







Geological activity shapes the microbiome in deep-subsurface aquifers by advection

Yuran Zhang (张毓然)^{a,1,2} , Roland N. Horne^{a,2} , Adam J. Hawkins^{a,b} , John Carlo Primo^c, Oxana Gorbatenko^d, and Anne E. Dekas^{e,2} 

Edited by Bo Barker Jorgensen, Aarhus Universitet, Aarhus C, Denmark; received July 29, 2021; accepted May 9, 2022

Subsurface environments host diverse microorganisms in fluid-filled fractures; however, little is known about how geological and hydrological processes shape the subterranean biosphere. Here, we sampled three flowing boreholes weekly for 10 mo in a 1478-m-deep fractured rock aquifer to study the role of fracture activity (defined as seismically or aseismically induced fracture aperture change) and advection on fluid-associated microbial community composition. We found that despite a largely stable deep-subsurface fluid microbiome, drastic community-level shifts occurred after events signifying physical changes in the permeable fracture network. The community-level shifts include the emergence of microbial families from undetected to over 50% relative abundance, as well as the replacement of the community in one borehole by the earlier community from a different borehole. Null-model analysis indicates that the observed spatial and temporal community turnover was primarily driven by stochastic processes (as opposed to deterministic processes). We, therefore, conclude that the observed community-level shifts resulted from the physical transport of distinct microbial communities from other fracture(s) that outpaced environmental selection. Given that geological activity is a major cause of fracture activity and that geological activity is ubiquitous across space and time on Earth, our findings suggest that advection induced by geological activity is a general mechanism shaping the microbial biogeography and diversity in deep-subsurface habitats across the globe.

deep subsurface | microbial community | microbial biogeography | fractured aquifers | microbial transport

The terrestrial subsurface holds immense amounts of Earth's groundwater resource and is where a significant fraction of global prokaryotic life resides (1–3). Adapted to the relatively stable subsurface environmental conditions with continuous darkness and limited nutrients (3, 4), subsurface prokaryotes have slow metabolic activity and associated turnover times from months to thousands of years (1) and tend to be resilient to natural disturbance such as recharge from heavy rainfalls (3, 5, 6). Fractures are ubiquitous in subsurface environments at multiple scales (7), providing conduits for fluid flow and, thus, spaces and nutrients for microorganisms to thrive. Previous studies, albeit scarce, suggested that the microbial communities in fractured aquifers are shaped by local lithology and the resulting geochemical conditions and are, therefore, heterogeneous (4, 8–12). Besides the close interaction with mineral surfaces, however, fractured aquifer habitats are also characterized by the restriction in space imposed by the discrete narrow fracture conduits in hard-rock formations. The microbial communities in fractured aquifers are, therefore, likely shaped additionally by a combination of geological and hydrological processes, which are poorly understood due to the challenging nature of performing in situ experiments in the inaccessible subsurface.

Deep-subsurface fracture networks are often dynamic. The aperture and geometry of fractures can change due to deformation (13, 14), clogging, or unclogging (14–17) induced by seismic (14–16) or aseismic (13, 17, 18) processes. Consequently, fracture permeability changes, leading to variation of groundwater flow (14, 19, 20). Here, we refer to the aforementioned mechanisms that cause changes in fracture networks collectively as fracture activity, notably triggerable by both geological activity and anthropogenic processes (Table 1). Fracture activity may, by altering groundwater flow, facilitate fast transport (or isolation) of microbes from distant locations with limited interaction time with surrounding mineral surfaces. As a result, the microbial community at a certain location may resemble (or diverge from) the dispersed colonizers instead of reflecting local environmental conditions.

Here, we show that advection induced by fracture activity allows replacement of the microbial community at a given location by distinct communities originating elsewhere in the subsurface. The effect of advection on microbial assemblages has recently received increased attention (21), especially in subsurface environments (4, 22–26),

Significance

Sheltered from radiation and asteroid strikes in Earth's early history and home to a significant portion of today's prokaryotic biomass, the deep subsurface could hold keys to understanding the early and continuing evolution of life. However, the processes that shape the distribution of subsurface microbial communities remain poorly understood due to sample inaccessibility. Here, at a deep-underground fractured hard-rock aquifer, we show that fracture activity leads to altered groundwater flow that drives profound changes in fluid-associated microbial communities by physical transport instead of environmental selection. We thereby identify advection induced by geological activity (a notable trigger for fracture activity) as a prominent yet overlooked mechanism shaping subsurface biogeography with potentially profound implications for life's evolutionary history.

Author contributions: Y.Z., R.N.H., and A.E.D. designed research; Y.Z., A.J.H., J.C.P., and O.G. performed research; Y.Z. analyzed data; R.N.H. and A.E.D. provided resources and supervised research implementation; and Y.Z. wrote the paper with the input of all coauthors.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹Present address: Department of Energy and Power Engineering, Tsinghua University, Beijing 100084, China

²To whom correspondence may be addressed. Email: yuranzhang4444@hotmail.com, horne@stanford.edu, or dekas@stanford.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2113985119/-DCSupplemental>.

Published June 13, 2022.

Table 1. Definitions of the terminologies throughout this article

Terminology	Definition
Advection	A specific type of microbial cell dispersal mediated by fluid flow; a stochastic process as opposed to deterministic process
Fractured aquifer	A major type of groundwater aquifer (besides porous aquifers) on Earth; characterized by discrete flow conduits through open fractures with negligible permeability in the rock matrix (Fig. 1B)
Fracture activity	The change in fracture aperture and/or geometry triggerable by both seismic and aseismic processes; results in altered groundwater flow
Shaft (mining)	A vertical or near-vertical tunnel excavated from the ground surface down a mine (the light gray vertical lines in Fig. 1A), through which people travel to certain underground levels via elevators
Drift (mining)	A horizontal passageway within a mine where people enter and operate (Fig. 1C)
Packer interval	A borehole segment hydraulically isolated from the rest of the borehole by a pair of straddle packers; allows flow measurement and sampling from only the fracture(s) covered by the packer interval
Deterministic process	One of the two processes (besides stochastic process) broadly recognized in microbial ecology to influence the assembly of species into communities (47, 50, 74); includes selection imposed by the abiotic environment (i.e., “environmental filtering”), which results from different organisms having different levels of fitness for a given set of environmental conditions* (51)
Stochastic process	One of the two processes (besides deterministic process) broadly recognized to influence the assembly of species into communities; nonselective, includes probabilistic dispersal, random birth-death events, and more (47)
β NTI	β NTI metric, uses phylogenetic information to generate random microbial communities in order to measure whether the observed communities are more or less different (phylogenetically) than expected by chance; differentiates deterministic and stochastic processes
RC_{bray}	Raup-Crick (Bray-Curtis) metric, probabilistically assembles null communities based on ASV abundances in a given pair of communities in order to determine whether the observed communities are more or less different (compositionally) than expected by chance; differentiates stochastic processes
Homogenizing dispersal	The scenario where high dispersal rate between a pair of communities (e.g., due to significant hydrological connectivity) is the primary cause for low compositional turnover; stochastic process
Dispersal limitation	The scenario where low dispersal rate between a pair of communities (e.g., due to the existence of a hydrological barrier) is the primary cause for high compositional turnover by enabling community compositions to drift apart; stochastic process
Ecological drift	Fluctuation in population sizes due to chance events, such as stochastic differences in birth and death rates, or random mutations; stochastic process
Homogenizing selection	The scenario where consistent environmental conditions (i.e., consistent selective pressure) between a pair of communities is the primary cause for low compositional turnover; deterministic process
Variable selection	The scenario where different environmental conditions (i.e., different selective pressure) between a pair of communities is the primary cause for high compositional turnover; deterministic process
Undominated	Indicates that no single ecological process could explain the observed community turnover

*Environmental condition (ecology) refers to the conditions with which a microorganism directly interacts (e.g., temperature, light, water content, salt content, acidity).

where physical and hydrological processes across space and time could play a larger role than environmental factors in driving community composition dynamics (22). However, existing studies were mostly conducted in static settings where hydrological processes shape the microbial communities indirectly by affecting local geochemical conditions through prolonged water residence times (23, 24), or under complex environmental gradients in shallow aquifers leading to convoluted effects from both hydrological and environmental factors on the observed communities (4, 22). In this study, we utilized an opportunity made possible by a 10-mo flow test at a deep-underground fractured rock formation to sample three flowing boreholes weekly for 282 d. The continuous water outflow from the boreholes, the fracture activity that occurred in the course of the campaign, and the detailed

characterization of the reservoir fracture network allowed the effect of advection, driven by fracture activity, on the subsurface microbiome to be observed directly.

Results

Dynamic Deep-Subsurface Permeable Fracture Network Analogous to Natural Geological Activity. The field site was located 1,478 m below ground surface at the Sanford Underground Research Facility (SURF), the former Homestake Gold Mine, in South Dakota (<https://www.sanfordlab.org/>). The study area was along the west mine drift (Fig. 1A and *SI Appendix*, Fig. S1 and Table 1) in the phyllite region of the Poorman formation. It is a naturally fractured aquifer with negligible matrix

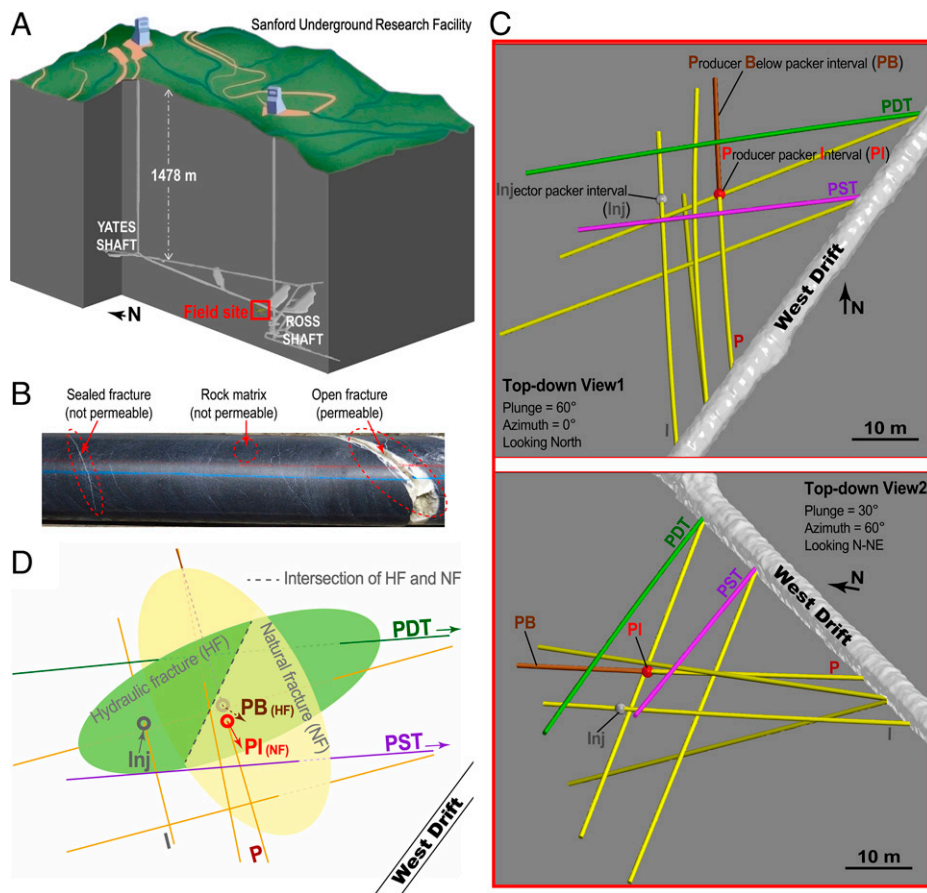


Fig. 1. Location, borehole configuration, and permeable fracture network at the deep-subsurface field site. (A) Field site located at the SURF along the west drift 1478 m below ground surface. (B) Example of a core log photo illustrating the difference among rock matrix, a sealed fracture, and an open (natural) fracture: The permeability of an open fracture is orders of magnitude larger than a sealed fracture/rock matrix. (C) Borehole configuration of the field site showing all boreholes at view-1 (Top) and view-2 (Bottom). (D) Simplified conceptual model of the fracture network in the crystalline-rock formation, modified from Wu et al. (72). Gray and red circles represent the packer intervals in boreholes I (“Inj”) and P (“PI”), whereas the brown circle represents the segment in borehole P below the packer interval (“PB”). Realistic representations of the natural and hydraulic fractures can be found in Zhang et al. (8) and Schoenball et al. (73), respectively.

permeability (Fig. 1B), representative of a major type of groundwater aquifer (besides porous aquifers) ubiquitous on Earth (*SI Appendix, Supplementary Text*). The site includes eight 60-m-long subhorizontal boreholes with nonintersecting trajectories (Fig. 1C) and one prominent fluid-filled natural fracture (8, 27). The rock was hydraulically stimulated in 2018, creating a permeable hydraulic fracture that intersected the permeable natural fracture (Fig. 1D and *SI Appendix, Supplementary Text*). Thorough reservoir characterization efforts at the testbed allowed the delineation of a conceptual model of the permeable fracture network in Fig. 1D, including fracture location and orientation. Note that the actual flow pathways were likely more complicated than the model in Fig. 1D, although the model does provide an easy layout of connected groundwater flowpaths that could explain most of the field data (e.g., flow, tracer, and geophysical data; see details in *SI Appendix, Supplementary Text*). Unknowns about the fracture flow do exist due to the inaccessible nature of interborehole flowpaths: for example, how far each permeable fracture extends, whether any other natural permeable fracture connects to the known fractures, and how exactly each permeable fracture contributes to the outflow from each sampling port (see details in *SI Appendix, Supplementary Text*).

During the 10-mo flow test, industrial water (sourced from a shallow dolomitic limestone karst) was injected at a constant 400 mL/min (Fig. 2A) into the packer interval of borehole I at a pressure up to 34.5 MPa. The injection pressurized the formation and maintained outflow, the majority of which was

formation water displaced from deeper within the reservoir, as suggested by chemical tracer (28, 29) analyses (see *SI Appendix, Supplementary Text* and Fig. S2 for details), as well as the distinct composition of the injectate communities compared with those of the produced fluids (Fig. 2 and *SI Appendix, Fig. S4B*). Outflow was produced via borehole P within the packer interval (referred to henceforth as PI) and below the packer interval (referred to henceforth as PB), as well as boreholes PST and PDT. A packer isolates different segments of a borehole in order to monitor/sample fluid flow separately (i.e., segments PI and PB in borehole P; see *SI Appendix, Supplementary Text*). Boreholes P, PST, and PDT have different locations and trajectories designed based on hydraulic stimulation purposes, although for this study, they all serve the same purpose of being simply separate ports producing outflow from different locations of the fractured aquifer. Respective flow rate history in each outflow port is shown in Fig. 2B. The injected water was ambient temperature ($\sim 20^\circ\text{C}$) before 8 May 2019 and cooled to $\sim 12^\circ\text{C}$ thereafter, although the outflow remained around 30°C in PB and PI with negligible temperature decrease ($<1^\circ\text{C}$) observed (27) or predicted (30). Operational disturbance was minimized throughout the 10-mo flow test. Outflows from PDT, PST, PI, and PB, as well as the injectate, were sampled and filtered ($0.22\ \mu\text{m}$) roughly weekly to capture the fluid-associated microbial cells. The filters were frozen on site and later subjected to high-throughput 16S ribosomal

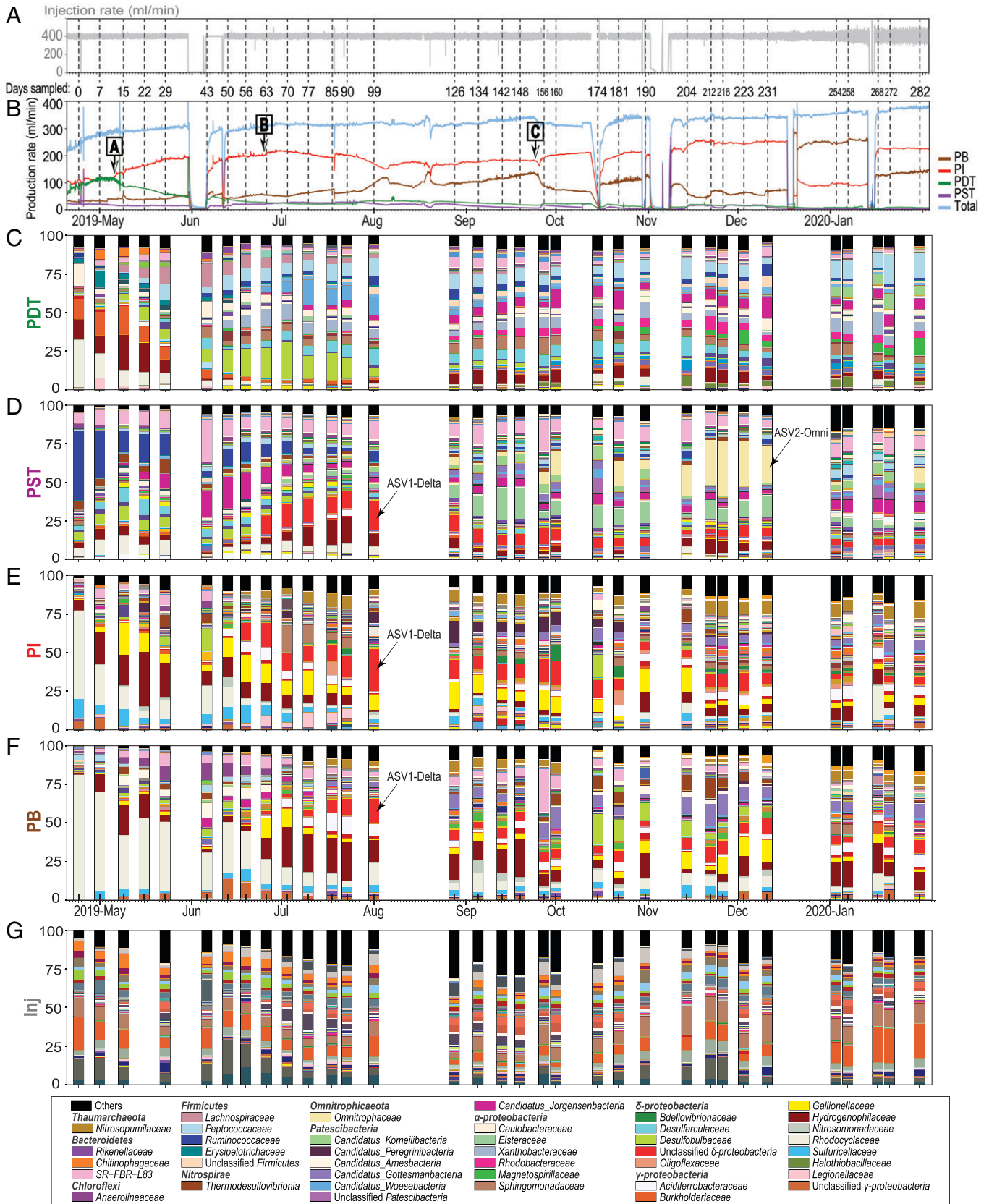


Fig. 2. Changes in microbial community composition in the outflow from boreholes PDT, PST, and P over the 282-d sampling, with the injectate community composition included as a reference. (A) Industrial water injection into borehole I at a constant volumetric rate of 400 mL/min (except in the case of field operational problems, which paused the injection briefly). (B) Volumetric flow rate produced from each of the four sampling locations—PDT, PST, PI (borehole P within packer interval), and PB (borehole P below packer interval)—along with the total production rate record. “A,” “B” and “C” refer to the spontaneous flow rate change events on day 13, day 62, and day 154, as described in the *Results*. (C–F) Temporal dynamics of microbial community composition in produced fluids from PDT (C), PST (D), PI (E), and PB (F). (G) The microbial community composition in the injectate taken every day that a set of produced-fluid samples were obtained. Bar plots show the finest classification possible down to the family level. The major taxa (i.e., taxa that were within the top 10 most abundant in at least one sample) are shown in color. Legend is simplified to annotate only a subset of taxa in the produced fluids. See full legend in *SI Appendix, Fig. S3*.

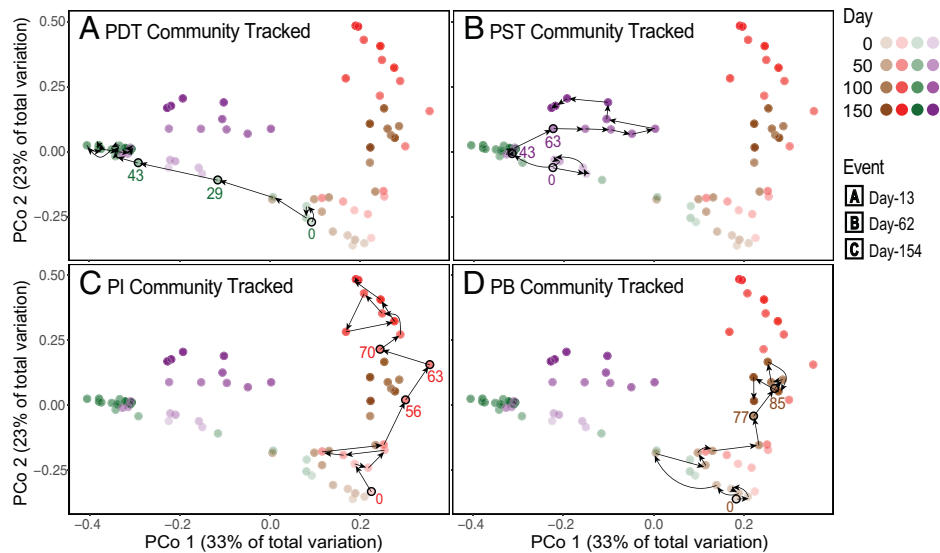


Fig. 3. PCoA on the microbial community data in the producing boreholes from day 0 to day 148, based on weighted Unifrac distance. (A) PCoA plot of the microbial community in produced fluids, with the PDT trajectory highlighted with black arrows. (B–D) The same PCoA plot as in (A) but highlighting the trajectory of PST (B), PI (C), and PB (D) using black arrows. Day 0 and the sampling dates revealing abrupt changes in microbial community composition are indicated next to the corresponding marker, highlighted with a black outline. Only data of the first 148 d and from the producing boreholes are included in this PCoA for ease of visualization. PCoA of the entire 282-d sample set along with the injectate is shown in *SI Appendix, Fig. S4*. Visual proximities of points are consistent with the optimal number of clusters defined using R function `NbClust()` (see *Materials and Methods* for details). PERMANOVA showed significant differences among the three defined clusters (pseudo-F = 46.66, $R^2 = 0.42$, $P < 0.001$), as shown in *SI Appendix, Fig. S5*.

RNA (rRNA) gene amplicon sequencing (8) to reveal the temporal dynamics of the microbial community composition in the produced fluids (see *Materials and Methods* for details).

Incidents occurred during the 10-mo flow test that are classifiable into two categories: 1) spontaneous, significant changes in production flow rate without known cause (e.g., events A–C in Fig. 2B), indicative of an altered permeable fracture network (in terms of aperture or geometry, i.e., fracture activity); and 2) operational issues such as pump restart or failure that halted the injection (and, hence, the outflow as well) for several minutes up to several days, likely resulting in an altered fracture network (i.e., fracture activity) when injection was restarted and the fractures repressurized. The occurrence of fracture activity at our testbed was confirmed by data from four tracer tests performed 2 to 3 mo apart throughout the long-term fluid sampling, which revealed significantly altered relative connectivities among sampling ports over time (*SI Appendix, Fig. S2 and Supplementary Text*). We consider such changed permeable fracture networks analogous to natural scenarios where geological activity (e.g., earthquakes) (14–16, 19), heavy rainfall (31), anthropogenic pumping (13, 32, 33), and so forth lead to subsurface fracture activity that alter groundwater flow because they all involve the same series of consequences regardless of the trigger: changes in fracture aperture/geometry leading to changed relative connectivity among flowpaths, which then leads to altered mixing of groundwater (and the microbial communities therein). Natural groundwater velocities in fractured aquifers are typically fast, ranging from several to hundreds of meters per day (34, 35), and are similar to the estimated flow velocities in this study (16.5 to 235.7 m/day). When it comes to variations in flow, geological activity such as earthquakes can induce variations in groundwater flow rate orders of magnitude larger than the variations occurred in this study, hence causing altered mixing of groundwater even more significantly and rapidly. For example, a spring in Italy disappeared completely after a 1979 earthquake and was reactivated to reach a discharge value of 1.5 m³/s (i.e., 1,500 L/s) after another earthquake in 2017 (36, 37). Changes

in river/spring flow on the order of several m³/s have been documented for other earthquake events as well (19, 37, 38).

Significant Shifts in Microbial Community Composition Coincided with Apparent Changes in Fracture Network. Similar to observations in previous subsurface microbial studies, highly stable microbial communities have been observed in borehole or fracture fluids at SURF on a 1-to-4-y time scale (8, 39), as well as in the PDT community of this study from June 2019 to January 2020 (Figs. 2C and 3A and *SI Appendix, Fig. S4*). A number of taxa were present in multiple boreholes throughout this time-series study, including Firmicutes families *Peptococcaceae* (up to 18.4%) and *Ruminococcaceae* (up to 45.4%); Deltaproteobacterial families *Desulfobulbaceae* (up to 24.6%) and *Desulfarculaceae* (up to 12.4%); Nitrospirae class *Thermodesulfovibronia* (up to 13.6%); Bacteroidetes family *SR-FBR-L83* (up to 27.8%); and Gammaproteobacterial families *Sulfuricellaceae* (up to 12.6%), *Rhodocyclaceae* (up to 80.4%), *Hydrogenophilaceae* (up to 35.3%), and *Gallionellaceae* (up to 20.7%), which have been reported previously in studies on oligotrophic groundwater environments (4, 8, 11, 26, 39, 40). Archaea were also present (e.g., *Nitrosopumilaceae* belonging to *Thaumarchaeota*) with generally low relative abundance (up to 12.4% in a sample), similar to what was observed in previously studied terrestrial groundwater sites as well (2, 22, 41, 42).

Despite a generally stable subsurface microbiome, prominent but different temporal community dynamics were observed in the sampling locations of this study at certain time points (Figs. 2 and 3 and *SI Appendix, Figs. S4 and S5 and SI Appendix, Supplementary Text*), even though the injection rate was kept constant (except during pump failure) and no further perturbation was applied to the system throughout the course of the time-series sampling. Specifically, three pronounced features in the community dynamics are evident in this 282-d dataset, showing signs of an advective driving force:

- 1) The PDT community was highly similar to the PI and PB communities during the first 3 wk of sampling, dominated by Gammaproteobacterial *Rhodocyclaceae* (up to 30.6%),

Hydrogenophilaceae (up to 23.5%), and Alphaproteobacterial *Caulobacteraceae* (up to 14.5%). However, at around day 29, the PDT community switched rapidly to a distinct community and thereafter remained stable at what resembled the day 43 to day 56 PST community, dominated by Deltaproteobacterial *Desulfobulbaceae* and *Desulfarculaceae*; Firmicutes families *Peptococcaceae* and *Ruminococcaceae*; and members of the *Cand. Patescibacteria* (43, 44) including *Candidatus Jorgensenbacteria*, *Candidatus Komeilibacteria*, and *Candidatus Woesebacteria* (Figs. 2 and 3). This visually evident switch in relative similarities is further confirmed by the combined use of data clustering and silhouette analysis, as shown in *SI Appendix, Fig. S5* (see *Materials and Methods*). The abrupt community shift in PDT coincided with “event A” in Fig. 2*B* (6 May 2019) when the flow rate of PDT spontaneously started to decline from roughly 120 mL/min to below 50 mL/min. This could indicate that PDT was initially producing primarily from the same flowpath(s) feeding PI/PB at the time (likely the hydraulic fracture), but later produced mainly from the flowpath(s) feeding PST in June 2019 (likely the natural fracture).

- 2) The “event B” in Fig. 2*B* coincided with the emergence of an unclassified Deltaproteobacterial amplicon sequence variant (ASV) (45) (“ASV1-Delta” in Fig. 2*D–F*), increasing in relative abundance from undetected to up to 23.8% in PI, followed by its similar emergence in PST (up to 18.2%) and PB (up to 14.9%) about 1 wk later. Additionally, between “event B” and “event C,” when multiple changes in outflow rate occurred spontaneously in all boreholes/intervals despite an invariant injection rate, an ASV belonging to *Omnitrophaceae* emerged in PST from undetected to up to 35.4% (“ASV2-Omni” in Fig. 2*D*). One explanation for the appearance of ASV1-Delta is that the 1-wk injection pause in early June 2019 may have caused the hydraulic fracture to partially close. When injection was restarted, the hydraulic fracture reopened to a state different from its previous form, leading to production from previously isolated fracture zones carrying distinct microbes. Another possibility is that the June 2019 injection halt depressurized the local fracture network at the time, allowing regional groundwater carrying distinct microbes to intrude and later be displaced when the system was repressurized. The emergence of ASV2-Omni occurred without an apparent preceding incident and could signify the arrival of distant *Omnitrophaceae*-bearing groundwater at borehole PST.
- 3) Principal coordinate analysis (PCoA) on the time-series sample set (Fig. 3 and *Movie S1*) shows that the PI and PB communities were closely clustered during the initial 50 d, after which notable changes in community structure occurred in PI, followed by PB with similar changes 2 to 3 wk later. For both PI and PB, the change in community structure was characterized by the appearance of several previously nonexistent taxa, including *Nitrosopumilaceae* (*Thaumarchaeota*), *Acidiferrobacteraceae* (*Gammaproteobacteria*), *Candidatus Gotesmanbacteria* (*Patescibacteria*), and ASV1-Delta (Fig. 2), summing up to as much as 32.3 and 27.6% in PI and PB, respectively, after their emergence. The two packer intervals PI and PB, although isolated in the borehole, were producing from adjacent locations along the permeable fracture network (Fig. 1*D*); therefore, it is not surprising that the microbial community in PI and PB was overall relatively similar. The temporal correlation between PI and PB community (i.e., PB community at a certain time resembling the PI community at an earlier time point) implies that the local

flow direction near borehole P was from PI to PB (i.e., fluid-associated community reaching PI first, then PB).

Evaluation at the finest taxonomic resolution (i.e., differentiation between DNA sequences that vary by only a single nucleotide) provides further evidence of advection-driven community dynamics. The percentage of ASVs detected in PST during period 2 (days 40 to 60, three samples per port) that also appeared in PDT increased from 23.1 to 35.4% between periods 1 (days 0 to 20, three samples per port) and 2 and continued to increase thereafter to 56.9% over time (into period 2+, days 40 to 282, 28 samples per port) (*SI Appendix, Fig. S6A*). Similarly, the percentage of ASVs detected in PI during period 1' (days 60 to 70, two samples per port) that also appeared in PB increased from 27.6 to 39.0% between periods 1' and 2' (days 70 to 85, three samples per port) and continued to increase thereafter to 45.2% over time (into period 2'+, days 70 to 130, six samples per port) (*SI Appendix, Fig. S6B*). The increasing percentage of shared ASVs over the defined time periods suggests that advection was the driving force in the converging community compositions rather than environmental selection (46) because similar environments may select for genetically similar microorganisms with similar functionality, but not necessarily the same exact DNA sequences (8, 39).

Stochastic Ecological Processes Dominate >80% of the Observed Spatial and Temporal Microbial Community Turnover. In order to quantitatively infer the ecological factor that is primarily responsible for the observed community turnover between each sample pair, a sequential phylogenetic- and abundance-based null-modeling analysis was performed (4, 12, 47, 48). This method uses ecological patterns in microbial community 16S rRNA gene sequencing data to identify which environmental and spatial aspects of a system primarily impose the community assembly processes. In our case, we were most interested in differentiating between hydraulic advection, a stochastic process, and environmental selection, a deterministic process, in shaping community composition. β -nearest taxon index (β NTI) and Raup-Crick (Bray-Curtis) (RC_{bray}) metrics were used to quantify whether the measured communities are more similar/dissimilar than expected by chance. Pairwise community comparisons with significantly greater phylogenetic turnover than the null expectation (β NTI > 2) occur due to variable selection, which can result from distinct environmental conditions between sampling sites (12) or fluctuating geochemical conditions such as the influx of organic carbon (49). Communities more phylogenetically similar than expected by chance (β NTI < -2) would result from homogenizing selection, which occurs under identical/stable environmental conditions (50). If the observed phylogenetic turnover is not significantly different from the null expectation ($|\beta$ NTI| < 2), a stochastic process explains most of the observed community difference. Stochastic processes include homogenizing dispersal (less than expected compositional turnover, or $RC_{\text{bray}} < -0.95$), dispersal limitation (greater than expected compositional turnover, or $RC_{\text{bray}} > 0.95$), or undominated (no single ecological process could explain the observed community turnover, or $|RC_{\text{bray}}| < 0.95$) (12, 47, 48). Specifically, homogenizing dispersal corresponds to the scenario where high dispersal rate between a pair of communities (e.g., due to significant hydrological connectivity) is the primary cause for low compositional turnover. Dispersal limitation, on the other hand, corresponds to the scenario where low dispersal rate between a pair of communities (e.g., due to the existence of a hydrological barrier) is the major cause for high compositional turnover by enabling community compositions to drift apart (47, 48, 51). Both homogenizing

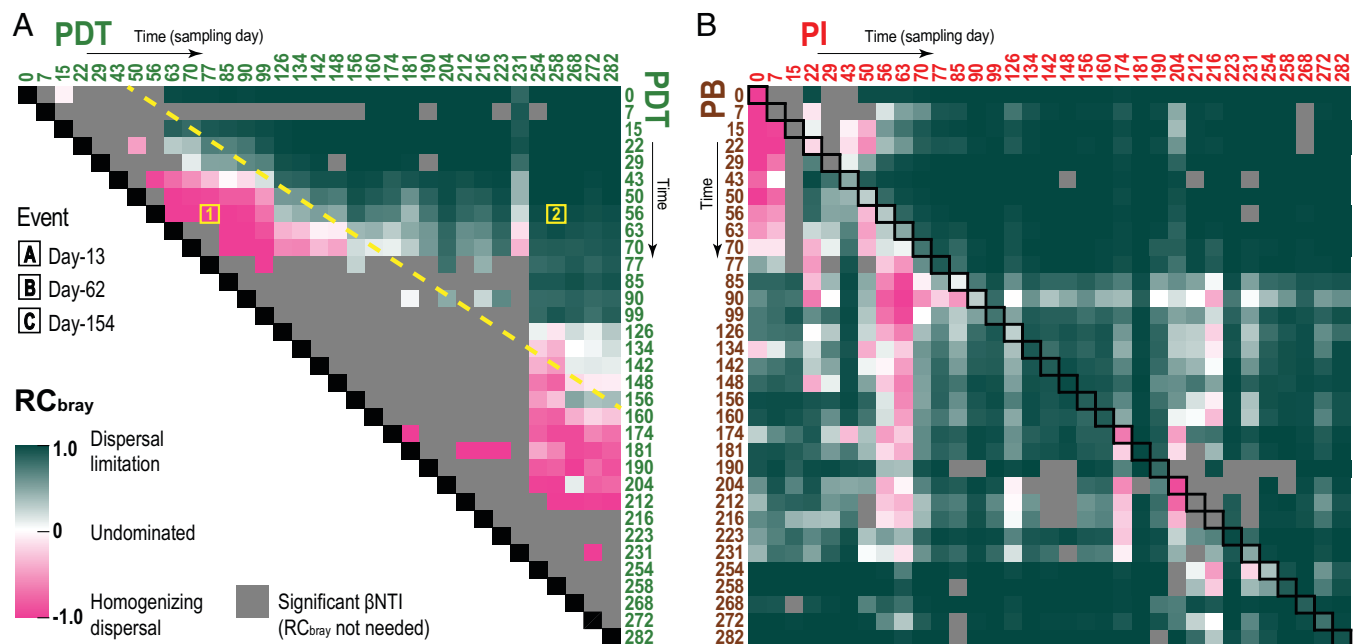


Fig. 4. Heatmaps representing RC_{bray} values for single-port (PDT) and cross-port (PI-PB) pairwise sample comparisons. (A) RC_{bray} values for pairwise comparisons among PDT samples, revealing the switch in assembly mechanism from homogenizing dispersal (pink) to dispersal limitation (green) over time (e.g., box 1 to box 2), consistent with advective mixing during the sampling period due to fracture activity. The yellow dashed line represents the rough time point at which the assembly mechanism switches for a given row/reference sample. For sample comparisons in a single port, the RC_{bray} heatmap is symmetrical with respect to the main diagonal; therefore, only the upper triangle is displayed. (B) RC_{bray} values for cross-port sample comparisons between PI and PB. In this context, homogenizing dispersal (pink) indicates strong hydraulic connectivity between ports, consistent with the close proximity between PI and PB along the permeable fracture network. Boxes along the main diagonal (i.e., comparison between samples from different ports on the same day) have black boundaries for clarity. RC_{bray} values for the rest of the single-/cross-port pairwise comparisons not shown here are available in *SI Appendix, Fig. S14*.

dispersal and dispersal limitation are subcategories of the advection mechanism proposed in this study (Table 1). At our testbed, if advection was indeed the primary driver for the observed patterns, the identified community assembly processes should be mostly stochastic rather than deterministic. Specifically, one would expect the community turnover within a single port or between highly connected ports to be driven mostly by homogenizing dispersal. In contrast, community turnover between ports with low connectivity (more commonly encountered at our site given the rock characteristics) should be driven primarily by dispersal limitation. Fracture activity changes the relative connectivities among sampling ports over time and, hence, may result in the identified assembly mechanism to switch over time.

The null modeling reveals that overall, 82.7% of the 8,646 pairwise comparisons among all 132 produced-fluid samples were primarily governed by stochastic processes (e.g., advection/dispersal) rather than deterministic (e.g., environmental selection), direct support for an advection-driven community assembly mechanism dominating the fractured aquifer (*SI Appendix, Fig. S14*, lower triangle). Within those governed by stochastic processes, 69.1% were dispersal limitation, 2.1% were homogenizing dispersal, and 28.8% were undominated (*SI Appendix, Fig. S14*, upper triangle). To further interpret the null-modeling results, we classify the pairwise comparisons in this time-series sample set into two categories—single port (i.e., temporal turnover) and cross port (i.e., spatial turnover)—and identify the dominant assembly processes in each case.

For single-port comparisons in all four producing ports, and within a relatively short time span, homogenizing dispersal and homogenizing selection were the major drivers for the (low) community turnover (Fig. 4A and *SI Appendix, Fig. S14*). This is not surprising because a sampling port must be well connected to

itself, and the environmental conditions in the deep-subsurface are typically stable. However, in all cases, when the time span between sample comparisons is long enough (1 to 5 mo), dispersal limitation became the primary driver for the (high) community turnover (Fig. 4A and *SI Appendix, Fig. S14*). Since a sampling location cannot be isolated from itself, one explanation is most reasonable for dispersal limitation to dominate the community turnover in the same port: fracture activity drove distinct microbial communities from a previously isolated hydrological compartment to arrive at the port at later times, by advection.

For most of the cross-port comparisons, the high community turnover was primarily driven by dispersal limitation (*SI Appendix, Fig. S14*), consistent with the extremely limited hydraulic conductivity among hydrological compartments in the rock formation. A notable exception, though, lies in the cross-port comparisons between PI and PB, where a number of instances of community turnover were identified to be driven by homogenizing dispersal (i.e., the pink boxes in Fig. 4B). In other words, besides confirming the dominant role of stochastic processes (i.e., advection) in driving the spatial community turnover, the null modeling further differentiated the stochastic processes: it identified the existence of a highly connected flowpath between PI and PB among the vast majority of impermeable rock mass and poorly connected flowpaths. The pink boxes being concentrated on the lower triangle of the heatmap in Fig. 4B (where a PB sample is always compared with an earlier PI sample) corroborate the observation that PB community at a certain time resembled the PI community at an earlier time point (Fig. 3) and confirm that the driver was, indeed, a local groundwater flow direction from PI to PB rather than convergence due to environmental selection.

Discussion

Geological Activity Drives Rapid Advection of Fluid Microbial Community Outpacing Environmental Selection. The paradigm that “everything is everywhere but the environment selects” (52) emphasizes the important role of environmental selection on a global scale. In subsurface environments, particularly deep crystalline-rock aquifers with discrete fractures, environmental selection takes effect by selecting for certain microorganisms along regional-scale flowpaths with geochemical gradients (24). When geological activity occurs, however, rapid hydrological changes are induced by the deformation or clearing/clogging of fracture conduits that lead to a significantly altered state of groundwater mixing/isolation (13–20). Under such nonselective driving forces, fracture conduits allow the rapid transport (or isolation) of distant and likely distinct fluid microbial communities along the altered pressure gradient with limited time for interaction with surrounding minerals. Regardless of the degree of environmental filtering between the source and the new location, the newly arrived (or residual) microbial community could play a deterministic role on the composition and functions of the final community (ref. 53 and references therein).

In this study, the environmental conditions of the fractured reservoir remained largely unchanged throughout the 282-d sampling campaign: 1,478 m away from surface effects (e.g., diel and seasonal), constant volumetric injection rate, and minimal operational disturbance. Without any notable environmental stimuli, the prominent community shifts in all sampling locations associated with the altered outflow rates are most likely attributed primarily to the physical translocation of fluid-associated microorganisms from elsewhere in the formation driven by fracture activity and are confirmed by the β NTI-RC_{bray} results. This suggests that besides transporting nutrients and some specific microbial taxa (25), advection driven by fracture activity in hard-rock aquifers may be able to drive the establishment of brand new microbiomes. Data interpretation in previous (and future) groundwater studies—especially those that observed surprisingly high community turnover rates in deep, oligotrophic environments (26)—could, therefore, consider advection (or variations of advection) as a possible explanation for the observed patterns. From a practical standpoint, specifically in subsurface resource exploitation and carbon sequestration, our findings imply that time-series microbial community data in reservoir fluids could serve as an indicator for fracture activities that are otherwise undetectable. This potential application of deep-subsurface microbial community data in long-term reservoir monitoring, together with its use for identifying natural interwell connectivity as demonstrated in our recent study (8), could reform the current reservoir characterization technologies and, hence, improve global energy efficiency.

Additionally, and with broader implications, our findings identify advection driven by geological activity as a general mechanism shaping the microbial biogeography and diversity in deep-subsurface hard-rock aquifers. The limited hydraulic communication between heterogeneous geological compartments (e.g., isolated fractures in crystalline rocks) allows the formation of distinct microbial communities (8); geological activity can then induce rapid advection outpacing environmental selection, exposing translocated microbial communities to new environmental conditions and/or disparate biological communities, both of which could promote community diversification (11, 54, 55). Given that geological activity is a ubiquitous process across space and time not only on Earth, but on other planets as well (56, 57), this mechanism may have fundamental implications for understanding the evolution and history of life.

Materials and Methods

Fluid Sampling, DNA Extraction, Library Preparation, and High-Throughput 16S rRNA Amplicon Sequencing. Fluid samples were filtered onto 0.22- μ m Sterivex Durapore filters (EMD Millipore, 0.5 ~ 3.9 L per filter) using sterile or sterilized supplies. The filters were frozen on site in dry ice and stored at -80°C until analysis. Genomic DNA extraction and 16S rRNA amplicon library preparation was conducted according to the protocols detailed in Zhang et al. (8) and included in *SI Appendix, Supplementary Text*. Briefly, genomic DNA was extracted from the filters using the Qiagen AllPrep DNA/RNA Mini Kit (cat# 80204) according to manufacturer’s recommendations. DNA yield was recorded for quality assurance of subsequent steps (*SI Appendix, Supplementary Text and Fig. S7*). PCR was performed on the extracted DNA using universal 16S primers 515F-Y and 926R (58). A second-round PCR was performed to add unique barcodes to the first-round PCR amplicons of each sample for sequencing (dual indexing strategy). Then, 4 μ L PCR product of each sample was loaded onto a 1% agarose gel after each round of PCR to check amplification. No amplification was found in PCR blanks or extraction blanks, confirming the absence of contaminants. PCR products were cleaned using magnetic beads. The cleaned-up samples were pooled in equimolar concentrations, purified again, and sequenced at the UC Davis Genome Center (Illumina MiSeq, 2 \times 250 bp paired-end sequencing).

High-Throughput Sequencing Data Processing. Primer sequences were trimmed from the raw sequencing reads of each sample using cutadapt (59). Bioinformatics packages in R, including Dada2 (60) and phyloseq (61), were used to analyze the sequencing data. The sequences were quality filtered by truncating all reads beyond 220 bases (to discard bases with quality scores <30) and discarding sequences that did not exactly match proximal primers, had more than two expected errors, or contained ambiguous bases (Ns). The Dada2 method (45) was used to infer ASVs from the quality-filtered reads, remove sequencing errors, merge forward and reverse reads with no mismatches allowed in the overlap region, remove chimeras, and generate a sequence table. Evaluations of sequencing data quality and sampling adequacy are provided in *SI Appendix, Supplementary Text* as well as *SI Appendix, Figs. S8 and S9*. The ASVs were assigned taxonomy using the Silva v132 database implemented in Dada2 (62). A phylogenetic tree of all the sequences was constructed using R packages DECIPHER (63) and phangorn (64) in order to enable the calculation of phylogeny-aware distances between communities among the samples obtained in this study.

Diversity and Statistical Analyses. The diversity and statistical analyses were performed using R packages phyloseq (61), microbiome (65), NbClust (66), and vegan (67). Alpha diversity metrics (*SI Appendix, Fig. S10*) of the dataset were calculated using the alpha() function (microbiome package). For beta diversity analyses, community similarities were calculated using the weighted Unifrac distance metric on the 16S sequencing data down to the ASV level (function Unifrac() in phyloseq). The weighted Unifrac distance matrix was visualized using PCoA (functions ordinate() and plot_ordination() in phyloseq). The visual proximities of datapoints on the PCoA plot as described in the main text were further confirmed by using “ward.D2” (68) as the clustering method and using silhouette analysis (69) to define the optimal number of clusters in the weighted Unifrac dissimilarity matrix (4), implemented using function NbClust() (min.nc = 2, max.nc = 15, method = “ward.D2”, index = “silhouette”) from NbClust package. Permutational multivariate analysis of variance (PERMANOVA) showed significant differences among the three defined clusters (adonis() function in vegan, permutations = 999), as shown in *SI Appendix, Fig. S5*.

Note that the 16S sequencing data were not rarefied in all analyses in this study (unless stated otherwise for testing purposes) because rarefaction throws away many rare taxa and could reduce detection power (4, 70). However, to confirm sampling adequacy and ensure our observations were not affected by imbalanced sample sizes, we explored the effect of rarefaction by rarefying the data down to the minimum number of ASVs (3,632) across all produced fluid samples (function rarefy_even_depth() in phyloseq). Nearly identical alpha diversities (*SI Appendix, Fig. S10*) and the same trends in beta diversity (*SI Appendix, Fig. S4 A and C*) were recovered with both unrarefied and rarefied data, confirming the adequacy of sampling and the robustness of our observations.

Ecological Modeling. The two-step null-model-based analyses were performed according to the protocol detailed in Stegen et al. (47, 48). Per this protocol, significant phylogenetic signal across relatively short phylogenetic distances was first found in the produced fluid communities whereby habitat preferences of closely related taxa are more similar to each other than to the habitat preferences of distant relatives (see *SI Appendix, Fig. S11*). The existence of such phylogenetic signal allows the use of phylogenetic turnover to make ecological inferences at our testbed. A phylogenetic-based null-model analysis was then performed to determine whether the phylogenetic turnover between a given pair of samples (quantified by the phylogeny-aware beta mean nearest taxon distance [β MNTD]) is more or less than what would be expected by chance. β MNTD was calculated using function "comdistnt()" in R package "picante" (abundance.weighted = TRUE). Randomization of ASV position on the phylogenetic tree that includes all the ASVs across all the produced fluid samples was repeated 999 times to generate the null distribution of β MNTD values. The deviation of the observed β MNTD from the mean of the null distribution is quantified in units of SD of the null distribution (referred to as β NTI). $|\beta$ NTI| > 2 means that the observed phylogenetic turnover is significantly greater or less than the null expectation, indicating that environmental selection (variable selection or homogenizing selection) is the primary driver of community assembly. If $|\beta$ NTI| < 2, the observed phylogenetic turnover is not significantly different from the null expectation, meaning that a stochastic process explains most of the observed community difference.

In the following step, an abundance-based null-model analysis was performed in order to examine which stochastic process primarily drives the compositional turnover of those sample pairs with nonsignificant β NTI. Sample pairs with nonsignificant β NTI were further evaluated by comparing the observed Bray-Curtis values (BC_{obs}) to the Bray-Curtis expected under randomization (BC_{null} , generated using 999 iterations per pairwise comparison). The value of Bray-Curtis-based Raup-Crick (RC_{bray} , $\in[-1,1]$) characterizes the deviation between BC_{obs} and BC_{null} . $|RC_{bray}| > 0.95$ is considered significant, suggesting the observed turnover is driven by dispersal limitation ($RC_{bray} > 0.95$) or homogenizing dispersal ($RC_{bray} < -0.95$). If $|RC_{bray}| < 0.95$, the comparison is interpreted to be the result of undominated processes.

For a summary of the definitions of relevant terminologies, see Table 1.

Water Chemistry Analysis. The pH of each fluid sample was measured on site with pH strips, and all samples were within the range of 7 to 8. Prior to filtering, each fluid sample was subsampled into a sterile 50-mL Falcon tube and frozen on site for geochemistry analysis. Each geochemistry sample was analyzed via inductively coupled plasma mass spectrometer (XSERIES 2, Thermo Scientific) for cations after acidification with nitric acid and via ion chromatograph (Dionex ICS 6000) for anions. Geochemistry data of each port are plotted against

sampling time as shown in *SI Appendix, Fig. S12*. Electrical conductivity values of the samples are estimated from the total ionic concentrations (see details in *SI Appendix, Fig. S13*) using the method described in McCleskey et al. (71).

Data Availability. Sequencing data have been deposited in the European Nucleotide Archive at the European Bioinformatics Institute under accession no. PRJEB44691. Raw flow rate data for Fig. 2 and raw tracer data for *SI Appendix, Fig. S2* are publicly available through the US Department of Energy's Geothermal Data Repository at: <https://gdr.openei.org/submissions/1254> (flow rate data) and <https://gdr.openei.org/submissions/1193> (tracer data).

ACKNOWLEDGMENTS This material was based upon work supported by the US Department of Energy, Office of Energy Efficiency and Renewable Energy, Office of Technology Development, Geothermal Technologies Office, under Award Number DE-AC02-05CH11231 and subcontract agreement 7348695 with Stanford University. Additional support was from TomKat Center for Sustainable Energy at Stanford University under Award DekasHorne2018. Support from the Shuimu Tsinghua Scholar Program at Tsinghua University (No. 2020SM014) and the Chinese International Postdoctoral Exchange Fellowship Program (Talent-Introduction Program) No. YJ20200102 for the first author during manuscript writing phase is acknowledged. The research supporting this work took place in part at the SURF in Lead, SD. The assistance of the SURF and its personnel in providing physical access and general logistical and technical support is acknowledged. This work is one of the contributions of the Enhanced Geothermal Systems (EGS) Collab Team. We would like to specifically thank Timothy Kneafsey, Paul Schwering, Megan Smith, Parker Sprinkle, and the entire EGS Collab team for making the sampling possible. Thank you to Cynthia Anderson and Dave Bergmann from Black Hills State University for providing additional resources near the site. Thank you to Ghanashyam Neupane for the assistance with 3D model generation. Many thanks to Guangchao Li at Stanford Environmental Measurements Facility for assistance in water chemistry analysis. The earth model output of the fractures was generated using Leapfrog Software. Copyright © Aranz Geo Limited. Leapfrog and all other Aranz Geo Limited product or service names are registered trademarks or trademarks of Aranz Geo Limited. We sincerely thank the anonymous reviewers and the editor for their thoughtful reviews and suggestions.

Author affiliations: ^aDepartment of Energy Resources Engineering, Stanford University, Stanford, CA 94305; ^bRobert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853; ^cLawrence Berkeley National Laboratory, Berkeley, CA 94720; ^dSchool of Natural Sciences, Black Hills State University, Spearfish, SD 57799; and ^eDepartment of Earth System Science, Stanford University, Stanford, CA 94305

1. Y. M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6506–6511 (2018).
2. C. Magnabosco et al., The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* **11**, 707–717 (2018).
3. C. Griebler, T. Lueders, Microbial biodiversity in groundwater ecosystems. *Freshw. Biol.* **54**, 649–677 (2009).
4. L. Yan et al., Environmental selection shapes the formation of near-surface groundwater microbiomes. *Water Res.* **170**, 115341 (2020).
5. A. H. Farnleitner et al., Bacterial dynamics in spring water of alpine karst aquifers indicates the presence of stable autochthonous microbial endokarst communities. *Environ. Microbiol.* **7**, 1248–1259 (2005).
6. M. Pronk, N. Goldscheider, J. Zopfi, Microbial communities in karst groundwater and their potential use for biomonitoring. *Hydrogeol. J.* **17**, 37–48 (2008).
7. E. Bonnet et al., Scaling of fracture systems in geological media. *Rev. Geophys.* **39**, 347–383 (2001).
8. Y. Zhang et al., Microbial community composition in deep-subsurface reservoir fluids reveals natural interwell connectivity. *Water Resour. Res.* **56**, e2019WR025916 (2020).
9. J. K. Fredrickson, D. L. Balkwill, Geomicrobial processes and biodiversity in the deep terrestrial subsurface. *Geomicrobiol. J.* **23**, 345–356 (2007).
10. M. R. Osburn, D. E. LaRowe, L. M. Mopper, J. P. Amend, Chemolithotrophy in the continental deep subsurface: Sanford underground research facility (SURF), USA. *Front. Microbiol.* **5**, 610 (2014).
11. M. Nyssönen et al., Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. *ISME J.* **8**, 126–138 (2014).
12. R. E. Danczak, M. D. Johnston, C. Kenah, M. Slattery, M. J. Wilkins, Microbial community cohesion mediates community turnover in unperturbed aquifers. *mSystems* **3**, e00066-18 (2018).
13. P. Jeanne et al., Seismic and aseismic deformations and impact on reservoir permeability: The case of EGS stimulation at The Geysers, California, USA. *J. Geophys. Res. Solid Earth* **120**, 7863–7882 (2015).
14. D. R. Montgomery, M. Manga, Streamflow and water well responses to earthquakes. *Science* **300**, 2047–2049 (2003).
15. N. D. Barbosa, J. Hunziker, S. Lissa, E. H. Saenger, M. Lupi, Fracture unclogging: A numerical study of seismically induced viscous shear stresses in fluid-saturated fractured rocks. *J. Geophys. Res. Solid Earth* **124**, 11705–11727 (2019).
16. N. D. Barbosa, S. G. Solazzi, M. Lupi, Seismically induced unclogging in fluid-saturated faults. *J. Geophys. Res. Solid Earth* **125**, e2020JB020152 (2020).
17. S. G. Prejean, D. P. Hill, "Earthquakes, dynamic triggering of" in *Encyclopedia of Complexity and Systems Science*, R. A. Meyers, Ed. (Springer New York, 2009), pp. 2600–2621, 10.1007/978-0-387-30440-3_157.
18. A. Gualandi, Z. Liu, C. Rollins, Post-large earthquake seismic activities mediated by aseismic deformation processes. *Earth Planet. Sci. Lett.* **530**, 115870 (2020).
19. M. R. Rosen et al., Mechanisms of earthquake-induced chemical and fluid transport to carbonate groundwater springs after earthquakes. *Water Resour. Res.* **54**, 5225–5244 (2018).
20. C.-y. Wang, Y. Chia, P.-I. Wang, D. Dreger, Role of S waves and Love waves in coseismic permeability enhancement. *Geophys. Res. Lett.* **36**, L09404 (2009).
21. D. Wilkins, E. van Sebille, S. R. Rintoul, F. M. Lauro, R. Cavicchioli, Advection shapes Southern Ocean microbial assemblages independent of distance and environment effects. *Nat. Commun.* **4**, 2457 (2013).
22. X. Lin et al., Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. *ISME J.* **6**, 1665–1676 (2012).
23. O. Bochet et al., Iron-oxidizer hotspots formed by intermittent oxic-anoxic fluid mixing in fractured rocks. *Nat. Geosci.* **13**, 149–155 (2020).
24. S. Ben Maamar et al., Groundwater isolation governs chemistry and microbial community structure along hydrologic flowpaths. *Front. Microbiol.* **6**, 1457 (2015).
25. G. Borgonie et al., New ecosystems in the deep subsurface follow the flow of water driven by geological activity. *Sci. Rep.* **9**, 3310 (2019).
26. V. Hubalek et al., Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield. *ISME J.* **10**, 2447–2458 (2016).

27. T. J. Kneafsey *et al.*, "The EGS collab project - Learnings from experiment 1" in *45th Workshop on Geothermal Reservoir Engineering* (Stanford University, Stanford, CA, 2020).
28. Y. Zhang *et al.*, DNA tracer transport through porous media - The effect of DNA length and adsorption. *Water Resour. Res.* **57**, e2020WR028382 (2021).
29. G. Neupane, E. Mattson, M. Plummer, R. Podgorney, "Results of multiple tracer injections into fractures in the EGS collab testbed-1" in *45th Workshop on Geothermal Reservoir Engineering* (Stanford University, Stanford, CA, 2020).
30. P. Fu, H. Wu, X. Ju, J. Morris, "Analyzing fracture flow channel area in EGS collab experiment 1 testbed" in *45th Workshop on Geothermal Reservoir Engineering* (Stanford University, Stanford, CA, 2020).
31. H. Wakita, Y. Nakamura, I. Kita, N. Fujii, K. Notsu, Hydrogen release: New indicator of fault activity. *Science* **210**, 188-190 (1980).
32. R. B. Neumann *et al.*, Anthropogenic influences on groundwater arsenic concentrations in Bangladesh. *Nat. Geosci.* **3**, 46-52 (2009).
33. M. L. Polizzotto, B. D. Kocar, S. G. Benner, M. Sampson, S. Fendorf, Near-surface wetland sediments as a source of arsenic release to ground water in Asia. *Nature* **454**, 505-508 (2008).
34. G. Medici, L. J. West, S. A. Banwart, Groundwater flow velocities in a fractured carbonate aquifer-type: Implications for contaminant transport. *J. Contam. Hydrol.* **222**, 1-16 (2019).
35. L. D. McKay, W. E. Sanford, J. M. Strong, Field-scale migration of colloidal tracers in a fractured shale saprolite. *Ground Water* **38**, 139-147 (2000).
36. D. Valigi *et al.*, "Discharge variations of springs induced by strong earthquakes: The case of the Mw 6.5 Norcia event (Italy, October 30th 2016)" in *3rd National Meeting on Hydrogeology* (Cagliari, Italy, 2017).
37. L. Di Matteo *et al.*, Effects of earthquakes on the discharge of groundwater systems: The case of the 2016 seismic sequence in the Central Apennines, Italy. *J. Hydrol. (Amst.)* **583**, 124509 (2020).
38. C.-y. Wang, C.-H. Wang, M. Manga, Coseismic release of water from mountains: Evidence from the 1999 (Mw = 7.5) Chi-Chi, Taiwan, earthquake. *Geology* **32**, 769-772 (2004).
39. M. R. Osburn *et al.*, Contrasting variable and stable subsurface microbial populations: An ecological time series analysis from the Deep Mine Microbial Observatory, South Dakota, USA. bioRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.09.15.298141>. Accessed 21 March 2021.
40. P. Geesink *et al.*, Genome-inferred spatio-temporal resolution of an uncultivated *Roizmanbacterium* reveals its ecological preferences in groundwater. *Environ. Microbiol.* **22**, 726-737 (2020).
41. L. Purkamo *et al.*, Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids. *Biogeosciences* **13**, 3091-3108 (2016).
42. J. Detmers *et al.*, FISH shows that *Desulfotomaculum* spp. are the dominating sulfate-reducing bacteria in a pristine aquifer. *Microb. Ecol.* **47**, 236-242 (2004).
43. R. Tian *et al.*, Small and mighty: Adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome* **8**, 51 (2020).
44. M. Herrmann *et al.*, Predominance of *Cand.* Patescibacteria in groundwater is caused by their preferential mobilization from soils and flourishing under oligotrophic conditions. *Front. Microbiol.* **10**, 1407 (2019).
45. B. J. Callahan, P. J. McMurdie, S. P. Holmes, Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639-2643 (2017).
46. J. A. Fuhrman, Microbial community structure and its functional implications. *Nature* **459**, 193-199 (2009).
47. J. C. Stegen, X. Lin, A. E. Konopka, J. K. Fredrickson, Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* **6**, 1653-1664 (2012).
48. J. C. Stegen *et al.*, Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7**, 2069-2079 (2013).
49. J. C. Stegen *et al.*, Groundwater-surface water mixing shifts ecological assembly processes and stimulates organic carbon turnover. *Nat. Commun.* **7**, 11237 (2016).
50. F. Dini-Andreote, J. C. Stegen, J. D. van Elsas, J. F. Salles, Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1326-E1332 (2015).
51. J. C. Stegen, X. Lin, J. K. Fredrickson, A. E. Konopka, Estimating and mapping ecological processes influencing microbial community assembly. *Front. Microbiol.* **6**, 370 (2015).
52. L. G. M. Baas-Becking, *Geobiologie Of Inleiding Tot De Milieukunde* (W.P. Van Stockum & Zoon, 1934).
53. L. Fillinger, Y. Zhou, C. Kellermann, C. Griebler, Non-random processes determine the colonization of groundwater sediments by microbial communities in a pristine porous aquifer. *Environ. Microbiol.* **21**, 327-342 (2019).
54. S. M. Gibbons, J. A. Gilbert, Microbial diversity-Exploration of natural ecosystems and microbiomes. *Curr. Opin. Genet. Dev.* **35**, 66-72 (2015).
55. M. J. Benton, B. C. Emerson, How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. *Palaeontology* **50**, 23-40 (2007).
56. S. McMahon, J. Parnell, N. J. Blamey, Evidence for seismogenic hydrogen gas, a potential microbial energy source on earth and mars. *Astrobiology* **16**, 690-702 (2016).
57. M. Panning, V. Lekic, M. Manga, F. Cammarano, B. Romanowicz, Long-period seismology on Europa: 2. Predicted seismic response. *J. Geophys. Res.* **111**, E12008 (2006).
58. A. E. Parada, D. M. Needham, J. A. Fuhrman, Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403-1414 (2016).
59. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10-12 (2011).
60. B. J. Callahan, K. Sankaran, J. A. Fukuyama, P. J. McMurdie, S. P. Holmes, Bioconductor workflow for microbiome data analysis: From raw reads to community analyses. *F1000 Res.* **5**, 1492 (2016).
61. P. J. McMurdie, S. Holmes, phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
62. B. J. Callahan, Silva taxonomic training data formatted for DADA2 (Silva version 132) [Data set]. *Zenodo*. doi.org/10.5281/zenodo.1172783 (2018).
63. E. S. Wright, Using DECIPHER v2.0 to analyze big biological sequence data in R. *R. J.* **8**, 352-359 (2016).
64. K. P. Schliep, phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592-593 (2011).
65. L. Lahti, S. Shetty, Microbiome R package (2012). [microbiome.github.io](https://github.com/microbiome/microbiome).
66. M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: AnRPackage for determining the relevant number of clusters in a data set. *J. Stat. Softw.* **61**, 1-36 (2014).
67. J. Oksanen *et al.*, The vegan package. *Community Ecol. Package* **10**, 631-637 (2007).
68. F. Murtagh, P. Legendre, Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *J. Classif.* **31**, 274-295 (2014).
69. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53-65 (1987).
70. P. J. McMurdie, S. Holmes, Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
71. R. B. McCleskey, D. K. Nordstrom, J. N. Ryan, Comparison of electrical conductivity calculation methods for natural waters. *Limnol. Oceanogr. Methods* **10**, 952-967 (2012).
72. H. Wu *et al.*, Characterization of flow and transport in a fracture network at the EGS Collab field experiment through stochastic modeling of tracer recovery. *J. Hydrol. (Amst.)* **593**, 125888 (2021).
73. M. Schoenball *et al.*, Creation of a mixed-mode fracture network at mesoscale through hydraulic fracturing and shear stimulation. *J. Geophys. Res. Solid Earth* **125**, e2020JB019807 (2020).
74. J. M. Chase, J. A. Myers, Disentangling the importance of ecological niches from stochastic processes across scales. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2351-2363 (2011).