



HAL
open science

Stochastic and deterministic assembly processes of microbial communities in relation to natural attenuation of black stains in Lascaux Cave

Zélia Bontemps, Yvan Moëgne-Loccoz, Mylène Hugoni

► **To cite this version:**

Zélia Bontemps, Yvan Moëgne-Loccoz, Mylène Hugoni. Stochastic and deterministic assembly processes of microbial communities in relation to natural attenuation of black stains in Lascaux Cave. *mSystems*, 2024, 9 (2), 10.1128/msystems.01233-23 . hal-04523218

HAL Id: hal-04523218

<https://hal.inrae.fr/hal-04523218>

Submitted on 27 Mar 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Stochastic and deterministic assembly processes of microbial communities in relation to natural attenuation of black stains in Lascaux Cave

Zélia Bontemps,^{1,2} Yvan Moëgne-Loccoz,¹ Mylène Hugoni^{1,3,4}

AUTHOR AFFILIATIONS See affiliation list on p. 15.

ABSTRACT Community assembly processes are complex and understanding them represents a challenge in microbial ecology. Here, we used Lascaux Cave as a stable, confined environment to quantify the importance of stochastic vs deterministic processes during microbial community dynamics across the three domains of life in relation to an anthropogenic disturbance that had resulted in the side-by-side occurrence of a resistant community (unstained limestone), an impacted community (present in black stains), and a resilient community (attenuated stains). Metabarcoding data showed that the microbial communities of attenuated stains, black stains, and unstained surfaces differed, with attenuated stains being in an intermediate position. We found four scenarios to explain community response to disturbance in stable conditions for the three domains of life. Specifically, we proposed the existence of a fourth, not-documented yet scenario that concerns the always-rare microbial taxa, where stochastic processes predominate even after disturbance but are replaced by deterministic processes during post-disturbance recovery. This suggests a major role of always-rare taxa in resilience, perhaps because they might provide key functions required for ecosystem recovery.

IMPORTANCE The importance of stochastic vs deterministic processes in cave microbial ecology has been a neglected topic so far, and this work provided an opportunity to do so in a context related to the dynamics of black-stain alterations in Lascaux, a UNESCO Paleolithic cave. Of particular significance was the discovery of a novel scenario for always-rare microbial taxa in relation to disturbance, in which stochastic processes are replaced later by deterministic processes during post-disturbance recovery, i.e., during attenuation of black stains.

KEYWORDS community assembly processes, disturbance, rare biosphere, microbial ecology

The relevance of community assembly rules for predicting and modeling how communities face different environmental scenarios, based on events controlling the presence and abundance of species in a community, is a key issue in ecology. According to the niche-based theory, deterministic processes linked to abiotic and biotic factors dictate community assembly, making environmental filtering and/or niche selection the prevalent mechanisms underpinning ecological selection (1, 2). Conversely, the neutral theory emphasizes the role of stochastic processes (1–5), with a major part played by ecological drift, dispersal limitation, and probabilistic dispersal (2, 6). While the consensus is that both deterministic and stochastic processes simultaneously influence the assembly of communities, along a continuum (2, 5), the challenge is now to quantify their relative importance in various ecosystems. While these ecological concepts have

Editor Xiao-Hua Zhang, Ocean University of China, Qingdao, China

Address correspondence to Mylène Hugoni, mylene.hugoni@univ-lyon1.fr.

The authors declare no conflict of interest.

See the funding table on p. 15.

Received 27 November 2023

Accepted 28 December 2023

Published 30 January 2024

Copyright © 2024 Bontemps et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

been historically applied to macroorganisms (7–12), they have been much less considered for microbial communities (13–16).

In natural ecosystems, microorganisms are highly diversified, growing and evolving fast, depending on horizontal gene transfers or mutation (17), and they form complex assemblages of Archaea, Bacteria, and Microeukaryota. Most studies have focused on a single domain of life instead of taking into account the entire microbiome (16, 18–21). As key contributors to biogeochemical cycles, microorganisms represent a biosphere compartment that must be considered to model and predict global changes (22). Most often, this has been done with dominant microorganisms or entire communities, but in recent years, the functional importance of less prevalent taxa has been increasingly recognized (23). This led to the formal definition of two types of rare taxa, namely the always-rare taxa (i.e., with very small relative abundance in all samples) and the conditionally rare taxa (i.e., with a relative abundance that is low in some samples and very low in others) (24–26). Indeed, members of the rare biosphere can be metabolically active and contribute significantly to ecosystem functioning (23, 27). Some of these rare taxa may become occasionally abundant under particular environmental conditions (28), accounting for large temporal shifts in microbial structure (29, 30).

Microbial communities exposed to disturbance may display (i) resistance, involving plasticity (31); (ii) resilience, based on functional redundancy (32, 33), and/or (iii) an alternative state where new taxa harboring different functions arise, inducing a new community functioning (33, 34). Many studies focusing on microbial responses to disturbance have used experimental setups that lacked realism and complexity (i.e., lab conditions and simplified consortia), but stable ecosystems experiencing ecological disturbance are more relevant to assessing community assembly patterns in natural conditions.

Cave ecosystems provide a unique set of stable environmental variables, including the absence of light, high relative humidity and CO₂ concentrations, almost constant temperature, and oligotrophy (0.5 mg organic carbon/L) (35–44). Interestingly, many are rather confined, implying slow exchanges with surrounding karst (45). This is the case today for Lascaux Cave (Montignac-Lascaux, Dordogne, France), which was discovered in 1940 and opened to the public in 1948. Infrastructure work, high visitor numbers, and excessive chemical treatments led to cave imbalance and artwork conservation issues, and the cave was closed in 1963 (46). From this time, the cave has been monitored from a climatic, hydrogeological, and microbiological standpoint (45–49). Within caves, terminal rooms are particularly stable, such as Lascaux's Chamber of Felines. During the 2000s, Lascaux Cave was the theater of chemical treatments [especially benzalkonium chloride; (50)] to prevent fungal outgrowth problems related to past tourism activities (46, 51). This anthropization constituted a disturbance for microbial communities initially established, prompted the development of melanin-producing fungi on cave surfaces, and resulted in the formation of black stains, many still present (46, 48–50, 52). Lascaux's Chamber of Felines was never visited by the public, yet it underwent black stain formation and was treated with chemical biocides (but less than elsewhere).

In the Chamber of Felines, three rock surface conditions (Fig. 1) are present on the same walls, i.e., (i) limestone that has never been stained (microbial community considered resistant to anthropization; control), (ii) black stains present since 2002 (impacted community), and (iii) former black stains, now grayish, where black pigmentation has progressively disappeared since 2007. These grayish, attenuated stains differ little from unstained limestone and correspond to a resilient status, probably involving particular microbial dynamics for all three domains of life. Here, we aimed to test the hypothesis of similar community assembly patterns for Archaea, Bacteria, and Microeukaryota retrieved during the resistant, impacted, and resilient states. By assessing the abundant and rare biospheres, we demonstrated that always-rare taxa, shaped by stochastic processes in control and impacted communities, were the only microbial fraction selected among the three domains of life during the resilient state.

A

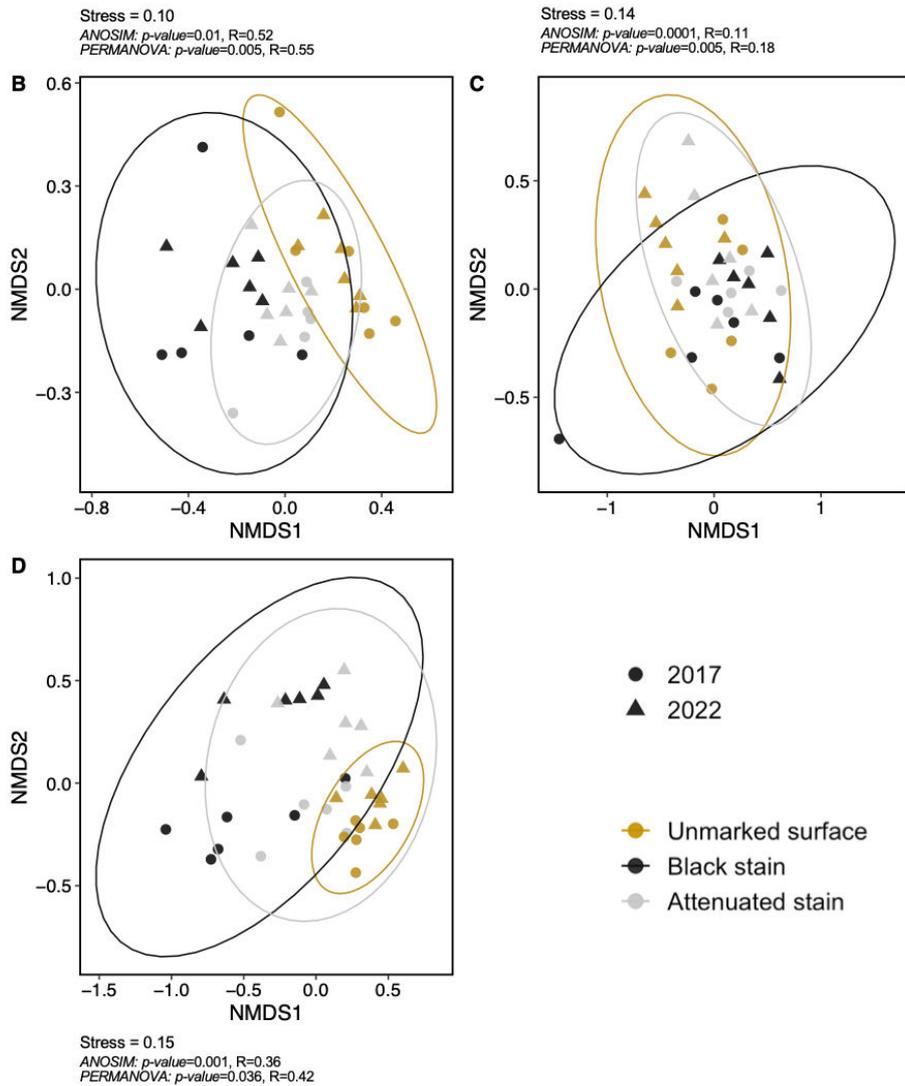
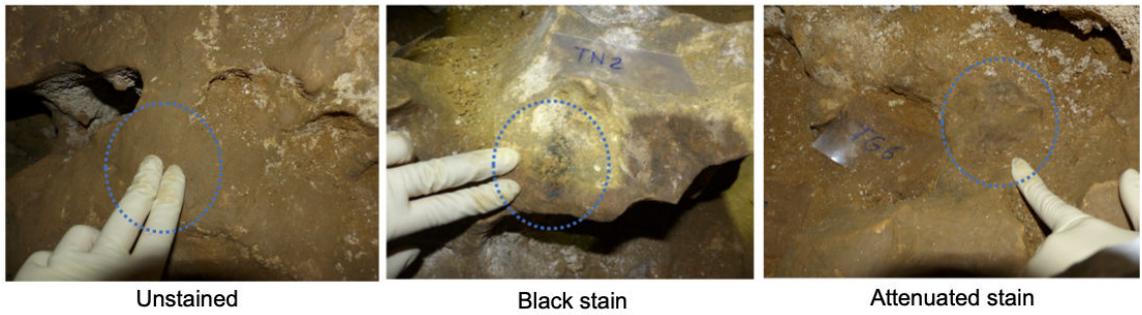


FIG 1 Comparison of rock surface conditions in Lascaux’s Chamber of Felines. (A) Photographs of unstained limestone (resistant microbial community), black stain (impacted community), and attenuated stain (resilient status) (credit: DRAC Nouvelle Aquitaine). (B–D) Differences in bacterial (B), archaeal (C), and microeukaryotic communities (D) indicated by non-metric dimensional scaling (NMDS) ordination based on Bray-Curtis dissimilarity index calculated on abundance matrices at the operational taxonomic unit level. Permutational analysis of variance (PERMANOVA), stress value, and analysis of similarities (ANOSIM) are presented in each case.

RESULTS

Comparison of black stains, attenuated stains and unstained surfaces

Non-metric multidimensional scaling (NMDS) (Fig. 1) and permutational multivariate analysis of variance (Table S1) showed that microbial communities in the three rock surface conditions differed, but the differences between sampling dates were not significant. Based on community composition data for phyla, attenuated stains displayed an intermediate situation between black stains and unstained surfaces, but with variability between 2017 and 2022 (Fig. 2). For example, Gammaproteobacteria, Vicinamibacteria, and Actinobacteria represented, respectively, 4.2%, 6.0%, and 14.5% of the bacterial sequences on unstained surfaces and 10.2%, 3.5%, and 9.4% in black stains, with intermediate values of, respectively, 5.2%, 4.8%, and 11.1% in attenuated stains (Fig. 2). The same trend was observed for the microeukaryotic community, especially for the Sordariomycetes, Agaromycetes, and Letiomycetes with, respectively, 43.5%, 33.6%, and 8.9% of sequences on unstained surfaces and 82.9%, 14.2%, and 0.5% in black stains vs 61.7%, 30.5%, and 3.3% in attenuated stains, respectively (Fig. 2). It was also the case for composition data at genus level for bacteria and microeukaryotes, and variability was again evidenced when comparing sampling dates (Fig. S1).

For each taxonomic marker, differences in diversity indices (Chao-1, Shannon, and Simpson) between sampling dates were not significant (multiple comparisons with Tukey tests: all $P > 0.1$) (Table S2). However, each of the three rock surface conditions was significantly distinct based on the Chao-1 index for the bacterial data set ($P = 0.013$) and all diversity indices for the microeukaryotic data set (all $P < 0.05$). No differences were observed for the archaeal data set (all $P > 0.1$).

Quantification of the relative importance of assembly processes at whole community level

Community assembly patterns were evaluated to understand the relative importance of stochastic vs deterministic processes using neutral-community model (NCM). The NCM fitted well for archaeal, bacterial, and microeukaryotic communities (Fig. S2), outperforming binomial and Poisson distribution models (Table S3). This points to the prevalence of stochastic processes (passive dispersal or ecological drift) independently of the random sampling performed within each of the three rock surface conditions. For the three domains of life combined, the community migration rate (m) reached an average of 0.47 in resistant communities vs 0.06 in impacted communities and 0.19 in resilient communities (Fig. 3), pointing to more extensive dispersal and ecological drift in resistant communities compared with the others.

We quantified the relative contribution of stochastic and deterministic processes shaping microbial community assembly by calculating the β -nearest taxon index (β NTI) (Fig. S3) and Bray-Curtis-based Raup-Crick (RC_{bray}) (Fig. S4) from the abundance of operational taxonomic units (OTUs) and phylogenetic distances. The β NTI is the difference between the observed β -mean nearest taxon index (β MNTD) and the mean of the null distribution of β MNTD normalized by its standard deviation. β NTI values $> +2$ indicate significantly more phylogenetic turnover than expected (heterogeneous selection), and β NTI values < -2 indicate less phylogenetic turnover than expected (homogeneous selection) (53). As the observed β MNTD value did not significantly deviate from the null β MNTD distribution (i.e., $|\beta\text{NTI}| < 2$), it meant stochastic turnover in phylogenetic composition. Since this stochasticity could entail dispersal limitation, homogenizing dispersal, or the lack of a single dominant process (termed the undominated fraction) (13), this was clarified by calculating the RC_{bray} on pairwise comparisons with $|\beta\text{NTI}| < 2$ as follows: $|RC_{\text{bray}}| > +0.95$ means dispersal limitation, $|RC_{\text{bray}}| < -0.95$ means homogenizing dispersal, and $-0.95 < |RC_{\text{bray}}| < 0.95$ means the undominated fraction (2, 13, 54). The identified assembly processes showed contrasted patterns among the three domains of life. Indeed, the contribution of stochastic processes to bacterial community assembly amounted to 93.3% on unstained limestone and 80.0%

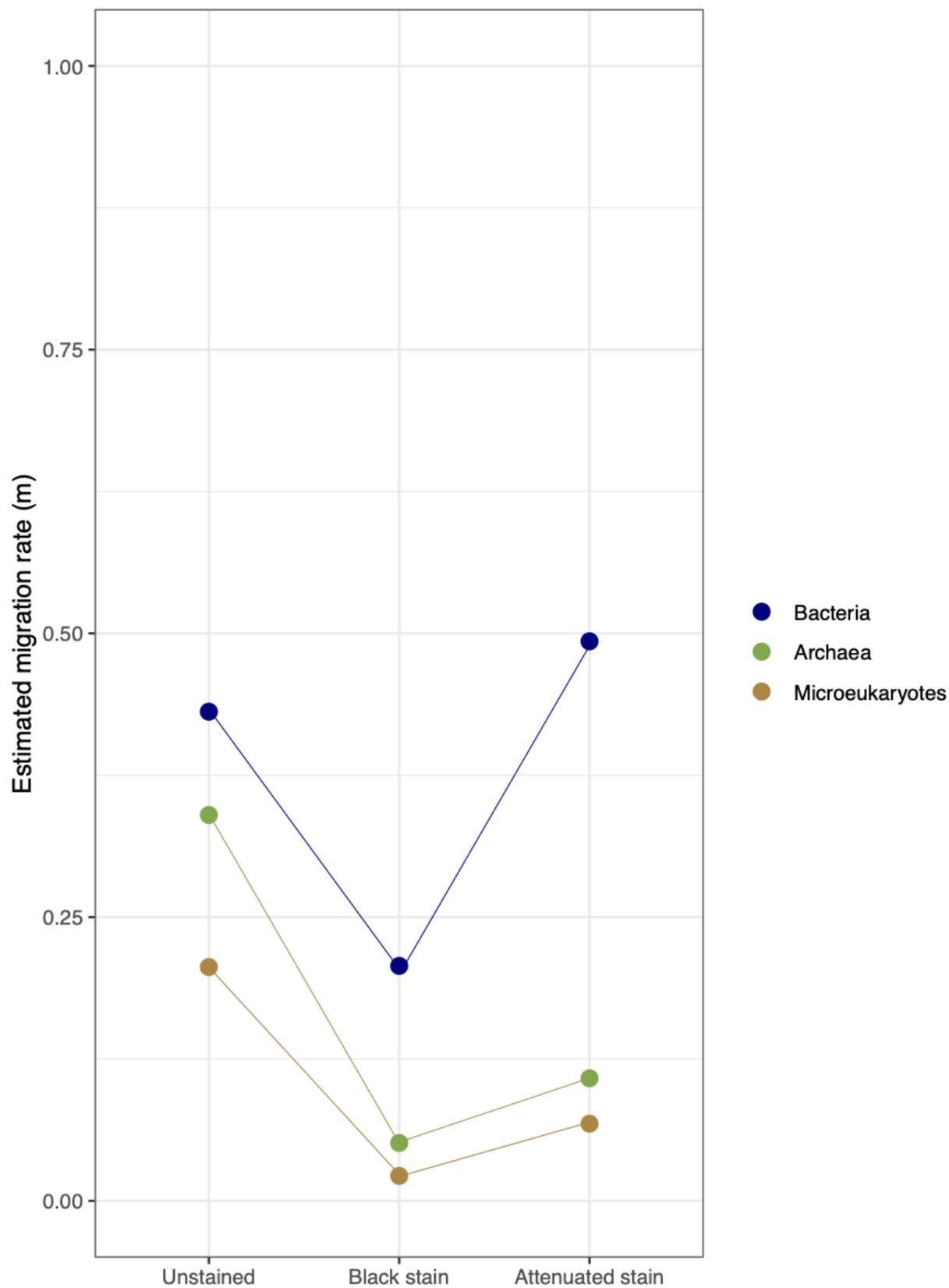


FIG 3 Estimated migration rate (m) of the neutral model during microbial dynamics for the three domains of life. The goodness-of-fit of the neutral model and the comparison of the maximum likelihood fit of the neutral, binomial, and Poisson models for each community can be found in Fig. S1 and Table S1.

in attenuated stains vs only 53.3% in black stains (Fig. 4). More precisely, the entire bacterial community showed a high level of heterogeneous selection in black stains (48.8%) but exhibited dispersal limitation in attenuated stains (69.2%) and on unstained surfaces (78.3%). For the archaeal community, in contrast, stochastic processes accounted for 100.0% of community assembly on unstained limestone and black stains with 24.5% and 28.0% of dispersal limitation, respectively. Whereas heterogeneous selection processes accounted for 20.8% in attenuated stains. Assembly patterns were exclusively stochastic for the microeukaryotic community, regardless of rock surface condition with 100% of undominated processes (i.e., selection was too weak and dispersal rates not strong enough for either process to drive compositional turnover) in both unstained and black stain conditions and 80.4% of dispersal limitation in attenuated stain rock surface condition.

Assembly processes for abundant vs rare biosphere

To go beyond analyses considering the microbial community as a whole, the abundance of taxa (i.e., abundant vs rare biosphere) was also taken into account for each domain of life to decipher community assembly patterns in more detail. To this end, OTUs were classified using four established abundance categories advocated for that purpose (55, 56), i.e., the “always-rare” taxa, the “conditionally rare” taxa, the “moderate” taxa, and the “dominant” taxa (Fig. S5), and assembly processes were reassessed for each taxa category. Bacterial dominant taxa showed a high level of heterogeneous selection in black stains (66.7%) but exhibited dispersal limitation in attenuated stains and on unstained surfaces. In contrast, microeukaryotic dominant taxa mostly displayed stochastic properties, with mainly undominated processes on the unstained surfaces and in black stains but only dispersal limitation in the attenuated stains. However, heterogeneous selection also occurred in impacted communities (33.3%), as for Bacteria. Analysis was not possible with Archaea, as only two dominant taxa were found.

The moderate taxa could not be analyzed because the number of OTUs for this category was less than six in each of the three microbial communities (see Materials and Methods).

In conditionally rare taxa, dispersal limitation predominated for Bacteria (75.5%) and microeukaryotes (60.0%) in all three surface conditions, whereas such taxa were rare in Archaea. Heterogeneous selection also contributed to the assembly of Bacteria (in all conditions) and microeukaryotes (but only in black stains).

On unstained surfaces and in black stains, always-rare taxa from the three domains of life were associated to stochastic assembly, mostly by undominated processes. The latter means that the underlying assembly mechanisms are not quantifiable yet, probably because the rare biosphere of caves remains very poorly documented. The same trend was found in the attenuated stains, with also a contribution of deterministic processes amounting to 6.7%, 10.0%, and 26.7% for heterogeneous selection in microeukaryotic, archaeal and bacterial communities, respectively (Fig. 4). This result clearly shows that always-rare taxa from the three domains of life were selected only in the resilient status, suggesting their importance in ecosystem functioning.

Always-rare taxa composition across rock surface conditions

We investigated the taxonomic identity of the 286 always-rare taxa since they were the only OTU category that had undergone deterministic assembly processes in attenuated stains. Among them, 150 OTUs displayed the always-rare status in all three rock surface conditions and 136 in one or two rock surface conditions (Fig. 5). Of these 286 OTUs, 15 were retrieved only in attenuated stains and they belonged to Acidobacteria (three OTUs), Actinobacteria (three OTUs), Chloroflexi, Alphaproteobacteria, Planctomycetota, Thaumarchaeota, and Ascomycota. In addition, 42 other OTUs, including Actinobacteria (12 OTUs), Alphaproteobacteria (5 OTUs), and Ascomycota (20 OTUs), were common to attenuated stains and unstained surfaces. Among these 42 rare microbial members, *Mycobacterium*, *Sphingobium*, *Phyllobacterium*, and *Vicinamibacter* (known as degraders

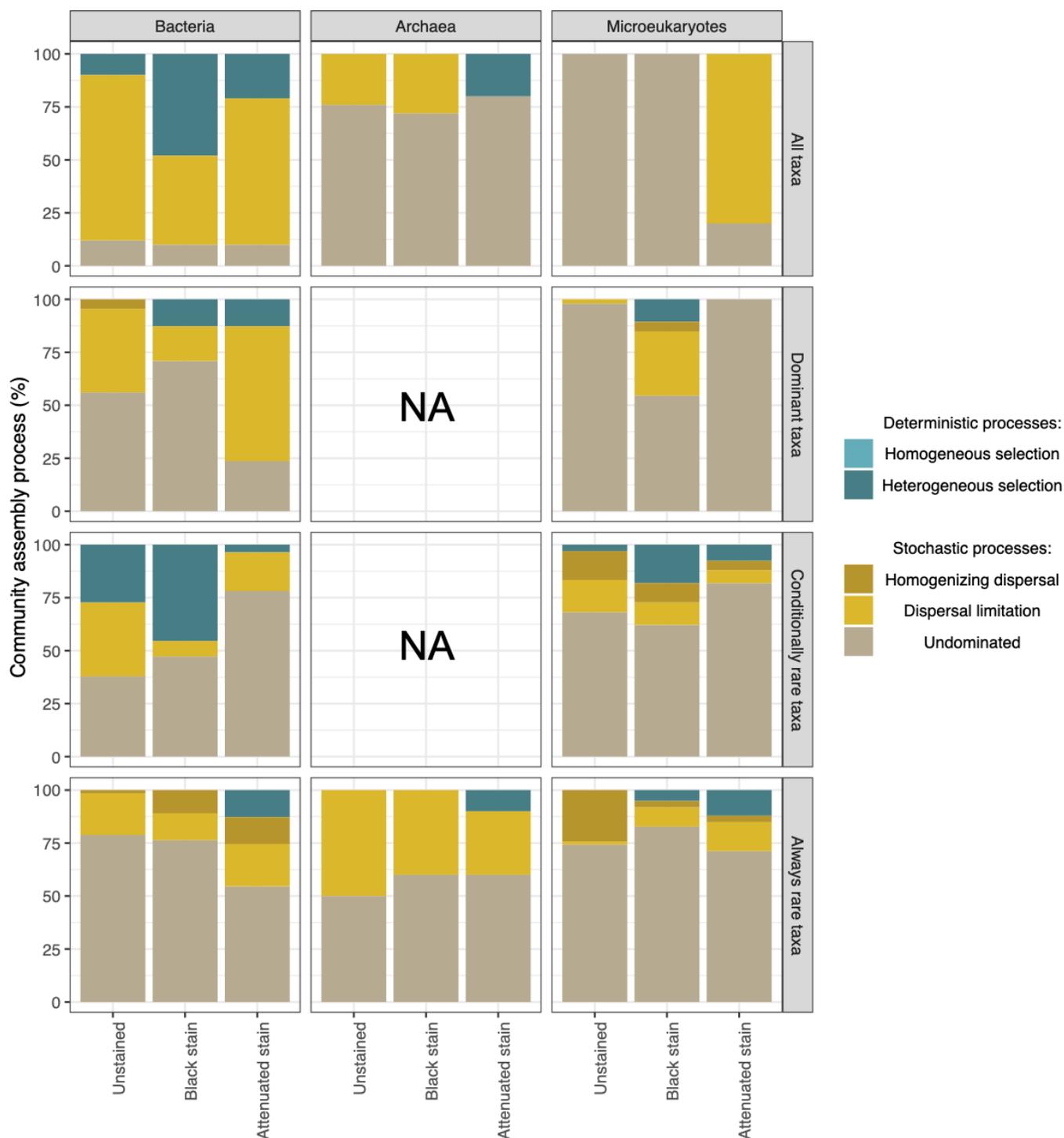


FIG 4 Contribution of deterministic and stochastic processes for all taxa, dominant taxa, conditionally rare taxa, and always-rare taxa of the three microbial communities. Moderate taxa were not assessed because the number of OTUs was too small. Community assembly processes were calculated for deterministic processes (including homogeneous and heterogeneous selection) and stochastic processes (including homogenizing dispersal, dispersal limitation, and undominated processes). NA, not available.

of polycyclic aromatic hydrocarbons or chloroaromatic compounds [(57–59)] have been identified, which might be relevant for the biodegradation of black melanin pigments in stains.

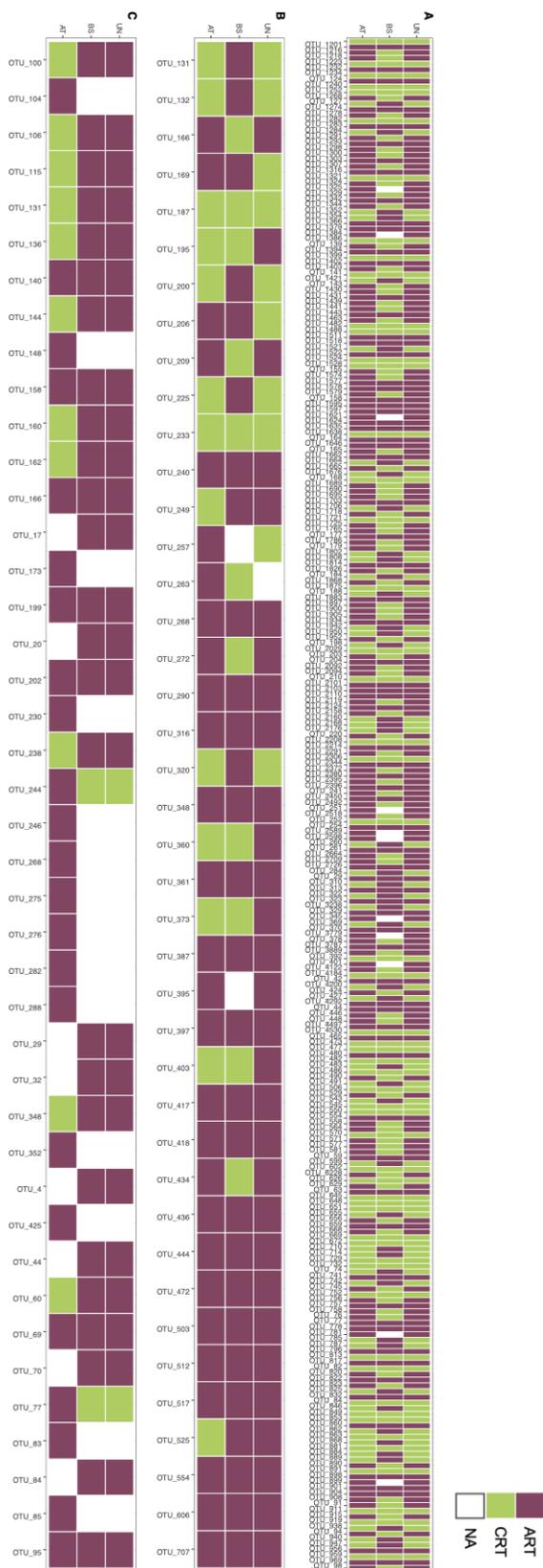


FIG 5 Distribution of always-rare taxa affiliated with Bacteria (A), Archaea (B), or microeukaryotes (C) for each rock surface condition. The heatmap was performed using OTU abundance of always-rare taxa (represented with a yellow-red gradient) and conditionally rare taxa (in blue).

DISCUSSION

In the present work, metabarcoding data showed that the different visual states for rock surfaces corresponded to distinct microbial communities. Therefore, these rock surface conditions were appropriate to assess microbial particularities in relation to perturbation (leading to stain formation) and resilience (leading to stain attenuation). Thus, the response of microbial communities to disturbance was assessed in relation to predictive scenarios by explicitly considering the ecological differences between conventionally defined taxa categories (55, 56, 60), i.e., dominant taxa, moderate taxa, conditionally rare taxa, and always-rare taxa, and the assembly processes identified (Fig. 6). The thresholds for the operational definition of these categories are arbitrary, but the use of the same thresholds in different research is important to compare microbial assembly processes in different environments and investigations (27). Interestingly, several scenarios proved relevant to describe microbial responses in the current work.

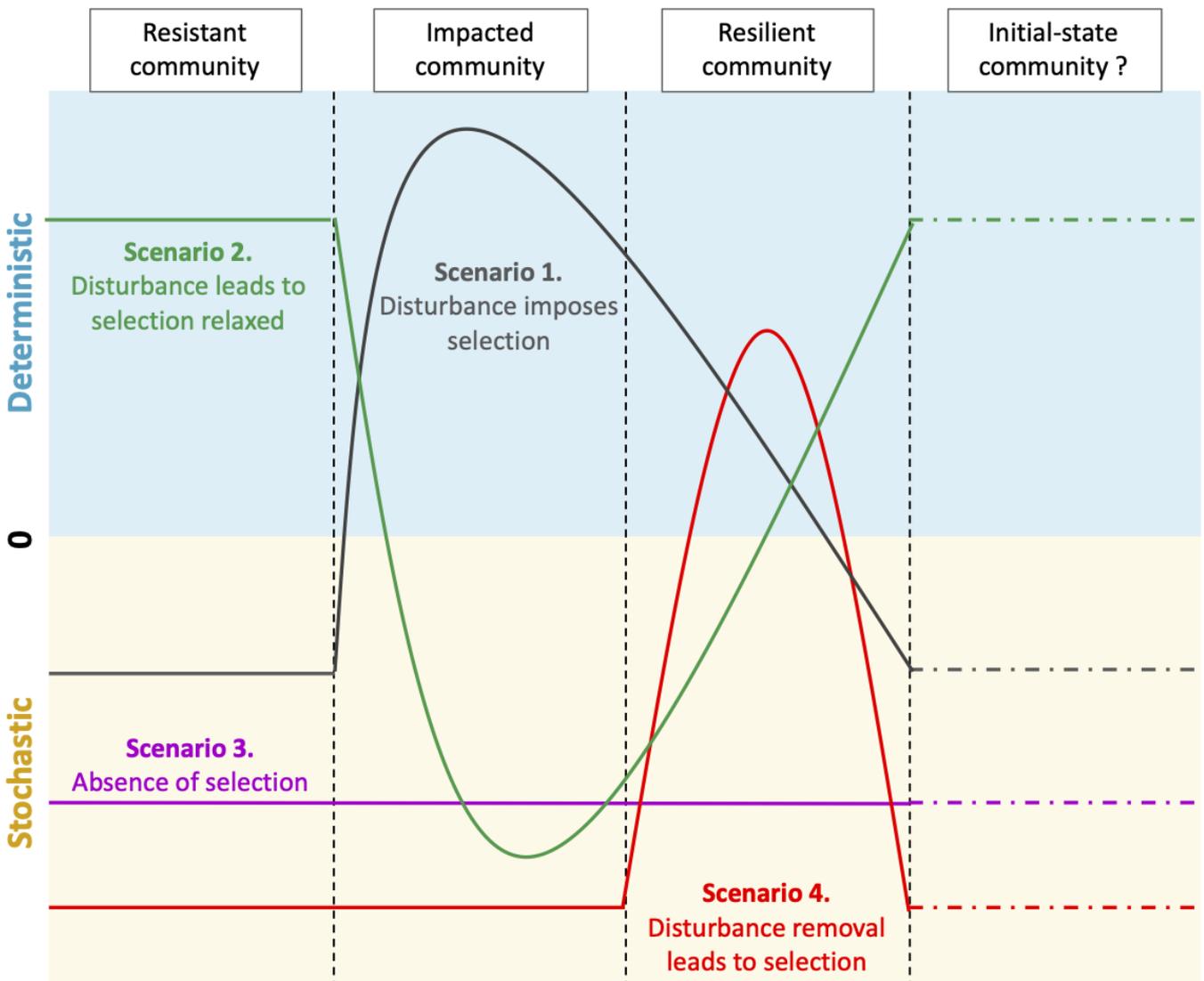


FIG 6 Four-scenario conceptual model integrating the three domains of life along the stochastic-deterministic continuum. The neutral hypothesis model (stochastic processes) and niche-based theory (deterministic processes) are represented in yellow and blue, respectively. Scenarios 1, 2, 3, and 4 are represented in black, green, purple, and red, respectively.

Scenario 1: disturbance imposes selection

In the resistant state (unstained surface), stochastic assembly processes (neutral theory) dominated. In the impacted state (black stains), deterministic assembly processes became predominant but lessened in the resilient state (attenuated stains). This scenario was demonstrated for Bacteria (considering all taxa together, or only the dominant taxa or the conditionally rare taxa) and microeukaryotes (for dominant taxa and conditionally rare taxa). These results are conceptually consistent with previous findings on macroorganisms (in the cases of drought in ponds [7] and forest evolution after fire disturbance [61]) and microorganisms ([e.g., bacterial succession within pH soil gradients or across microbial desiccation processes [16, 62]).

Scenario 2: disturbance leads to selection relaxed

Here, resistant communities (unstained surface) have undergone strong variable selection. Indeed, the disturbance weakens selection, which will lead to stochastic community assembly. As resilience develops (from black stains to attenuated stains), deterministic processes will increase in importance, as found for Archaea (all taxa together). The relevance of this scenario is supported by the analysis of forest harvesting (8, 63), which leads to the depletion of soil organic matter and soil compaction. Hartmann et al. (8) found that soil microbial communities were significantly different between such disturbed forests and non-harvested forests, even 15 years after harvest. Such a scenario was also observed for bacterial communities in agricultural soils affected by tillage (63).

Scenario 3: absence of selection

This scenario reflects the neutral theory and highlights that some microbial systems are constantly dominated by stochastic processes (15). This might be linked to abundant resource supply and high levels of organismal dispersal (2). In the present work, this was the case for microeukaryotic dominant taxa. This concept has been previously reported in fluidic systems with constant resource supply (63), where only a few microbial taxa were prevented from growing.

Scenario 4: disturbance removal leads to selection

Unexpectedly, the first three scenarios proved insufficient to depict the diversity of microbial responses documented. Therefore, it resulted in the proposition of a novel scenario applicable to the always-rare taxa. In this fourth scenario, the resistant and impacted states (unstained surface and black stain, respectively) displayed strong stochastic signatures, probably because perturbation was either minor for all taxa including always-rare taxa (in the resistant state) or strong for dominant taxa, thereby alleviating competition toward always-rare taxa (in the impacted state). In the resilient state, selection processes are important for always-rare taxa, perhaps because dominant taxa become better adapted and impose selection pressures. Alternatively, certain always-rare taxa may be positively selected by trophic opportunities, such as the fungus *Ochroconis lascauxensis* able to degrade benzalkonium chloride previously used in chemical treatments in Lascaux Cave (38, 64–66). Probably, this also includes *Mycobacterium*, *Sphingobium*, *Phyllobacterium*, and *Vicinamibacter*, which can degrade polycyclic aromatic hydrocarbons or chloroaromatic compounds that have a chemical structure similar to those of benzalkonium chloride and the benzene core of melanins. Therefore, these taxa may be involved in the biodegradation of, respectively, the chemical biocides used in Lascaux and the black pigments present in the stains (57–59). A previous rhizosphere study showed that rare bacterial taxa had undergone strong selection processes as well (20). Importantly, we found that scenario 4 was relevant for always-rare taxa from the three domains of life, suggesting common ecological behaviors as well as cooperative relations promoted by disturbance. It is probably of major significance for ecosystem functioning.

To conclude, we propose a novel scenario to predict the dynamics of always-rare taxa, which was validated in the case of Lascaux's Chamber of Felines. We showed the necessity to consider the rare biosphere (and its functions) to fully assess microbial responses to environmental change and their involvement in ecosystem functioning. The conceptual model reported in this work (scenario 4) needs now to be tested in diverse environments to better predict and model ecosystem functioning.

MATERIALS AND METHODS

Sampling

Lascaux Cave is located in the south-west of France (N 45°03'13.087" and E 1°10'12.362"). It received up to 1,800 visitors daily, until the cave closure in 1963 (49). Samples were collected in February 2017 and June 2022 from the walls of the Chamber of Felines (never visited by tourists), in areas devoid of Paleolithic paintings and engravings, with special authorization (from the Ministry of Culture, DRAC Nouvelle-Aquitaine). Six black stains, six control zones (i.e., unstained surface), and six attenuated stains were sampled using swabs gently rubbed against the rock wall, so as to target wall surface microorganisms. For each replicate, samples for the three conditions were taken in the same location, at about 10–20 cm from one another, and the six locations were 1–2 m from one another. Strain attenuation did not proceed further between 2017 and 2022. Samples were conserved at –80°C until DNA extraction.

Illumina sequencing and bioinformatics analysis

Total genomic DNA extraction was performed using the FastDNA SPIN Kit for Soil (MP Biomedicals, Illkirch, France), following the manufacturer's instructions. DNA concentration was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's preconization.

Three gene markers were analyzed in each individual sample. The V3–V4 region of the 16S rRNA genes was amplified in triplicate for Bacteria and Archaea using the universal primers 341F (CCTACGGGNGGCWGCAG) and 805R (GACTACHVGGGTATCTAATCC) (67), and 515F (CAGCCGCCGCGTAA) and 915R (GTGCTCCCCGCCAATTCCT) (68), respectively. For Eukaryota, the V4 region of the 18S rRNA genes was amplified in triplicate using the universal primers 0067a_deg (AAGCCATGCATGYCTAAGTATMA) and NSR399 (TCTCAGGCTCCYTCTCCGG) (69). The PCR mix consisted of 5 µL of 5× Hot BioAmp Blend Master Mix RTL (Biofidal, Vaux-en-Velin, France), 0.1 µM of the two primers, 0.1× GC-rich-enhancer (Biofidal), 0.2 ng·µL⁻¹ of bovine serum albumin (Promega, Madison, WI, USA), and 0.2–1.0 ng DNA. All amplifications were performed in a Bio-Rad T1000 thermal cycler (Bio-Rad, Hercules, CA, USA). The PCR program for Bacteria was 3 min at 95°C, 28 cycles of 45 s at 95°C, 45 s at 50°C, and 90 s at 72°C, followed by 7 min at 72°C. For Archaea, it consisted of 10 min at 94°C, 30 cycles of 1 min at 94°C, 1 min at 58°C, and 90 s at 72°C, followed by 10 min at 72°C. For the 18S rRNA gene, it consisted of 10 min at 95°C, 30 cycles of 1 min at 94°C, 1 min at 55°C, and 90 s at 72°C, followed by 10 min at 72°C. For the ITS2 region, it consisted of 10 min at 95°C, 28 cycles of 20 s at 94°C, 30 s at 47°C, and 20 s at 72°C, followed by 7 min at 72°C. All primers were tagged with the Illumina adapter sequences (TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG and GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G), allowing the construction of amplicon libraries by a two-step PCR. DNA extraction was also carried out from tubes without any biological matrix, and this negative control was used to evaluate and subtract ambient and kit products' contamination. Amplification signals were checked by electrophoresis on 1.5% agarose gel. High-throughput sequencing was achieved after pooling PCR triplicates, using Illumina MiSeq with 2 × 300 bp, paired-end chemistry v3 (aiming at 40,000 sequences for Bacteria and microeukaryotes, and 70,000 sequences for Archaea).

For each of the three datasets (i.e., bacterial 16S rRNA genes, archaeal 16S rRNA genes, and eukaryotic 18S rRNA genes), paired-end reads were merged with a maximum

of 10% mismatches in the overlap region using Fast Length Adjustment of Short reads (70). Denoising procedures were carried out by discarding reads without the expected length (200–500 bp) or containing any ambiguous bases (N). After the dereplication of sequences, the clusterization tool was run with Swarm (71), which uses a local clustering threshold rather than a global threshold and an aggregation distance of 3 for the identification of OTUs. Chimeric OTUs were removed using VSEARCH (72). Moreover, low-abundance sequences were filtered to keep OTUs representing at least 0.005% of all sequences (73). Taxonomic affiliation was performed with both RDP Classifier and BLASTn (74) against the 138.1 SILVA database (75) for bacterial 16S rRNA genes, archaeal 16S rRNA genes, and microeukaryotic 18S rRNA genes and against the 8.2 UNITE database for fungal ITS markers (76). This procedure was automated in the FROGS pipeline (77). Contaminant OTUs identified from the negative control samples were removed. To compare samples, a normalization procedure, using phyloseq package, was applied by randomly resampling down to 12,105, 6,209, 31,981, and 25,821 sequences in the bacterial, archaeal, microeukaryotic, and fungal data sets, respectively (78).

Rarefaction curves were calculated to assess sequencing effort, using Paleontological Statistics software v4.02 (79). OTU richness and diversity were estimated using Chao 1 index (80), Shannon's H' (81), and Simpson (1-D) index (82). Communities were primarily compared with NMDS, using vegan and stats packages in R (83–85). The procedure computes a stress value, which measures the difference between the ranks on the ordination configuration and the ranks in the original similarity matrix for each replicate (86). Analysis of similarity was conducted using the vegan package in R (83) to test the differences ($P < 0.05$) in overall community composition between the three rock surface conditions, i.e., resistant, impacted, and resilient, and to further confirm the results observed in the NMDS plot. Multiple comparisons of treatments by Tukey's test were performed using agricolae package. A Bonferroni correction was applied to P values to lower alpha risk. All analyses were based on abundance dissimilarity matrices with the Bray-Curtis index (87) using R.

Definition of abundant and rare taxa

In this study, OTUs were classified, after combining data from both sampling dates, into four categories according to their relative abundance, as proposed by Liu et al. (55), used in subsequent works (24–27, 56, 60) and described in Fig. S5: (i) always-rare taxa, with an abundance $< 0.01\%$ in all samples; (ii) conditionally rare taxa, with an abundance $< 1\%$ in all samples and $< 0.01\%$ in some samples (maximum 33% of samples); (iii) moderate taxa, with an abundance $< 1\%$ and $> 0.01\%$ in all samples; and (iv) dominant taxa, with an abundance $\geq 1\%$ of all sequences in all samples or that fluctuates strongly, i.e., $< 0.01\%$ in certain samples and $\geq 1\%$ in others. Metagenomic data for other Lascaux rooms suggest that artificial over-representation of certain microbial groups with the molecular approach followed could concern taxa already prominent and was very unlikely to have concerned rare taxa.

Assessment of community dispersal rate: neutral model

Sloan's NCM was used to estimate the importance of passive dispersal on community assembly by assessing the relationship between the frequency at which taxa occur in a set of local communities and their abundance across the wider meta-community (4). In this model, the migration rate (m) is a parameter used to evaluate the probability of a random loss of an individual in a local community, to be replaced by an immigrant from the meta-community (3). It is calculated as follows:

$$\text{Freq}_i = 1 - I(1 \div N / N \times m \times p_i, N \times m \times (1 - p_i)),$$

where Freq_i is the occurrence frequency of taxa i across communities; N is the number of individuals per community; m is the estimated dispersal rate; p_i is the average relative

abundance of taxa i across communities; and $l()$ is the probability density function of β -distribution (21). This analysis was performed using nonlinear least-square fitting, and calculation of 95% confidence interval for the model predictions was conducted using the Wilson score interval (“minpack.lm” and “Hmisc” package in R [88, 89]). The overall fit of the model to observed data was done by comparing the sum of squares of residuals (SS_{err}) with the total sum of squares (SS_{total}), with model fit = $1 - SS_{err}/SS_{total}$ (generalized R-squared). The Akaike information criterion and Bayesian information criterion (BIC) were calculated based on 1,000 bootstrap replicates for the neutral, binomial, and Poisson distribution models to determine the best model fit and whether the model was based on only the random sampling of the source meta-community.

Phylogenetic analysis and estimation of assembly community processes

In order to determine the ecological processes controlling the assembly of all taxa and the different abundance categories of taxa described above (dominant taxa, moderate taxa, conditionally rare taxa, and always-rare taxa), phylogenetic trees were built independently for each of the five categories and for each taxonomic marker (in total, 15 different phylogenetic trees). Briefly, for each taxonomic marker, sequences were aligned using MAFFT (90), then gaps and poorly conserved regions were eliminated using Gblocks (91, 92). Finally, phylogeny was reconstructed using IQ-TREE-2 (93) (default parameters), with the best model being automatically chosen by the program according to the BIC. Node support was estimated using the bootstrap method (1,000 replicates).

To quantify the relative importance of stochastic and deterministic processes that drive microbial community assembly, β NTI and RC_{bray} were calculated using ade4, picante, and iCamp packages and scripts from Stegen et al. in R software v4.0.2 (13, 94–96), based on the phylogenetic distance and OTU abundance ($n > 6$ OTUs) (13). β NTI is the number of standard deviations of the β -mean nearest taxon distance from the mean of the null distribution (13). The β NTI is the difference between the observed β -mean nearest taxon index and the mean of the null distribution of β MNTD normalized by its standard deviation. The β NTI values between -2 and 2 indicate the dominance of the stochastic processes, whereas β NTI values smaller than -2 or larger than 2 indicate that deterministic processes (i.e., homogeneous selection and heterogeneous selection) play a more important role in community assembly than stochastic processes (2). Indeed, β NTI values $> +2$ indicate significantly more phylogenetic turnover than expected (heterogeneous selection), and β NTI values < -2 less indicate phylogenetic turnover than expected (homogeneous selection) (53). As the observed β MNTD value did not significantly deviate from the null β MNTD distribution (i.e., $|\beta$ NTI| < 2), it meant stochastic turnover in phylogenetic composition. Since this stochasticity could entail dispersal limitation, homogenizing dispersal, or the lack of a single dominant process (termed the undominated fraction) (13, 97), this was clarified by calculating the RC_{bray} on pairwise comparisons with $|\beta$ NTI| < 2 , as it enables to further partition pairwise comparisons that were assigned to stochastic processes (2). Thus, for $|\beta$ NTI| < 2 , $RC_{bray} < -0.95$ and $RC_{bray} > 0.95$ indicate the relative dominant influence of homogenizing dispersal and dispersal limitation, respectively, and $|RC_{bray}| < 0.95$ indicates a crucial role of “undominated” assembly, including weak selection, weak dispersal, diversification, and/or drift (2, 13, 14, 54). The major factors that influence the assembly of dominant taxa, conditionally rare taxa, and always-rare taxa were explored separately.

ACKNOWLEDGMENTS

Funding was provided by DRAC Nouvelle Aquitaine (Bordeaux, France).

We are very grateful to S. Géraud, J.C. Portais, and M. Mauriac (DRAC Nouvelle Aquitaine) for key information and help, D. Henry - Lormelle and colleagues for technical help with Lascaux sampling, Lascaux Scientific Board for helpful discussions, and L. Alonso for help in technical analysis.

Y.M.-L. obtained funding and coordinated the project, Y.M.-L. and M.H. designed the research, Z.B. analyzed the data, and Z.B., Y.M.-L., and M.H. wrote the paper. All authors were involved in writing and approving the manuscript.

AUTHOR AFFILIATIONS

¹UMR 5557 Ecologie Microbienne, Université Claude Bernard Lyon 1, CNRS, INRAE, VetAgro Sup, Villeurbanne, France

²Department of Medical Biochemistry and Microbiology, Science for Life Laboratories, Uppsala University, Uppsala, Sweden

³UMR 5240 Microbiologie Adaptation et Pathogénie, INSA Lyon, CNRS, Université Claude Bernard Lyon 1, Villeurbanne, France

⁴Institut Universitaire de France (IUF), France

AUTHOR ORCID*s*

Zélia Bontemps  <http://orcid.org/0009-0002-6785-9197>

Yvan Moënne-Loccoz  <http://orcid.org/0000-0002-9817-1953>

Mylène Hugoni  <http://orcid.org/0000-0002-2430-1057>

FUNDING

Funder	Grant(s)	Author(s)
DRAC Nouvelle Aquitaine		Yvan Moënne-Loccoz

AUTHOR CONTRIBUTIONS

Zélia Bontemps, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing | Yvan Moënne-Loccoz, Conceptualization, Data curation, Formal analysis, Investigation, Supervision, Validation, Writing – original draft, Writing – review and editing | Mylène Hugoni, Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All data that support the findings of this study have been deposited in EBI under reference [PRJEB46483](https://doi.org/10.1093/bioinformatics/btad001), [PRJEB46545](https://doi.org/10.1093/bioinformatics/btad002), and [PRJEB46546](https://doi.org/10.1093/bioinformatics/btad003) for bacterial 16S rRNA genes, archaeal 16S rRNA genes, and microeukaryotic 18S rRNA genes, respectively. The authors declare that the R (R 4.0.2) codes used to generate the results reported in this study are available in this paper. The R code supporting the finding presented here is available from the GitHub Repository https://github.com/LascauxZelia/Bontemps_et_al_2022.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (mSystems01233-23-s0001.pdf). Supplemental figures and tables.

REFERENCES

- Vellend M. 2010. Conceptual synthesis in community ecology. *Q Rev Biol* 85:183–206. <https://doi.org/10.1086/652373>
- Zhou J, Ning D. 2017. Stochastic community assembly: does it matter in microbial ecology? *Microbiol Mol Biol Rev* 81:e0002-17. <https://doi.org/10.1128/MMBR.00002-17>
- Hubbell S. 2001. The unified neutral theory of biodiversity and biogeography. Monographs in population biology. Princeton University Press.
- Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* 8:732–740. <https://doi.org/10.1111/j.1462-2920.2005.00956.x>

5. Chase JM, Myers JA. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. *Phil Trans R Soc B* 366:2351–2363. <https://doi.org/10.1098/rstb.2011.0063>
6. Corline NJ, Peek RA, Montgomery J, Katz JVE, Jeffres CA. 2021. Understanding community assembly rules in managed floodplain food webs. *Ecosphere* 12:e03330. <https://doi.org/10.1002/ecs2.3330>
7. Chase JM. 2007. Drought mediates the importance of stochastic community assembly. *Proc Natl Acad Sci U S A* 104:17430–17434. <https://doi.org/10.1073/pnas.0704350104>
8. Hartmann M, Howes CG, VanInsberghe D, Yu H, Bachar D, Christen R, Henrik Nilsson R, Hallam SJ, Mohn WW. 2012. Significant and persistent impact of timber harvesting on soil microbial communities in northern coniferous forests. *ISME J* 6:2199–2218. <https://doi.org/10.1038/ismej.2012.84>
9. Kardol P, Souza L, Classen AT. 2013. Resource availability mediates the importance of priority effects in plant community assembly and ecosystem function. *Oikos* 122:84–94. <https://doi.org/10.1111/j.1600-0706.2012.20546.x>
10. Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, Bohannan BJ. 2016. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J* 10:655–664. <https://doi.org/10.1038/ismej.2015.142>
11. Mi X, Swenson NG, Jia Q, Rao M, Feng G, Ren H, Bebbler DP, Ma K. 2016. Stochastic assembly in a subtropical forest chronosequence: evidence from contrasting changes of species, phylogenetic and functional dissimilarity over succession. *Sci Rep* 6:32596. <https://doi.org/10.1038/srep32596>
12. Ning D, Yuan M, Wu L, Zhang Y, Guo X, Zhou X, Yang Y, Arkin AP, Firestone MK, Zhou J. 2020. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat Commun* 11:4717–4719. <https://doi.org/10.1038/s41467-020-18560-z>
13. Stegen JC, Lin X, Konopka AE, Fredrickson JK. 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* 6:1653–1664. <https://doi.org/10.1038/ismej.2012.22>
14. Stegen JC, Lin X, Fredrickson JK, Konopka AE. 2015. Estimating and mapping ecological processes influencing microbial community assembly. *Front Microbiol* 6:370. <https://doi.org/10.3389/fmicb.2015.00370>
15. Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. 2015. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci U S A* 112:E1326–E1332. <https://doi.org/10.1073/pnas.1414261112>
16. Tripathi BM, Stegen JC, Kim M, Dong K, Adams JM, Lee YK. 2018. Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J* 12:1072–1083. <https://doi.org/10.1038/s41396-018-0082-4>
17. Prescott L-M, Willey JM, Sherwood LM, Woolwerton CJ. 2003. *Microbiology*. 5th ed. DeBoeck.
18. Ferrenberg S, O'Neill SP, Knelman JE, Todd B, Duggan S, Bradley D, Robinson T, Schmidt SK, Townsend AR, Williams MW, Cleveland CC, Melbourne BA, Jiang L, Nemergut DR. 2013. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *ISME J* 7:1102–1111. <https://doi.org/10.1038/ismej.2013.11>
19. Gao C, Montoya L, Xu L, Madera M, Hollingsworth J, Purdom E, Singan V, Vogel J, Hutmacher RB, Dahlberg JA, Coleman-Derr D, Lemaux PG, Taylor JW. 2020. Fungal community assembly in drought-stressed sorghum shows stochasticity, selection, and universal ecological dynamics. *Nat Commun* 11:34. <https://doi.org/10.1038/s41467-019-13913-9>
20. Jiao S, Lu Y. 2020. Soil pH and temperature regulate assembly processes of abundant and rare bacterial communities in agricultural ecosystems. *Environ Microbiol* 22:1052–1065. <https://doi.org/10.1111/1462-2920.14815>
21. Luan L, Jiang Y, Cheng M, Dini-Andreote F, Sui Y, Xu Q, Geisen S, Sun B. 2020. Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. *Nat Commun* 11:6406. <https://doi.org/10.1038/s41467-020-20271-4>
22. Madsen EL. 2011. Microorganisms and their roles in fundamental biogeochemical cycles. *Curr Opin Biotechnol* 22:456–464. <https://doi.org/10.1016/j.copbio.2011.01.008>
23. Lynch MDJ, Neufeld JD. 2015. Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* 13:217–229. <https://doi.org/10.1038/nrmicro3400>
24. Zhang Z-F, Pan J, Pan Y-P, Li M. 2021. Biogeography, assembly patterns, driving factors, and interactions of archaeal community in mangrove sediments. *mSystems* 6:e0138120. <https://doi.org/10.1128/mSystems.01381-20>
25. Zhang H, Zhu Y, Yu C, Li Y, Long M, Li W, Liao Y, Wang W, Wen X. 2023. Conservation tillage and moderate nitrogen application changed the composition, assembly pattern and interaction network of abundant and rare microbial community on straw surface. *Appl Soil Ecol* 191:105060. <https://doi.org/10.1016/j.apsoil.2023.105060>
26. Yang Y, Cheng K, Li K, Jin Y, He X. 2022. Deciphering the diversity patterns and community assembly of rare and abundant bacterial communities in wetland system. *Sci Total Environ* 838:156334. <https://doi.org/10.1016/j.scitotenv.2022.156334>
27. Chen W, Ren K, Isabwe A, Chen H, Liu M, Yang J. 2019. Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons. *Microbiome* 7:138. <https://doi.org/10.1186/s40168-019-0749-8>
28. Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, Gilbert JA. 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio* 5:e01371-14. <https://doi.org/10.1128/mBio.01371-14>
29. Nuppenen-Puputti M, Purkamo L, Kietäväinen R, Nyysönen M, Itävaara M, Ahonen L, Kukkonen I, Bomberg M. 2018. Rare biosphere archaea assimilate acetate in Precambrian terrestrial subsurface at 2.2 km depth. *Geosciences* 8:418. <https://doi.org/10.3390/geosciences8110418>
30. Liu X, Li M, Castelle CJ, Probst AJ, Zhou Z, Pan J, Liu Y, Banfield JF, Gu JD. 2018. Insights into the ecology, evolution, and metabolism of the widespread Woesearchaeotal lineages. *Microbiome* 6:102. <https://doi.org/10.1186/s40168-018-0488-2>
31. Maurice CF, Haiser HJ, Turnbaugh PJ. 2013. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* 152:39–50. <https://doi.org/10.1016/j.cell.2012.10.052>
32. Allison SD, Martiny JBH. 2008. Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A* 105 Suppl 1:11512–11519. <https://doi.org/10.1073/pnas.0801925105>
33. Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JBH, Matulich KL, Schmidt TM, Handelsman J. 2012. Fundamentals of microbial community resistance and resilience. *Front Microbiol* 3:417. <https://doi.org/10.3389/fmicb.2012.00417>
34. Zaneveld JR, McMinds R, Vega Thurber R. 2017. Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nat Microbiol* 2:17121. <https://doi.org/10.1038/nrmicrobiol.2017.121>
35. Di Russo C, Carchini G, Rampini M, Lucarelli M, Sbordoni V. 1997. Long term stability of a terrestrial cave community. *Int J Speleol* 26:75–88. <https://doi.org/10.5038/1827-806X.26.1.7>
36. Cañaveras JC, Sanchez-Moral S, Sloer V, Saiz-Jimenez C. 2001. Microorganisms and microbially induced fabrics in cave walls. *Geomicrobiol J* 18:223–240. <https://doi.org/10.1080/01490450152467769>
37. Barton HA, Northup DE. 2007. Geomicrobiology in cave environments: past, current and future perspectives. *J Cave Karst Stud* 69:163–178.
38. Bastian F, Alabouvette C, Jurado V, Saiz-Jimenez C. 2009. Impact of biocide treatments on the bacterial communities of the Lascaux Cave. *Naturwissenschaften* 96:863–868. <https://doi.org/10.1007/s00114-009-0540-y>
39. Cueva S, Sanchez-Moral S, Saiz-Jimenez C, Cañaveras JC. 2009. Microbial communities and associated mineral fabrics in Altamira Cave, Spain. *Int J Speleol* 38:83–92. <https://doi.org/10.5038/1827-806X.38.1.9>
40. Engel AS. 2010. Microbial diversity of cave ecosystems, p 219–238. In Barton LL, Mandl M, Loy A (ed), *Geomicrobiology: molecular and environmental perspective*. Springer Netherlands, Dordrecht.
41. Pasić L, Kovce B, Sket B, Herzog-Velikonja B. 2010. Diversity of microbial communities colonizing the walls of a Karstic Cave in Slovenia. *FEMS Microbiol Ecol* 71:50–60. <https://doi.org/10.1111/j.1574-6941.2009.00789.x>
42. De Leo F, Iero A, Zammit G, Urzi C. 2012. Chemoorganotrophic bacteria isolated from biodeteriorated surfaces in cave and catacombs. *Int J Speleol* 41:125–136. <https://doi.org/10.5038/1827-806X.41.2.1>

43. Hüppop K. 2012. Adaptation to low food, p 1–9. In White WB, Culver DC (ed), *Encyclopedia of caves*, 2nd ed. Academic Press, Amsterdam.
44. Banerjee S, Joshi SR. 2013. Insights into cave architecture and the role of bacterial biofilm. *Proc Natl Acad Sci U S A* 83:277–290. <https://doi.org/10.1007/s40011-012-0149-3>
45. Bourges F, Genthon P, Genty D, Lorblanchet M, Mauduit E, D'Hulst D. 2014. Conservation of prehistoric caves and stability of their inner climate: lessons from Chauvet and other French caves. *Sci Total Environ* 493:79–91. <https://doi.org/10.1016/j.scitotenv.2014.05.137>
46. Bastian F, Jurado V, Nováková A, Alabouvette C, Saiz-Jimenez C. 2010. The microbiology of Lascaux Cave. *Microbiology (Reading)* 156:644–652. <https://doi.org/10.1099/mic.0.036160-0>
47. Lacanette D, Malaurent P, Caltagirone J-P, Brunet J. 2007. Étude des transferts de masse et de chaleur dans la grotte de Lascaux: le suivi climatique et le simulateur. *Karstologia* 50:19–30. <https://doi.org/10.3406/karst.2007.2608>
48. Alabouvette C, Saiz-Jiménez C. 2011. Écologie microbienne de la Grotte de Lascaux. CSIC-Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS).
49. 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* 6:216. <https://doi.org/10.1186/s40168-018-0599-9>
50. Martin-Sanchez PM, Miller AZ, Saiz-Jimenez C. 2015. Lascaux cave: an example of fragile ecological balance in subterranean environments. *De Gruyter*.
51. Lefèvre M. 1974. La 'Maladie Verte' de Lascaux. *Stud Conserv* 19:126–156. <https://doi.org/10.1179/sic.1974.013>
52. Martin-Sanchez PM, Nováková A, Bastian F, Alabouvette C, Saiz-Jimenez C. 2012. Two new species of the genus *Ochroconis*, *O. lascauxensis* and *O. anomala* isolated from black stains in Lascaux Cave, France. *Fungal Biol* 116:574–589. <https://doi.org/10.1016/j.funbio.2012.02.006>
53. Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, Rockhold ML, Konopka A. 2013. Quantifying community assembly processes and identifying features that impose them. *ISME J* 7:2069–2079. <https://doi.org/10.1038/ismej.2013.93>
54. Chase JM, Kraft NJB, Smith KG, Vellend M, Inouye BD. 2011. Using null models to disentangle variation in community dissimilarity from variation in α -diversity. *Ecosphere* 2:art24. <https://doi.org/10.1890/ES10-00117.1>
55. Liu L, Yang J, Yu Z, Wilkinson DM. 2015. The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J* 9:2068–2077. <https://doi.org/10.1038/ismej.2015.29>
56. Dai T, Zhang Y, Tang Y, Bai Y, Tao Y, Huang B, Wen D. 2016. Identifying the key taxonomic categories that characterize microbial community diversity using full-scale classification: a case study of microbial communities in the sediments of Hangzhou Bay. *FEMS Microbiol Ecol* 92:fw150. <https://doi.org/10.1093/femsec/fw150>
57. Teng Y, Feng S, Ren W, Zhu L, Ma W, Christie P, Luo Y. 2017. Phytoremediation of diphenylarsinic-acid-contaminated soil by *Pteris vittata* associated with *Phyllobacterium myrsinacearum* RC6b. *Int J Phytoremediation* 19:463–469. <https://doi.org/10.1080/15226514.2016.1244166>
58. Vieira S, Luckner M, Wanner G, Overmann J. 2017. *Luteitalea pratensis* gen. nov., sp. nov. a new member of subdivision 6 acidobacteria isolated from temperate grassland soil. *Int J Syst Evol Microbiol* 67:1408–1414. <https://doi.org/10.1099/ijsem.0.001827>
59. Dudhagara DR, Dave BP. 2018. *Mycobacterium* as polycyclic aromatic hydrocarbons (PAHs) degrader, p 349–361. In Ribón W (ed), *Mycobacterium-research and development*. IntechOpen, London.
60. Chen W, Pan Y, Yu L, Yang J, Zhang W. 2017. Patterns and processes in marine microeukaryotic community biogeography from Xiamen coastal waters and intertidal sediments, Southeast China. *Front Microbiol* 8:1912. <https://doi.org/10.3389/fmicb.2017.01912>
61. Myers JA, Harms KE. 2011. Seed arrival and ecological filters interact to assemble high-diversity plant communities. *Ecology* 92:676–686. <https://doi.org/10.1890/10-1001.1>
62. Valverde A, Makhallanyane TP, Cowan DA. 2014. Contrasting assembly processes in a bacterial metacommunity along a desiccation gradient. *Front Microbiol* 5:668. <https://doi.org/10.3389/fmicb.2014.00668>
63. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD, Yang Y, He Z, Wu L, Stahl DA, Hazen TC, Tiedje JM, Arkin AP. 2014. Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc Natl Acad Sci U S A* 111:836–845. <https://doi.org/10.1073/pnas.1324044111>
64. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A* 106:22427–22432. <https://doi.org/10.1073/pnas.0908284106>
65. Martin-Sanchez PM, 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. *Environ Sci Technol* 46:3762–3770. <https://doi.org/10.1021/es2040625>
66. Pedrós-Alió C. 2012. The rare bacterial biosphere. *Ann Rev Mar Sci* 4:449–466. <https://doi.org/10.1146/annurev-marine-120710-100948>
67. Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic sea. *ISME J* 5:1571–1579. <https://doi.org/10.1038/ismej.2011.41>
68. Herfort L, Kim J, Coolen M, Abbas B, Schouten S, Herndl G, Sinninghe Damsté J. 2009. Diversity of archaea and detection of crenarchaeotal *amoA* genes in the rivers Rhine and Têt. *Aquat Microb Ecol* 55:189–201. <https://doi.org/10.3354/ame01294>
69. Dollive S, Peterfreund GL, Sherrill-Mix S, Bittinger K, Sinha R, Hoffmann C, Nabel CS, Hill DA, Artis D, Bachman MA, Custers-Allen R, Grunberg S, Wu GD, Lewis JD, Bushman FD. 2012. A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol* 13:R60. <https://doi.org/10.1186/gb-2012-13-7-r60>
70. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
71. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. 2014. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* 2:e593. <https://doi.org/10.7717/peerj.593>
72. 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>
73. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* 10:57–59. <https://doi.org/10.1038/nmeth.2276>
74. Zhang J, Madden TL. 1997. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res* 7:649–656. <https://doi.org/10.1101/gr.7.6.649>
75. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>
76. Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K. 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res* 47:D259–D264. <https://doi.org/10.1093/nar/gky1022>
77. Escudé F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S, Pascal G. 2018. FROGS: find, rapidly, OTUs with galaxy solution. *Bioinformatics* 34:1287–1294. <https://doi.org/10.1093/bioinformatics/btx791>
78. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>
79. Hammer Ø, Harper D, Ryan P. 2001. PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electr* 4:1–9.
80. Chao A. 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43:783–791. <https://doi.org/10.2307/2531532>
81. Shannon CE. 1948. A mathematical theory of communication. *Bell Labs Tech J* 27:623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>
82. Simpson EH. 1949. Measurement of diversity. *Nature* 163:688–688. <https://doi.org/10.1038/163688a0>

83. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara RB, Simpson G, Solymos P, Stevens MHH, Szöcs E, Wagner H. 2020. *vegan*: community ecology package
84. Bolar K. 2019. *STAT*: interactive document for working with basic statistical analysis
85. R Core Team. 2020. *R*: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. <https://www.R-project.org/>.
86. Clarke KR. 1993. Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* 18:117–143. <https://doi.org/10.1111/j.1442-9993.1993.tb00438.x>
87. Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 27:325–349. <https://doi.org/10.2307/1942268>
88. Elzhov TV, Mullen KM, Spiess A-N, Bolker B. 2016. Package *minpack.lm*: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds
89. Harrell F, Dupont C. 2018. *Hmisc*: Harrell miscellaneous
90. Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. *aLeaves* facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced Interactivity. *Nucleic Acids Res*. 41:W22–W28. <https://doi.org/10.1093/nar/gkt389>
91. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17:540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
92. Talavera G, Castresana J, Kjer K, Page R, Sullivan J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 56:564–577. <https://doi.org/10.1080/10635150701472164>
93. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R, Teeling E. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>
94. Ning D. 2021. iCAMP: infer community assembly mechanisms by phylogenetic-bin-based null model analysis.
95. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010. *Picante*: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464. <https://doi.org/10.1093/bioinformatics/btq166>
96. Dray S, Dufour A-B. 2008. The *ade4* package: implementing the duality diagram for ecologists. *J Stat Soft* 22:1–20. <https://doi.org/10.18637/jss.v022.i04>
97. Hardy OJ. 2008. Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *J Ecol* 96:914–926. <https://doi.org/10.1111/j.1365-2745.2008.01421.x>