role of Actinobacteria in Altamira Cave, Spain. *FEMS Microbiology Ecology*, *81*(1), 281–290. <https://doi.org/10.1111/j.1574-6941.2012.01391.x>

Cuezva, S., Sanchez-Moral, S., Saiz-Jimenez, C., & Cañaveras, J. C. (2009). Microbial communities and associated mineral fabrics in Altamira Cave, Spain. *International Journal of Speleology*, *38*(1), 83–92. <https://doi.org/10.5038/1827-806x.38.1.9>

- De, K. A. (2019). Profiling bacterial diversity of B2 Cave, a limestone cave of Baratang, Andaman and Nicobar islands, India. *Proceedings of the Indian National Science Academy*, 85(4), 853–862. <https://doi.org/10.16943/ptinsa/2019/49589>
- De La Rosa, J. M., Martin-Sanchez, P. M., Sanchez-Cortes, S., Hermosin, B., Knicker, H., & Saiz-Jimenez, C. (2017). Structure of melanins from the fungi *Ochroconis lascauxensis* and *Ochroconis anomala* contaminating rock art in the Lascaux Cave. *Scientific Reports*, 7(1), 1–11. <https://doi.org/10.1038/s41598-017-13862-7>
- De Mandal, S., Chatterjee, R., & Kumar, N. S. (2017). Dominant bacterial phyla in caves and their predicted functional roles in C and N cycle. *BMC Microbiology*, 17(1), 1–9. <https://doi.org/10.1186/s12866-017-1002-x>
- Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome *c* oxidase subunit I marker: Not a perfect match. *Biology Letters*, 10(9), 20140562. <https://doi.org/10.1098/rsbl.2014.0562>
- Duchêne, A.-M., Kieser, H. M., Hopwood, D. A., Thompson, C. J., & Mazodier, P. (1994). Characterization of two *groEL* genes in *Streptomyces coelicolor* A3 (2). *Gene*, 144(1), 97–101. [https://doi.org/10.1016/0378-1119\(94\)90210-0](https://doi.org/10.1016/0378-1119(94)90210-0)
- Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10(10), 996–998. <https://doi.org/10.1038/nmeth.2604>
- El-Sayed, M. H. (2012). Di-(2-ethylhexyl) Phthalate, a major bioactive metabolite with antimicrobial and cytotoxic activity isolated from the culture filtrate of newly isolated soil *Streptomyces* (*Streptomyces mirabilis* strain NSQu-25). *World Applied Sciences Journal*, 20(9), 1202–1212. doi: 10.5829/idosi.wasj.2012.20.09.2868
- Fang, B.-Z., Salam, N., Han, M.-X., Jiao, J.-Y., Cheng, J., Wei, D.-Q., Xiao, M., & Li, W.-J. (2017). Insights on the Effects of Heat Pretreatment, pH, and Calcium Salts on Isolation of Rare Actinobacteria from Karstic Caves. *Frontiers in Microbiology*, 8. <https://doi.org/10.3389/fmicb.2017.01535>

- Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
- Fernandez-Cortes, A., Cuezva, S., Sanchez-Moral, S., Cañaveras, J. C., Porca, E., Jurado, V., Martin-Sanchez, P. M., & Saiz-Jimenez, C. (2011). Detection of human-induced environmental disturbances in a show cave. *Environmental Science and Pollution Research*, 18(6), 1037–1045. <https://doi.org/10.1007/s11356-011-0513-5>
- Flinspach, K., Rückert, C., Kalinowski, J., Heide, L., & Apel, A. K. (2014). Draft genome sequence of *Streptomyces niveus* NCIMB 11891, producer of the aminocoumarin antibiotic novobiocin. *Genome announcements*, 2(1), e01146-13. doi: 10.1128/genomeA.01146-13.
- Fox, G. E., Wisotzkey, J. D., & Jurtshuk, P. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.*, 42(1), 166–170. <https://doi.org/10.1099/00207713-42-1-166>
- Frey-Klett, P., Burlinson, P., Deveau, A., Barret, M., Tarkka, M., & Sarniguet, A. (2011). Bacterial-fungal interactions: Hyphens between agricultural, clinical, environmental, and food microbiologists. *Microbiology and Molecular Biology Reviews : MMBR*, 75(4), 583–609. <https://doi.org/10.1128/MMBR.00020-11>
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129–1164. <https://doi.org/10.1002/spe.4380211102>
- Galand, P. E., Lucas, S., Fagervold, S. K., Peru, E., Pruski, A. M., Vétion, G., Dupuy, C., & Guizien, K. (2016). Disturbance increases microbial community diversity and production in marine sediments. *Frontiers in Microbiology*, 7(DEC), 1950. <https://doi.org/10.3389/fmicb.2016.01950>
- Gao, B., & Gupta, R. (2012). Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiology and Molecular Biology Reviews*, 76(1), 66–112.
- Gonzalez-Pimentel, J. L., Miller, A. Z., Jurado, V., Laiz, L., Pereira, M. F. C., & Saiz-Jimenez, C. (2018). Yellow coloured mats from lava tubes of La Palma (Canary Islands, Spain) are dominated by

- metabolically active Actinobacteria. *Sci. Rep.*, 8(1), 1–11. <https://doi.org/10.1038/s41598-018-20393-2>
- Groth, I., Vettermann, R., Schuetze, B., Schumann, P., & Saiz-Jimenez, C. (1999). Actinomycetes in Karstic caves of northern Spain (Altamira and Tito Bustillo). *Journal of Microbiological Methods*, 36(1–2), 115–122. [https://doi.org/10.1016/S0167-7012\(99\)00016-0](https://doi.org/10.1016/S0167-7012(99)00016-0)
- Hall, T., BioSciences, I., & Carlsbad, C. (2011). BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*, 2(June), 60–61. <https://doi.org/10.1002/prot.24632>
- Hamedi, J., Kafshnouchi, M., & Ranjbaran, M. (2019). A Study on actinobacterial diversity of Hampoel cave and screening of their biological activities. *Saudi Journal of Biological Sciences*, 26(7), 1587–1595. <https://doi.org/10.1016/j.sjbs.2018.10.010>
- Hamm, P. S., Caimi, N. A., Northup, D. E., Valdez, E. W., Buecher, D. C., Dunlap, C. A., Labeda, D. P., Lueschow, S., & Porras-Alfaro, A. (2017). Western bats as a reservoir of novel *Streptomyces* species with antifungal activity. *Applied and Environmental Microbiology*, 83(5), e03057-16. <https://doi.org/10.1128/AEM.03057-16>
- Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). Past: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1), 1–9.
- Herlemann, D. P. R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., & Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME Journal*, 5(10), 1571–1579. <https://doi.org/10.1038/ismej.2011.41>
- Hirsch, P. R., Mauchline, T. H., & Clark, I. M. (2010). Culture-independent molecular techniques for soil microbial ecology. *Soil Biol. Biochem.*, 42(6), 878–887. <https://doi.org/10.1016/j.soilbio.2010.02.019>
- Hoyos, M., Cañaveras, J. C., Sánchez-Moral, S., Sanz-Rubio, E., & Soler, V. (1998). Microclimatic characterization of a karstic cave: Human impact on microenvironmental parameters of a prehistoric rock art cave (Candamo Cave, northern Spain). *Environmental Geology*, 33(4), 231–242. <https://doi.org/10.1007/s002540050242>

- Johnston, M. D., Muench, B. A., Banks, E. D., & Barton, H. A. (2012). Human urine in Lechuguilla cave: The microbiological impact and potential for bioremediation. *Journal of Cave and Karst Studies*, 74(3), 278–291. <https://doi.org/10.4311/2011MB0227>
- Jurado, V., Boiron, P., Kroppenstedt, R. M., Laurent, F., Couble, A., Laiz, L., Klenk, H.-P., Gonzalez, J. M., Saiz-Jimenez, C., Mouniee, D., Bergeron, E., & Rodriguez-Nava, V. (2008). *Nocardia altamirensis* sp. Nov., isolated from Altamira cave, Cantabria, Spain. *International Journal of Systematic and Evolutionary Microbiology*, 58(9), 2210–2214. <https://doi.org/10.1099/ij.s.0.65482-0>
- Jurado, V., Fernandez-Cortes, A., Cuezva, S., Laiz, L., Cañaveras, J. C., Sanchez-Moral, S., & Saiz-Jimenez, C. (2009). The fungal colonisation of rock-art caves: Experimental evidence. *Naturwissenschaften*, 96(9), 1027–1034. <https://doi.org/10.1007/s00114-009-0561-6>
- Jurado, V., Kroppenstedt, R. M., Saiz-Jimenez, C., Klenk, H. P., Mouniée, D., Laiz, L., Couble, A., Pötter, G., Boiron, P., & Rodríguez-Nava, V. (2009). *Hoyosella altamirensis* gen. Nov., sp. Nov., a new member of the order Actinomycetales isolated from a cave biofilm. *International Journal of Systematic and Evolutionary Microbiology*, 59(12), 3105–3110. <https://doi.org/10.1099/ij.s.0.008664-0>
- Jurado, V., Laiz, L., Rodriguez-Nava, V., Boiron, P., Hermosin, B., Sanchez-Moral, S., & Saiz-Jimenez, C. (2010). Pathogenic and opportunistic microorganisms in caves. *International Journal of Speleology*, 39 (1), 15–24.
- Kageyama, A., Yazawa, K., Kudo, T., Taniguchi, H., Nishimura, K., & Mikami, Y. (2004). First isolates of *Nocardia abscessus* from humans and soil in Japan. *Nippon Ishinkin Gakkai Zasshi*, 45(1), 17–21. <https://doi.org/10.3314/jjmm.45.17>
- Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120
- Kominek, L. A. (1972). Biosynthesis of novobiocin by *Streptomyces niveus*. *Antimicrobial Agents and Chemotherapy*, 1(2), 123–134. doi: 10.1128/AAC.1.2.123

- Kopecky, J., Kyselkova, M., Omelka, M., Cermak, L., Novotna, J., Grundmann, G., Moënné-Loccoz, Y., & Sagova-Mareckova, M. (2011). Environmental mycobacteria closely related to the pathogenic species evidenced in an acidic forest wetland. *Soil Biology and Biochemistry*, *43*(3), 697–700. <https://doi.org/10.1016/J.SOILBIO.2010.11.033>
- Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.*, *79*(17), 5112–5120. <https://doi.org/10.1128/AEM.01043-13>
- Kumar, C. M. S., Mande, S. C., & Mahajan, G. (2015). Multiple chaperonins in bacteria—Novel functions and non-canonical behaviors. *Cell Stress & Chaperones*, *20*(4), 555–574. <https://doi.org/10.1007/s12192-015-0598-8>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, *35*(6), 1547. doi: 10.1093/molbev/msy096
- Kuo, C. H., & Ochman, H. (2009). Inferring clocks when lacking rocks: The variable rates of molecular evolution in bacteria. *Biology Direct*, *4*(1), 35. <https://doi.org/10.1186/1745-6150-4-35>
- Lavoie, K. H., Winter, A. S., Read, K. J. H., Hughes, E. M., Spilde, M. N., & Northup, D. E. (2017). Comparison of bacterial communities from lava cave microbial mats to overlying surface soils from Lava Beds National Monument, USA. *PLoS One*, *12*(2). <https://doi.org/10.1371/journal.pone.0169339>
- Long, Y., Jiang, J., Hu, X., Zhou, J., Hu, J., & Zhou, S. (2019). Actinobacterial community in Shuanghe Cave using culture-dependent and -independent approaches. *World Journal of Microbiology and Biotechnology*, *35*(10). <https://doi.org/10.1007/s11274-019-2713-y>
- Maciejewska, M., Adam, D., Naômé, A., Martinet, L., Tenconi, E., Calusinska, M., Delfosse, P., Hanikenne, M., Baurain, D., Compère, P., Carno I, M., Barton, H. A., & Rigali, S. (2017).

- Assessment of the potential role of *Streptomyces* in cave moonmilk formation. *Front. Microbiol.*, 8(JUN), 1–18. <https://doi.org/10.3389/fmicb.2017.01181>
- Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
- Makovcova, J., Babak, V., Slany, M., & Slana, I. (2015). Comparison of methods for the isolation of mycobacteria from water treatment plant sludge. *Antonie van Leeuwenhoek*, 107(5), 1165–1179. doi: 10.1007/s10482-015-0408-4.
- Mammola, S. (2019). Finding answers in the dark: Caves as models in ecology fifty years after Poulson and White. *Ecography*, 42(7), 1331–1351. <https://doi.org/10.1111/ecog.03905>
- Martín, M. C., Manteca, A., Castillo, M. L., Vázquez, F., & Méndez, F. J. (2004). *Streptomyces albus* isolated from a human actinomycetoma and characterized by molecular techniques. *Journal of clinical microbiology*, 42(12), 5957–5960. doi: 10.1128/JCM.42.12.5957-5960.2004.
- Martin-Sanchez, P. M., Miller, A. Z., & Saiz-Jimenez, C. (2015). Lascaux Cave: An example of fragile ecological balance in subterranean environments. *Microbial Life of Cave Systems*, AS, De Gruyter (editors), Berlin/Boston, 279–301. <https://doi.org/10.1515/9783110339888-015>
- Martin-Sanchez, P. M., Nováková, A., Bastian, F., Alabouvette, C., & Saiz-Jimenez, C. (2012). Use of biocides for the control of fungal outbreaks in subterranean environments: The case of the Lascaux Cave in France. *Environmental Science & Technology*, 46(7), 3762–3770. <https://doi.org/10.1021/es2040625>
- Miller, A. Z., Garcia-Sanchez, A. M., Martin-Sanchez, P. M., Costa Pereira, M. F., Spangenberg, J. E., Jurado, V., Dionísio, A., Afonso, M. J., Iglésias Chaminé, H. I., Hermosin, B., & Saiz-Jimenez, C. (2018). Origin of abundant moonmilk deposits in a subsurface granitic environment. *Sedimentology*, 65(5), 1482–1503. <https://doi.org/10.1111/sed.12431>
- Modra, H., Bartos, M., Hribova, P., Ulmann, V., Hubelova, D., Konecny, O., Gersl, M., Kudelka, J., Voros, D., & Pavlik, I. (2017). Detection of mycobacteria in the environment of the Moravian Karst (Bull Rock Cave and the relevant water catchment area): The impact of water sediment,

earthworm castings and bat guano. *Veterinárni Medicina*, 62 (3), 153–168.

<https://doi.org/10.17221/126/2016-VETMED>

Moradi, S., Nasiri, M. J., Pourahmad, F., & Darban-Sarokhalil, D. (2019). Molecular characterization of nontuberculous mycobacteria in hospital waters: a two-year surveillance study in Tehran, Iran.

Journal of water and health, 17(2), 350-356. doi: 10.2166/wh.2019.294.

Mrlik, V., Slany, M., Kubecka, J., Seda, J., Necas, A., Babak, V., Slana, I., Kriz, P., & Pavlik, I. (2012).

A low prevalence of mycobacteria in freshwater fish from water reservoirs, ponds and farms. *Journal of fish diseases*, 35(7), 497-504. doi: 10.1111/j.1365-2761.2012.01369.

Mulec, J. (2014). Human impact on underground cultural and natural heritage sites, biological parameters of monitoring and remediation actions for insensitive surfaces: Case of Slovenian show caves.

Journal for Nature Conservation, 22(2), 132–141. <https://doi.org/10.1016/J.JNC.2013.10.001>

Myronovskiy, M., Tokovenko, B., Manderscheid, N., Petzke, L., & Luzhetskyy, A. (2013). Complete genome sequence of *Streptomyces fulvissimus*. *Journal of Biotechnology*, 168(1), 117–118. doi:

10.1016/j.jbiotec.2013.08.013

Neral, A., Rajput, Y., & Rai, V. (2015). Genotypic and serotypic confirmations of bacterial community to Kotumsar cave for occupational safety of cave workers and visitors from pathogenic threats.

International Journal of Occupational Safety and Health, 5(1), 22–27.

<https://doi.org/10.3126/ijosh.v5i1.16631>

Nosanchuk, J. D., & Casadevall, A. (2003). The contribution of melanin to microbial pathogenesis. In *Cellular Microbiology* (Vol. 5, Issue 4).

Ogier, J. C., Pagès, S., Galan, M., Barret, M., & Gaudriault, S. (2019). RpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiology*,

19(171), 2–16. <https://doi.org/10.1186/s12866-019-1546-z>

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Maintainer, & Wagner,

- H. (2018). *Package 'vegan': Community Ecology Package. R package version 2.5-2 Vienna: R Foundation for Statistical Computing.*
- Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*(1), 385. <https://doi.org/10.1186/1471-2105-12-385>
- Ortiz, M., Legatzki, A., Neilson, J. W., Fryslie, B., Nelson, W. M., Wing, R. A., Soderlund, C. A., Pryor, B. M., & Maier, R. M. (2014). Making a living while starving in the dark: Metagenomic insights into the energy dynamics of a carbonate cave. *The ISME Journal*, *8*(2), 478–491. <https://doi.org/10.1038/ismej.2013.159>
- Porca, E., Jurado, V., Žgur-Bertok, D., Saiz-Jimenez, C., & Pašić, L. (2012). Comparative analysis of yellow microbial communities growing on the walls of geographically distinct caves indicates a common core of microorganisms involved in their formation. *FEMS Microbiology Ecology*, *81*(1), 255–266. <https://doi.org/10.1111/j.1574-6941.2012.01383.x>
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, *5*(3), e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. <https://doi.org/10.1093/nar/gks1219>
- R Core Team. (2018). R Core Team, R : A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*, R Foundation for Statistical Computing.
- Rajput, Y., Rai, V., & Biswas, J. (2012). Screening of bacterial isolates from various microhabitat sediments of Kotumsar cave: A cogitation on their respective benefits and expected threats for complete biosphere and tourists. *Research Journal of Environmental Toxicology*, *6*(1), 13–24.
- Rangseekaew, P., & Pathom-Aree, W. (2019). Cave actinobacteria as producers of bioactive metabolites. *Frontiers in Microbiology*, *10*, 387. <https://doi.org/10.3389/fmicb.2019.00387>
- Rodríguez-Nava, V., Couble, A., Devulder, G., Flandrois, J.-P., Boiron, P., & Laurent, F. (2006). Use of PCR-restriction enzyme pattern analysis and sequencing database for hsp65 gene-based

- identification of *Nocardia* species. *Journal of Clinical Microbiology*, *44*(2), 536–546.
<https://doi.org/10.1128/JCM.44.2.536-546.2006>
- Rodriguez-Nava, V., Khan, Z. U., Pötter, G., Kroppenstedt, R. M., Boiron, P., & Laurent, F. (2007). *Nocardia coubleae* sp. Nov., isolated from oil-contaminated Kuwaiti soil. *International Journal of Systematic and Evolutionary Microbiology*, *57*(7), 1482–1486.
<https://doi.org/10.1099/ijms.0.64815-0>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. <https://doi.org/10.7717/peerj.2584>
- Roxburgh, S. H., Shea, K., & Wilson, J. B. (2004). The intermediate disturbance hypothesis: Patch dynamics and mechanisms of species coexistence. *Ecology*, *85*(2), 359–371.
<https://doi.org/10.1890/03-0266>
- Saberi-Riseh, R., & Moradi-Pour, M. (2021). A novel encapsulation of *Streptomyces fulvissimus* Uts22 by spray drying and its biocontrol efficiency against *Gaeumannomyces graminis*, the causal agent of take-all disease in wheat. *Pest Management Science*. doi: 10.1002/ps.6469.
- Sagova-Mareckova, M., Ulanova, D., Sanderova, P., Omelka, M., Kamenik, Z., Olsovska, J., & Kopecky, J. (2015). Phylogenetic relatedness determined between antibiotic resistance and 16S rRNA genes in actinobacteria. *BMC Microbiology*, *15*(1), 81. <https://doi.org/10.1186/s12866-015-0416-6>
- Sahraoui, N., Ballif, M., Zelleg, S., Yousfi, N., Ritter, C., Friedel, U., Amstutz, B., Yala, Y., Boulahbal, F., Guetarni, D., Zinsstag, J., & Keller, P. M. (2011). *Mycobacterium algericum* sp. nov., a novel rapidly growing species related to the *Mycobacterium terrae* complex and associated with goat lung lesions. *International journal of systematic and evolutionary microbiology*, *61*(8), 1870-1874. . doi: 10.1099/ijms.0.024851-0
- Saitou N. & Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* *4*:406-425.
- Sánchez-Moral, S., Soler, V., Cañaveras, J. C., Sanz-Rubio, E., Van Grieken, R., & Gysels, K. (1999). Inorganic deterioration affecting the Altamira Cave, N Spain: Quantitative approach to wall-

- corrosion (solutional etching) processes induced by visitors. *Science of The Total Environment*, 243–244, 67–84. [https://doi.org/10.1016/S0048-9697\(99\)00348-4](https://doi.org/10.1016/S0048-9697(99)00348-4)
- Schabereiter-Gurtner, C., Saiz-Jimenez, C., Pinar, G., Lubitz, W., & Rölleke, S. (2002). Altamira cave Paleolithic paintings harbor partly unknown bacterial communities. *FEMS Microbiology Letters*, 211(1), 7–11. <https://doi.org/10.1111/j.1574-6968.2002.tb11195.x>
- Schlaberg, R., Huard, R. C., & Della-Latta, P. (2008). *Nocardia cyriacigeorgica*, an emerging pathogen in the United States. *Journal of clinical microbiology*, 46(1), 265-273. doi: 10.1128/JCM.00937-07.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 75(23), 7537–7541. <https://doi.org/10.1128/AEM.01541-09>
- Schütze, E., Klose, M., Merten, D., Nietzsche, S., Senfleben, D., Roth, M., & Kothe, E. (2014). Growth of streptomycetes in soil and their impact on bioremediation. *Journal of Hazardous Materials*, 267, 128–135. <https://doi.org/10.1016/j.jhazmat.2013.12.055>
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6), R60. <https://doi.org/10.1186/gb-2011-12-6-r60>
- Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., Bürgmann, H., Huber, D. H., Langenheder, S., Lennon, J. T., Martiny, J. B. H., Matulich, K. L., Schmidt, T. M., & Handelsman, J. (2012). Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology*, 3, 417. <https://doi.org/10.3389/fmicb.2012.00417>
- Syiemiong, D., & Jha, D. K. (2019). Antibacterial potential of Actinobacteria from a Limestone Mining Site in Meghalaya, India. *Journal of Pure and Applied Microbiology*, 13(2), 789–802. <https://doi.org/10.22207/JPAM.13.2.14>

- Taj-Aldeen, S. J., Deshmukh, A., Doiphode, S., Wahab, A. A., Allangawi, M., Muzrkchi, A. Al, Klaassen, C. H., & Meis, J. J. (2013). Molecular identification and susceptibility pattern of clinical *Nocardia* species: Emergence of *Nocardia crassostreae* as an agent of invasive nocardiosis. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *24*(2), e33–e38.
- Takeda, K., Kang, Y., Yazawa, K., Gono, T., & Mikami, Y. (2010). Phylogenetic studies of *Nocardia* species based on *gyrB* gene analyses. *Journal of Medical Microbiology*, *59*(2), 165–171. <https://doi.org/10.1099/jmm.0.011346-0>
- Tamura K, Stecher G, Peterson D, Filipiński A, & Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, *30*(12):2725-9. doi: 10.1093/molbev/mst197
- Telenti, A., Marchesi, F., Balz, M., Bally, F., Bottger, E. C., & Bodmer, T. (1993). Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. *Journal of Clinical Microbiology*, *31*(2), 175–178.
- Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, (1), 2-3. doi: 10.1002/0471250953.bi0203s00
- Tomczyk-Żak, K., & Zielenkiewicz, U. (2016). Microbial diversity in caves. *Geomicrobiology Journal*, *33*(1), 20–38. <https://doi.org/10.1080/01490451.2014.1003341>
- Van Elsas, J. D., Chiurazzi, M., Mallon, C. A., Elhottova, D., Křišťůfek, V., & Salles, J. F. (2012). Microbial diversity determines the invasion of soil by a bacterial pathogen. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1159–1164. <https://doi.org/10.1073/pnas.1109326109>
- Vautrin, F., Pujic, P., Paquet, C., Bergeron, E., Mounié, D., Marchal, T., ... & Rodriguez-Nava, V. (2021). Microbial risk assessment of *Nocardia cyriacigeorgica* in polluted environments, case of urban rainfall water. *Computational and Structural Biotechnology Journal*, *19*, 384–400. doi: 10.1016/j.csbj.2020.12.017

- Větrovský, T., & Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE*, *8*(2), e57923. <https://doi.org/10.1371/journal.pone.0057923>
- Vos, M., Quince, C., Pijl, A. S., de Hollander, M., & Kowalchuk, G. A. (2012). A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. *PLoS ONE*, *7*(2), e30600. <https://doi.org/10.1371/journal.pone.0030600>
- Watanabe, K., Shinagawa, M., Amishima, M., Iida, S., Yazawa, K., Kageyama, A., Ando, A., & Mikami, Y. (2006). First clinical isolates of *Nocardia carnea*, *Nocardia elegans*, *Nocardia paucivorans*, *Nocardia puris* and *Nocardia takedensis* in Japan. *Nippon Ishinkin Gakkai Zasshi*, *47*(2), 85–89. <https://doi.org/10.3314/jjmm.47.85>
- White, J. R., Nagarajan, N., & Pop, M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS computational biology*, *5*(4), e1000352. [doi:10.1371/journal.pcbi.1000352](https://doi.org/10.1371/journal.pcbi.1000352)
- Yasir, M. (2018). Analysis of bacterial communities and characterization of antimicrobial strains from cave microbiota. *Brazilian Journal of Microbiology*, *49*(2), 248–257. <https://doi.org/10.1016/j.bjm.2017.08.005>
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., & Glöckner, F. O. (2014). The SILVA and ‘all-species Living Tree Project (LTP)’ taxonomic frameworks. *Nucleic Acids Research*, *42*(D1), 643–648. <https://doi.org/10.1093/nar/gkt1209>
- Zhang, Y. Y., Li, Y. B., Huang, M. X., Zhao, X. Q., Zhang, L. S., Liu, W. E., & Wan, K. L. (2013). Novel species including *Mycobacterium fukienense* sp. is found from tuberculosis patients in Fujian Province, China, using phylogenetic analysis of *Mycobacterium chelonae/abscessus* complex. *Biomedical and Environmental Sciences*, *26*(11), 894–901. doi: 10.1128/JCM.42.12.5957-5960.2004

- Zhang, B., Wu, X., Tai, X., Sun, L., Wu, M., Zhang, W., Chen, X., Zhang, G., Chen, T., Liu, G., & Dyson, P. (2019). Variation in actinobacterial community composition and potential function in different soil ecosystems belonging to the arid Heihe River Basin of northwest China. *Frontiers in Microbiology*, *10*, 2209. <https://doi.org/10.3389/fmicb.2019.02209>
- Zhu, H. Z., Zhang, Z. F., Zhou, N., Jiang, C. Y., Wang, B. J., Cai, L., & Liu, S. J. (2019). Diversity, distribution and co-occurrence patterns of bacterial communities in a karst cave system. *Frontiers in Microbiology*, *10*, 1726. <https://doi.org/10.3389/fmicb.2019.01726>
- Zoropogui, A., Pujic, P., Normand, P., Barbe, V., Belli, P., Graindorge, A., Roche, D., Vallenet, D., Mangenot, S., Boiron, P., Rodriguez-Nava, V., Ribun, S., Richard, Y., Cournoyer, B., & Blaha, D. (2013). The *Nocardia cyriacigeorgica* GUH-2 genome shows ongoing adaptation of an environmental Actinobacteria to a pathogen's lifestyle. *BMC genomics*, *14*(1), 1-18. doi: 10.1186/1471-2164-14-286

Figure Legends

FIGURE 1. (A). Map of Dordogne area indicating the cave locations (Lascaux, Rouffignac, Reille, and Mouflon; white squares – pristine, gray squares – anthropized). (B). Map of Lascaux Cave (entrance Sas-1, Passage, Apse, and Diaclase).

FIGURE 2. (A). Rarefaction curves for 37 common samples of the *hsp65* and 16S rRNA genes at the 97 and 99% OTU cutoffs. The X-axis denotes the number of samples, and the Y-axis denotes the number of OTUs. (B). Pairwise molecular distances between sequences of the *hsp65* (Y axes) and 16S rRNA gene (X-axes) genes among *Actinobacteria* species from different taxonomical levels with the correlation coefficient (R^2) for each equation.

FIGURE 3. Sammon projection of nonmetric multidimensional scaling (NMDS) based on the Bray-Curtis distance matrices of the *hsp65* and 16S rRNA gene markers for different caves (A) and Lascaux Cave locations, and marked/unmarked areas (B). The F and P values of overall AMOVA (97% OTU cutoff) are indicated in Table A7.

FIGURE 4. Significantly different OTUs between pairs of caves (Metastats, $P < 0.005$). For each cave, the number of OTUs that differed proportionally from other caves and taxonomically assigned OTUs proportionally the most abundant in the respective cave are indicated (99% OTU cutoff).

FIGURE 5. Average proportions of *Actinobacteria* families based on *hsp65* (A) and 16S rRNA gene (B) sequence libraries from different caves (Rouffignac, Lascaux, Mouflon, and Reille), Lascaux Cave locations (Sas-1, Passage B, Passage IP, Apse, and Diaclase) and marked (S)/unmarked (U) areas within Lascaux.

FIGURE 6. Co-occurrence networks of *Actinobacteria* OTUs: (A) *hsp65* marker, (B) 16S rRNA gene marker; OTUs differed significantly between marked (black) and unmarked (white) areas in Lascaux Cave, and those that did not differ between these areas (gray) using Metastats ($P < 0.05$). The letters indicate OTUs that were specific for the respective Lascaux Cave locations (S, Sas-1; PB, Passage banks; PI, Passage inclined planes; A, Apse; and D, Diaclase) using Lefse ($P < 0.03$). Strong significant connections (Spearman's correlation > 0.8 and < 0.5 for 16S rRNA gene and < 0.35 for *hsp65*) are displayed (99% OTU cutoff). Red lines indicate negative correlations and black lines indicate positive correlations.

APPENDIX

Appendix Figures

FIGURE A1. Venn diagrams showing the number of different Actinobacteria taxa recovered with *hsp65* (blue) or 16S rRNA gene (red) at different taxonomic ranks. (A) Class, (B) Order, (C) Family, (E) Genus.

FIGURE A2. Indices of alpha diversity (richness: Chao 1, evenness: Simpson evenness, diversity: inverse Simpson) for *Actinobacteria* communities among caves, different locations within Lascaux Cave, and marked/unmarked areas. Significant differences between groups are shown with letters (ANOVA and Tukey post hoc test; $P < 0.05$; 97% OTU cutoff).

FIGURE A3. Venn diagrams showing unique and shared OTUs between (A) different caves and (B) different locations within Lascaux Cave based on *hsp65* and 16S rRNA gene markers (99% OTU cutoff).

FIGURE A4. Taxonomic compositions of the core microbiome between (A) caves and (B) different locations within Lascaux Cave based on the *hsp65* and 16S rRNA gene markers (99% OTU cutoff).

FIGURE A5. Phylogenetic tree showing the taxonomic diversity and allocation of *hsp65* sequences from caves retrieved by a metabarcoding approach. OTUs were used to build this figure. The evolutionary history was inferred using the Neighbor-Joining method (Saitou et al., 1987) and reveals the relationship of sequences from caves obtained in this work with sequences belonging to type and reference strains classified within the *Actinobacteria* class. The tree is based on the Kimura two-parameter method (Kimura, 1980) with the confidence values of the branches determined by bootstrap analyses (Felsenstein, 1985) based on 1000 replicates. Only values $> 80\%$ are shown at the nodes. This analysis involved 116 nucleotide

sequences. The scale bar represents 0.020 substitutions per nucleotide position. Evolutionary analyses were conducted in MEGA 6 (Kumar et al., 2018).

FIGURE A6. Average proportions of species within the genus *Streptomyces* in the *Actinobacteria hsp65* amplicon sequence libraries significantly different in marked (S) and unmarked areas (U) of Lascaux Cave (Metastats, $P < 0.05$).

FIGURE A7. Average proportions of genus *Mycobacterium* in the *Actinobacteria hsp65* gene amplicon sequence libraries in different caves (Rouffignac, Lascaux, Mouflon, and Reille) and different locations within Lascaux Cave (Sas-1, Passage-B, Passage-IP, Apse, and Diaclase).

FIGURE A8. Average proportions of genus *Nocardia* in the *Actinobacteria hsp65* gene amplicon sequence libraries in different caves (Rouffignac, Lascaux, Mouflon, and Reille) and different locations within Lascaux Cave (Sas-1, Passage-B, Passage-IP, Apse, and Diaclase).

Appendix Tables

Table A1. Location of cave samples and list of gene markers sequenced using Illumina MiSeq.

Table A2. Genomes of *Actinobacteria* reference strains from the orders *Streptomyetales* and *Corynebacteriales* retrieved from the Integrated Microbial Genome and Microbiome database and used for *in silico* analyses: comparison of the *hsp65* and 16S rRNA gene partial sequences.

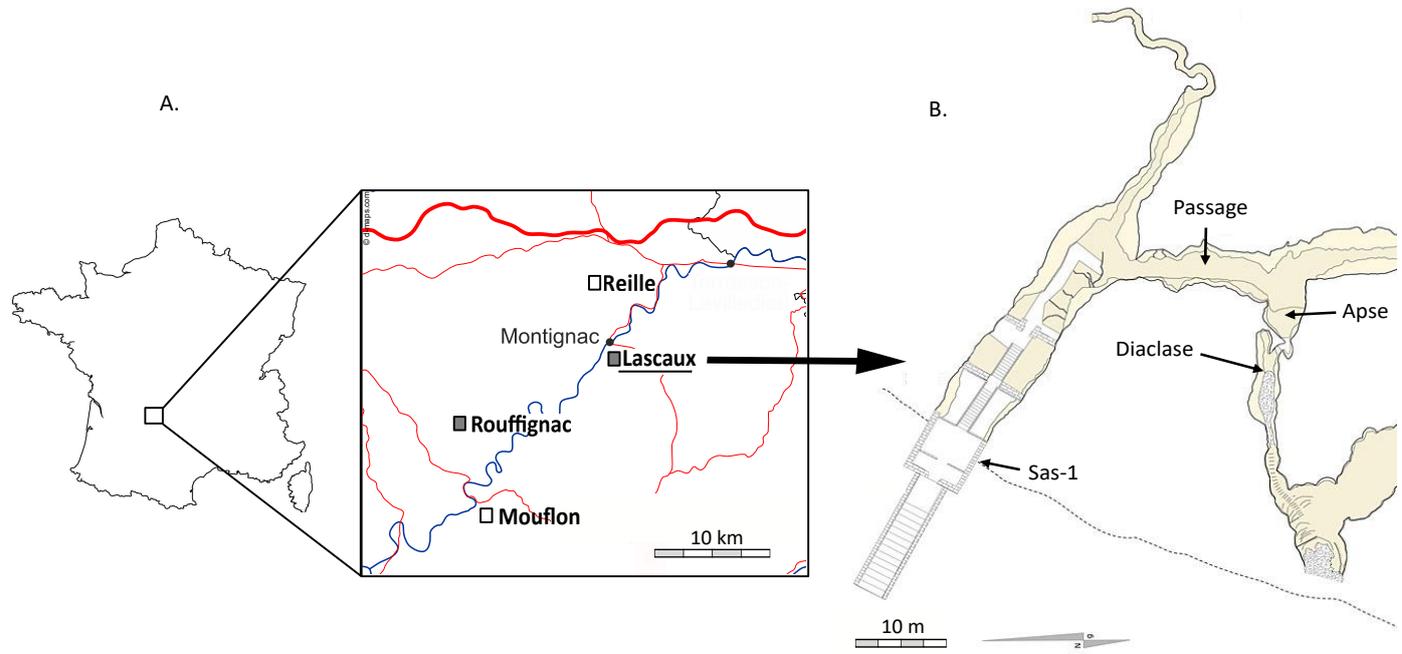
Table A3. Percentage complementarity between the *hsp65*-specific forward (TB11) and reverse (TB12) primers and the *hsp65* sequence database (*hsp65* sequences and GROESL sequences, paralogs to *hsp65*) with respect to the number of mismatches.

Table A4. Percent coverage of primers to the *Actinobacteria* RefSeq representative genomes from NCBI of different classes and orders. Only matches with no mismatches on the last three nucleotides of the 3' end and no more than four overall mismatches were included.

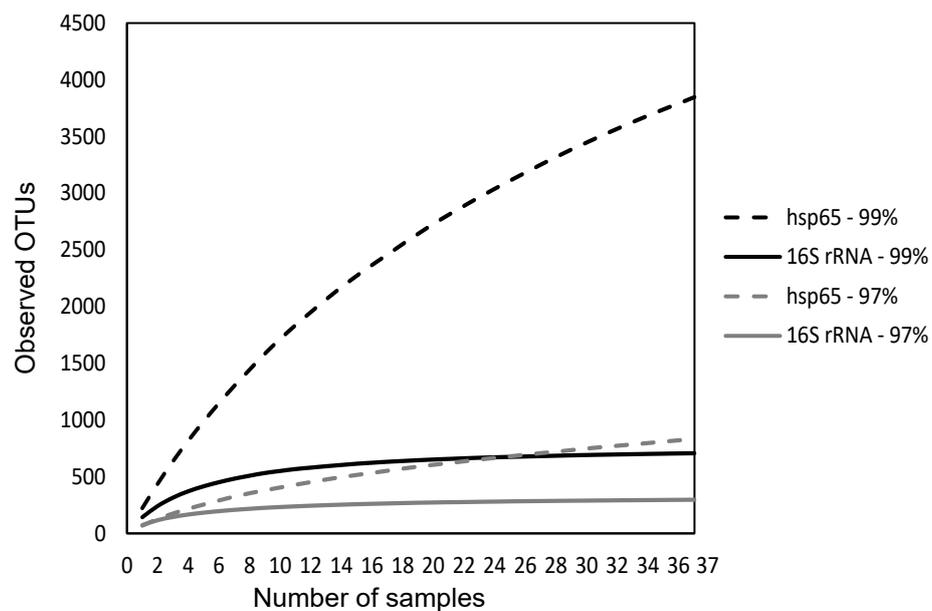
Table A5. The number of *Actinobacteria* species identified by the *hsp65* marker for each respective genus.

Table A6. T and F tests to compare means and variation of diversity indices for 37 common samples of *hsp65* and 16S rRNA gene markers at 97 and 99% OTUs cutoffs. For each marker, averages and standard deviations are indicated.

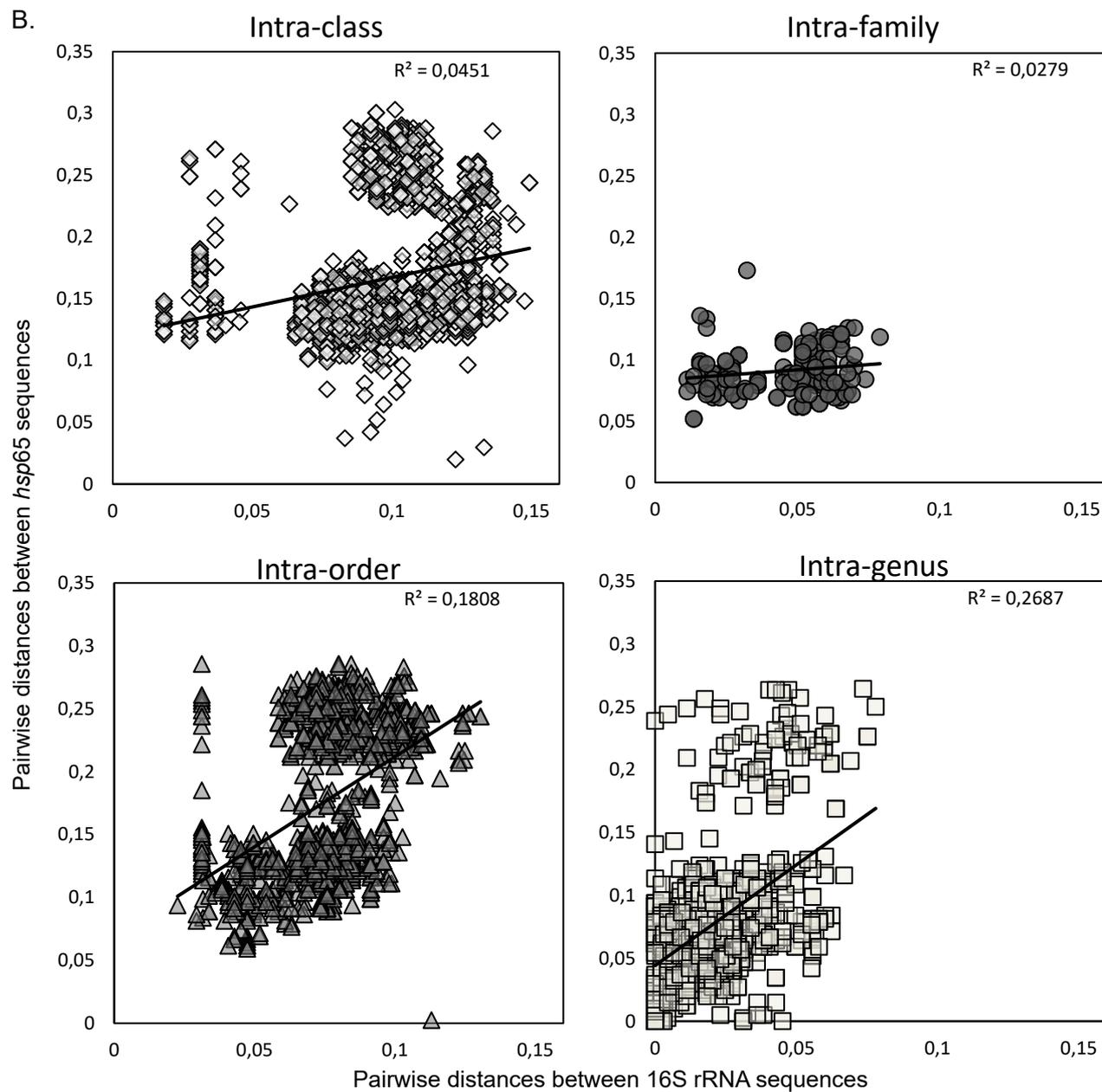
Table A7. Differences between *Actinobacteria* communities between (A) caves, (B) different locations within Lascaux Cave, and (C) marked (S) and unmarked (U) areas calculated by analysis of molecular variance (AMOVA) and homogeneity of molecular variance (HOMOVA) for the *hsp65* and 16S rRNA gene markers (97% OTUs cutoff).

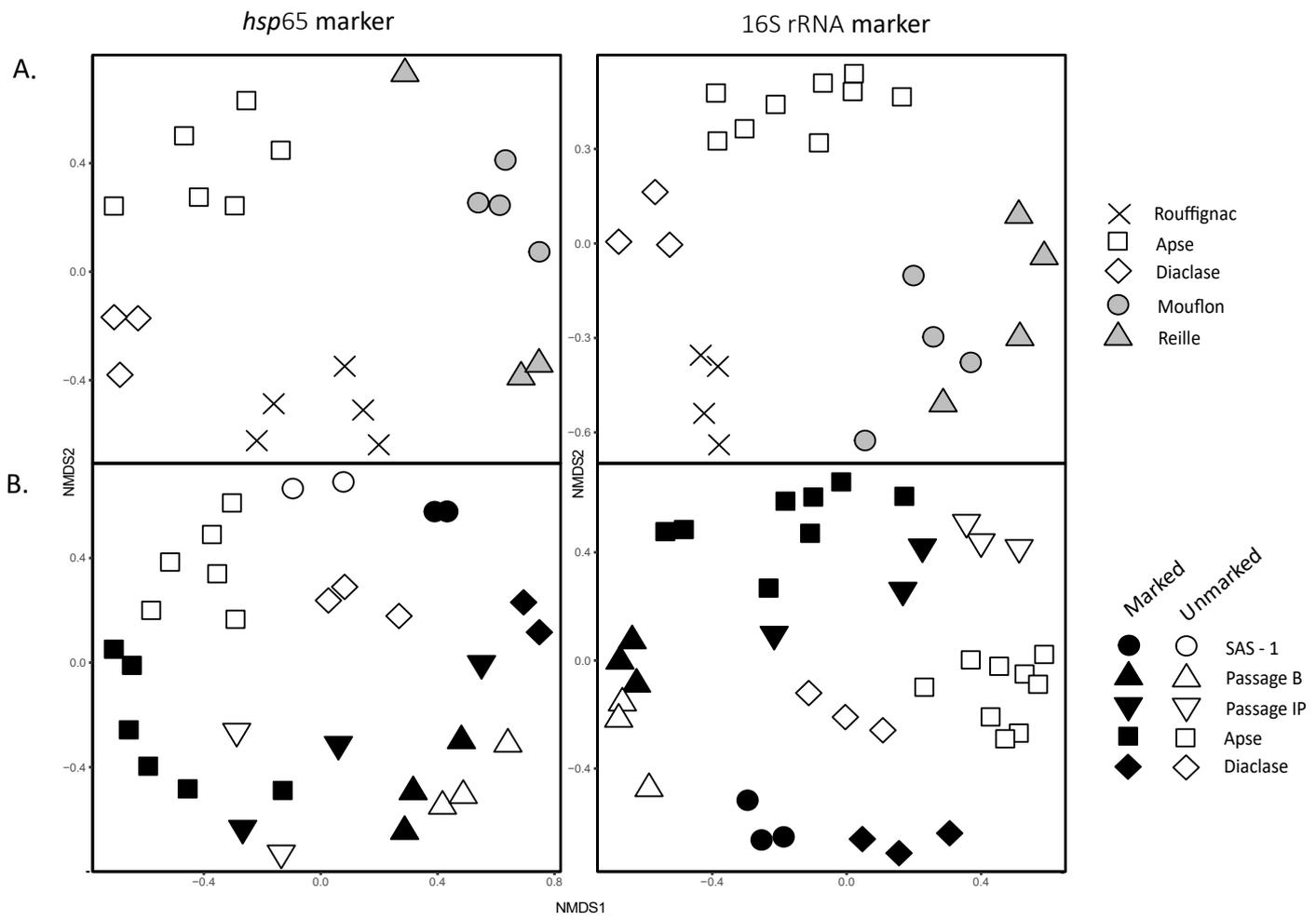


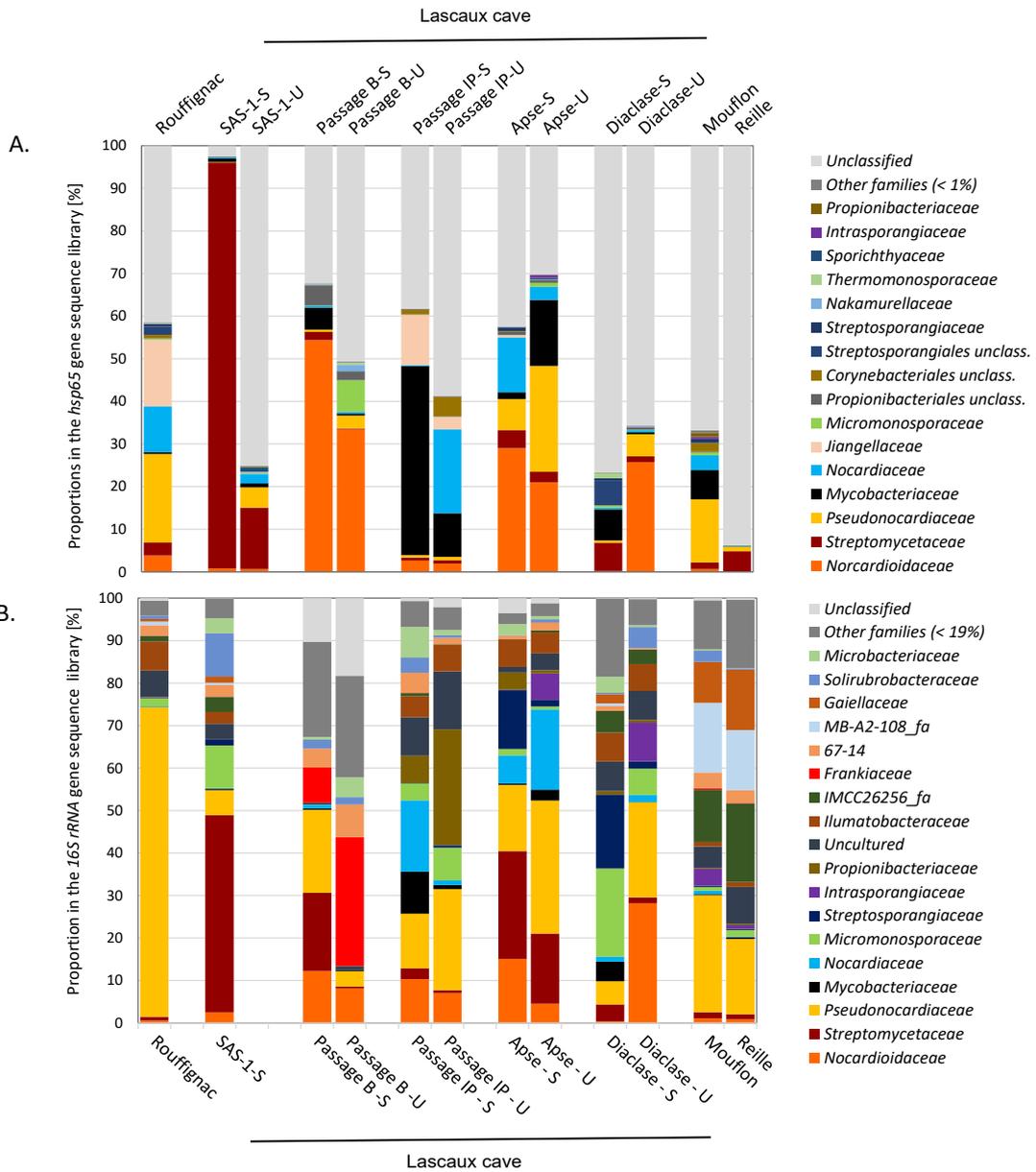
A.



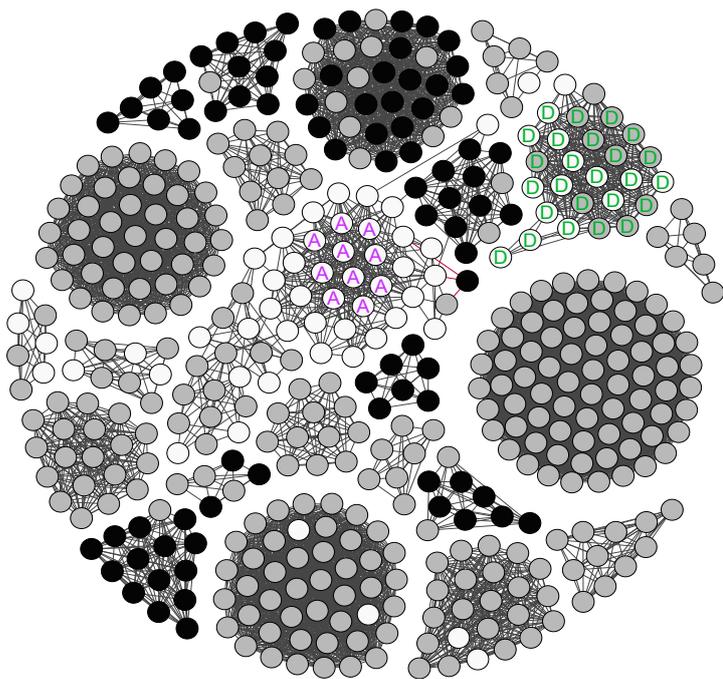
B.



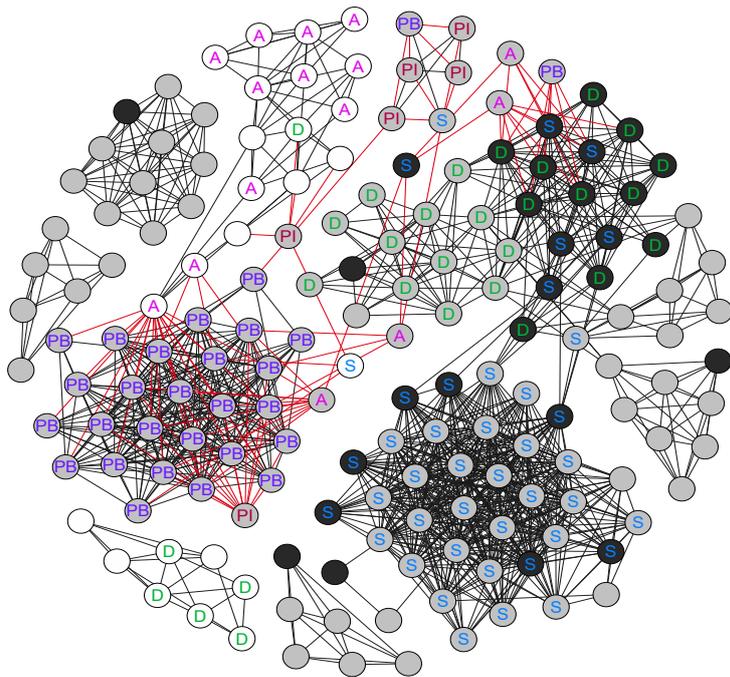


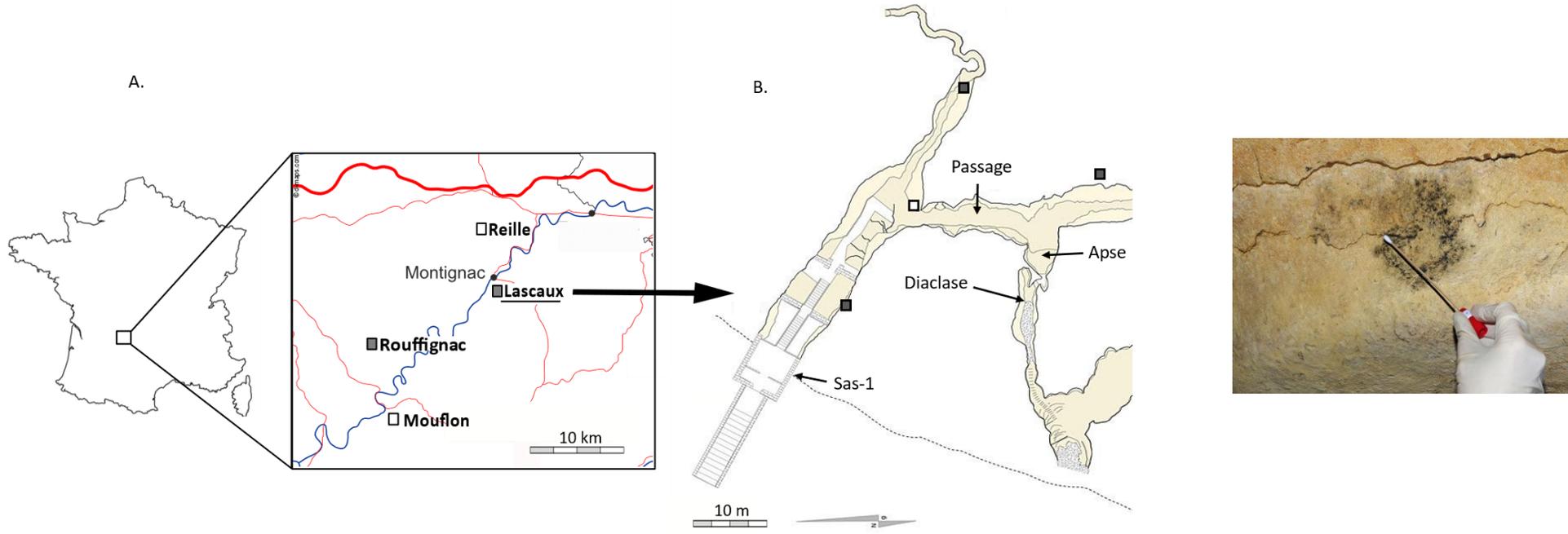


hsp65 marker

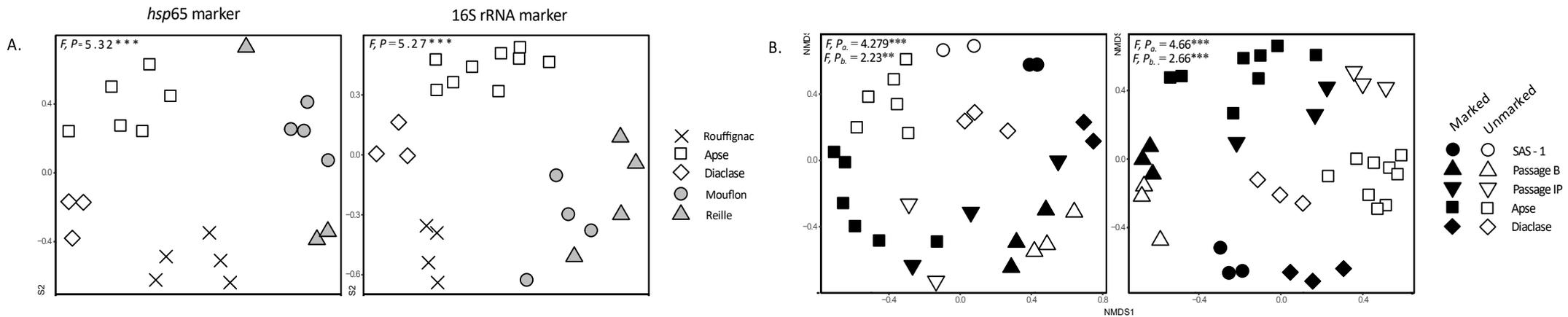


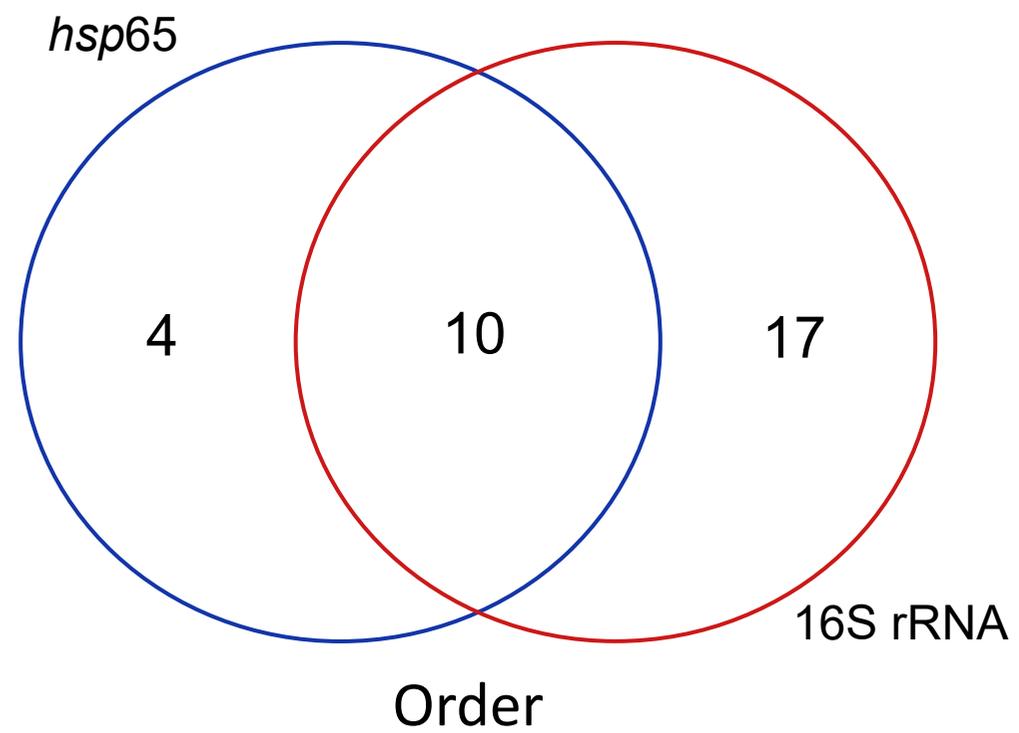
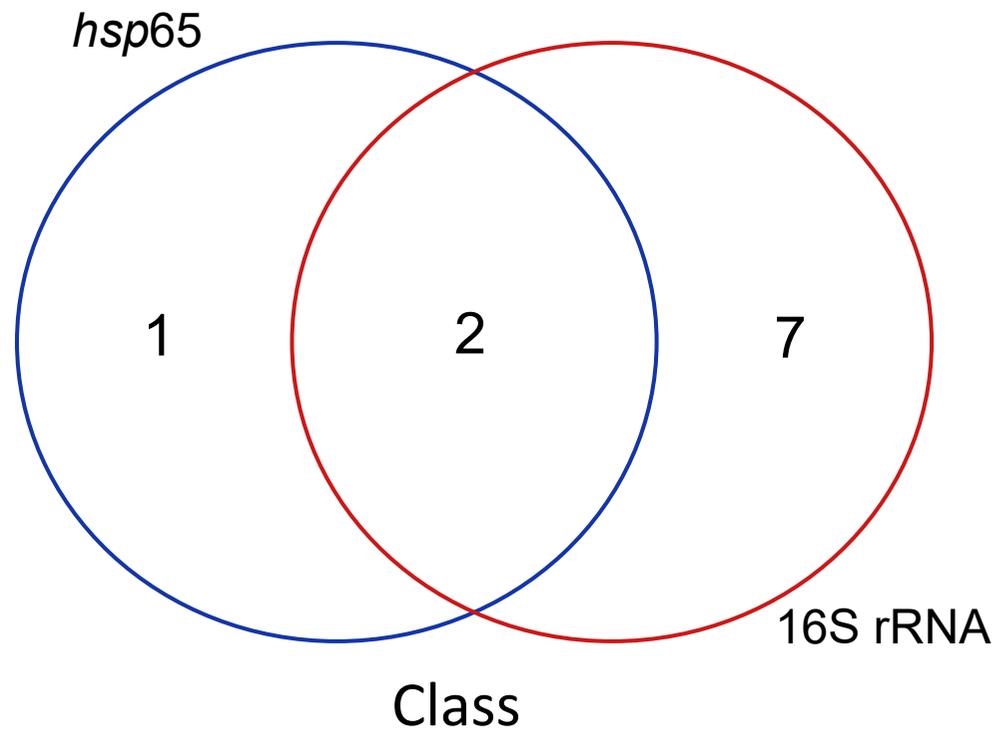
16S rRNA marker



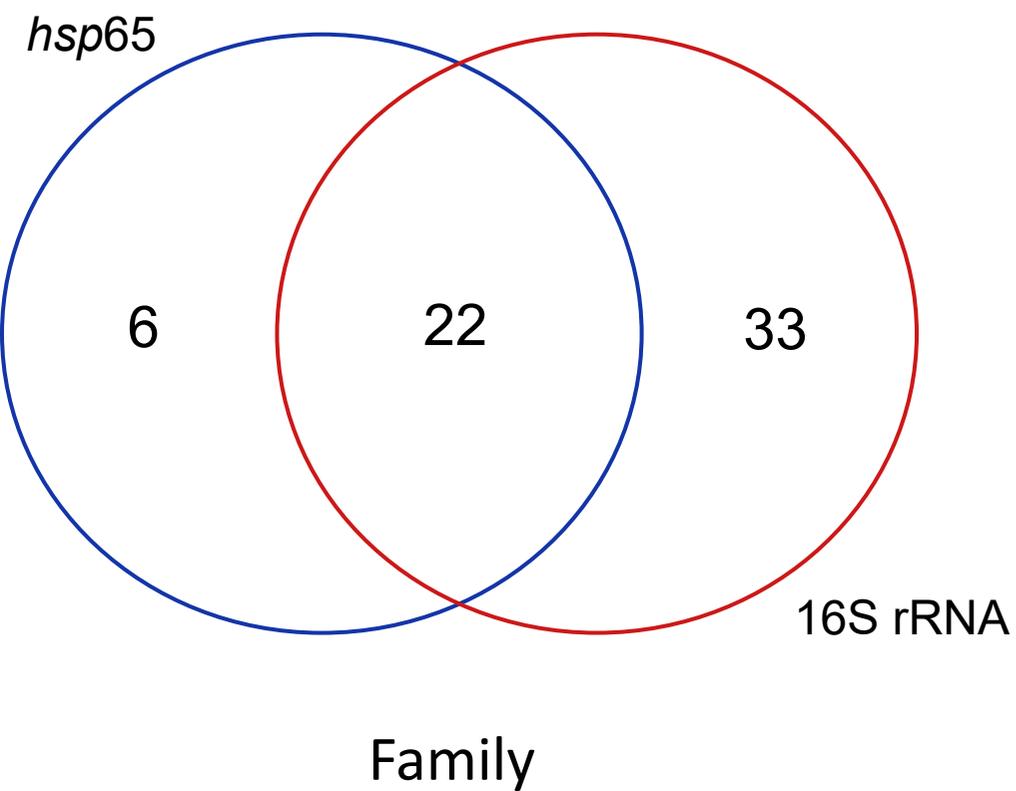


Actinobacteria communities

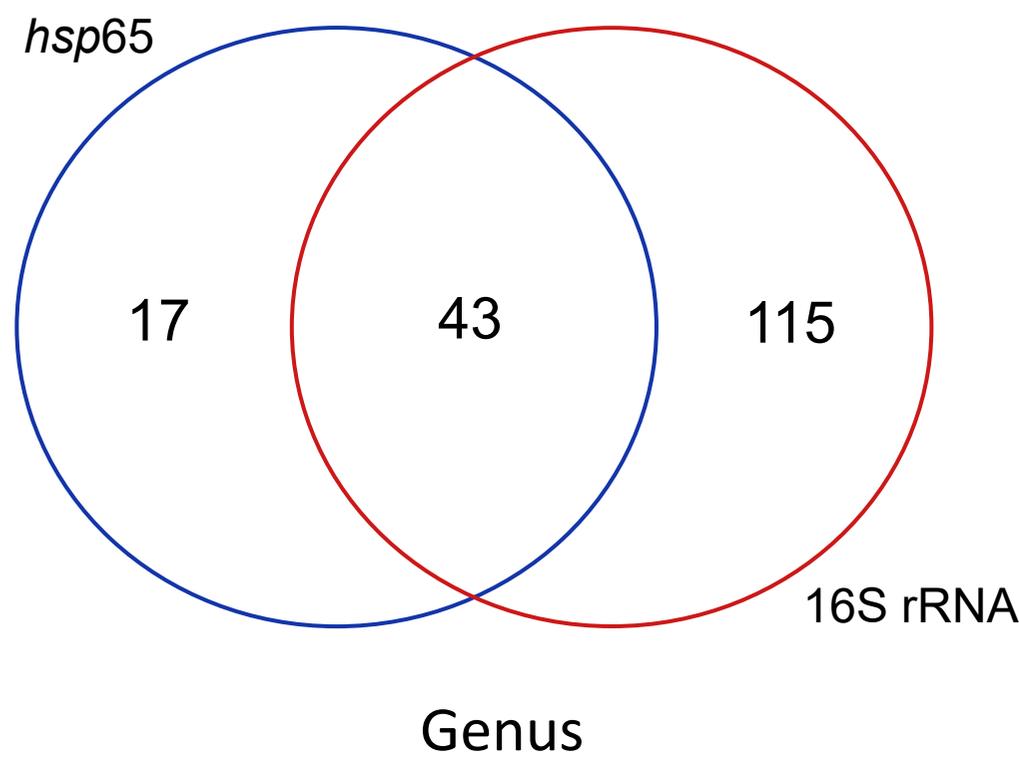


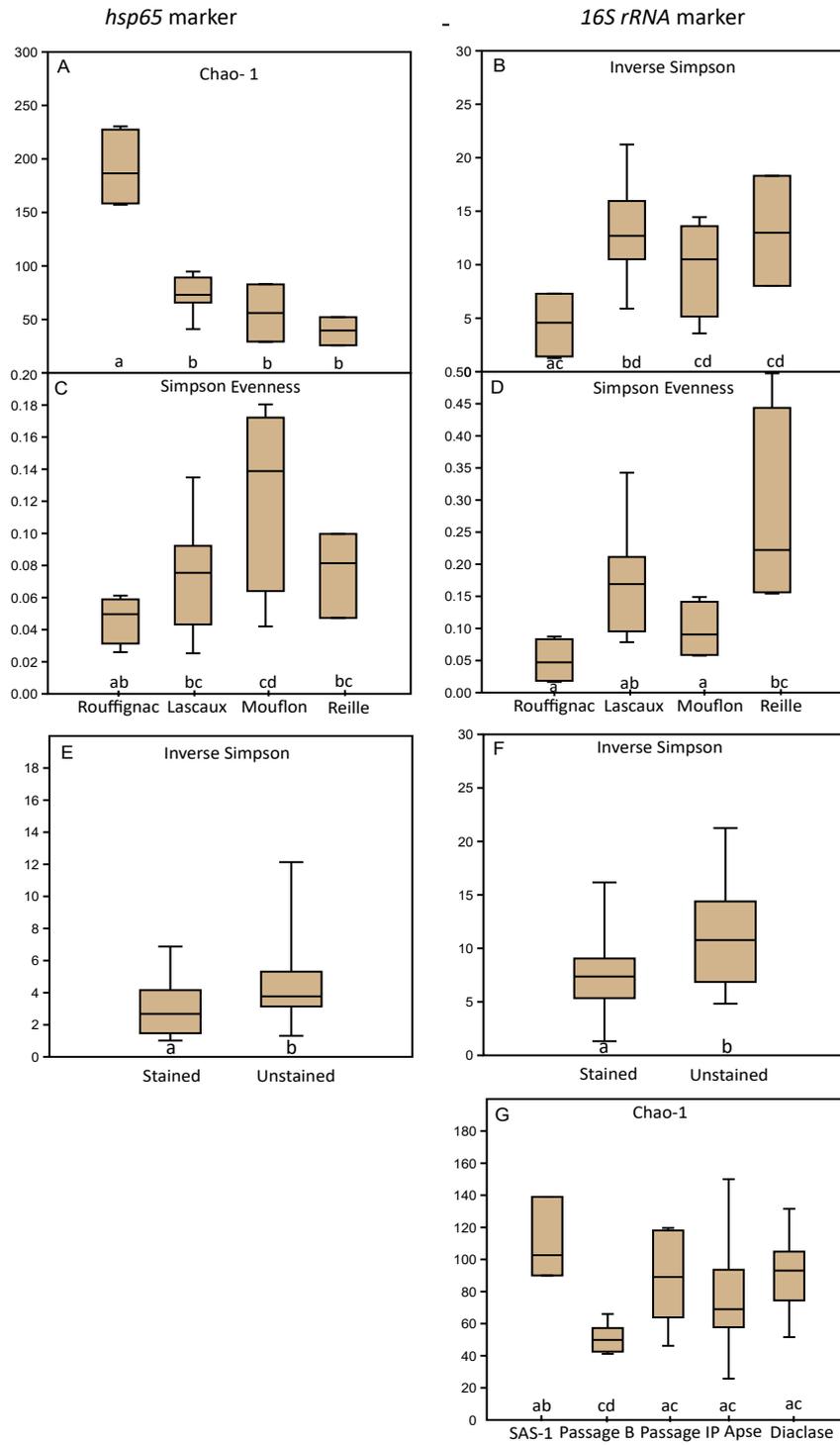


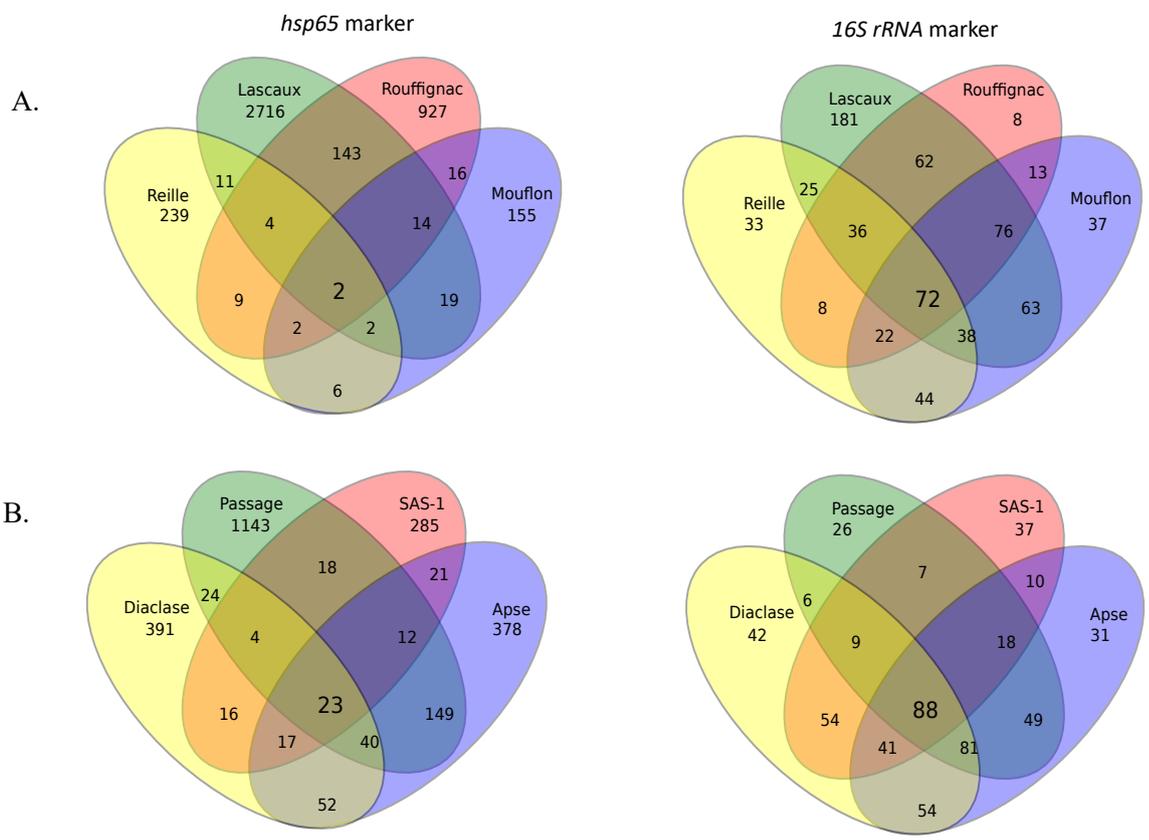
C.



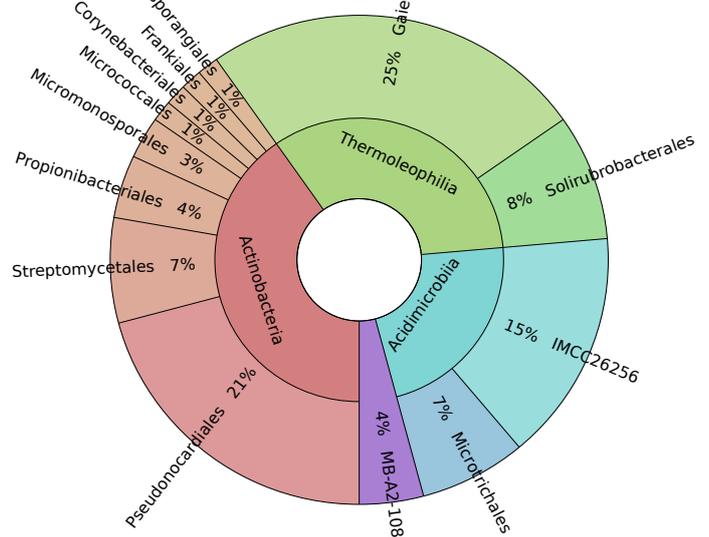
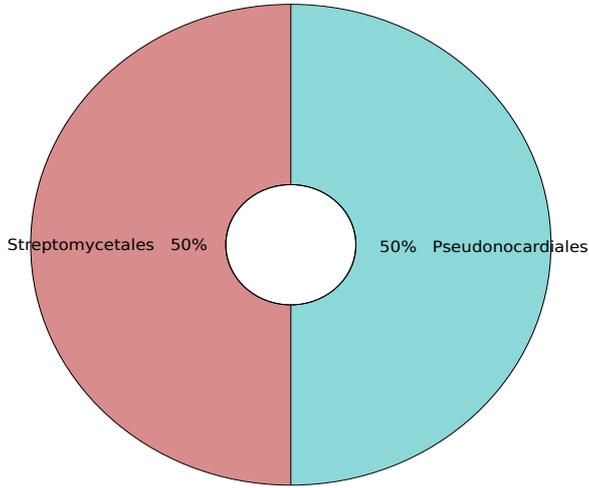
D.



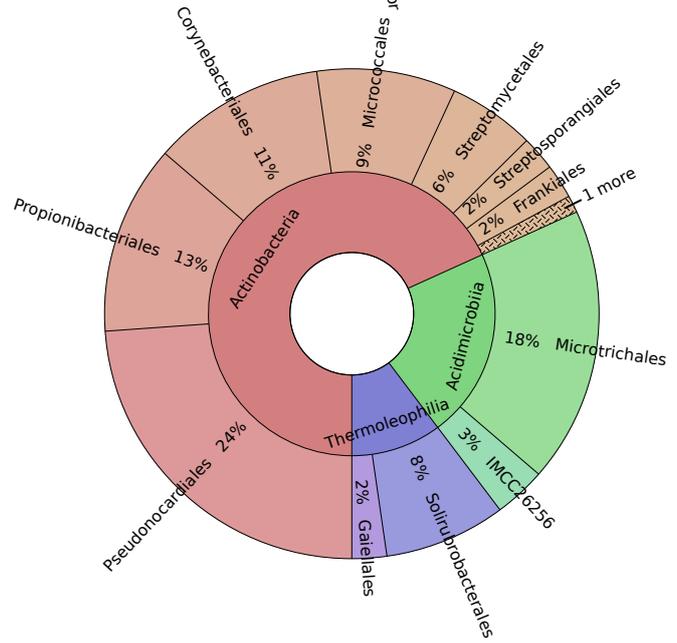
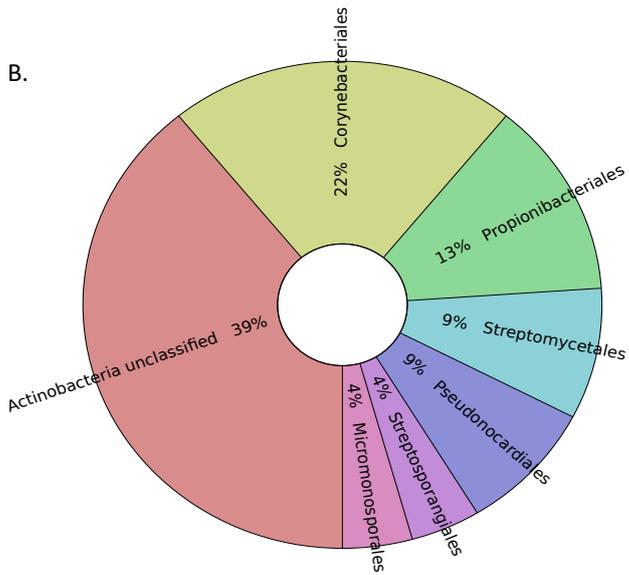


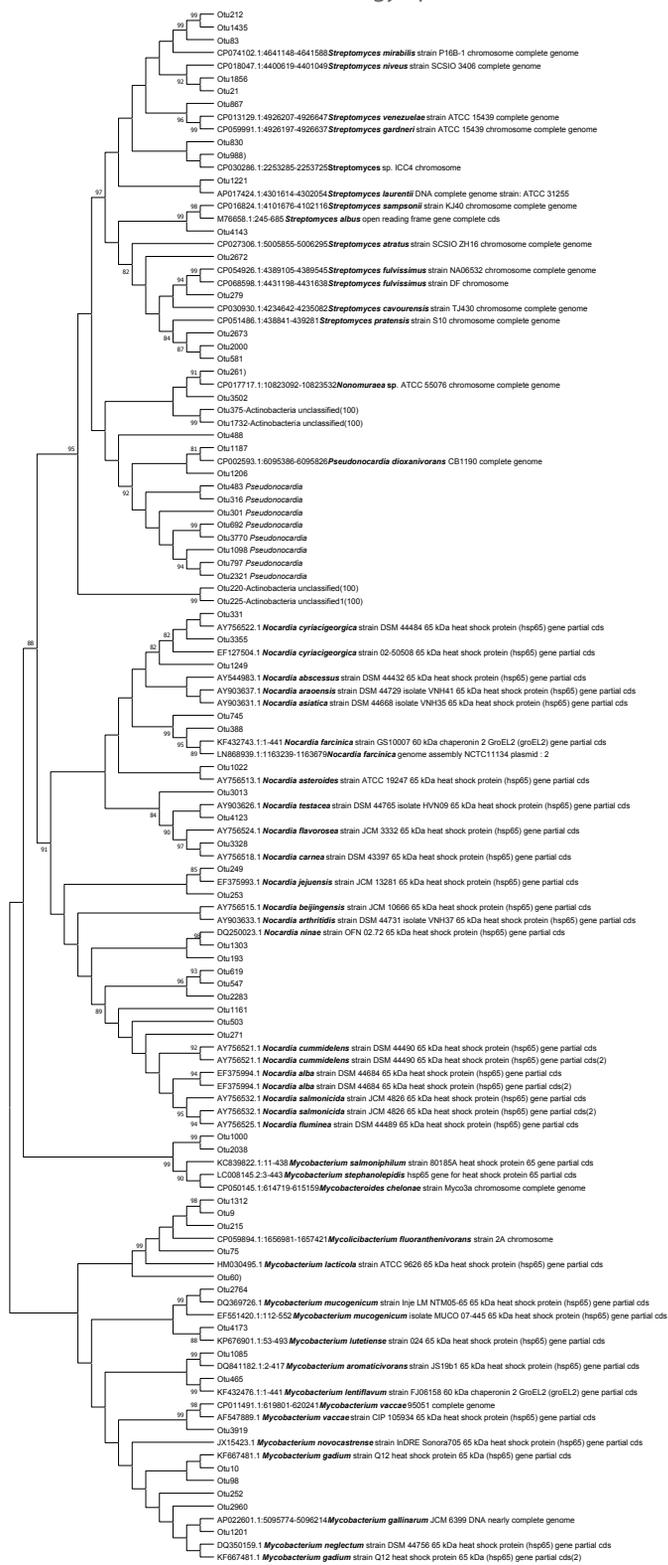


A.



B.





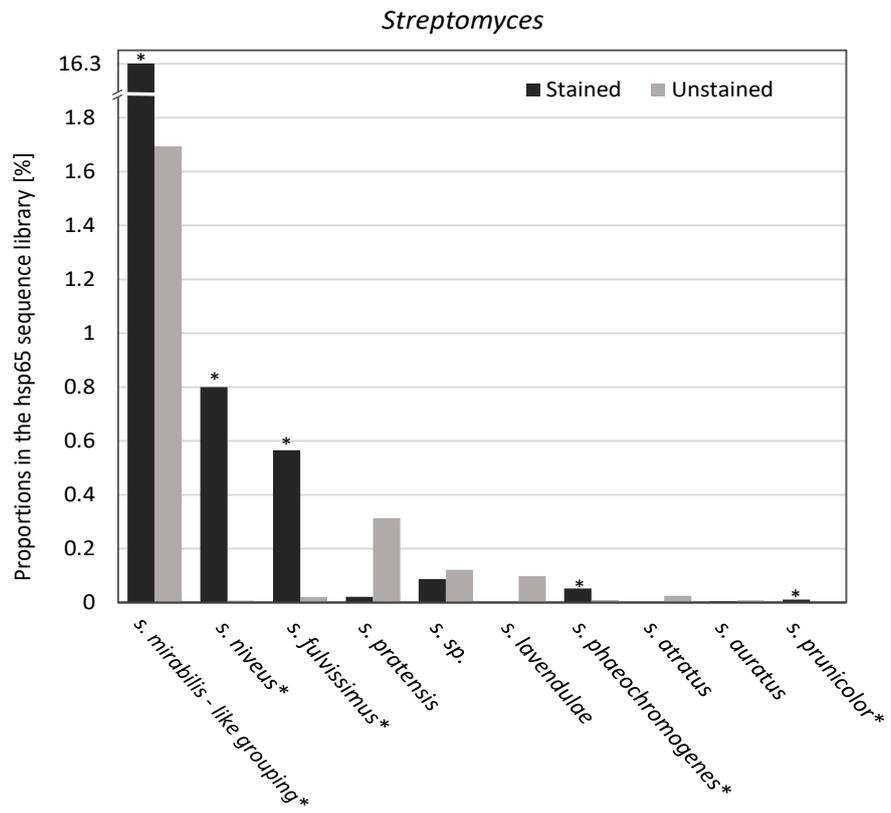


Table A1. Location of cave samples and list of gene markers sequenced using Illumina MiSeq

Sample ID	Sample Number	Cave name	Room	Area	Obtained sequences
Apse-S1	La568	Lascaux	Apse		16S rRNA gene
Apse-S2	La572	Lascaux	Apse	Visual mark	<i>hsp65</i> , 16S rRNA gene
Apse-S3	La574	Lascaux	Apse		<i>hsp65</i> , 16S rRNA gene
Apse-U1	La479	Lascaux	Apse		<i>hsp65</i> , 16S rRNA gene
Apse-U2	La480	Lascaux	Apse	Unmarked area	<i>hsp65</i> , 16S rRNA gene
Apse-U3	La481	Lascaux	Apse		<i>hsp65</i> , 16S rRNA gene
Apse-DZ1	La464	Lascaux	Apse		16S rRNA gene
Apse-DZ2	La466	Lascaux	Apse		<i>hsp65</i>
Apse-DZ3	La468	Lascaux	Apse	Visual mark	<i>hsp65</i> , 16S rRNA gene
Apse-DZ4	La470	Lascaux	Apse		16S rRNA gene
Apse-DZ5	La472	Lascaux	Apse		<i>hsp65</i> , 16S rRNA gene
Apse-DZ6	La474	Lascaux	Apse		<i>hsp65</i> , 16S rRNA gene
Apse-UDZ1	La583	Lascaux	Apse		16S rRNA gene
Apse-UDZ2	La584	Lascaux	Apse		16S rRNA gene
Apse-UDZ3	La585	Lascaux	Apse	Unmarked area	<i>hsp65</i> , 16S rRNA gene
Apse-UDZ4	La589	Lascaux	Apse		<i>hsp65</i> , 16S rRNA gene
Apse-UDZ5	La590	Lascaux	Apse		16S rRNA gene
Apse-UDZ6	La591	Lascaux	Apse		<i>hsp65</i> , 16S rRNA gene
Dia-S1	La440	Lascaux	Diaclase		<i>hsp65</i> , 16S rRNA gene
Dia-S2	La443	Lascaux	Diaclase	Visual mark	<i>hsp65</i> , 16S rRNA gene
Dia-S3	La444	Lascaux	Diaclase		16S rRNA gene
Dia-U1	La433	Lascaux	Diaclase		<i>hsp65</i> , 16S rRNA gene
Dia-U2	La434	Lascaux	Diaclase	Unmarked area	<i>hsp65</i> , 16S rRNA gene
Dia-U3	La435	Lascaux	Diaclase		<i>hsp65</i> , 16S rRNA gene
Passage B-S1	La457	Lascaux	Passage banks		<i>hsp65</i> , 16S rRNA gene
Passage B-S2	La458	Lascaux	Passage banks	Visual mark	<i>hsp65</i> , 16S rRNA gene
Passage B-S3	La459	Lascaux	Passage banks		<i>hsp65</i> , 16S rRNA gene
Passage B-U1	La445	Lascaux	Passage banks		<i>hsp65</i> , 16S rRNA gene
Passage B-U2	La446	Lascaux	Passage banks	Unmarked area	<i>hsp65</i> , 16S rRNA gene
Passage B-U3	La447	Lascaux	Passage banks		<i>hsp65</i> , 16S rRNA gene
Passage IP-S1	La555	Lascaux	Passage inclined planes		<i>hsp65</i> , 16S rRNA gene
Passage IP-S2	La556	Lascaux	Passage inclined planes	Visual mark	<i>hsp65</i> , 16S rRNA gene
Passage IP-S3	La558	Lascaux	Passage inclined planes		<i>hsp65</i> , 16S rRNA gene
Passage IP-U1	La549	Lascaux	Passage inclined planes		<i>hsp65</i> , 16S rRNA gene
Passage IP-U2	La550	Lascaux	Passage inclined planes	Unmarked area	<i>hsp65</i> , 16S rRNA gene
Passage IP-U3	La551	Lascaux	Passage inclined planes		16S rRNA gene
SAS-1-S1	La530	Lascaux	Sas-1 (airlock-1 entrance zone, compartment 2)		<i>hsp65</i> , 16S rRNA gene
SAS-1-S2	La531	Lascaux	Sas-1 (airlock-1 entrance zone, compartment 2)	Visual mark	<i>hsp65</i> , 16S rRNA gene
SAS-1-S3	La532	Lascaux	Sas-1 (airlock-1 entrance zone, compartment 2)		<i>hsp65</i> , 16S rRNA gene
SAS-U1	La542	Lascaux	Sas-1 (airlock-1 entrance zone, compartment 2)		<i>hsp65</i>
SAS-U3	La544	Lascaux	Sas-1 (airlock-1 entrance zone, compartment 2)	Unmarked area	<i>hsp65</i>
Rouf1	Ro10	Rouffignac			<i>hsp65</i> , 16S rRNA gene
Rouf2	Ro11	Rouffignac			<i>hsp65</i>
Rouf3	Ro12	Rouffignac	x	x	<i>hsp65</i>
Rouf4	Ro13	Rouffignac			<i>hsp65</i> , 16S rRNA gene
Rouf5	Ro14	Rouffignac			<i>hsp65</i> , 16S rRNA gene
Rouf6	Ro15	Rouffignac			16S rRNA gene
Mouf1	Mf20	Mouflon			16S rRNA gene
Mouf2	Mf22	Mouflon			<i>hsp65</i> , 16S rRNA gene
Mouf3	Mf23	Mouflon	x	x	<i>hsp65</i>
Mouf4	Mf24	Mouflon			<i>hsp65</i> , 16S rRNA gene
Mouf5	Mf25	Mouflon			<i>hsp65</i>
Mouf6	Mf26	Mouflon			16S rRNA gene
Reil1	Re10	Reille			<i>hsp65</i>
Reil2	Re11	Reille			16S rRNA gene
Reil3	Re13	Reille	x	x	<i>hsp65</i> , 16S rRNA gene
Reil4	Re9	Reille			<i>hsp65</i> , 16S rRNA gene
Reil5	Re14	Reille			16S rRNA gene

Table A3. Percentage of complementarity between the *hsp 65*-specific forward (TB11) and reverse (TB12) primers and the *hsp 65* sequence database (*hsp 65* sequences and GROESL sequences, the paralogs to *hsp 65*) with respect to the number of mismatches.

	Number of mismatches	Match to <i>hsp65</i> [%]	Match to GROESL [%]
Forward primer	0	30,8	0
	1	74,8	0
	2	86	4,2
	3	98,1	56,2
	4	99,1	85,4
	5	100	100
Reverse primer	0	6,5	0
	1	44,9	0
	2	89,7	52,1
	3	96,3	64,6
	4	100	81,2
	5	100	89,6

Table A4. Percent coverage of the primers to the *Actinobacteria* RefSeq representative genomes from NCBI of different classes and orders. Only targets with no mismatches on the last three nucleotides of the 3' end and no more than 4 overall mismatches were included.

Class	Order	Numb. of RefSeq representative genomes	Percentual primer matches
Acidimicrobiia		8	38
Coriobacteriia		63	10
Nitriliruptoria		5	0
Rubrobacteria		7	43
Thermoleophilia		7	14
Actinobacteria			
	Acidothermales	1	0
	Actinomycetales	75	27
	Actinopolysporales	8	100
	Bifidobacteriales	87	68
	Catenulisporales	3	100
	Corynebacteriales	473	83
	Frankiales	18	94
	Geodermatophilales	34	100
	Glycomycetales	13	77
	Kineosporiales	5	80
	Micrococcales	482	53
	Micromonosporales	106	100
	Nakamurellales	5	20
	Propionibacteriales	120	63
	Pseudonocardiales	151	74
	Streptomycetales	348	95
	Streptosporangiales	112	88

Table A5. Number of identified Actinobacteria species by hsp65 marker for respective genus

Family	Genus (number of recovered species)			
<i>Nocardiaceae</i>	<i>Nocardia</i> (24)	<i>Rhodococcus</i> (10)		
<i>Streptomycetaceae</i>	<i>Streptomyces</i> (30)	<i>Kitasatospora</i> (1)		
<i>Mycobacteriaceae</i>	<i>Mycobacterium</i> (26)			
<i>Pseudonocardiaceae</i>	<i>Amycolatopsis</i> (7)	<i>Pseudonocardia</i> (4)	<i>Lentzea</i> (2)	<i>Actinomycetospora</i> (1)
	<i>Kibdelosporangium</i> (1)	<i>Thermocrispum</i> (1)	<i>Saccharothrix</i> (1)	<i>Alloactinosynnema</i> (1)
	<i>Sciscionella</i> (1)	<i>Actinoalloteichus</i> (1)	<i>Actinokineospora</i> (1)	
<i>Norcardioidaceae</i>	<i>Nocardioides</i> (6)	<i>Kribbella</i> (2)	<i>Pimelobacter</i> (1)	
	<i>Marmoricola</i> (1)	<i>Aeromicrobium</i> (1)		
<i>Streptosporangiaceae</i>	<i>Microbispora</i> (1)	<i>Microtetraspora</i> (1)	<i>Nonomuraea</i> (1)	<i>Planomonospora</i> (1)
	<i>Sinosporangium</i> (1)	<i>Streptosporangium</i> (1)		
<i>Micromonosporaceae</i>	<i>Micromonospora</i> (3)	<i>Longispora</i> (1)	<i>Catelliglobospora</i> (1)	<i>Plantactinospora</i> (1)
<i>Intrasporangiaceae</i>	<i>Janibacter</i> (3)	<i>Knoellia</i> (1)		
<i>Actinomycetaceae</i>	<i>Actinomyces</i> (3)			
<i>Jiangellaceae</i>	<i>Jaingella</i> (3)			
<i>Microbacteriaceae</i>	<i>Agromyces</i> (1)	<i>Microcella</i> (1)		
<i>Propionibacteriaceae</i>	<i>Cutibacterium</i> (1)	<i>Propionibacterium</i> (1)		
<i>Geodermatophilaceae</i>	<i>Geodermatophilus</i> (1)	<i>Modestobacter</i> (1)		
<i>Micrococcaceae</i>	<i>Micrococcus</i> (1)	<i>Rothia</i> (1)		
<i>Dermabacteraceae</i>	<i>Brachybacterium</i> (2)			
<i>Brevibacteriaceae</i>	<i>Brevibacterium</i> (2)			
<i>Corynebacteriaceae</i>	<i>Corynebacterium</i> (2)			
<i>Catenulisporaceae</i>	<i>Catenulispora</i> (1)			
<i>Dermacoccaceae</i>	<i>Kytococcus</i> (1)			
<i>Actinobacteria_incertae_sedis</i>	<i>Thermobispora</i> (1)			
<i>Nakamurellaceae</i>	<i>Nakamurella</i> (1)			
<i>Sporichthyaceae</i>	<i>Sporichthya</i> (1)			
<i>Nocardiopsaceae</i>	<i>Thermobifida</i> (1)			

Table A6. *t* test and F test for comparison of means and variation of diversity indices for 37 common samples of *hsp 65* and 16S rRNA markers at 97% and 99% OTUs cutoffs. For each marker averages and standard deviations are indicated.

	<i>hsp 65</i>	16S rRNA	t test		F test		
			t	P-value	F	P-value	
99% OTU							
Chao-1	260.78 ± 137.05	182.17 ± 82.02	2.98	0.003	2.75	0.017	
Simpson evenness	0.29 ± 0.25	0.089 ± 0.05	4.72	<0.001	26.81	<0.001	
Inverse Simpson	6.57 ± 4.97	14.07 ± 6.29	5.66	<0.001	1.6	0.153	
97% OTU							
Chao-1	75.74 ± 42.26	89.12 ± 36.35	1.45	0.15	1.32	0.588	
Simpson evenness	0.38 ± 0.26	0.15 ± 0.11	4.91	<0.001	6.28	0.002	±
Inverse Simpson	4.09 ± 2.74	9.11 ± 4.44	5.82	<0.001	2.64	0.013	

Table A7. Differences between Actinobacteria communities between A. caves, B. different locations within Lascaux Cave, and marked and unmarked areas calculated by analysis of molecular variance (AMOVA) and homogeneity of molecular variance (HOMOVA) for the hsp65 and 16S rRNA gene markers (97% OTUs cutoff).

A. AMOVA											
<i>hsp 65</i>											
	Cave			Lascaux cave location			Surface area				
	Lascaux	Rouffignac	Reille	Sas-1	Passage B	Passage IP	Apse	Marked	Unmarked	Df	F, P
Rouffignac	13	5.72***		10	7.858**					32	2.23**
Reille	11	3.614**	7	9	3.978	10	4.419**				
Moufflon	12	6.265**	8	16	4.505***	17	5.903***	16	2.572***		
				9	4.518	10	5.635**	9	2.786	16	3.193***
Overall	20	5.317***		32	4.279***						
16S rRNA											
	Lascaux	Rouffignac	Reille	Sas-1	Passage B	Passage IP	Apse	Marked	Unmarked	Df	F, P
Rouffignac	15	5.79***		8	10.063					37	2.66***
Reille	15	5.43***	7	8	4.863	11	9.013**				
Moufflon	15	5.36**	7	19	4.09***	22	8.115***	22	2.788**		
				8	2.723	11	5.79**	11	2.902**	22	3.005***
Overall	23	5.274***		37	4.656***						
Bonferonmi pair-wise err. rate: 0.008											
0.008 > * P > 0.005 > ** P > 0.001 > P***											
Bonferonmi pair-wise err. rate: 0.005											
0.005 > * P > 0.0025 > ** P > 0.001 > P***											
Exp.-wise err. rate: 0.05											
0.05 > * P > 0.01 > ** P > 0.001 > P											
B. HOMOVA											
<i>hsp 65</i>											
	Cave			Lascaux cave location			Surface area				
	Rouffignac	Lascaux	Reille	Sas-1	Passage B	Passage IP	Apse	Marked	Unmarked	Df	F, p
Rouffignac (0.193)											
Lascaux (0.303)	0.235*			0.006							
Reille (0.342)	0.193	0.010		0.246	0.367						
Moufflon (0.160)	0.026	0.359	0.295	0.270*	0.433*	0.005					
				0.126	0.206	0.020	0.010				
Overall	0.573*			Overall	0.693*						
16S rRNA											
	Rouffignac	Lascaux	Reille	Sas-1	Passage B	Passage IP	Apse	Marked	Unmarked	Df	F, p
Rouffignac (0.145)											
Lascaux (0.276)	0.380			0.012							
Reille (0.23)	0.133	0.034		0.166	0.531*						
Moufflon (0.199)	0.062	0.105	0.014	0.364	1.227*	0.079					
				0.415	1.166*	0.134*	0.031				
Overall	0.432*			Overall	1.689*						
Bonferonmi pair-wise err. rate: 0.008 > P*											
Bonferonmi pair-wise err. rate: 0.005 > P*											
Exp.-wise err. rate: 0.05 > P*											