Genetic diversity in terrestrial subsurface ecosystems impacted by geological degassing

Till L. V. Bornemann1, Panagiotis S. Adam1, Victoria Turzynski1, Ulrich Schreiber2, Perla Abigail Figueroa-Gonzalez1, Janina Rahlff1,7, Daniel Köster3, Torsten C. Schmidt3,4, Ralf Schunk5, Bernhard Krauthausen6 & Alexander J. Probst1,4

Earth’s mantle releases 38.7 ± 2.9 Tg/yr CO2 along with other reduced and oxidized gases to the atmosphere shaping microbial metabolism at volcanic sites across the globe, yet little is known about its impact on microbial life under non-thermal conditions. Here, we perform comparative metagenomics coupled to geochemical measurements of deep subsurface fluids from a cold-water geyser driven by mantle degassing. Key organisms belonging to uncultivated Candidatus Altiarchaeum show a global biogeographic pattern and site-specific adaptations shaped by gene loss and inter-kingdom horizontal gene transfer. Comparison of the geyser community to 16 other publicly available deep subsurface sites demonstrate a conservation of chemolithoautotrophic metabolism across sites. In silico replication measures suggest a linear relationship of bacterial replication with ecosystems depth with the exception of impacted sites, which show near surface characteristics. Our results suggest that subsurface ecosystems affected by geological degassing are hotspots for microbial life in the deep biosphere.

https://doi.org/10.1038/s41467-021-27783-7

1Environmental Microbiology and Biotechnology, Faculty of Chemistry, University Duisburg-Essen, Essen, Germany. 2Department of Geology, University Duisburg-Essen, Essen, Germany. 3Instrumental Analytical Chemistry and Centre for Water and Environmental Research (ZWU), University of Duisburg-Essen, Essen, Germany. 4Centre of Water and Environmental Research (ZWU), University of Duisburg-Essen, Universitätsstraße 5, Essen, Germany. 5Geyser-Center, Andernach, Germany. 6Institute of Applied Geosciences, Karlsruhe Institute of Technology, Karlsruhe, Germany. 7Present address: Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Department of Biology and Environmental Science, Linnaeus University, Kalmar, Sweden.

✉email: alexander.probst@uni-due.de
The continental subsurface is a huge reservoir for life, hosting about 60% of all microorganisms on Earth. Carbon, nitrogen, and sulfur turnover by these microorganisms have a vast contribution to all biogeochemical cycles on the planet. In addition to the great number of microorganisms, subsurface ecosystems can accommodate a large diversity of different bacteria and archaea, with even single ecosystems containing representatives of almost all known bacterial phyla. Subsurface ecosystems are categorized as either detrital or productive, depending on whether buried organic carbon or inorganic carbon are the main carbon sources of the community. Since no light is available as an energy source in the deep biosphere, alternative electron donors to water like hydrogen (H₂) or sulfide (H₂S) are used to fuel mostly anaerobic carbon fixation pathways such as the Wood–Ljungdahl pathway. Subsurface lithoautotrophic microbial communities have been reported for many terrestrial ecosystems including the Fennoscandian Shield, the Columbia River Basalt, the Witwatersrand Basin, and subsurface fluids discharged by Crystal Geyser. While these subsurface ecosystems are usually dominated by bacteria, one exception are archaea belonging to the Alti-1 clade of the Ca. Altiaarchaeota. Alti-1 form biofilms using their characteristic nano-grappling hooks (hami). The other clade, Alti-2, is more widespread and diverse but found at lower abundances in their ecosystems. Ca. Altiaarchaeota live autotrophically using the Wood-Ljungdahl carbon fixation pathway, which was the most dominant carbon fixation pathway prior to the evolution of photosynthesis.

Chemolithoautotrophic life in subsurface ecosystems necessitates the presence of adequate electron donors like hydrogen, hydrogen sulfide, or methane. One source of such gases can be Earth’s mantle, which also releases 38.7 ± 2.9 Tg/yr of oxidized carbon, mainly in form of carbon dioxide (CO₂), into the crust and the atmosphere. This process, also termed mantle degassing, is the transition of volatiles from the mantle (super-)critical to the subcritical zone of the upper crust fueled by lower degassing, is the transition of volatiles from the mantle (super-)critical to the subcritical zone of the upper crust fueled by lower pressure of volatiles near the surface compared to the mantle. Modern Earth has few areas with active mantle degassing, which are usually restricted to terrestrial volcanoes, subduction zones, or hydrothermal vents in oceans. Modern Earth has few areas with active mantle degassing, which are usually restricted to terrestrial volcanoes, subduction zones, or hydrothermal vents in oceans. At hydrothermal vents, chemolithoautotrophs initiate the microbial trophic network and proliferate at high rates leading to high microbial cell numbers. While volcanic sites and hydrothermal vent fields have been studied fairly thoroughly regarding both their microbial community composition and activity, little is known about deep subsurface ecosystems with low temperatures and predicted impacts on microbial communities. Beulig and co-workers reported an increase in deep carbon fixation and found evidence that the CO₂ from the degassing is indeed incorporated into biomass based on 13C-labeled CO₂. Along with fermentation processes, the pathways for the turnover of organic carbon were similar in both systems, while the microbial diversity of soils in mofettes was lower compared to controls. Carbon and sulfate respiration were enriched during degassing, while aerobic respiration declined and acetogenesis were suggested to play a critical role in these systems. However, these studies were limited to the upper 50-cm of Earth’s critical zone, and the influence of mantle degassing on mesophilic microbial communities in the deep subsurface including their metabolic capacity and activity has not been investigated so far.

The cold-water (291 K) Geyser Andernach is located in the Rhine Valley near Koblenz in western Germany and is driven by gases discharged from the mantle. Since 2001 the geyser has had intact tubing, thus tapping into a unique ecosystem. Once released by a mechanical shutter, the gases from the mantle (mainly CO₂) permeating the groundwater cause the eruption of cold subsurface fluids sourced from a uniform aquifer system. Thus, Geyser Andernach is an ideal ecosystem to investigate how mantle degassing shapes mesophilic microbial life in the subsurface.

Here, we used a combination of long-term geochemical characterization coupled to genome-resolved metagenomics to investigate the geyser’s microbial community. To analyze how mantle degassing impacts mesophilic microbial communities, we set the bacterial replication index values, minimal generation times, and microbial metabolism abundances in Geyser Andernach into relation to 16 other deep continental subsurface ecosystems across the globe. We identified a pattern of decreasing replication indices but shorter minimal generation times with increasing depth. Sites impacted by mantle degassing showed similar replication indices and generation times as near-surface sites, rendering them hotspots for microbial activity in the subsurface. Comparative genomics applied to a key player at sites impacted by geological degassing (Ca. Altiaarchaeum sp.), revealed that the slow evolutionary rate present in this phylum might be counteracted by horizontal gene transfer (HGT) and gene loss events in this organism group.

Results

Geyser Andernach provides access to a stable ecosystem impacted by mantle degassing. Geyser Andernach was drilled to a depth of 351 m in 1903 tapping into a shale-hosted aquifer with quartz veins. Its eruptions are driven by mantle degassing and can be controlled via mechanical shutters (a diagram of the plumbing system is provided in Supplementary Fig. 1). Geochemical measurements averaged over 14 years have demonstrated that the subsurface fluids provide a constant environment (Supplementary Table 1). The gaseous and ionic composition of the geyser showed the predominance of CO₂ in the system and previously reported traces of hydrogen and hydrogen sulfide. Prominent electron donors and acceptors were determined to be hydrogen and ferric iron as well as sulfate, respectively. To investigate the microbial community in subsurface fluids impacted by mantle degassing, we sampled two eruptions of Geyser Andernach and collected the planktonic fraction of microorganisms onto three individual 0.1-µm filters. Metagenomic sequencing of the community resulted in ~7 billion bp per sample (5% SD), covering about 80% of the microbial diversity as estimated by Nonpareil (Supplementary Fig. 2). Reads were assembled into 921,520 scaffolds on average (20% SD, for further statistics please see Supplementary Table 2). Approximately 75% of the reads (2.6% SD) mapped back the assembly providing evidence that the reconstructed metagenome is representative of the planktonic community at the time of sampling. The community composition based on ribosomal protein S3 (rpS3) sequences assembled from the metagenome displayed a fairly restricted diversity consisting of 52 organisms, which spanned twelve phyla (Fig. 1). The core community was composed of 15 organisms detected via rpS3 across all three metagenomes (Fig. 1), and they accounted for 42.8% (1.3% SD) of the total relative abundance of the community. For 20 of these 52 microorganisms, we reconstructed high-quality genomes with at least 70% estimated completeness (and less than 10% estimated contamination, details in Supplementary Data 1). The most
abundant species recruited 42.8% (1.3% SD) of the metagenomic reads and belonged to the Ca. phylum Altiarchaeota\textsuperscript{5} (in the following denoted as Ca. Altiarchaeum GA) and specifically grouped within the Alti-1\textsuperscript{14} clade. The second most abundant organism was classified as Caldiserica, which were originally known to inhabit hot springs\textsuperscript{38} but were recently also detected in subsurface ecosystems populated by mesophiles\textsuperscript{5,11}.

We verified that bacteria in this community were replicating at the time of sampling using in situ replication index values. Replication index values are calculated from the difference of sequencing coverage between the origin of replication and terminus of replication. Proliferating organisms replicate their genomes with multiple replication forks starting at the replication origin and thus contributing more to sequencing reads. In our study, these index values ranged between 1.4 and 1.5, indicating that 40–50% of those microbial populations, whose iRep values were calculated, underwent genome replication at the time of sampling. Microscopic cell counts of organisms from the subsurface fluids ranged from $2.7 \times 10^6$ to $4.2 \times 10^6$ (average $3.5 \times 10^6$) cells ml\textsuperscript{−1} (Supplementary Fig. 3) and displayed...
Replication index values and maximal growth rates across multiple deep continental subsurface ecosystems. To investigate if mantle degassing has an impact on microbial replication in the continental subsurface, we used in situ replication index values (iRep) of bacterial genomes and maximal growth rate estimates of bacterial and archaeal genomes. We first investigated if iRep can be used as a measure of replication by comparing groundwater with sediments, as microbes in sediments are known to be more active\(^5\). Indeed, iRep suggested a significantly higher replication of microbes in sediments than groundwater (\(p\)-value < 10\(^{-3}\)). Replication measures from Geyser Andernach were then compared with those from other public datasets from deep subsurface environments of varying depth (overview of samples and ecosystems is provided in Supplementary Table 4). The sampling depth varied from 0 m below ground (cave systems) to 3140 m depth. We reconstructed genomes of previously unbinned metagenomes resulting in 560 newly assembled and classified prokaryotes (Supplementary Data 1) representing 415 different organisms after dereplication. Combined with genomes and iRep results from previous studies\(^4,5,13\), we leveraged in situ replication measures for 895 bacteria (Supplementary Data 2) spanning the vast majority of all known bacterial phyla (see Supplementary Data 5). The average iRep value of bacteria of the individual ecosystems correlated negatively and highly significantly with sample depth across all individual iRep values (Pearson’s test, \(p\)-value < 10\(^{-8}\)) and across median per sampled ecosystem (\(p\)-value < 0.0007, Fig. 2, Supplementary Table 5). In other words, the deeper the origin of the retrieved sample, the lower the genome replication measure.

In particular, organisms with the capacity of carbon fixation (cor = −0.47), sulfur oxidation (cor = −0.46), or of metabolizing hydrogen (cor = −0.45) contributed to this observation (correlations are summarized in Supplementary Table 6). Samples impacted by high CO\(_2\) concentrations, either solely from mantle degassing (this study) or from both mantle degassing and thermal activity\(^3\), were outliers in this correlation analysis. In fact, iRep measures of bacteria in these samples were significantly higher than iRep measures of other subsurface samples (\(p\)-value < 10\(^{-15}\)) and nearly reached values of samples that are close to Earth’s surface (Fig. 2). When excluding these samples from the correlation analysis with depth, the respective correlation coefficient decreased from −0.20 to −0.28 (\(p\)-value < 10\(^{-8}\)). We also tested how the availability of oxygen influences genome replication measures of bacteria in the continental subsurface. iRep values were on average 0.09 higher for bacteria in oxygenic samples (\(p\)-value < 10\(^{-8}\)) meaning that about 9% more of the bacteria were undergoing genome replication.

While iRep values indicated that there is less ongoing replication in deeper regions of the subsurface, they do not allow any inference about the speed at which organisms are replicating. Thus, we also calculated maximal possible growth rates, i.e., minimal generation times, based on the codon usage bias between constitutionally expressed ribosomal proteins and the rest of the genes per genome using growthtp\(^{40}\). Correlation analyses of these maximal growth rates with the sampling depth revealed that the maximally possible replication speed increases, i.e., shorter doubling times, with increasing depth (\(p < 0.0011\), cor = −0.143, Supplementary Fig. 4).

Conserved chemolithoautotrophic metabolism of subsurface microbial communities. Since bacterial replication is predicted to differ between sites impacted by mantle degassing and reference sets, we investigated if the general metabolism for carbon, nitrogen, and sulfur turnover of entire communities is adapted to high-CO\(_2\) subsurface environments. We searched for key enzymes for metabolic pathways across our entire metagenomic assemblies (Supplementary Table 2) and used the abundance of scaffolds that carried a key enzyme as a relative abundance measure of the respective metabolism (Fig. 3, Supplementary Fig. 5, Supplementary Fig. 6). The core metabolism remained relatively stable across all tested ecosystems. We performed both Student’s \(t\)-tests and Kruskal–Wallis tests along with equivalence testing to determine whether there was a significant difference between high-CO\(_2\) and non-high-CO\(_2\) metabolisms and could only detect a significant difference in the nitrite reduction metabolism (Kruskal–Wallis group comparison, \(p\)-value = 6 × 10\(^{-4}\), details on tests in Supplementary Table 7). Consequently, and in congruence with previous studies investigating the metabolic diversity in a subsurface aquifer\(^41\), little difference exists in the metabolic potential between regular subsurface microbial communities and those at sites impacted by mantle degassing, although the indigenous organisms at these sites appear to have higher replication index values.

Biogeography and functional adaptations of deep subsurface Ca. Altiaarchaeota. Key organisms in continental subsurface ecosystems impacted by geological degassing belong to the Ca. phylum Altiaarchaeota due to their high abundance. Ca. Altiaarchaeota can currently be divided into two clusters, Alti-1 and Alti-2, with the latter having a broader metabolic variability than Alti-1\(^{14}\). In the following, we are going to refer to Alti-1 Altiaarchaeota as Ca. Altiaarchaeota. However, organisms of the Ca. Altiaarchaeota is one that can dominate entire ecosystems, as shown for multiple sites across the globe\(^5,12,13\). Nearly all of the ecosystems are dominated by Ca. Altiaarchaeota has all been reported to have high CO\(_2\) partial pressure or great amounts of carbonate deposits\(^32\). The average nucleotide (ANI) and amino acid (AAI) identity of all so-far recovered Ca. Altiaarchaeota genomes indicated that they belong to the same genus (Supplementary Fig. 7), although 16S ribosomal RNA gene similarity suggested the same species. When correlating the genomic differences based on ANI to the geographical distance between sampling sites of the Ca. Altiaarchaeota genomes, a highly significant negative correlation (Pearson, cor = −0.77, \(p = 9 \times 10^{-4}\)) could be observed, indicating that a greater distance led to greater dissimilarity (Supplementary Fig. 7). We challenged this observation by using robust phylogenetic analyses based on a supermatrix of 30 ribosomal proteins and found that Ca. Altiaarchaeota cluster based on geographical sampling site going all the way to continent-scale (Fig. 4C, Supplementary Fig. 8). However, we did not observe any biogeographic pattern for Ca. Altiaarchaeota of the Alti-2 clade, which mainly occurs in ocean sediments\(^14\). Based on Hidden Markov Model (HMM) profiles of key chemolithoautotrophic genes of Alti-2 and Alti-1 genomes, some of which we newly reconstructed from public datasets, we identified substantial differences particularly in the hydrogen metabolism (Fig. 4B, details on Ca. Altiaarchaeota genomes in Supplementary Table 3). However, Alti-2 showed a significantly smaller minimal generation time than Alti-1 (\(U\)-test \(p < 0.0024\); Supplementary Fig. 9).
Since Ca. Altiiarchaea showed a strict biogeographic pattern, we further investigated their differences in metabolic capacities in depth using a genome model published previously12 (Fig. 5). We identified that all Ca. Altiiarchaea share a central NAD(P)H-based Wood–Ljungdahl pathway for carbon fixation and carbon monoxide utilization. The main difference of Ca. Altiiarchaeum GA to the reference genome Ca. Altiiarchaeum hamiconexum12 was the presence of genes for a NiFe hydrogenase (Fig. 4B), which seems to be a specific adaptation to hydrogen-containing gases from the mantle. Indeed, we identified that this NiFe hydrogenase existed in multiple other Ca. Altiiarchaea and was lost in Ca. Altiiarchaeum hamiconexum from IMS. The phylogenetic relatedness revealed that NiFe-hydrogenases of Altii-1 were sister to those of Altii-2 suggesting a conservation of this key enzyme in their last common ancestor (tree is provided in Supplementary Data 7). Other genes are affected by gene loss across Ca. Altiiarchaea encoded for proteins, which function as mechanosensitive channels, desulfurodoxin, polysaccharide biosynthesis enzymes, and some peptidases and glycosylhydrolases (Supplementary Data 8–15). By contrast, ruberythrine and multiple peptidases spanning the families C44 (precursor of amidophosphoribosyltransferase), M06 (metalloendopeptidases), and C01b (endo- and exo-peptidases) were horizontally

Fig. 2 In situ bacterial replication rates across subsurface ecosystems ordered by ecosystem depth. The figure depicts a beeswarm plot of iRep values of genomes (x-axis) across ecosystems (y-axis) with genomes colored according to their predicted metabolic potential and the black dot representing the median iRep value (individual iRep values in Supplementary Data 2). C represents carbon, N2 nitrogen, H2 hydrogen, O2 oxygen, and S sulfur. Colored squares depict the sample type. Samples impacted by geological degassing and a sediment sample along with the respective aquifer sample are plotted separately. The top y-axis shows the sampling depth of the different ecosystems (Supplementary Table 5). In total, 895 genomes were used for this analysis with ≥ 70% completeness and ≤ 10% contamination based on 51 bacterial and 38 archaeal single-copy genes. The order of samples is given in Supplementary Table 5. p-Values are derived from two-sided student’s t-tests. The exact p-values from top to bottom are \( p < 2.2 \times 10^{16} \) (minimal value in R) and \( p = 0.0003934 \), respectively.
acquired by Ca. Altiarchaea species, mostly from the bacterial domain (Supplementary Data 16–19).

This indicates an extreme degree of biogeographic provincialism across Earth. The small genetic divergence of Ca. Altiarchaeota of the Alti-1 clade reach high cell densities in the CO₂ subsurface ecosystem and represent the main primary producers similar to the other high-CO₂ aquifer system Crystal Geyser, which additionally harbors a tremendous amount of bacterial diversity but also taps into three different aquifer ecosystems⁵,¹¹. The predicted higher minimal determined constant cell division¹² implies a very slow evolutionary rate of these organisms. However, gene loss and HGT in Ca. Altiarchaea suggests compensation for these slow evolutionary rates potentially providing a substantial advantage over other organisms in deep subsurface environments.

**Discussion**

Modeling of current cell counts estimates the number of prokaryotic microorganisms in the continental subsurface to 2 to 6 × 10^29,¹ which amounts to 60% of the prokaryotic life on our planet². The diversity of microorganisms declines with sampling depth in the continental subsurface³. Our metagenome assemblies showed the same trend in diversity change (based on the rps3 marker gene, cor = −0.40, p-value = 0.021, Supplementary Fig. 10). This indicates that they are representative of general subsurface microbial communities and were consequently used to establish a genome database to calculate genome replication index values and minimal generation times across various subsurface ecosystems. These metrics revealed an apparent contradiction, with both replication index values and minimal generation times decreasing, thus indicating that organisms in the deep biosphere can replicate faster though they replicated less at the time of sampling. Prior studies⁴³,⁴⁴ observed a reduction in microbial load with marine sediment depth and age, indicating that communities in older sediments were probably formed by members of surface communities that have a higher degree of persistence compared to others. Thus, subsurface communities would not be formed by actively replicating organisms but instead be shaped by the differing mortality of surface community members⁴³,⁴⁴. The upper ten centimeters of sediment were found to be an exception showing active proliferation⁴⁵. Although we analyzed many different ecosystems, our data do not allow drawing conclusions about the impact of mortality shaping subsurface microbial communities as they originate from different geologic formations.

However, our observed decrease in replication measures with sampling depth does agree with these prior observations of a reduction of microbial load with depth and indicates that replication is occurring, albeit with fewer replication forks in the subsurface compared to near-surface ecosystems. On the other hand, the genome structures indicated a faster ability to replicate for organisms in the deep subsurface. This faster possible generation times with depth can be explained by the strategy employed by subsurface microorganisms recently termed as "halt and catch fire"⁴⁶. This strategy refers to an adaptation to nutrient-poor environments like the deep subsurface, where organisms need to adapt to utilize short bursts of available nutrients and thus replicate fast during times when nutrients are available. Sites impacted by geological degassing showed a similar pattern compared to surface samples, both in terms of replication index values and minimal generation time estimates. This could be caused by the unique geology of sites impacted by geological and thermal degassing. In these fracture-controlled aquifers, which are characterized by solid rock formation-embedded channels, flows can reach up to multiple magnitudes greater speeds than flows in comparable sediment-hosted aquifers. Thus, the availability of reduced mantle gases like H₂ and H₂S as microbial electron donors highlights the absence of nutrient bursts and the presence of a continuous nutrient flow similar to biomes on Earth’s surface.

At Geyser Andernach, Ca. Altiarchaeota of the Alti-1 clade reach high cell densities in the CO₂ subsurface ecosystem and represent the main primary producers similar to the other high-CO₂ aquifer system Crystal Geyser, which additionally harbors a tremendous amount of bacterial diversity but also taps into three different aquifer ecosystems⁵,¹¹. The predicted higher minimal
generation time for the Alti-1 clade compared to their sister clade Alti-2 is likely caused by their higher costs of living. In contrast to their sister clade, *Ca*. Altiarchaeota (Alti-1) live in biofilms, likely granting them increased survivability against a multitude of biotic and abiotic factors (see Olsen 2015 for a review on biofilm resistance). But this increased resistance also comes with a cost of requiring the synthesis of hundreds of their characteristic cell surface appendages called hamis as well as other materials making up the extracellular polymeric substances matrix. In addition, *Ca*. Altiarchaeota all need to assimilate CO₂ via the Wood–Ljungdahl pathway instead of also supplementing their carbon compounds by taking up organic carbon compounds as only gases can freely penetrate the biofilms. Thus, their proliferation would presumably be much more expensive than for their planktonic sister clade. This leads to the hypothesis that not replication speed but energy requirements limit *Ca*. Altiarchaeota proliferation, making an optimization of the codon code to increase replication speed unnecessary.

The abovementioned hypothesis regarding the replication speed of *Ca*. Altiarchaeota would also align well with their strict
biogeography. The clustering by continent of origin (North America, Europe, Asia), also reproducible in ANI and AAI (Supplementary Fig. 7), indicates strict provincialism. As dispersal via the surface is unlikely due to the high oxygen sensitivity of Ca. Altiarchaea\textsuperscript{12}, plate tectonics could have been a viable alternative dispersal route providing ample opportunities for the common ancestor to distribute to North America and Europe. Plate tectonics has recently been implicated as the potential dispersal route for Ca. Desulforudis audaxviator to Africa, North America, and Eurasia between 55 and 165 Myr\textsuperscript{48}. The dispersal of Ca. Altiarchaea could have occurred within the Phanerozoic, starting with the early Devonian (~400 Myr), when the continental margins Laurentia and Baltica, which form today’s North America and Europe, respectively, collided to form Laurasia\textsuperscript{49,50}. The dispersal of Ca. Altiarchaea could have occurred within the Phanerozoic, starting with the early Devonian (~400 Myr), when the continental margins Laurentia and Baltica, which form today’s North America and Europe, respectively, collided to form Laurasia\textsuperscript{49,50}.Japan, on the other hand, has not been in contact with those margins since the break-up of Rodinia 750–600 Myr ago\textsuperscript{51}, thus making dispersal to Japan during the Phanerozoic unlikely. As European and Japanese Ca. Altiarchaea is indicated to have a common ancestor, one possible route of dispersal from Europe to Japan could be across the Siberian plate through China in the early Mesozoic and then transferal to Japan during the plate processes, which uplifted the Japanese islands from the sea 25 Myr ago. Future studies are necessary to recover Ca. Altiarchaea genomes from Asia further underpin this hypothesis of dispersal since current public datasets from this continent are substantially underrepresented in databases.

The strict biogeography of the Ca. Altiarchaea is reflected by the conserved core metabolism, with most pathways being present in every Ca. Altiarchaea genome and indicate a slow evolving genus. However, observed putative gene loss and gene transfer events in investigated Ca. Altiarchaea populations indicate a compensatory strategy to counteract the slow evolutionary rate. This observed gene loss and transfer might be exuberated by the exclusive living in biofilms, which have generally been known as hotspots of HGT for Bacteria\textsuperscript{52}. The genes in Ca. Altiarchaea acquired via HGT are mainly from the bacterial domain, an evolutionary process frequently occurring in nature\textsuperscript{53}. This HGT likely took place in the subsurface due to the immobility of Ca.

**Fig. 5 Metabolic capacities of Ca. Altiarchaeum pangenome.** Previously identified genes in Ca. Altiarchaeum hamiconexum IMS\textsuperscript{12} was used as the basis to query the other genomes of known Altiarchaea clade members (see Fig. 4 for all members used in this analysis). To expand the predictable metabolic capacity of the genomes, METABOLIC\textsuperscript{86} was used to annotate genes, which mainly resulted in peptidases and glycosylhydrolases. If multiple genomes copies per site were available, they were all used to query for the respective genes. All gene functions are listed in Supplementary Data 3.
Altiaarchae is mediated by the anchoring of cells via their hami. Consequently, our analyses provide evidence that subsurface ecosystems impacted by geological degassing can be hotspots of microbial life and of increased evolutionary rates bolstered by lateral gene transfer across domains.

Methods

Geological setting. The cold-water Geyser Andernach is located 2 km downstream of Andernach (Rhine kilometer 615) on a 0.21 km² peninsula called Namedyer. The small peninsula is part of the Pleistocene terrace which is covered by a thin layer of fluvial Holocene deposits. The thickness of the Quaternary layer varies from 14 m (drilling 2001) to 20.75 m (drilling 1903) and 24.2 m (drilling 1955) in the vicinity of the cold-water geyser. Beneath the Quaternary deposits follow Devonian rock formations of low metamorphic shale, such as clayish shale and intercalated minor layers of quartzitic sandstones; the thickness of these series is up to 5000 m.

The peninsula is located in the Middle Rhine Valley, which is a part of the European Cenozoic Rift System. This rift system runs between the cities Bingen and Bonn in SE–NW direction and crosses the Variscan massif of any Rhenish Massif. Located at the SE edge of the lower Middle Rhine Valley, Geyser Andernach is situated on the intersection of two major fault structures: about one km to the NW the Variscian Sieg thrust fault running SW–NE crosses the Rhine Valley and can be traced for over 100 km from the Eifel area to the Westerwald. This fault shows a vertical displacement of several thousand meters and occurred during the Variscan orogenesis, thus bringing rocks of the middle Siennien stage in lateral contact with the lower Emsian stage. About 2 km to the SE the lower Middle Rhine valley is morphologically separated from the adjacent intraplate Tertiary Neuwied basin by an approx. 100 m vertical displacement caused by the SW–NE trending Andernach fault.

The Andernach fault and the Sieg thrust fault were in post–Variscian time intersected and 200–300 m displaced by a SE–NW trending dextral strike-slip fault. The fault is supposed in the river Rhine bed and covered by Quaternary deposits. The horizontal movement was probably combined with shear strain and cataclastic rocks in the vicinity of the fault. This fault is the cause for pathways of mantle gases to reach the subsurface aquifers and ultimately the atmosphere.

Starting in the Tertiary, a mantle plume under the Eifel area caused an uplift of the Rhenish massif during the last two million years and is the driving force for the volcanic activity in the Quaternary Eifel area since 700 k years.

The mantle plume is the basic requirement for the rise of magma under and into the crust, whereby magmatic gases are released.

Sampling and geochemical measurements. The mesophilic and CO₂-driven Geyser Andernach (50.448888°, 7.73555°E) in western Germany was sampled on 21 February 2018 in a collection of erupting water in sterile, DNA-free water. The metagenomes were extracted from the sampled eruptions and the tu...
Metabolic potential predictions. A set of HMM with respective score thresholds for chemolithoautotrophic key enzymes was used to predict the metabolic potential of recovered genomes and overall in entire assemblies (see Supplementary material for more detailed information).

Biogeographical analysis. The R package sp2 was used to calculate the geographical elliptical distance between two sampling sites (based on longitude/latitude), in which putative genomes of the Ca. Altiarchaeaeas subclade Alti-1 was identified. The average nucleotide identities (ANI) between all available putative genomes of the Ca. Altiarchaeaeas subclade Alti-1 was calculated using the ANI calculator with default parameters. Correlations between geographical distance and ANI were done using Pearson’s r4.

Genome comparison of Ca. Altiarchaeota. Genes of all Ca. Altiarchaeota genomes were blasted against each other (E-value: $10^{-5}$) and matches were filtered to matches with the similarity (alignment length x density)/query length thresholds of 2400, 60%, 2000, and 3.72$^{100}$ was used to visualize the networks at the respective similarity thresholds.

Metabolic network of Ca. Altiarchaeota (Alti-1). The annotated genes from Probst et al. were used as the basis to identify homologs in other Alti-1 genomes using an E-value of $10^{-5}$ as the cutoff. If multiple versions of a genome were available, their results were concatenated. In addition, genomes were annotated using METABOLIC4, mainly incorporating annotations for glycosyl hydrolases, peptidases, and aminotransferases.

Phylogenomic analysis of Ca. Altiarchaeotina. Amino acid sequences and annotations for Alti-1 ORFs plus one Alti-2 serving as outgroup were predicted using HMMER 3.2.188 for homologs of 30 enzymes, and a set of HMM with respective score thresholds were used as the basis to identify homologs in other Alti-1 genomes using the ANI calculator. The average nucleotide identities (ANI) between all available putative genomes of the Ca. Altiarchaeaeas subclade Alti-1 was calculated using the ANI calculator with default parameters. Correlations between geographical distance and ANI were done using Pearson’s r4.

Data availability

Raw sequencing data and MAGs from Geyser Anndernach have been deposited at SRA under Genbank, respectively, and are available under the BioProject PRJNA627655. MAGs binned from additional ecosystems have been deposited at Genbank in the BioProject PRJNA767587. Individual BioSample IDs of all MAGs are listed in Supplementary Data 1 and individual SRA accession codes are listed in Supplementary Table 4.

References

1. Magnabosco, C. et al. The biomass and biodiversity of the continental subsurface. Nat. Geosci. 11, 707–717 (2018).
2. Fleming, H.-C. & Vuets, S. Bacteria and archaea on Earth and their abundance in biofilms. Nat. Rev. Microbiol. 17, 247–260 (2019).
3. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive Earth’s biogeochemical cycles. Science 320, 1034–1039 (2008).
4. Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. 7, 1–11 (2016).
5. Probst, A. J. et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat. Microbiol. 3, 328–336 (2018).
6. Castelle, C. J. et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr. Biol. 25, 690–701 (2015).
7. Stevens, T. Lithoautotrophy in the subsurface. FEMS Microbiol. Rev. 20, 327–337 (1997).
8. Stevens, T. O. & McKinley, J. P. Abiotic controls on H2 production from basalt–water reactions and implications for aquifer biogeochemistry. Environ. Sci. Technol. 34, 826–831 (2000).
9. Nyssönen, M. et al. Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. ISME J. 8, 126–138 (2014).
10. Lai, M. C. Y. et al. An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur–fumarole autotrophic denitrifiers. Proc. Natl Acad. Sci. USA 113, E7927–E7936 (2016).
11. Probst, A. J. et al. Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2 concentrations. Environ. Microbiol. 19, 459–474 (2017).
12. Probst, A. J. et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. Nat. Commun. 5, 5497–5504 (2014).
13. Hornsdorf, A. W. et al. Potential for microbial H2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. ISME J. 11, 1915–1929 (2017).
14. Bird, J. T., Baker, B. J., Probst, A. J., Podar, M. & Lloyd, K. G. Culture-independent genomic comparisons reveal environmental adaptations for Altiarchaeotes. Front. Microbiol. 7, 1221 (2016).
15. Moisil, C., Rachel, R., Briegel, A., Engelhardt, H. & Huber, R. The unique structure of archaeal ‘hami’, highly complex cell appendages with nanoparticle grappling hooks: unique structure of archaeal ‘hami’. Mol. Microbiol. 56, 361–370 (2005).
16. Wood, H. G. Life with CO or CO2 and H2 as a source of carbon and energy. FASEB J. 5, 156–163 (1991).
17. Gutiérrez-Preciado, A. et al. Functional shifts in microbial mats recapitulate early Earth metabolic transitions. Nat. Ecol. Evol. 2, 1700–1708 (2018).
18. Adam, P. S., Borrell, G. & Gribaldo, S. An archaeal origin of the Wood–Ljungdahl H4 MPT branch and the emergence of bacterial methylotrophy. Nat. Microbiol. 4, 2155–2163 (2019).
19. Aiuppa, A., Fischer, T. P., Plank, T. & Bani, P. CO2 flux emissions from the Earth’s most actively degassing volcanoes, 2005–2015. Sci. Rep. 9, 5442 (2019).
20. Bräuer, K., Kämpf, H., Niedermann, S. & Strauch, G. Indications for the existence of different magmatic reservoirs beneath the Eifel area (Germany): A multi-isotope (C, N, He, Ne, Ar) approach. Deep Carbon. 336 (2018).
21. Werner, C. et al. Carbon dioxide emissions from subaerial volcanic regions: two decades in review. in Deep Carbon (eds Orcutt, B. N., Daniel, I. & Shannon–Wiener index3, and equivalence testing using TOSTER94. As the upper and lower equivalence boundaries for equivalence testing of two groups, we used the effect size the CO2-poor sample group had a 33% power to detect as recommended previously95. Results were visualized using ggplot296.

Methods for DAPI staining, cell counting, geochemical measurements are provided in the Supplementary Methods.
24. Loreto, M. F., Italiano, F., Deponte, D., Facchin, L. & Zgur, F. Mantle degassing along strike-slip faults in the Southeastern N. Geophys. Res. Solid Earth 120, 2200–2211 (2015).

26. Lee, H. et al. Mantle degassing along strike-slip faults in the Southeastern N. Geophys. Res. Solid Earth 120, 2200–2211 (2015).

27. Fullerton, K. M. et al. Plate Tectonics Drive Deep Biosphere Microbial Community Composition. https://doi.org/10.31225/osf.io/g5y79 (2019).

28. Hedrick, D. B., Pledger, R. D., White, D. C. & Baross, J. A. In situ microbial ecology of hydrothermal vent sediments. FEMS Microbiol. Lett. 101, 1–10 (1992).

29. Schrenk, M. O., Holden, J. F. & Baross, J. A. Magma-to-microbe networks in the context of sulfide-hosted microbial ecosystems. Wash. DC Am. Geophys. Union Geophys. Monogr. Ser. 178, 233–258 (2008).

30. Ding, J. et al. Microbial community structure of deep-sea hydrothermal vents on the ultrashallow spreading Southwest Indian Ridge. Front. Microbiol. 8, 1012 (2017).

31. Tu, T.-H. et al. Microbial community composition and functional capacity in a terrestrial ferruginous, sulfate-depleted mud volcano. Front. Microbiol. 8, 2137 (2017).

32. Galambos, D., Anderson, B. E., Revelleaud, J. & Huber, J. A. Genome-resolved metagenomics and metatranscriptomics reveal niche differentiation in functionally redundant microbial communities at deep-sea hydrothermal vents. Environ. Microbiol. 21, 4395–4410 (2019).

33. Rodríguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R. & Konstantinidis, K. T. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. mSystems 3, e00393–18 (2018).

34. Mori, K., Yamaguchi, K., Sakiyama, Y., Urabe, T. & Suzuki, K. Caldisericum nov. and Caldisericia classis nov. in the Eurosiberian foreland lithosphere. Terra Nova 29, 693–713 (1999).

35. Nelson-Sathi, S. et al. Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature 517, 77–80 (2015).

36. Beulig, F. et al. Altered carbon turnover processes and microbiomes in soils altered by oil seepage and subsea community with high functional redundancy inhabits the cold, oxic floor aquifer. Int. J. Syst. Evol. Microbiol. 64, 2299–2311 (2014). https://doi.org/10.1099/ijs.0.078926-0.

37. Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R. & Konstantinidis, K. T. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. mSystems 3, e00393–18 (2018).

38. Mori, K., Yamaguchi, K., Sakiyama, Y., Urabe, T. & Suzuki, K. Caldisericum nov. and Caldisericia classis nov. in the Eurosiberian foreland lithosphere. Terra Nova 29, 693–713 (1999).

39. Nelson-Sathi, S. et al. Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature 517, 77–80 (2015).

40. Vieira-Silva, S. & Rocha, E. P. C. The systemic imprint of growth and its uses in ecological (meta)genomics. Microb. Ecol. 72, 1343–1353 (2016).
80. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a
toolkit to classify genomes with the Genome Taxonomy Database.
Bioinformatics 36, 1923–1927 (2020).
81. Wu, Z., Liu, J., Yang, H. & Xiang, H. DNA replication origins in archaea.
Front. Microbiol. 5, 179 (2014).
82. Pebesma, E. & Bivand, R. Classes and Methods for Spatial Data in R. R News
5, 9–13 (2005).
83. Rodriguez, R. L. M. & Konstantinidis, K. T. The Enveomics Collection: A
Toolbox for Specialized Analyses of Microbial Genomes and Metagenomes.
https://doi.org/10.7287/peerj.preprints.1900v1 (2016).
84. R Core Team. R: A Language and Environment for Statistical Computing. (R
Foundation for Statistical Computing, Vienna, Austria, 2008).
85. Shannon, P. et al. Cytoscape: a software environment for integrated models of
biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
86. Zhou, Z., Tran, P., Liu, Y., Kieft, K. & Anantharaman, K. METABOLIC: a
scalable high-throughput metabolic and biogeochemical functional trait
profiler based on microbial genomes. Preprint at bioRxiv https://doi.org/
10.1101/761643 (2019).
87. Seemann, T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30,
2068–2069 (2014).
88. Eddy, S. R. Accelerated profile HMM Searches. PLoS Comput. Biol. 7, e1002195 (2011).
89. Darling, A. E. et al. PhyloSift: phylogenetic analysis of genomes and
metagenomes. PeerJ 2, e243 (2014).
90. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin,
L. S. ModelFinder: fast model selection for accurate phylogenetic estimates.
Nat. Methods 14, 587–589 (2017).
91. Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C. & D
L. S. ModelFinder: fast model selection for accurate phylogenetic estimates.
Nat. Methods 14, 587–589 (2017).
92. Haynes, W. Tukey’s Test. In Encyclopedia of Systems Biology (eds Dubitzky,
W., Wolkhenuater, O., Cho, K.-H. & Yokota, H.) 2303–2304 (Springer, New
York, 2013).
93. Shannon, C. E. A mathematical theory of communication. Bell Syst. Tech. J.
27, 379–423 (1948).
94. Lakens, D., Scheel, A. M. & Isager, P. M. Equivalence testing for psychological
research: a tutorial. Adv. Methods Pract. Psychol. Sci. 1, 259–269 (2018).
95. Simonsohn, U. Small telescopes: detectability and the evaluation of replication
results. Psychol. Sci. https://doi.org/10.1177/0956797614567341 (2015).
96. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag,
2009).
97. Anisimova, M., Gil, M., Dufayard, J.-F., Desimoz, C. & Gascuel, O. Survey of
branch support methods demonstrates accuracy, power, and robustness of fast
likelihood-based approximation schemes. Syst. Biol. 60, 685–699 (2011).

Acknowledgements
We thank Hubert Müller for technical assistance, Sabrina Eisfeld for laboratory main-
tenance, Ken Dreger for server administration and maintenance, and Karen L. lloyd for
scientific discussions.

Author contributions
T.L.V.B. performed the main bioinformatics analysis, P.S.A. performed phylogenomics. V.T.
and A.J.P. performed microscopy. U.S., R.S., and B.K. performed geological analyses and
geological data interpretation. T.L.V.B. and P.A.F.G. analyzed genomes. T.L.V.B., J.R., and
A.J.P. took samples. D.K. and T.C.S. performed geochemical analyses. A.J.P. conceptualized
the study. T.L.V.B. and A.J.P. wrote the paper with revisions from all co-authors.

Funding
This study was funded by the Ministerium für Kultur und Wissenschaft des Landes
Nordrhein-Westfalen (Nachwuchsgruppe Dr. Alexander Probst). The Geyser Andernach
metagenomes were sequenced within the Census of Deep Life Sequencing call 2017,
phase 13 project Microbial metabolism in a deep subsurface, shale-hosted aquifer of the
Volcanic Eifel (central Europe): a comparative analysis of two cold, high-CO₂ geysers. JR
received funding by the DFG (RA 3432/1-1) during revisions of the manuscript. Open Access
funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material
available at https://doi.org/10.1038/s41467-021-27783-7.

Correspondence and requests for materials should be addressed to Alexander J. Probst.

Peer review information Nature Communications thanks Steven D’Hondt, Beth Orcutt,
and the other, anonymous, reviewer(s) for their contribution to the peer review of this
work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in
published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons
Attribution 4.0 International License, which permits use, sharing,
adaptation, distribution and reproduction in any medium or format, as long as you give
appropriate credit to the original author(s) and the source, provide a link to the Creative
Commons license, and indicate if changes were made. The images or other third party
material in this article are included in the article’s Creative Commons license, unless
indicated otherwise in a credit line to the material. If material is not included in the
article’s Creative Commons license and your intended use is not permitted by statutory
regulation or exceeds the permitted use, you will need to obtain permission directly from
the copyright holder. To view a copy of this license, visit http://creativecommons.org/
licenses/by/4.0/.

© The Author(s) 2022