Bacterial Communities in Concrete Reflect Its Composite Nature and Change with Weathering

E. Anders Kiledal,[a] Jessica L. Keffer,[b]* Julia A. Maresca[b]

aDepartment of Biological Sciences, University of Delaware, Newark, Delaware, USA
bDepartment of Civil and Environmental Engineering, University of Delaware, Newark, Delaware, USA

ABSTRACT Concrete is an extreme but common environment and is home to microbial communities adapted to alkaline, saline, and oligotrophic conditions. Microbes inside the concrete that makes up buildings or roads have received little attention despite their ubiquity and capacity to interact with the concrete. Because concrete is a composite of materials which have their own microbial communities, we hypothesized that the microbial communities of concrete reflect those of the concrete components and that these communities change as the concrete ages. Here, we used a 16S amplicon study to show how microbial communities change over 2 years of outdoor weathering in two sets of concrete cylinders, one prone to the concrete-degrading alkali-silica reaction (ASR) and the other having the risk of the ASR mitigated. After identifying and removing taxa that were likely laboratory or reagent contaminants, we found that precursor materials, particularly the large aggregate (gravel), were the probable source of ~50 to 60% of the bacteria observed in the first cylinders from each series. Overall, community diversity decreased over 2 years, with temporarily increased diversity in warmer summer months. We found that most of the concrete microbiome was composed of *Proteobacteria*, *Firmicutes*, and *Actinobacteria*, although community composition changed seasonally and over multiyear time scales and was likely influenced by environmental deposition. Although the community composition between the two series was not significantly different overall, several taxa, including *Arcobacter*, *Modestobacter*, *Salinicoccus*, *Rheinheimera*, *Lawsonella*, and *Bryobacter*, appear to be associated with ASR.

IMPORTANCE Concrete is the most-used building material in the world and a biologically extreme environment, with a microbiome composed of bacteria that likely come from concrete precursor materials, aerosols, and environmental deposition. These microbes, though seeded from a variety of materials, are all subject to desiccation, heating, starvation, high salinity, and very high pH. Microbes that survive and even thrive under these conditions can potentially either degrade concrete or contribute to its repair. Thus, understanding which microbes survive in concrete, under what conditions, and for how long has potential implications for biorepair of concrete. Further, methodological pipelines for analyzing concrete microbial communities can be applied to concrete from a variety of structures or with different types of damage to identify bioindicator species that can be used for structural health monitoring and service life prediction.

KEYWORDS concrete, alkali-silica reaction, low biomass, built environment, microbial communities, bioindicators

Concrete is a uniquely harsh environment characterized by high alkalinity and salinity and low water activity. In outdoor structures, it is also subject to large fluctuations in temperature and moisture. Due to its strength, resistance to weathering, and

Address correspondence to Julia A. Maresca, jmaresca@udel.edu.

* Present address: Jessica L. Keffer, Department of Earth Sciences, University of Delaware, Newark, Delaware, USA.

The microbial community in concrete, though small, comes from the components of concrete - gravel, sand, cement powder, and water. This community changes as concrete ages, and can tell us about the structural health of concrete.

low cost, concrete is the most widely used building material in the world (1). It is, therefore, a very common environment, and despite the tough conditions, bacteria are known to live in and on concrete (2–11).

When concrete is poured, its pH is ~12.5, higher than most known naturally alkaline environments (12, 13) and comparable to that of leachate from steel slag or bauxite and Solvay wastes (14). Highly alkaline environments like concrete present metabolic and physical challenges to the microbes inhabiting them, including a reversed proton gradient, with implications for ATP production, enzyme inactivation, and instability of membranes and DNA, among others (15). Microbe-concrete relationships have been extensively studied in the specific contexts of concrete degradation (3–9, 16–18) and biorepair (19). In these cases, the microbes that alter the concrete structure are introduced from outside after the structure has been poured. However, very little is known about the microbes that inhabit ordinary concrete. We hypothesized that the concrete microbiome comes from the precursor materials and that it is similar to that of other high-pH environments, like alkaline soils and soda lakes.

We further predicted that concrete mixes with different chemical properties would have different microbial communities, so we compared the microbial communities of a concrete mix prone to alkali-silica reaction (ASR) and one for which the risk of ASR had been mitigated with fly ash. ASR is a concrete-degrading chemical reaction of global concern that occurs between alkali hydroxides from cement powder and silica in the fine and large aggregates that constitute most of concrete's mass (20). ASR results in a silicate gel that expands when hydrated, creating internal pressure and extensive map cracking that significantly shortens the life span of affected structures. To prevent ASR, the Delaware Department of Transportation adds fly ash to the concrete mix, a common practice (21). In addition to reducing the probability of ASR, fly ash reacts with lime, and the reaction products fill concrete pores, particularly larger ones most prone to water infiltration, decreasing concrete porosity (22, 23). Because ASR alters both the chemistry of the concrete and its structure, allowing more infiltration by water and waterborne microbes and chemicals (20, 24), while fly ash reduces porosity and excess alkalinity (21, 25), we expected the microbial communities in these two types of concrete to diverge from each other over time.

We expected analysis of these communities to be complicated, because concrete is a low-biomass environment (2, 26), and such environments are particularly susceptible to laboratory contamination. Contaminant DNA represents a higher proportion of total DNA in low-biomass systems than it does in a higher-biomass environment and thus has more power to obscure noncontaminant sequences or influence interpretation of the data (27). For this reason, it is important to use rigorous methods to identify and categorize as many potential contaminants and their likely sources as possible. An aggressive approach to identifying and removing contaminants increases confidence that the real concrete microbiome, not reagent or laboratory microbiomes, can be described.

Our previous work showed that cultivable bacteria are present in concrete and that microbial DNA can be extracted from concrete and analyzed, providing a snapshot of the bacteria living in and on concrete a year after pouring (2). Here, we investigated how precursor materials, probability of ASR, time, weather, and laboratory contamination influence concrete microbial communities. We addressed these questions using 16S amplicon sequencing to characterize, over 2 years, the microbial communities of two series of concrete cylinders made with different mix designs to confer or prevent ASR and the precursor materials used in their preparation.

RESULTS

Concrete cylinder preparation and sampling. To determine how microbial communities change as concrete weathers, and to test the effects of ASR reactivity on community composition, we prepared two series of concrete cylinders which were weathered outdoors on a rooftop for ~2 years. One cylinder from each series was collected approximately every 6 weeks
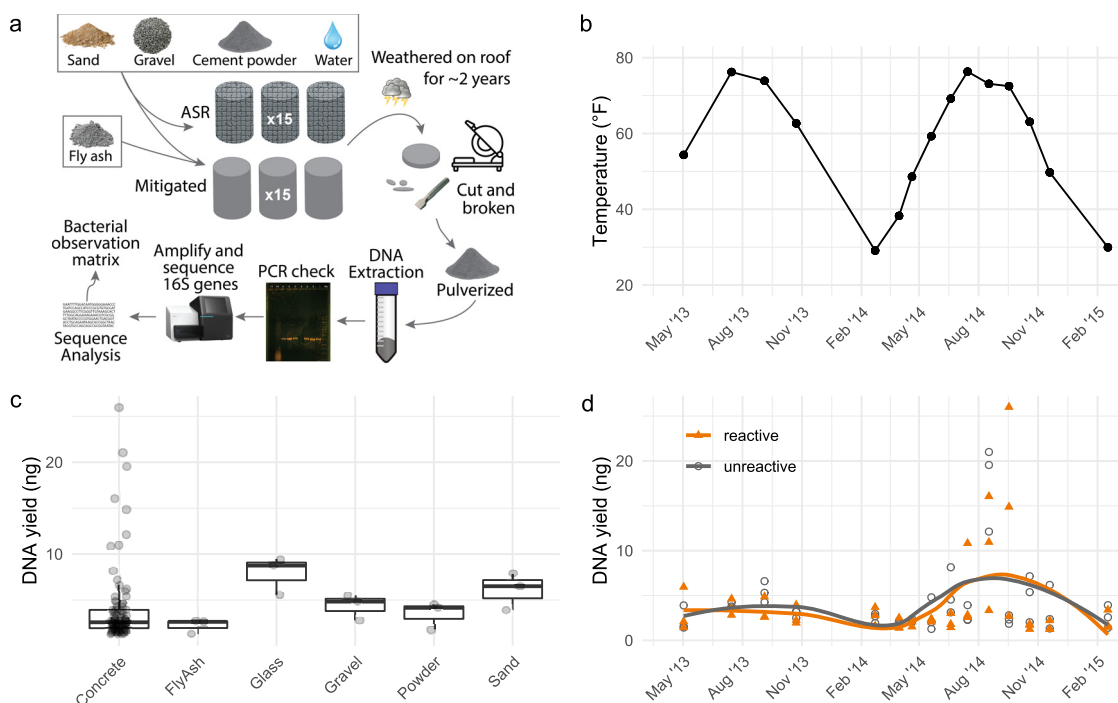
FIG 1 Experimental design, sample collection, and DNA extraction. (a) Two series of concrete cylinders, one prone to the ASR and the other having the risk of the ASR mitigated with fly ash, were placed on a campus roof for 2 years. Every 4 to 8 weeks, one cylinder of each type was archived. DNA was extracted from triplicate samples of pulverized concrete and precursor materials and sent for 16S amplicon sequencing with primers 357F and 806R (29, 30). (b) Temperatures throughout the sampling period. Each point represents a sample collection date, and reported temperatures are the mean temperature of the 30 days preceding collection. (c) DNA extraction yields from 5 g of material obtained using the protocol from reference 2. Triple-sterilized glass beads served as a negative control. Quantification was performed with a Qubit double-stranded-DNA high-sensitivity fluorometric assay. Differences in extraction efficiency may contribute to apparent differences between materials. (d) DNA yield from 5 g of concrete for each concrete sample in the series. DNA yields were higher for samples collected during the second summer.

and sampled by slicing an internal section from each (Fig. 1a). Although 2 years is short on the time scale of ASR development, this time frame allows the microbial communities to be exposed to multiple annual temperature and precipitation cycles (Fig. 1b) and is long enough to see the early effects of ASR damage (28). DNA was extracted from the concrete samples, the precursor materials used in the concrete mixes, and triple sterilized glass beads, as the negative control (Fig. 1c). DNA yields ranged from 1.25 ng to 26 ng of DNA from 5 g of starting material, with the highest yields obtained during the second summer (Fig. 1d).

16S amplification, sequencing, and quality filtering. Using primers 357F and 806R (29, 30), the V3-V4 region of the 16S gene was amplified from 105 samples, representing 90 concrete samples (15 time points × 2 ASR conditions × 3 replicates) and triplicate samples of each constituent material and negative-control glass beads (see table S1 at doi.org/10.6084/m9.figshare.14211038). A total of 9.4 million 300-bp paired-end Illumina MiSeq amplicon reads were obtained (mean, 89,743 reads per sample). Following initial primer trimming and quality filtering with Cutadapt, 8.6 million high-quality reads were retained. DADA2 denoising, chimera filtering, and read joining performed with the QIIME 2 platform (31, 32) resulted in 5.1 million observations (mean, 48,600 per sample) of 6,924 unique amplicon sequence variants (ASVs) (see table S1 at doi.org/10.6084/m9.figshare.14211038). ASVs that were identical in their overlapping region but differed in length were merged with vsearch clustering at 100% similarity, resulting in 6,898 unique ASVs (33). Finally, 31 ASVs of unexpectedly short length (<400 bp) were discarded, resulting in 48,526 mean reads per sample of 6,867 unique ASVs. For phylogenetic analyses, ASV representative sequences were inserted into the SILVA reference phylogeny (release 128) (34–36). All sequences that passed the >400-bp length filter were successfully inserted into the reference tree.

Contaminant identification and removal. Contaminant ASVs are present in nearly all 16S samples (27, 37) and can dominate results in low-biomass environments (38), so several established and custom techniques were used to classify the observed taxa as concrete-associated or contaminant. The *prevalence* method of the *decontam* R package (39) was used to identify 181 ASVs statistically more likely to be found in negative controls (glass beads) than samples (see table S2 at doi.org/10.6084/m9.figshare.14211038; also, see Fig. S1a in the supplemental material). The *frequency* method of *decontam*, which identifies ASVs whose relative abundance is inversely correlated with sample DNA concentration, was not applied because it is less reliable for very-low-biomass samples (39) and because our DNA concentrations were correlated with other sample metadata, such as season/temperature.

Reagent contaminants were identified with a correlation approach. ASV-ASV interaction networks were determined with SPARCC (40) implemented in FastSpar (41), SPIEC-EASI (42), and the *propr* R package (43). In this approach, core groups of highly intercorrelated contaminants were first identified for two types of contamination: reagent and laboratory environment. The core group of reagent contaminants was identified as highly intercorrelated ASVs present in negative-control samples with more strong positive than strong negative correlations with other negative-control ASVs (scripts available at github.com/MarescaLab/concrete_series). Reagent contaminants determined with SPARCC were also required to have a mean correlation greater than 0.3 with other negative-control ASVs. All other ASVs were then screened for correlations with this list of reagent contaminants and classified as contaminants if they exceeded cutoffs for the net number of positive correlations and/or mean correlation with the core reagent contaminants (see "Pairwise ASV comparisons" and "Contaminant identification"). Six hundred fifty-four ASVs were identified as suspected reagent contaminants with this method, of which 359 were uniquely identified by this method (see table S2 at doi.org/10.6084/m9.figshare.14211038; Fig. S1b). The 149 ASVs observed only in negative controls were also classified as reagent contaminants. The most abundant reagent contaminants belong to the *Burkholderiaceae* (*Betaproteobacteria*), while ASVs from *Enterococcus*, *Methylobacterium*, *Sphingomonas*, and *Bradyrhizobium* were also abundant reagent contaminants (Fig. S2 and S3).

Contaminants introduced from nonreagent laboratory sources like air, surfaces, or human handling were identified in a similar manner. Several strains researched in our laboratory, such as *Rhodoluna lacicola*, were unexpectedly observed in the data and are likely contaminants. Nineteen ASVs with >99% similarity to 5 likely lab contaminants (as determined via BLAST alignment [44]) were identified as contaminants (see table S2 at doi.org/10.6084/m9.figshare.14211038). We identified 167 additional contaminants by correlation with these 19 ASVs (scripts available at github.com/MarescaLab/concrete_series). Of these 167 contaminants, 147 were uniquely identified by strong correlation with known contaminants while 20 were also identified by other methods. The most abundant lab contaminants were enterobacteria in the genera *Escherichia-Shigella* and *Pantoea*. Other putative abundant lab contaminants included ASVs from *Microbacteriaceae*, *Pseudomonas*, *Sediminibacterium*, *Pedobacter*, *Exiguobacterium*, *Planococcus*, and *Bacillus*.

Finally, remaining ASVs were screened against a master list of reagent and laboratory contaminants identified up to this point, identifying 98 additional contaminants using correlation metric-specific cutoffs. (see table S2 at doi.org/10.6084/m9.figshare.14211038; Fig. S1b). Many of these putative contaminants belong to taxa identified as common contaminants in other studies (45), such as *Sphingomonas*, *Burkholderiaceae*, *Enterobacteriaceae*, *Acidibacter*, *Planococcaceae*, *Bacillus*, and *Micrococcaceae* (Fig. S3). Certain species of *Bacteroidia*, *Pedobacter*, and *Vibrionimonas* were also identified as contaminants this way.

In total, 1,112 of 6,867 ASVs were determined to be contaminants. The most abundant contaminants were *Proteobacteria* (*Beta-*, *Gamma-*, and *Alphaproteobacteria*, in decreasing order of abundance) and *Firmicutes* (Fig. S2; see table S3 at doi.org/10.6084/m9.figshare

.14211038). Removal of potential contaminant reads left 547,573 observations of 5,755 ASVs, an average of 5,368 observations per sample. Only ~15% of reads were retained because of our aggressive contaminant identification approach; however, there was little impact on the observed diversity because in about 85% of samples, >50% of the removed reads were from only a few (<5) high-abundance contaminants.

Prior to decontamination, we observed low replicate similarity (1 − generalized [0.5] UniFrac distance). The mean similarity within concrete replicates was 0.480, not significantly higher than the mean between-sample similarity of 0.466 (one-sided Welch's t test; $P = 0.166$). This was also true following decontamination where the mean within-replicate similarity was 0.340 and the between-replicate similarity was 0.332 (one-sided Welch's t test; $P = 0.111$). Decontamination decreased within-replicate similarity (one-sided Welch's t test; $P = 2.634 \times 10^{-15}$).

Concrete microbial community composition. To broadly characterize the microbial communities in concrete, all concrete cylinder samples were pooled after removal of contaminants. Overall, 50% of sequences were classified at the phylum level as *Proteobacteria*, 19% as *Firmicutes*, 14% as *Actinobacteria*, 7% as *Cyanobacteria*, 5% as *Bacteroidetes*, ~1% as *Acidobacteria*, and ~0.5% as *Planctomycetes*, with other phyla each representing <0.5% of observations (Fig. 2). No *Archaea* were observed in this data set. All percentages referenced in this section are of the total ASV observation count (547,573) after decontamination for all taxa.

Of the *Proteobacteria*, *Gammaproteobacteria* were the most abundant (~34% of all reads), in part due to the pseudomonad genera *Acinetobacter* (9%, the most abundant genus observed), *Pseudomonas* (~2%), and *Psychrobacter* (~1.7%). Other groups of *Gammaproteobacteria* were also abundant, particularly the families *Enterobacteriaceae* (~5%), *Xanthomonadaceae* (~2%), and *Halomonadaceae* (~1%). Nearly 9% of sequences were classified as *Betaproteobacteriales*, of which most belong to the family *Burkholderiaceae* (~8%). *Alphaproteobacteria* (~15%) were also observed, particularly those belonging to the orders *Rhizobiales* (~5%) and *Sphingomonadales* (~4%).

Gram-positive taxa account for ~33% of all reads. Almost 19% of sequences were classified as *Firmicutes*, which can form spores, potentially allowing them to survive in dormant states in the harsh concrete environment. Most of these belong to the order *Bacillales* (~10%), such as members of the genera *Bacillus* (~4%) and *Staphylococcus* (~2%). Nearly 5% of sequences also belonged to the order *Lactobacillales*. Actinobacteria account for ~14% of sequences, spread across several families: *Nocardiaceae* (2.0%), *Dietziaceae* (2.0%), *Corynebacteriaceae* (1.4%), *Microbacteriaceae* (1.4%), *Micrococcaceae* (1.2%), *Nocardiaceae* (0.99%), and *Geodermatophilaceae* (0.83%).

Three additional phyla each represented more than 1% of all sequences: *Cyanobacteria* (~7%), *Bacteroidetes* (~5%), and *Acidobacteria* (~1%). Many of the cyanobacteria were classified as chloroplasts (~6%) and were likely introduced from pollen or other plant material.

Precursor community composition. Prior to decontamination, higher diversity (Faith's phylogenetic diversity [PD]) was observed in sand and fly ash than concrete, while the diversity of negative controls (glass), cement powder, and gravel was similar to the median diversity of concrete samples (Fig. S4). Diversity patterns after decontamination were similar, with generally greater bacterial diversity observed in fly ash and sand samples than in concrete and diversity similar to that in concrete observed in gravel and powder samples (Fig. 3). Shannon and Faith's diversity metrics generally showed similar differences between groups, although decontaminated gravel samples and cement powder samples before decontamination had comparatively low Shannon values.

Each of the precursor materials had indicator taxa that were exclusive to that material. Indicator values (IndVal), a measure of exclusiveness, were calculated for these taxa as the product of relative observation frequency (specificity) and likelihood of observation (sensitivity) with a correction for unequal group sizes (46). Gravel communities, the least diverse of the precursors, were composed primarily of *Acinetobacter*, *Pelotomaculum*, *Bacillus*, *Aeromicrobium*, and *Burkholderia* (see table S4 at doi.org/10
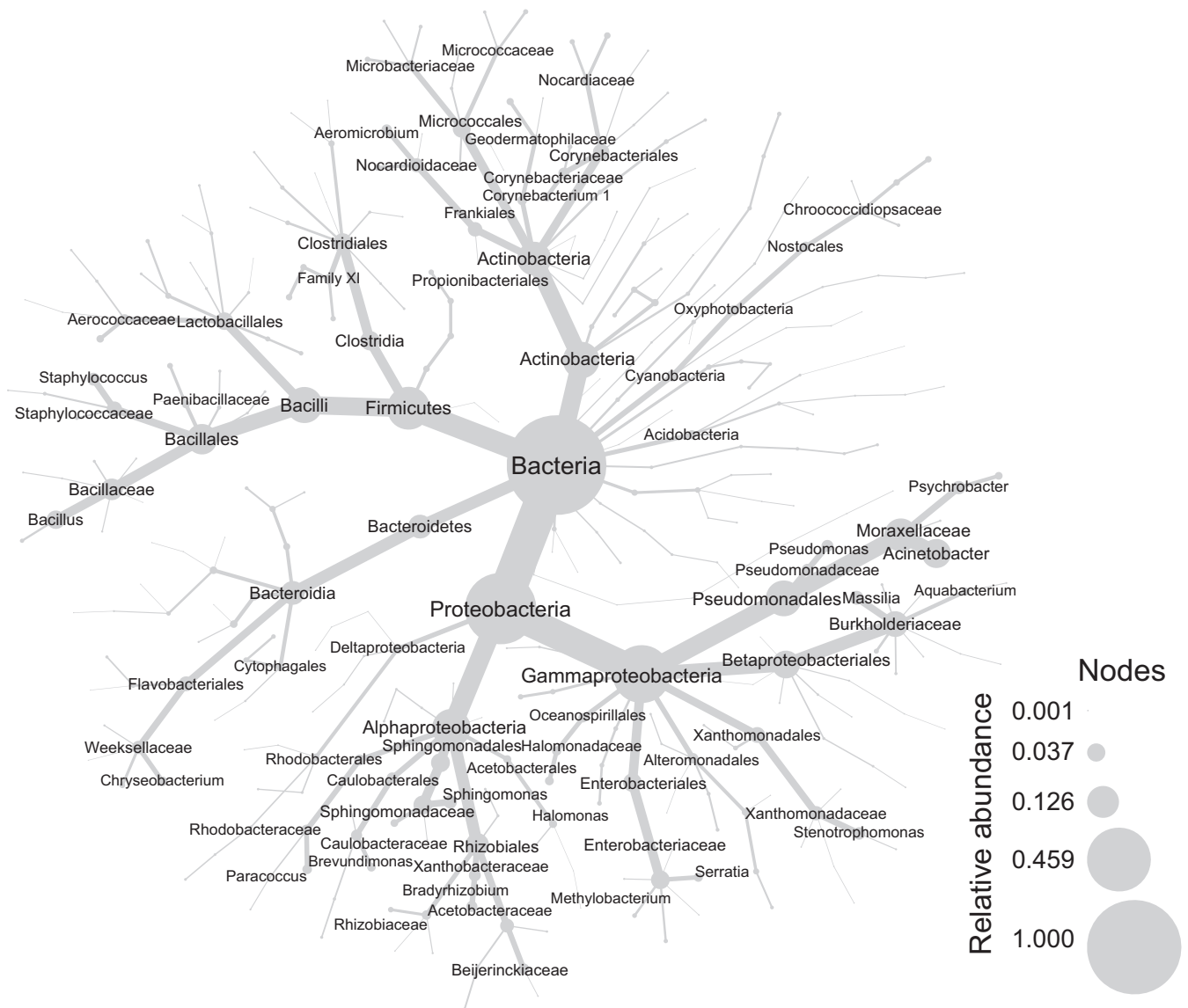
FIG 2 Concrete bacterial community relative abundance. The relative abundances of bacterial taxa observed in all concrete samples from the time series are indicated by node size and edge thickness. Higher phylogenetic ranks are more central in the tree, and lower ranks, such as species, are found at the periphery. *Proteobacteria*, *Firmicutes*, and *Actinobacteria* are the most abundant phyla observed. ASVs determined to be contaminants have been removed.

.6084/m9.figshare.14211038). *Nosocomiicoccus* was an indicator for gravel samples (IndVal = 0.82, P = 0.02), although its relative abundance—the proportion of all observations—was low (6×10^{-5}). Cement powder bacterial communities consisted primarily of *Pantoea*, *Burkholderia*, *Acinetobacter*, *Psychrobacter*, *Streptococcus*, *Staphylococcus*, *Chryseobacterium*, and *Corynebacterium* (see table S4 at doi.org/10.6084/m9.figshare .14211038). *Nocardia* was an indicator of cement powder samples (IndVal = 0.996, P = 0.001, rel.abund ~ 0.0003). Fly ash microbial communities consisted primarily of *Paracoccus*, *Hydrogenophaga*, *Bacillus*, *Thiobacillus*, and *Nocardioides* (see table S4 at doi.org/10.6084/m9.figshare.14211038). Several genera were indicators of fly ash, including *Thermithiobacillus* (IndVal = 1, P = 0.001, rel.abund ~ 0.005), *Meiothermus* (IndVal = 0.998, P = 0.001, rel.abund ~ 0.02), *Truepera* (IndVal = 0.988, P = 0.005, rel.abund ~ 0.02), and uncultured members of "*Candidatus* Kaiserbacteria" (IndVal = 1, P = 0.001, rel.abund ~ 0.01).

Of the precursors, sand microbial communities were the most different from those of concrete. We observed more *Alphaproteobacteria* in sand samples (see table S4 at doi.org/
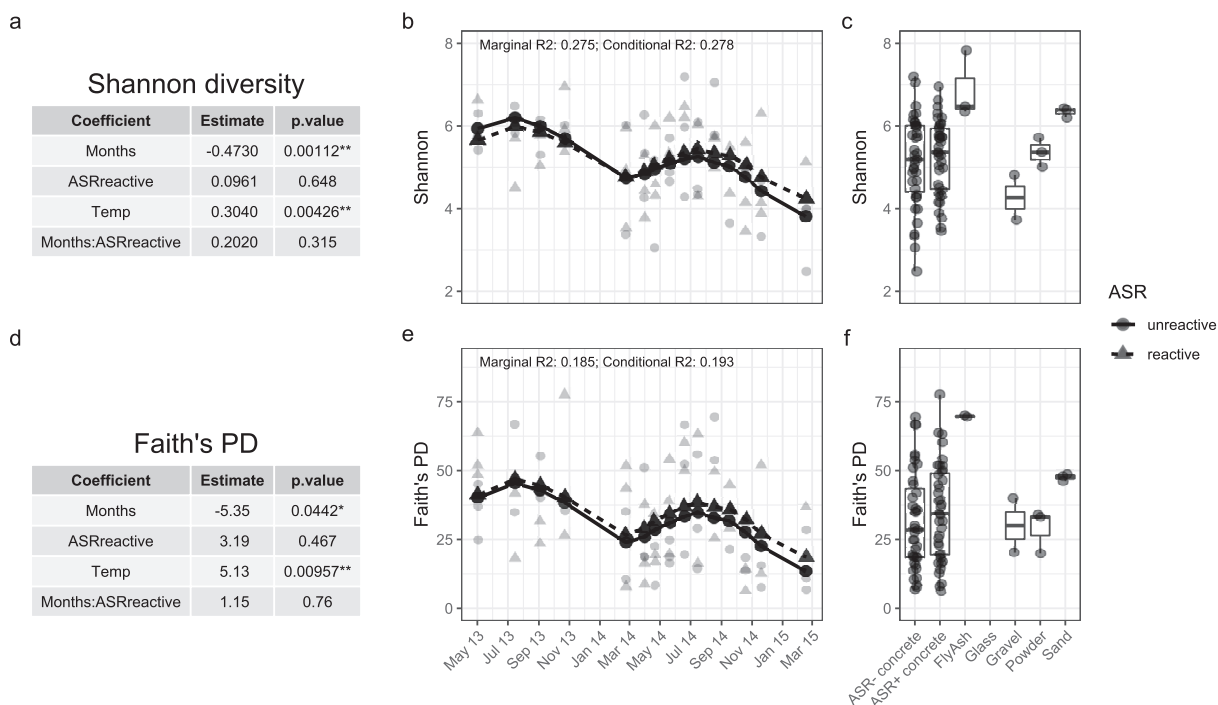
FIG 3 Within-sample diversity over time. Two metrics of bacterial diversity, Shannon diversity (a to c) and Faith's phylogenetic diversity (d to f), show similar trends in both decontaminated cylinder series. In panels b and e, light points are the observed sample values, while connected dark points were predicted with a linear mixed-effect model. (a and d) Coefficients for both models show that time (months) has a significant negative effect on diversity, while temperature has a significant positive effect (*, $P \leq 0.05$; **, $P \leq 0.01$). ASR reactivity appears to have little effect on bacterial diversity, although reactive concrete is associated with (nonsignificant) increased diversity for both metrics. The interaction effect of time and ASR reactivity indicates that over time, more diversity is associated with ASR reactivity; however, this cannot be considered significant, since the effect of ASR reactivity alone was nonsignificant. (c and f) Concrete sample diversity compared to precursor material diversity.

10.6084/m9.figshare.14211038), specifically *Sphingomonadaceae* like *Rubritepida* (rel.abund ~0.05), *Porphyrobacter* (rel.abund ~0.03), *Sphingoaurantiacus* (rel.abund ~0.002), and *Sandaracinobacter* (rel.abund ~0.0004) (all indicator values > 0.99, $P < 0.01$). Other taxa with high indicator values for sand were *Bacteroidia*, including *Rhabdobacter* (IndVal = 0.97, $P = 0.02$, rel.abund ~0.06) and *Cnuella* (IndVal = 0.99, $P = 0.001$, rel.abund ~0.05), and the *Gammaproteobacteria Alishewanella* (IndVal = 0.99, $P = 0.001$, rel.abund ~0.07) and *Lysobacter* (IndVal = 0.98, $P = 0.007$, rel.abund ~0.02). *Gemmatirosa* (IndVal = 0.99, $P = 0.001$, rel.abund ~0.02) and several *Verrucomicrobia*, such as *Prosthecobacter* (IndVal = 1, $P = 0.001$, rel.abund ~0.004), also had high indicator values for sand.

Influence of precursor materials on concrete microbial communities. SourceTracker analysis (47) was used to assess the influence of the concrete constituent material microbial communities on bacteria in the first (t_0; May 2013) and last (t_{end}; February 2015) concrete samples (Fig. 4). Using microbial community data from concrete precursor materials (sources), SourceTracker identified the probable fraction of microbes in the concrete cylinders coming from each precursor material and also estimated the contribution of an "unknown" source (mean contribution to reactive and unreactive concrete: t_0 ~ 39%, t_{end} ~ 23%). Precursor sources considered in this analysis were fly ash (t_0 ~ 7%, t_{end} ~ 1%), the large aggregate gravel (t_0 ~ 32%, t_{end} ~ 41%), the fine aggregate sand (t_0 ~ 2%, t_{end} ~ 3%), and cement powder (t_0 ~ 14%, t_{end} ~ 30%). Amplicon sequence data were not obtained for the water used in cylinder preparation (Newark, DE, tap water) and was instead approximated with data from multiple water utilities in the eastern United States published in reference 48. Leave-one-out cross validation was performed (Fig. S5). This SourceTracker analysis suggests that precursor materials contribute a considerable (>50%) portion of the concrete microbiome in early
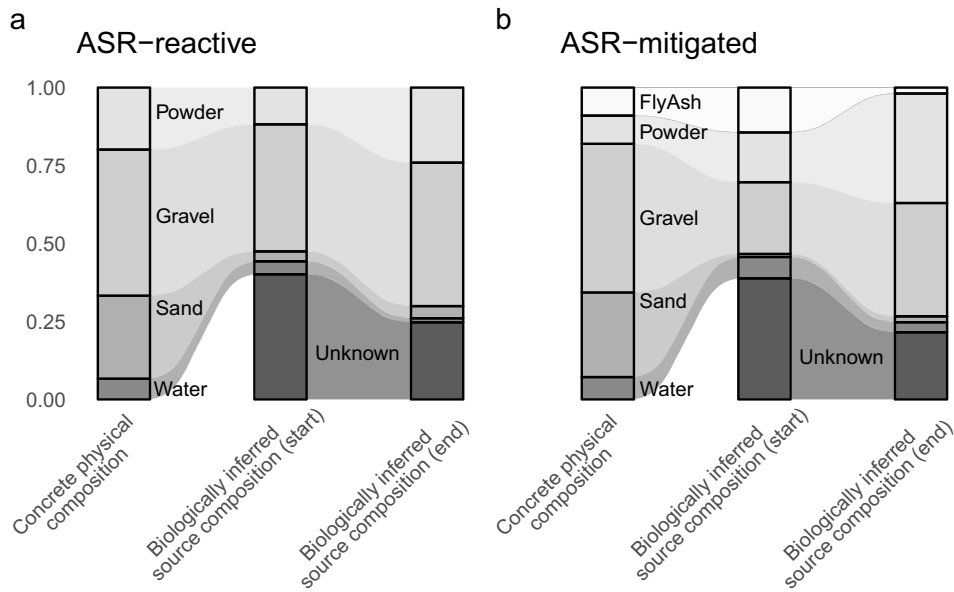
FIG 4 Influence of precursors on concrete bacterial communities. Concrete composition by weight (a and b, left) compared to the mixing proportion of source materials inferred by comparing concrete microbial communities to precursor material communities (a and b, right). Using sequenced microbial community information for concrete samples ("sinks") and precursor materials ("sources"), the SourceTracker2 tool was used to infer the proportion of bacteria in concrete sinks originating from each precursor source. SourceTracker also predicts a proportion of bacteria from an unknown source. Inferred proportions are for the first concrete samples in each series, archived prior to weathering. Unlike the other precursor materials, sequence data were not directly obtained from the water used for concrete production (Newark, DE, tap water). Instead, tap water samples from multiple utilities in the eastern United States were used (48).

samples, which represents an even larger portion of the bacterial community in later samples (Fig. 4). Gravel had the greatest contribution. Interestingly, sand was found to have little influence on the composition of the concrete microbial community, though it is a major constituent material of concrete and had high bacterial diversity (Fig. 3). In the ASR-mitigated concrete, gravel had a smaller contribution while fly ash and cement powder had larger predicted contributions to the microbial community (Fig. 4). Over time, the proportion of microbes predicted to be of unknown origin decreased, while similarity to gravel and cement powder communities increased.

Time and season affect diversity and community composition. We analyzed the bacterial diversity in concrete and precursor materials and how diversity in concrete changed over time. Within-sample (alpha) diversity metrics, Shannon and Faith's phylogenetic diversity, were calculated with QIIME 2 on decontaminated individual concrete samples subsampled to 1,000 observations (Fig. 3). Shannon diversity summarizes both richness and evenness; however, because it can obscure absolute richness, Faith's phylogenetic diversity was also included. Linear mixed-effect models were applied to both metrics, yielding similar results: diversity decreased over time (months of weathering) (Shannon's estimate $= -0.47$, $P = 0.001$; Faith's PD estimate $= -5.35$, $P = 0.04$), with seasonal increases associated with warmer temperatures (Shannon's estimate $= 0.3$, $P = 0.004$; Faith's PD estimate $= 5.13$, $P = 0.01$). Probability of ASR did not have a significant effect on observed diversity, although reactive samples had slightly higher diversity than mitigated samples with both metrics (Shannon's estimate $= 0.096$, $P = 0.6$; Faith's PD estimate $= 3.2$, $P = 0.5$) (Fig. 3). To understand the effects of decontamination, alpha diversity was also calculated from raw (before decontamination) data (Fig. S4). Removing laboratory contaminants strengthened the apparent effect of season on diversity, though the effect was present in the raw data.

Permutational multivariate analysis of variance (PERMANOVA) models (49, 50) were used to assess environmental influences on community composition. Like alpha

TABLE 1 PERMANOVA results for the generalized UniFrac metric[a]

Term testing	Coefficient	R^2	F	P value	Dispersion P value
Sequential	Temp	0.024	1.851	0.003**	0.891
	Mo	0.025	1.972	0.002**	0.891
	ASR	0.01	0.802	0.896	0.671
Marginal	Temp	0.02	1.607	0.015*	0.891
	Mo	0.025	1.939	0.001***	0.891
	ASR	0.01	0.802	0.898	0.671

[a]PERMANOVA models (distance ~ temperature + months + ASR) were computed with the *adonis2* function of the *vegan* R package for generalized (0.5) UniFrac distances. The effect of terms was tested both sequentially and marginally. For both models, temperature and time had significant (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.005$) associations with observed bacterial community differences. Pseudo-F ratios compare the total sum of squared dissimilarities between groups to those within groups, a measure of group separation, with larger F ratios indicating greater differences between groups. Dispersion was also assessed with the *betadisper* function of the *vegan* R package for each distance/term pair because of potential for confounding. No significant differences in dispersion were observed.

diversity, time (months of weathering) and temperature were associated with significant changes ($P < 0.05$) in overall community structure as determined with PERMANOVA performed on generalized (0.5) UniFrac distances (Table 1). This significant effect was observed when added sequentially to the model (distance ~ temperature + time + ASR) and also when the marginal effect of each term was considered. The marginal effect of time was greater than that of temperature, indicating that the time since pouring has a greater impact than temperature. No significant difference in overall community composition was observed between ASR-reactive and ASR-unreactive cylinders. Dispersion was also assessed, since it can confound PERMANOVA models, but no significant differences were found (Table 1). Other distance metrics yielded consistent results (Table S1), as did Mantel matrix correlation tests, although the Mantel tests did not find time alone to be significant (Table S2). PERMANOVA on pre-decontamination data also revealed significant time and temperature effects (Table S3).

Taxonomic changes over time. Large-scale changes in diversity and community composition are often driven by changes in the presence and/or abundance of particular taxa. Therefore, each pattern observed with alpha- and beta-diversity metrics should have groups of bacteria following the same pattern. The changes in diversity associated with time and temperature reflect changes in patterns of abundance of both common and rare taxa.

Time was associated with decreased diversity and changes in community composition. Similarly, generally decreasing diversity with seasonal peaks in warmer summer months was observed over the series in the most abundant phyla: *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, and *Cyanobacteria* (Fig. 5). Decreasing prevalence (the proportion of samples a group appears in) of the most abundant genera like *Acinetobacter* and *Bacillus* could help explain this trend, especially when coupled with taxa that became significantly less common, like *Beijerinckiaceae* (log odds = 0.24, P = 0.0004) (Fig. 5; see table S8 at doi.org/10.6084/m9.figshare.14211038). Many taxa were more frequently observed in the summer, such as *Ferruginibacter* (Fig. 5; see table S9 at doi.org/10.6084/m9.figshare.14211038), whose seasonal change in relative abundance was detected by fitting a 1-year cyclical spline as part of a general additive mixed model (R^2 = 0.8, P = 0.00014), used previously to detect long-term and seasonal trends (51).

While community-level analyses capture the largest changes, they also obscure smaller changes. We were particularly interested in bacteria capable of surviving the harsh conditions of concrete, which would increase in relative abundance and/or prevalence throughout the series as other species die off. *Psychrobacter* was identified as an example of a taxon that increased throughout the series in both prevalence (log odds = 1.83, P = 0.017) and relative abundance (Fig. 5; see table S10 at doi.org/10.6084/m9.figshare.14211038).
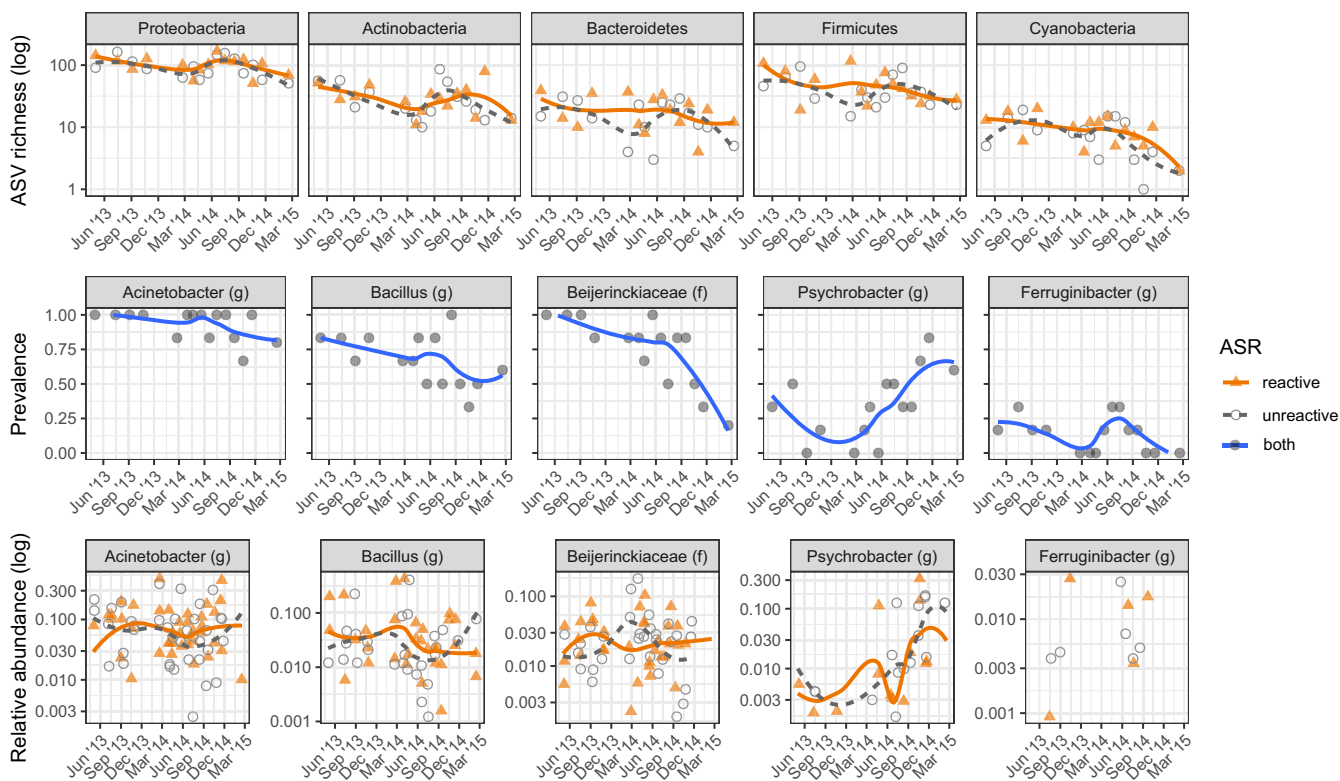
FIG 5 Taxon changes over time. Changes in the diversity (top), prevalence (middle), and relative abundance (bottom) of certain taxa observed over the 2 years during which concrete cylinders were weathered. The diversity (number of ASVs) of the five most abundant phyla is shown in the top row; they generally exhibit the same trends of decreasing diversity with seasonally increased diversity in summer seen for overall diversity (Fig. 3). Several examples of prevalence (the proportion of samples in which a group appears) at each time point are shown, including for the two most abundant genera, *Acinetobacter* and *Bacillus*, whose prevalence roughly mirrors prevailing diversity patterns. Logistic regression of presence/absence across the series was used to identify taxa with increasing (*Psychrobacter*) or decreasing (*Beijerinckiaceae*) prevalence. Seasonally associated taxa were detected with general additive mixed models, such as the summer-associated genus *Ferruginibacter*. The bottom row shows how relative abundance of the prevalence examples changes over time. While the same patterns are generally seen, they are more apparent for some taxa, like *Psychrobacter* and *Ferruginibacter*, and obscured for taxa like *Beijerinckiaceae* and *Acinetobacter*. Lines show local regression (locally estimated scatterplot smoothing [LOESS]).

Earth Microbiome Project comparison. The Earth Microbiome Project (EMP) is a data set of microbial communities from many different environments analyzed using a standardized protocol and classified in a standardized ontology (52). As little is known about the concrete microbiome, the EMP provides a simple comparison to similar environments and to potential influences such as aerosolic deposition or animal feces. Decontaminated concrete ASVs were trimmed to 90 bp and merged (vsearch at 99% similarity) with EMP ASVs to create a combined table of operational taxonomic units (OTUs).

Principal-coordinate analysis (PCoA) ordination of generalized (0.5) UniFrac distances shows that concrete communities overlap communities from aerosols, surfaces, negative controls (EMP sterile water blanks), sebum, dust, animal and corpus/surfaces, nonsaline waters, and hyperalkaline environments (Fig. 6a). Concrete samples are near the middle of PC1, between the major groups driving variance, with animal-associated communities found at high PC1 values and plant/soil communities at low PC1 values. However, concrete samples clearly have some spread along PC1, with most precursor samples found at lower, less animal-associated PC1 values. Concrete samples are also found at central PC2 values, above water samples. Concrete communities are more similar to nonsaline water and tap water communities than to saline water and plant corpus communities, which are separated along the PC3 axis.

Pairwise generalized (0.5) UniFrac similarities (similarity = 1 − distance) (53, 54) were used to assess the mean similarity of microbial communities for all concrete and EMP samples grouped at level 3 of the EMP ontology (EMPO) (55) (Fig. 6b). The
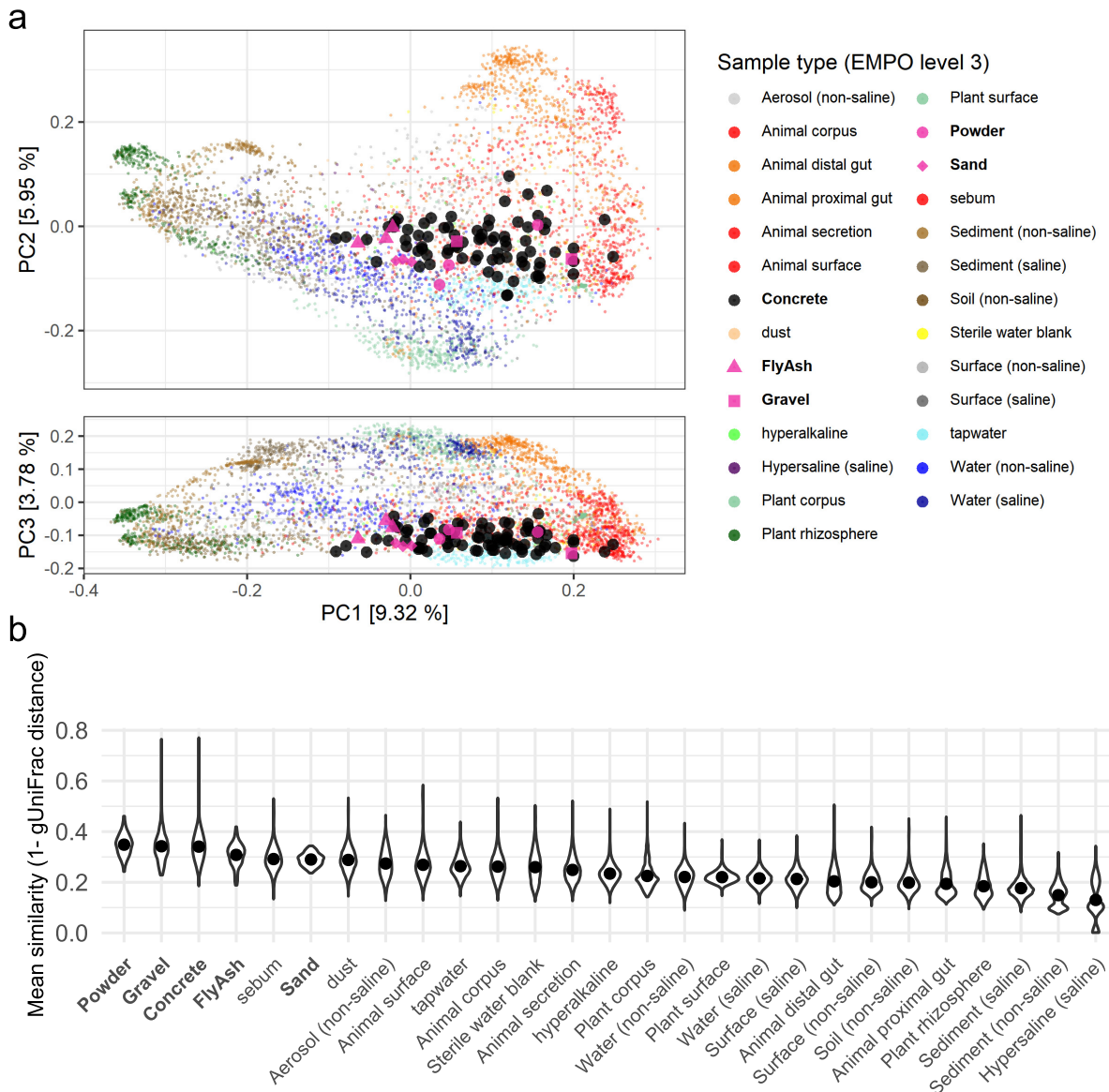
FIG 6 Concrete compared to Earth Microbiome Project samples. A merged OTU table was made with concrete ASVs clustered at 99% similarity with EMP ASVs (52) and tap water ASVs (48). (a) PCoA plot based on generalized (0.5) UniFrac distances. Concrete and precursor samples (bold) cluster together in the center of the PCoA plot, an area shared with a large diversity of other sample types that likely have many influences, such as animal surfaces, aerosols, surfaces, tap water, and alkaline environments. However, there is a clear spread from greater similarity to animal-related EMP samples (higher PC1 values) to environments like water and soils (lower PC1 values). (b) Similarity of concrete samples to precursor materials and EMPO3 groups shown with violin plots, where points represent the mean similarity.

communities observed in the cement powder and gravel used in the concrete mix were on average as similar to concrete communities as those of other concrete samples (~0.34), with fly ash communities being slightly less similar (~0.31). The most similar EMPO3 groups to concrete were sebum and dust, which were approximately as similar as the sand used in concrete production (~0.29). Aerosols, animal surfaces, tap water, and sterile water blanks were the next most similar (~0.26). Microbial communities from alkaline environments, plant and saline surfaces, and nonsaline water had a similarity of about 0.22, while sediment, soil, animal gut, and hypersaline environment microbial communities were the least similar to concrete communities. Changes in the similarity of concrete and EMP groups were also observed over time (Fig. S6).

To characterize the effects of decontamination, similarities of concrete before decontamination to the EMP were also computed. Removing contaminant taxa increased the

similarity of concrete sample communities to EMP samples from aerosol, plant rhizosphere, sebum, surface (saline, nonsaline, and plant), nonsaline water, dust, and nonsaline sediment communities (Fig. S7). It had little effect on similarity to saline water and sediment, hypersaline environments, hyperalkaline environments, animal guts and secretions, and tap water (Fig. S7). Decontamination also resulted in decreased similarity between concrete samples and all precursor materials except sand, which increased in similarity but remained the least similar precursor. Contaminant removal also decreased similarity to animal and plant corpuses, sterile water blanks, and animal surfaces, suggesting that they might be sources of contaminants.

Bacteria found in concrete were also found in many hyperalkaline and hypersaline EMP samples. The genus *Paenibacillus* was an indicator for hyperalkaline, plant rhizosphere, and aerosol EMP samples (IndVal = 0.63, *P* = 0.001), as was the genus *Bacillus* (IndVal = 0.60, *P* = 0.001) (see table S11 at doi.org/10.6084/m9.figshare.14211038). The genus *Halomonas* was an indicator for hypersaline, saline surface, and concrete samples (IndVal = 0.44, *P* = 0.001) (see table S11 at doi.org/10.6084/m9.figshare.14211038). At the ASV level, an *Exiguobacterium* sp. was an indicator of concrete and hyperalkaline samples (IndVal = 0.36, *P* = 0.001) while a *Halobacillus* indicated concrete and hypersaline samples (IndVal = 0.21, *P* = 0.009) and a member of *Rhodobacteraceae* found in concrete was an indicator for alkaline EMP samples (IndVal = 0.58, *P* = 0.001) (see table S4 at doi.org/10.6084/m9.figshare.14211038). We were also interested in bacteria uniquely associated with concrete and identified several indicators belonging to several genera: *Aliidiomarina* (IndVal = 0.44, *P* = 0.001), *Alishewanella* (IndVal = 0.34, *P* = 0.001), and *Chroococcidiopsis SAG 2023* (IndVal = 0.26, *P* = 0.005) (see table S11 at doi.org/10.6084/m9.figshare.14211038).

Microbial bioindicators of ASR. Taxa associated with ASR (Fig. 7) could potentially serve as bioindicators of the ASR. Several methods were used to identify bioindicator taxa grouped at the genus level, including indicator species analysis (46), smoothing spline analysis of variance (ssANOVA) from the *metagenomeSeq* R package (56), and logistic regression using presence/absence data. Indicator species analysis considers both the specificity and sensitivity of potential indicators, and it identified *Arcobacter* (IndVal = 0.54, *P* = 0.007) and *Bryobacter* (IndVal = 0.39, *P* = 0.023) as potential bioindicators of the ASR (Fig. 7).

In contrast, ssANOVA allows time to be explicitly considered in selecting indicators, as it finds time periods of differential abundance. As the ASR increases over time, we expected indicators to become or remain differentially abundant later in the series. ssANOVA identified the following potential indicators and their periods of differential abundance, with month 21 being the last in the series: *Bryobacter* (interval, 10 to 21 months; *P* = 0.004), *Sediminibacterium* (interval, 9 to 21 months; *P* = 0.009 [data not shown]), *Lawsonella* (interval, 16 to 21 months; *P* = 0.016), *Arcobacter* (interval, 16 to 21 months; *P* = 0.024), *Modestobacter* (interval, 12 to 21 months; *P* = 0.027), *Salinicoccus* (interval, 19 to 21 months; *P* = 0.031), and *Carnobacterium* (interval, 15 to 21 months; *P* = 0.036) (Fig. 7).

Finally, logistic regression was used on presence/absence data to find potential indicators with a significant (*P* < 0.05) effect for the interaction of time and ASR reactivity. Potential indicators of the ASR identified with logistic regression included *Rhodocyclaceae* (log odds = 5.19, *P* = 0.016), *Rheinheimera soli* (log odds = 5.5, *P* = 0.020), *Flavobacterium* (log odds = 4.62, *P* = 0.046), and *Lawsonella* (log odds = 5.94, *P* = 0.038). In total, *Arcobacter*, *Lawsonella*, and *Bryobacter* were identified with multiple methods as robust potential bioindicators. The potential bioindicator taxa have increased abundance in ASR-reactive samples but, more importantly for use as indicators, are also found in more ASR-reactive samples than mitigated samples. Further, the differentiating power generally increases with time, as do the effects of the ASR.

DISCUSSION

As the most-used building material in the world, concrete is a very common environment. However, relatively few microbes can survive this dry, salty, alkaline habitat
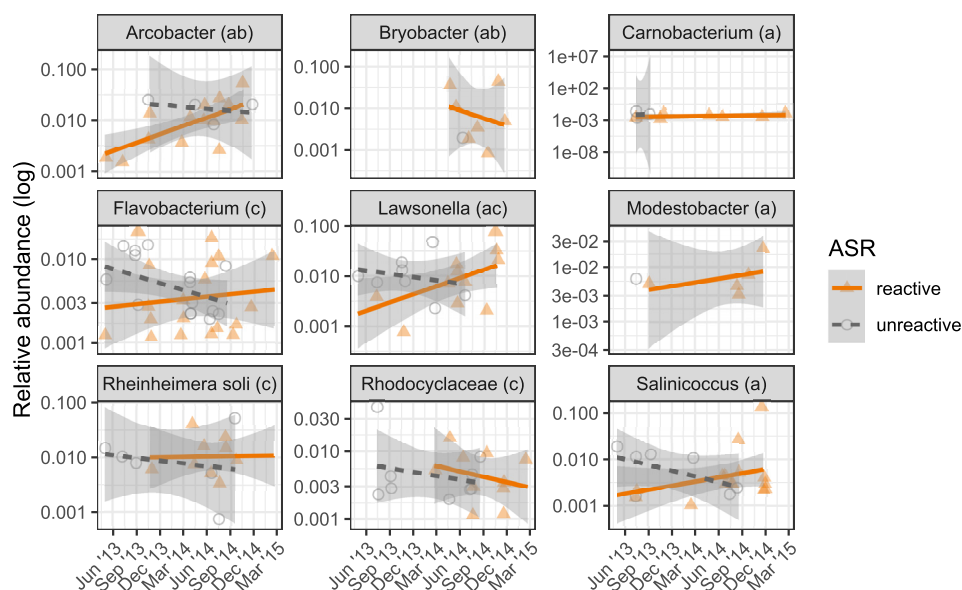
FIG 7 Potential bioindicators of the ASR. Bioindicators of the ASR determined with ssANOVA (indicated with the letter "a") in *metagenomeSeq* (which determined time periods of differential abundance), indicator species analysis (indicated with the letter "b") in *indicspecies* (which considers both specificity and sensitivity of the indicator), and logistic regression (indicated with the letter "c") for presence/absence. All taxa are summarized at the genus level except for *Rheinheimera soli* and the family *Rhodocyclaceae*. For ssANOVA, indicator taxa that remained differentially abundant through the end of the series were identified. The following indicators were identified with ssANOVA: *Bryobacter* (interval, 10 to 21 months; *P* value, 0.004), *Sediminibacterium* (interval, 9 to 21 months; *P* value, 0.009), *Lawsonella* (interval, 16 to 21 months; *P* value, 0.016), *Arcobacter* (interval, 16 to 21 months; *P* value, 0.024), *Modestobacter* (interval, 12 to 21 months; *P* value, 0.027), *Salinicoccus* (interval, 19 to 21 months; *P* value, 0.031), and *Carnobacterium* (interval, 15 to 21 months; *P* value, 0.036). The following indicators were identified with indicator species analysis: *Arcobacter* (IndVal = 0.54, *P* = 0.007) and *Bryobacter* (IndVal = 0.39, *P* = 0.023). Logistic regression, run at multiple taxonomic levels, identified the following indicators: *Rhodocyclaceae* (log odds = 5.19, *P* = 0.016), *Rheinheimera soli* (log odds = 5.5, *P* = 0.020), *Flavobacterium* (log odds = 4.62, *P* = 0.046), and *Lawsonella* (log odds = 5.94, *P* = 0.038).

with very low macronutrient concentrations. We predicted that the microbial community of concrete would be seeded from the component materials and that over time it would begin to resemble communities from other dry, saline, and/or alkaline environments. We further hypothesized that the physical and chemical changes due to ASR damage would be reflected in the microbial communities of ASR-reactive and ASR-mitigated concrete.

Here, we show that microbial communities in ASR-prone and ASR-mitigated concrete series reflected those of the components of concrete. The inclusion of fly ash to prevent ASR in the ASR-mitigated concrete mix, and the resulting difference in probability of ASR, did not have a statistically significant effect on overall community composition. However, individual taxa were identified with significantly different abundance patterns in the ASR-reactive and ASR-mitigated series, and these could potentially serve as microbial bioindicators of the ASR. As the cylinders in both series weathered, the diversity in their bacterial communities decreased, though we observed seasonal increases in diversity in the summer.

ASR generates a hygroscopic silica-rich gel, which, as it expands, cracks the concrete from the inside. We expected the ASR-reactive samples to be more alkaline than the ASR-mitigated samples, because fly ash reduces the pH (21). We also expected that because of the cracking, more water and nutrients would infiltrate the ASR-reactive samples. ASR reactivity, therefore, might increase selection for alkaliphiles while reducing selection for oligotrophs or xerophiles. In fact, the identification of *Bryobacter* as a potential bioindicator for the ASR was surprising, given its description as "mildly acidophilic" (57) and its previous association with wet environments: bogs (58), wetlands (59), and wastewater (60). Its presence may imply greater water infiltration into ASR-damaged concrete, though this genus

has also been observed in desert biocrusts (61, 62). Another putative ASR bioindicator, *Modestobacter* (family *Geodermatophilaceae*), is commonly associated with dry environments: stone surfaces, desert biocrusts, and endolithic biofilms (63, 64). *Salinicoccus* is a member of the *Staphylococcaceae* commonly found in high-salt environments, including desert soils (65, 66), saline soils (67), coastal water (68), salt mines (69, 70), salterns (71), and soda lakes (72–74). These potential bioindicator genera are not known to be alkaliphiles. Alternatively, because concrete is seeded from several different materials, we also expected to see generalist species capable of surviving a variety of different stresses, as previously observed in a cultivation campaign (2). Several of the other potential ASR indicators identified, including *Arcobacter*, *Flavobacterium*, *Rheinheimera*, and *Rhodocyclaceae*, are generalists found in a wide range of environments. Interestingly, *Rheinheimera* spp. are found in marine (75), coastal (76–78), freshwater (79–81), industrial waste (82, 83), and alkaline environments (84, 85). One strain produces calcite (86)—potentially applicable for concrete biorepair—and another has recently been associated with stainless steel corrosion (87). The *Flavobacterium* and *Rhodocyclaceae* groups are true generalists (88, 89). Since the ASR-mitigated and ASR-reactive cylinders were produced using the same starting materials, with the exception of the fly ash, these bioindicators suggest that differences due to physical and chemical effects of the ASR are reflected in their microbial communities. Further work using genomics and metagenomics to analyze microbial metabolic functions in different types of concrete may better elucidate the physiological reasons for the differences.

Concrete samples share similarities with microbial communities in the Earth Microbiome Project (EMP) (52) from physically and chemically similar environments such as alkaline soils and travertine, desert soils, stone surfaces, and hypersaline lakes. Indicator species analysis with the combined EMP data found several bacterial groups strongly associated with both concrete and saline EMP samples, including the gammaproteobacterial genus *Halomonas* and the *Firmicutes* genus *Halobacillus*, which include both halotolerant and alkalitolerant species (90–92). Other groups of bacteria in the concrete community were strongly associated with alkaline environments, including *Paenibacillus*, *Bacillus*, *Exiguobacterium*, and *Rhodobacteraceae*. Several taxa were indicators of concrete, including halotolerant and alkalitolerant *Aliidiomarina* (93–95), *Alishewanella* (96, 97), and the desiccation-resistant cyanobacterium *Chroococcidiopsis* (98–100). We also observed several taxa in concrete that are commonly found in desert soil crusts (101, 102), including the radiation-resistant genera *Rubrobacter* (103) and *Acinetobacter* (104), the nitrogen-fixing genus *Frankia* (105), and the halophile- and alkaliphile-containing genera *Halomonas* (91, 106), *Bacillus* (107), and *Stenotrophomonas* (108, 109). Photosynthetic *Chloroflexi* (110) and *Cyanobacteria* belonging to the genus *Nostoc* (111, 112) are also common in saline soil crusts and were present in concrete. Common epilithic bacteria were also found in concrete, including *Sphingomonas*, *Frankiales*, *Truepera*, *Hymenobacter*, *Sphingobacteriales*, *Massilia*, *Stenotrophomonas*, and *Paenibacillus* (113–115). In addition, several species belonging to groups common in alkaline soda lakes were detected, including *Dietzia*, *Halomonas*, *Paenibacillus*, *Exiguobacterium*, *Bacillus*, and *Rhodobacteraceae* (116, 117). Fundamentally, these groups all tolerate osmotic stress, which is likely necessary for survival in concrete, whether it is imposed by salt concentration, pH, or low water activity.

Although each batch of concrete is seeded from its materials, the chemical and physical characteristics of all concrete are similar, which may impose strong enough selective pressure that after a time, microbial communities in concrete become broadly similar. Thus, we expected concrete to have a community of generally resilient microbes, with a few (poly)extremophiles, rather than an endemic community of unique microbes that have adapted to harsh conditions over a long period of time. Characteristics of the gravel (large aggregate) typically determine the ASR reactivity of the concrete (24), and 30 to 40% of the microbial community in these cylinders likely came from gravel. The contribution from cement powder ($\sim$14% and 30% at the beginning and end of the study, respectively) was surprising, since it is a highly alkaline, oxidized material produced in

high-temperature (~1,500°C) furnaces (118). The gravel- and cement powder-associated microbes make up a larger portion of the microbial community in aged samples, suggesting that gravel and cement powder resemble concrete in terms of the constraints imposed on the microbial community. As the microbial community became less diverse, the survivors from these materials increased in relative abundance.

Unlike naturally occurring extreme environments, which are stable on time scales up to hundreds of thousands of years and allow ample time for evolution of specialists, concrete is unlikely to have endemic microbes. Instead, its microbial community is seeded from precursor materials in each batch and is also likely influenced and/or seeded by the surrounding environment, as suggested by summertime increases in diversity and DNA yield, a proxy for biomass (119). Seasonal changes in diversity and biomass are common in environments such as temperate soils and aerosols (120–122) and tend to be greater in stressed environments (123). In our samples, microbial growth or survival might have been enhanced in the summer due to greater precipitation, since infiltration of water would relieve xeric stress and could also bring dissolved nutrients into the concrete matrix. Further studies of older concrete, concrete exposed to different environmental conditions, or concrete subject to different types of damage will therefore have implications not just for bioindicators in the built environment but also for understanding microbial seeding of and adaptation to both concrete and other composite materials.

Conclusions. This work provides a comprehensive picture of microbial communities in concrete during the first 2 years of weathering. We discovered that most of the microbiome in concrete is seeded from the materials used to make it and that this microbial community is broadly similar to other communities exposed to different types of osmotic stress. We show that the community composition changes over the 2 years, decreasing in diversity over time, and that weather also impacted community composition, with increased diversity in summer months. Knowing how bacterial communities in and on concrete change over time potentially benefits several applications. Understanding the naturally resident bacteria and their temporal dynamics in concrete could improve ongoing efforts to repair concrete via microbially induced mineral deposition, and identification of indicator species of the concrete degrading alkali-silica reaction could enable earlier detection of the ASR and thus more effective damage mitigation.

MATERIALS AND METHODS

Preparation and exposure of test cylinders. Two series of concrete cylinders (Fig. 1), one prone to ASR and the other having the risk of ASR mitigated with fly ash, were prepared in the UD Structures Lab using materials obtained from the Delaware Department of Transportation (DelDOT). ASR-reactive concrete was prepared following the mix design for DelDOT class A-503 concrete and contained 20.14 kg of cement, 6.76 kg of tap water, 26.94 kg fine aggregate (sand), and 47.40 kg large aggregate for a total weight of 101.24 kg. ASR-mitigated concrete was prepared using the mix design for DelDOT class A-3-50 with the same materials but with replacement of half of the cement with fly ash. ASR-mitigated concrete contained 15.38 kg each of cement and fly ash, 12.25 kg tap water, 46.31 kg fine aggregates, and 81.37 kg large aggregates for a total weight of 170.69 kg. Additional water was added to improve workability.

Concrete cylinders were produced using cylindrical plastic forms, 4 in. in diameter by 8 in. high, agitated with a shaker table after pouring to remove air pockets, and then capped and allowed to cure at room temperature for 2 weeks before removal from forms. One cylinder from each series was archived on 2 May 2013, and the rest were placed on the roof of Colburn Laboratory at the University of Delaware for exposure to weather while protected from foot and car traffic. One cylinder from each series was then archived every 4 to 8 weeks until 17 February 2015.

Weather data. Weather data for the sampling period were downloaded from Weather Underground for the KILG (New Castle Airport) site with the *rwunderground* R package. Summary weather information was compiled for the 30 days prior to each sampling. The R script used to generate summary information is included in the GitHub repository (github.com/MarescaLab/concrete_series).

Sample processing, DNA extraction, 16S amplification, and sequencing. From each cylinder, slices perpendicular to the long axis were obtained using a saw (blade and platform cleaned with 70% ethanol); a chisel cleaned with 70% ethanol was then used to remove the outside edges and break apart the section. Subsamples were pulverized using a ring and puck mill cleaned with 70% ethanol, and the resulting powder was stored at −20°C until DNA extraction. DNA was extracted in triplicate as described by Maresca et al. (2): 5 g of sample was washed with 20 ml of 0.5 M EDTA to remove divalent cations, followed by suspension in a lysis buffer and lysozyme with agitation at 37°C for 30 min. Proteinase K and 20% SDS were then added and

incubated for 2.5 to 3 h with gentle agitation at 56°C, followed by chloroform (20 ml) extraction. One milliliter of 1.95 M sodium acetate was added to the aqueous phase of each sample and re-extracted with 0.8 volume of chloroform. Finally, DNA was precipitated with 1 volume of cold isopropanol and 0.1 volume 3 M sodium acetate, washed with 70% ethanol, allowed to dry, and resuspended in 200 μl Tris-EDTA (TE). DNA was reprecipitated with 13% polyethylene glycol (PEG) 8000, washed with 70% ethanol, dried, resuspended in 25 μl 10 mM Tris (pH 8.0), and stored at −20°C.

DNA was also extracted in triplicate from equivalent quantities (5 g) of precursor materials used in concrete cylinder production: large aggregates (gravel), fine aggregates (sand), Portland cement, and fly ash. Triplicate negative-control DNA extractions were also performed on glass beads (5 g) that had been sterilized by bleaching, UV irradiation, and autoclaving. Extracted DNA was quantified with an Invitrogen Qubit 2.0 fluorometer. PCR with general 16S primers 357F and 806R and visualization with gel electrophoresis were performed to check that no PCR inhibitors were present. 16S amplicon libraries were generated at the UD Sequencing and Genotyping Center following the Illumina 16S metagenomic sequencing library preparation protocol (124) using primers 357F (CCTACGGGNGGCWGCAG) and 806R (GACTACHVGGGTATCTAATCC) targeting the V3-V4 region of the 16S gene (29, 30). Libraries were sequenced with an Illumina MiSeq system using 2 × 300-bp paired-end reads. One outlier sample (021715U_6) was discarded at this point because only 1,313 reads were obtained, while all other samples had >50,000 reads.

Determination of ASVs, taxonomy assignment, and phylogenetic placement. Primers were trimmed from reads using Cutadapt (v. 1.18) (125) with the following parameters: -g, CCTACGGGNGGCWGCAG; -a, GGATTAGATACCCBDGTAGTC; -G, GACTACHVGGGTATCTAATCC; -A, CTGCWGCCNCCCGTAGG; -minimum-length, 50, -n, 2. Exact amplicon sequence variants were determined using the DADA2 denoiser (32) plugin for QIIME 2 (version 2019.10) (31) with the following parameters: p-trunc-len-f = 270, p-trunc-len-r = 191, and p-max-ee = 2. ASVs are more specific than commonly used operational taxonomic units (OTUs), which consist of sequences clustered by similarity (typically 97%). DADA2 infers exact ASVs, merges paired-end reads, removes chimeric sequences, and generates a per-sample count table of ASV observations.

Taxonomy was assigned to ASVs using the QIIME 2 naive Bayes machine-learning feature classifier (126) trained on the Greengenes (127) and SILVA (release 132) (34) databases trimmed to match the V3-V4 region sequenced. Trees were constructed using SEPP (35, 36) to place ASV representative sequences into the SILVA reference phylogeny (release 128) (34). This approach was developed to overcome difficulties in *de novo* tree construction from short sequencing fragments and facilitates comparison of amplicon data from different variable regions (36).

Pairwise ASV comparisons. Correlation, or analogous metrics in the case of SPIEC-EASI and *propr*, were computed using SPARCC (40) implemented in FastSpar (41) with the parameters iterations = 50 and permutations = 1,000, the SPIEC-EASI R package (42) with the parameters method = 'glasso', lambda. min.ratio = 1e−3, nlambda = 30, and rep.num = 50, and the Propr R package (43) with the parameters metric = "rho." and P = 999. SPARCC P values were computed by permutation, and only correlations with an absolute value greater than 0.35 (the default cutoff) and P values of <0.05 were considered. In addition to the inherent error control in SPIEC-EASI, pseudo-P values were computed as 1-edge stability across the sparsity path. Only edges with a pseudo-P of ≤0.5 were used. Only *propr* rho values of >0.65 and <−0.5 were considered, in line with recommended false discovery rate (FDR) control.

Levenshtein edit distances for sequence similarity were computed for each ASV pair using the *string-distmatrix* function of the *stringdist* R package (128).

Chi-square tests were used to compare distributions for each ASV pair using the *chisq.test* function in the R stats package (129). P values were simulated with 100,000 permutations.

Contaminant identification. Contaminant ASVs were identified with the *prevalence* method of the *decontam* package (39) with a threshold of 0.33, corresponding to a probability threshold below which the null hypothesis (noncontaminant) is rejected in favor of the alternative hypothesis (contaminant).

Sequences with a BLAST similarity of >99% to unique lab strains were identified as contaminants. Correlation analyses were used to identify additional suspected lab contaminants.

Using pairwise ASV correlations calculated as described above, ASVs highly correlated with identified contaminants were also deemed to be contaminants. Two groups of contaminants were used as starting points: a cluster of highly intercorrelated ASVs present in negative controls (calculated independently with each correlation-like metric [sparcc # of pos - neg >0 and mean cor > 0.3; spiec easi # pos - neg > 0; propr # pos - neg > 0]) and lab contaminants identified by BLAST similarity as described above. For both contaminant starting groups, highly correlated ASVs were separately identified as contaminants with each of the correlation-like metrics (reagent: sparcc # pos > 10 & mean cor > 0.25; spiec-easi # pos - neg >5; propr # pos - neg > 0; lab: sparcc # pos > 1 & mean > 0.3; spiec-easi # pos - neg > 5, propr # pos > 0). For each metric, a secondary search for highly correlated ASVs was then conducted against all contaminants identified up to this point (sparcc mean > 0.3 & # pos > 5; spiec-easi # pos > 5; propr mean > 0.3 & # pos > 1). Cutoffs were chosen based on plotted distributions of net positive correlations versus mean correlation. ASVs found only in negative-control samples were also classified as contaminants.

To account for residual sequencing errors (baseline Illumina error rates, ~0.0042 error per base; DADA2 residual error rate, ~2.5×10^{-8}), ASVs only one nucleotide edit away (Levenshtein edit distance, calculated as described above) (32, 130) from identified contaminants were also classified as contaminants and assigned the same detection method as their match.

Statistical analysis. Within sample (alpha) diversity metrics Faith's PD and Shannon were calculated with the QIIME 2 Diversity Plugin (31). Changes in alpha diversity over time were evaluated with linear mixed-effect models [alpha.div ~ scaled_months + scaled_avg_temperature + ASR_status + (1 | ASR_status)] using the *lmer* function of the *lme4* R package (131) and *lmerTest* package to obtain P values (132).

The between-sample (beta) diversity metrics Bray-Curtis, Jaccard, UniFrac, weighted UniFrac (53), and generalized UniFrac (54) were also calculated with the QIIME 2 diversity plugin (31). Permutational multivariate analysis of variance was performed with the *adonis2* function of the *vegan* R package (50).

Principal-coordinate analysis was performed using the QIIME 2 diversity plugin (31).

Indicator species analysis was used to determine indicators of ASR status, implemented in the *multipatt* function of the *indicspecies* R package (version 1.7.8) (46) using the IndVal.g association statistic and 1,000 permutations. Taxonomic agglomeration at the genus level was performed using the *tax_glom* function of the *speedyseq* (133) R package. Only genera with more than 5 observations in more than 5 samples were considered.

Leveraging this study's longitudinal design, the *fitMultipleTimeSeries* function of the *MetagenomeSeq* R package (version 1.28.0) (134, 135) applied ssANOVA to identify ASVs with differential abundance that increased throughout the experiment (formula, abundance $\sim$ time $\times$ ASR status; permutations, 1,000). Only genera with more than 5 observations in more than 5 samples were considered.

The ASV observation table was converted to presence/absence, and logistic regression was performed with the *glm* function of the R *stats* package using family = "binomial" and the formula presence/absence $\sim$ ASR $\times$ scale (months). This model was applied to taxa summarized at all levels that were observed in at least 5 samples.

Following methods described in reference 51, general additive mixed models were applied with the *mgcv* R package (136). Abundances were Hellinger transformed prior to running the model *scaled.abundance* $\sim$ s(DoY, bs = "cc," k = 4) + s(months, bs = "cr," k = 1). The cyclic cubic regression spline for day of year detects seasonal changes in abundance, limited in complexity to 4 knots. In addition, time of year with peak abundance can be identified by inspection of this spline. The cubic regression spline limited to 1 knot is essentially a linear regression, which models changes occurring over the entire length of the series. This model was applied to taxa summarized at all levels that were observed in at least 5 samples.

To understand potential sources of bacteria in concrete samples, SourceTracker2 (47, 137) was used to calculate source mixing proportions for each concrete sample (alpha1 = 0.01, alpha2 = 1). Sources were pooled and subsampled to a depth of 8,000, while sink samples were subsampled to a depth of 635.

Comparison to Earth Microbiome Project. Comparison to the EMP is limited to broad comparisons (36) due to different primer biases of the EMP V4 region primers 515F and 806R and the V3V4 primers 357F and 806R used here (138, 139) and the short amplicon length (90 bp) from early EMP samples. The 90-bp release 1 of the Earth Microbiome Project was downloaded from ftp://ftp.microbio.me/emp/release1/otu_tables/deblur/emp_deblur_90bp.release1.biom. From this, an expanded version of the standard 5k sample EMP subset was produced, with additional inclusion of sterile water blanks and samples from highly alkaline environments. Representative sequences were extracted from the OTU table for import into QIIME 2. Additional samples from a study of tap water (48) prepared with the EMP protocol were included; representative sequences and a BIOM table generated from reads trimmed to 90 bp and processed with deblur (140) were downloaded from Qiita (141) study 10251.

Representative sequences for ASVs from the current study were trimmed to the same 90 bp as EMP and tap water samples using the GetV4Region.py script released by the EMP authors. After tables and representative sequences were merged, the QIIME 2 VSEARCH plugin (33) was used to cluster reads from this study, tap water samples, and EMP samples at 99% nucleotide similarity. The 99% similarity threshold was chosen to account for slight variability but was set higher than the standard 97% due to the short sequence length and because both EMP reads and those from this study were previously denoised.

Combined EMP, tap water, and concrete tables were produced for raw concrete samples, "decontaminated" concrete samples, and only contaminants of concrete samples, allowing evaluation of concrete decontamination. Sequences from this combined data set were inserted into release 128 of the SILVA reference phylogeny using SEPP (35). Taxonomy was assigned to ASVs using a naive Bayes classifier trained on the 515F-806R region of the SILVA 132 release.

QIIME 2 was used to calculate distance tables based on several diversity metrics (Bray-Curtis, Jaccard, UniFrac, weighted UniFrac, and generalized [0.5] UniFrac) and corresponding PCoA ordinations.

Indicator species analysis was conducted with the *multipatt* function of the *Indicspecies* R package (46). EMPO level 3 groups were used for groupings with 999 permutations and max.order = 3. This analysis was also conducted at the genus level with prior taxonomic agglomeration using the *tax_glom* function of the *speedyseq* (133) R package.

The relative occurrence frequency was calculated as described by Thompson et al. (52) using custom R scripts available at github.com/MarescaLab/concrete_series. Per-ASV Shannon entropies of the relative occurrence frequencies were calculated with the *diversity* function of the *vegan* R package (50).

Figure plotting. Figures 1, 4, and 7 and many of the supplemental figures were produced with the ggplot2 (142) R package. Heat trees (Fig. 2; Fig. S2 and S3) were produced using the Metacoder (143) R package.

Data availability. Data processing and analysis scripts are available at github.com/MarescaLab/concrete_series. Demultiplexed 16S amplicon data sets trimmed of Illumina adapters, but otherwise unmodified, were deposited at the NCBI Sequence Read Archive (SRA) under the BioProject number PRJNA629592. For tables S1 to S4 and S8 to S11 (tsv files), see doi.org/10.6084/m9.figshare.14211038.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.2 MB.

FIG S2, PDF file, 0.3 MB.

FIG S3, PDF file, 0.2 MB.
FIG S4, PDF file, 0.1 MB.
FIG S5, PDF file, 0.1 MB.
FIG S6, PDF file, 0.3 MB.
FIG S7, PDF file, 0.2 MB.
TABLE S1, PDF file, 0.1 MB.
TABLE S2, PDF file, 0.1 MB.
TABLE S3, PDF file, 0.1 MB.

REFERENCES

1. Damtoft JS, Lukasik J, Herfort D, Sorrentino D, Gartner EM. 2008. Sustainable development and climate change initiatives. Cem Concr Res 38:115–127. https://doi.org/10.1016/j.cemconres.2007.09.008.
2. Maresca JA, Moser P, Schumacher T. 2017. Analysis of bacterial communities in and on concrete. Mater Struct 50:25. https://doi.org/10.1617/s11527-016-0929-y.
3. Parker CD. 1947. Species of sulphur bacteria associated with the corrosion of concrete. Nature 159:439. https://doi.org/10.1038/159439b0.
4. Ling AL, Robertson CE, Harris JK, Frank DN, Kotter CV, Stevens MJ, Pace NR, Hernandez MT. 2015. High-resolution microbial community succession of microbially induced concrete corrosion in working sanitary manholes. PLoS One 10:e0116400. https://doi.org/10.1371/journal.pone.0116400.
5. Gomez-Alvarez V, Revetta RP, Santo Domingo JW. 2012. Metagenome analyses of corroded concrete wastewater pipe biofilms reveal a complex microbial system. BMC Microbiol 12:122. https://doi.org/10.1186/1471-2180-12-122.
6. Grengg C, Mittermayr F, Baldermann A, Böttcher ME, Leis A, Koraimann G, Grunert P, Dietzel M. 2015. Microbiologically induced concrete corrosion: a case study from a combined sewer network. Cem Concr Res 77:16–25. https://doi.org/10.1016/j.cemconres.2015.06.011.
7. Okabe S, Odagiri M, Ito T, Satoh H. 2007. Succession of sulfur-oxidizing bacteria in the microbial community on corroding concrete in sewer systems. Appl Environ Microbiol 73:971–980. https://doi.org/10.1128/AEM.02054-06.
8. Nica D, Davis JL, Kirby L, Zuo G, Roberts DJ. 2000. Isolation and characterization of microorganisms involved in the biodeterioration of concrete in sewers. Int Biodeterior Biodegradation 46:61–68. https://doi.org/10.1016/S0964-8305(00)00064-0.
9. Monteny J, Vincke E, Beeldens A, De Belie N, Taerwe L, Van Gemert D, Verstraete W. 2000. Chemical, microbiological, and in situ test methods for biogenic sulfuric acid corrosion of concrete. Cem Concr Res 30:623–634. https://doi.org/10.1016/S0008-8846(00)00219-2.
10. Kim HJ, Eom HJ, Park C, Jung J, Shin B, Kim W, Chung N, Choi I-G, Park W. 2016. Calcium carbonate precipitation by Bacillus and Sporosarcina strains isolated from concrete and analysis of the bacterial community of concrete. J Microbiol Biotechnol 26:540–548. https://doi.org/10.4014/jmb.1511.11008.
11. Alshalif AF, Irwan JM, Othman N, Al-Gheethi A. 2018. New medium for isolation of bacteria from cement kiln dust with a potential to apply in bio-concrete. IOP Conf Ser Earth Environ Sci 140:e012155. https://doi.org/10.1088/1755-1315/140/1/012155.
12. Grant WD, Jones BE. 2016. Bacteria, archaea and viruses of soda lakes, p 97–147. In Schagerl M (ed), Soda lakes of East Africa. Springer International Publishing, Cham, Switzerland.
13. Yates EL, Detweiler AM, Iraci LT, Bebout BM, McKay CP, Schiro K, Sheffner EJ, Kelley CA, Tadić JM, Loewenstein M. 2013. Assessing the role of alkaline soils on the carbon cycle at a playa site. Environ Earth Sci 70:1047–1056. https://doi.org/10.1007/s12665-012-2194-x.
14. Gomes HI, Mayes WM, Rogerson M, Stewart DI, Burke IT. 2016. Alkaline residues and the environment: a review of impacts, management practices and opportunities. J Clean Prod 112:3571–3582. https://doi.org/10.1016/j.jclepro.2015.09.111.
15. Horikoshi K. 1999. Alkaliphiles: some applications of their products for biotechnology. Microbiol Mol Biol Rev 63:735–750. https://doi.org/10.1128/MMBR.63.4.735-750.1999.
16. Taylor CB, Hutchinson GH. 1947. Corrosion of concrete caused by sulphur-oxidising bacteria. J Chem Technol Biotechnol 66:54–57. https://doi.org/10.1002/jctb.5000660205.
17. Gomez-Alvarez V. 2014. Biofilm-growing bacteria involved in the corrosion of concrete wastewater pipes: protocols for comparative metagenomic analyses. Methods Mol Biol 1147:323–340. https://doi.org/10.1007/978-1-4939-0467-9_23.
18. Yamanaka T, Aso I, Togashi S, Tanigawa M, Shoji K, Watanabe T, Watanabe N, Maki K, Suzuki H. 2002. Corrosion by bacteria of concrete in sewerage systems and inhibitory effects of formates on their growth. Water Res 36:2636–2642. https://doi.org/10.1016/s0043-1354(01)00473-0.
19. De Muynck W, De Belie N, Verstraete W. 2010. Microbial carbonate precipitation in construction materials: a review. Ecol Eng 36:118–136. https://doi.org/10.1016/j.ecoleng.2009.02.006.
20. Folliard KJ, Thomas MDA, Kurtis KE. 2003. Guidelines for the use of lithium to mitigate or prevent ASR. FHWA-RD-03-047. Federal Highway Administration, Washington, DC.

21. Detwiler RJ. 2002. Substitution of fly ash for cement or aggregate in concrete: strength development and suppression of ASR. R&D Bulletin RD127. Portland Cement Association, Skokie, IL.

22. American Coal Ash Association. 2003. Fly ash facts for highway engineers. FHWA-IF-03-019. Federal Highway Administration, Washington, DC.

23. Cabrera JG. 1986. The use of pulverized fuel ash to produce durable concrete, p 29–57. *In* Improvement of concrete durability. Thomas Telford Publishing, London, United Kingdom.

24. Farny JA, Kerkhoff B. 2007. Concrete technology: diagnosis and control of alkali-aggregate reactions in concrete. Portland Cement Association, Skokie, IL.

25. Saha AK. 2018. Effect of class F fly ash on the durability properties of concrete. Sustain Environ Res 28:25–31. https://doi.org/10.1016/j.serj.2017.09.001.

26. Lee YS, Park W. 2018. Current challenges and future directions for bacterial self-healing concrete. Appl Microbiol Biotechnol 102:3059–3070. https://doi.org/10.1007/s00253-018-8830-y.

27. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12:87. https://doi.org/10.1186/s12915-014-0087-z.

28. Thomas MDA, Fournier B, Folliard KJ, Resendez YA. 2011. Alkali-silica reactivity field identification handbook. FHWA-HIF-12-022. Federal Highway Administration, Washington, DC.

29. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1. https://doi.org/10.1093/nar/gks808.

30. Illumina I. 2013. 16S sample preparation guide. Illumina, Inc., San Diego, CA.

31. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9.

32. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.

33. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. https://doi.org/10.7717/peerj.2584.

34. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: D590–D596. https://doi.org/10.1093/nar/gks1219.

35. Mirarab S, Nguyen N, Warnow T. 2012. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput 2012:247–258. https://doi.org/10.1142/9789814366496_0024.

36. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, Winker K, Kado DM, Orwoll E, Manary M, Mirarab S, Knight R. 2018. Phylogenetic placement of exact amplicon sequences improves associations with clinical information. mSystems 3:e00021-18. https://doi.org/10.1128/mSystems.00021-18.

37. Weyrich LS, Farrer AG, Eisenhofer R, Arriola LA, Young J, Selway CA, Handsley-Davis M, Adler CJ, Breen J, Cooper A. 2019. Laboratory contamination over time during low-biomass sample analysis. Mol Ecol Resour 19:982–996. https://doi.org/10.1111/1755-0998.13011.

38. Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J, McWeeney S. 2019. Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. mSystems 4:e00290-19. https://doi.org/10.1128/mSystems.00290-19.

39. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6:226. https://doi.org/10.1186/s40168-018-0605-2.

40. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687. https://doi.org/10.1371/journal.pcbi.1002687.

41. Watts SC, Ritchie SC, Inouye M, Holt KE. 2019. FastSpar: rapid and scalable correlation estimation for compositional data. Bioinformatics 35:1064–1066. https://doi.org/10.1093/bioinformatics/bty734.

42. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol 11:e1004226. https://doi.org/10.1371/journal.pcbi.1004226.

43. Quinn TP, Richardson MF, Lovell D, Crowley TM. 2017. propr: an R-package for identifying proportionally abundant features using compositional data analysis. Sci Rep 7:16252. https://doi.org/10.1038/s41598-017-16520-0.

44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

45. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. 2019. Contamination in low microbial biomass microbiome studies: issues and recommendations. Trends Microbiol 27:105–117. https://doi.org/10.1016/j.tim.2018.11.003.

46. De Caceres M, Jansen F, Dell N. 2020. CRAN indicspecies.

47. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, Bushman FD, Knight R, Kelley ST. 2011. Bayesian community-wide culture-independent microbial source tracking. Nat Methods 8:761–763. https://doi.org/10.1038/nmeth.1650.

48. Ji P, Parks J, Edwards MA, Pruden A. 2015. Impact of water chemistry, pipe material and stagnation on the building plumbing microbiome. PLoS One 10:e0141087. https://doi.org/10.1371/journal.pone.0141087.

49. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol 26:32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x.

50. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2019. vegan: community ecology package.

51. Cram JA, Chow C-ET, Sachdeva R, Needham DM, Parada AE, Steele JA, Fuhrman JA. 2015. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. ISME J 9:563–580. https://doi.org/10.1038/ismej.2014.153.

52. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Xu ZZ, Jiang L, Haroon MF, Kanbar J, Zhu Q, Song SJ, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, The Earth Microbiome Project Consortium. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457–463. https://doi.org/10.1038/nature24621.

53. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. 2011. UniFrac: an effective distance metric for microbial community comparison. ISME J 5:169–172. https://doi.org/10.1038/ismej.2010.133.

54. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, Collman RG, Bushman FD, Li H. 2012. Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics 28:2106–2113. https://doi.org/10.1093/bioinformatics/bts342.

55. Earth Microbiome Project. 2017. EMP Ontology (EMPO).

56. Paulson JN, Pop M, Bravo HC. 2013. metagenomeSeq: statistical analysis for sparse high-throughput sequencing. Bioconductor package 1:63.

57. Dedysh SN, Kulichevskaya IS, Huber KJ, Overmann J. 2017. Defining the taxonomic status of described subdivision 3 Acidobacteria: proposal of Bryobacteraceae fam. nov. Int J Syst Evol Microbiol 67:498–501. https://doi.org/10.1099/ijsem.0.001687.

58. Kulichevskaya IS, Suzina NE, Liesack W, Dedysh SN. 2010. Bryobacter aggregatus gen. nov., sp. nov., a peat-inhabiting, aerobic chemo-organotroph from subdivision 3 of the Acidobacteria. Int J Syst Evol Microbiol 60:301–306. https://doi.org/10.1099/ijs.0.013250-0.

59. Kulichevskaya IS, Suzina NE, Rijpstra WIC, Damsté JSS, Dedysh SN. 2014. Paludibaculum fermentans gen. nov., sp. nov., a facultative anaerobe capable of dissimilatory iron reduction from subdivision 3 of the Acidobacteria. Int J Syst Evol Microbiol 64:2857–2864. https://doi.org/10.1099/ijs.0.066175-0.

60. Allievi MJ, Silveira DD, Cantão ME, Filho PB. 2018. Bacterial community diversity in a full scale biofilter treating wastewater odor. Water Sci Technol 77:2014–2022. https://doi.org/10.2166/wst.2018.114.

61. Xue L, Ren H, Li S, Leng X, Yao X. 2017. Soil bacterial community structure and co-occurrence pattern during vegetation restoration in karst rocky desertification area. Front Microbiol 8:2377. https://doi.org/10.3389/fmicb.2017.02377.

62. Miralles I, Lázaro R, Sánchez-Marañón M, Soriano M, Ortega R. 2020. Biocrust cover and successional stages influence soil bacterial composition and diversity in semiarid ecosystems. Sci Total Environ 709:134654. https://doi.org/10.1016/j.scitotenv.2019.134654.

63. Gorbushina AA. 2007. Life on the rocks. Environ Microbiol 9:1613–1631. https://doi.org/10.1111/j.1462-2920.2007.01301.x.

64. Normand P, Daffonchio D, Gtari M. 2014. The family Geodermatophilaceae, p 361–379. In Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (ed), The prokaryotes: Actinobacteria. Springer, Berlin, Germany.

65. Wang X, Xue Y, Yuan S, Zhou C, Ma Y. 2008. Salinicoccus halodurans sp. nov., a moderate halophile from saline soil in China. Int J Syst Evol Microbiol 58:1537–1541. https://doi.org/10.1099/ijs.0.65467-0.

66. Zhang Y-Q, Yu L-Y, Liu H-Y, Zhang Y-Q, Xu L-H, Li W-J. 2007. Salinicoccus luteus sp. nov., isolated from a desert soil. Int J Syst Evol Microbiol 57:1901–1905. https://doi.org/10.1099/ijs.0.64967-0.

67. Chen Y-G, Cui X-L, Li W-J, Xu L-H, Wen M-L, Peng Q, Jiang C-L. 2008. Salinicoccus salitudinis sp. nov., a new moderately halophilic bacterium isolated from a saline soil sample. Extremophiles 12:197–203. https://doi.org/10.1007/s00792-007-0116-8.

68. Qu Z, Li Z, Zhang X, Zhang X-H. 2012. Salinicoccus qingdaonensis sp. nov., isolated from coastal seawater during a bloom of green algae. Int J Syst Evol Microbiol 62:545–549. https://doi.org/10.1099/ijs.0.030551-0.

69. Fariq A, Yasmin A, Jamil M. 2019. Production, characterization and antimicrobial activities of bio-pigments by Aquisalibacillus elongatus MB592, Salinicoccus sesuvii MB597, and Halomonas aquamarina MB598 isolated from Khewra Salt Range, Pakistan. Extremophiles 23:435–449. https://doi.org/10.1007/s00792-019-01095-7.

70. Rivadeneyra MA, Delgado G, Soriano M, Ramos-Cormenzana A, Delgado R. 1999. Biomineralization of carbonates by Marinococcus albus and Marinococcus halophilus isolated from the Salar de Atacama (Chile). Curr Microbiol 39:53–57. https://doi.org/10.1007/pl00006827.

71. Srinivas A, DivyASRee B, Tushar L, Suresh G, Sasikala C, Ramana CV. 2016. Salinicoccus amylolyticus sp. nov., isolated from a saltern. Int J Syst Evol Microbiol 66:3814–3820. https://doi.org/10.1099/ijsem.0.001270.

72. Ramana CV, Srinivas A, Subhash Y, Tushar L, Mukherjee T, Kiran PU, Sasikala C. 2013. Salinicoccus halitifaciens sp. nov., a novel bacterium participating in halite formation. Antonie Van Leeuwenhoek 103:885–898. https://doi.org/10.1007/s10482-012-9870-4.

73. Gao M, Wang L, Chen S-F, Zhou Y-G, Liu H-C. 2010. Salinicoccus kekensis sp. nov., a novel alkaliphile and moderate halophile isolated from Keke Salt Lake in Qinghai, China. Antonie Van Leeuwenhoek 98:351–357. https://doi.org/10.1007/s10482-010-9449-x.

74. Zhang W, Xue Y, Ma Y, Zhou P, Ventosa A, Grant WD. 2002. Salinicoccus alkaliphilus sp. nov., a novel alkaliphile and moderate halophile from Baer Soda Lake in Inner Mongolia Autonomous Region, China. Int J Syst Evol Microbiol 52:789–793. https://doi.org/10.1099/00207713-52-3-789.

75. Wang Q, Sun Y-W, Liu J, Zhang D-C. 2018. Rheinheimera marina sp. nov., isolated from a deep-sea seamount. Int J Syst Evol Microbiol 68:266–270. https://doi.org/10.1099/ijsem.0.002496.

76. Yu Q, Qi Y, Zhang H, Pu J. 2020. Rheinheimera sediminis sp. nov., a marine bacterium isolated from coastal sediment. Int J Syst Evol Microbiol 70:1282–1287. https://doi.org/10.1099/ijsem.0.003917.

77. Park S, Park J-M, Won S-M, Jung Y-T, Yoon J-H. 2014. Rheinheimera arenilitoris sp. nov., isolated from seashore sand. Int J Syst Evol Microbiol 64:3749–3754. https://doi.org/10.1099/ijs.0.067504-0.

78. Romanenko LA, Tanaka N, Svetashev VI, Kalinovskaya NI, Mikhailov VV. 2015. Rheinheimera japonica sp. nov., a novel bacterium with antimicrobial activity from seashore sediments of the Sea of Japan. Arch Microbiol 197:613–620. https://doi.org/10.1007/s00203-015-1095-2.

79. Chen W-M, Chen W-T, Young C-C, Sheu S-Y. 2019. Rheinheimera riviphila sp. nov., isolated from a freshwater stream. Arch Microbiol 201:919–926. https://doi.org/10.1007/s00203-019-01657-5.

80. Merchant MM, Welsh AK, McLean RJC. 2007. Rheinheimera texasensis sp. nov., a halointolerant freshwater oligotroph. Int J Syst Evol Microbiol 57:2376–2380. https://doi.org/10.1099/ijs.0.65045-0.

81. Chen W-M, Yang S-H, Young C-C, Sheu S-Y. 2013. Rheinheimera tilapiae sp. nov., isolated from a freshwater culture pond. Int J Syst Evol Microbiol 63:1457–1463. https://doi.org/10.1099/ijs.0.043836-0.

82. Kumar R, Bajaj A, Mathan Kumar R, Kaur G, Kaur N, Kumar Singh N, Manickam N, Mayilraj S. 2015. Taxonomic description and genome sequence of Rheinheimera mesophila sp. nov., isolated from an industrial waste site. Int J Syst Evol Microbiol 65:3666–3673. https://doi.org/10.1099/ijsem.0.000471.

83. Schröder J, Braun B, Liere K, Szewzyk U. 2016. Draft genome sequence of Rheinheimera sp. strain SA_1 isolated from iron backwash sludge in Germany. Genome Announc 4:e00853-16. https://doi.org/10.1128/genomeA.00853-16.

84. Liu Y, Jiang J-T, Xu C-J, Liu Y-H, Song X-F, Li H, Liu Z-P. 2012. Rheinheimera longhuensis sp. nov., isolated from a slightly alkaline lake, and emended

description of genus Rheinheimera Brettar et al. 2002. Int J Syst Evol Microbiol 62:2927–2933. https://doi.org/10.1099/ijs.0.036020-0.

85. Panda AN, Ray L, Mishra SR, Raina V. 2020. Rheinheimera pleomorphica sp. nov., a novel alkali-tolerant bacteria isolated from Chilika Lake, India. Curr Microbiol 77:158–165. https://doi.org/10.1007/s00284-019-01802-9.

86. Hatayama K, Saito K. 2019. Calcite formation induced by Ensifer adhaerens, Microbacterium testaceum, Paeniglutamicibacter kerguelensis, Pseudomonas protegens and Rheinheimera texasensis. Antonie Van Leeuwenhoek 112:711–721. https://doi.org/10.1007/s10482-018-1204-8.

87. Garrison CE, Field EK. 2020. Introducing a "core steel microbiome" and community functional analysis associated with microbially influenced corrosion. FEMS Microbiol Ecol 97:fiaa237. https://doi.org/10.1093/femsec/fiaa237.

88. Bernardet J-F, Bowman JP. 2006. The genus Flavobacterium, p 481–531. In Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (ed), The prokaryotes, vol 7. Proteobacteria: delta, epsilon subclass. Springer, New York, NY.

89. Oren A. 2014. The family Rhodocyclaceae, p 975–998. In Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (ed), The prokaryotes: Alphaproteobacteria and Betaproteobacteria. Springer, Berlin, Germany.

90. Wang K, Zhang L, Yang Y, Pan Y, Meng L, Liu H, Hong S, Huang H, Jiang J. 2015. Halobacillus andaensis sp. nov., a moderately halophilic bacterium isolated from saline and alkaline soil. Int J Syst Evol Microbiol 65:1908–1914. https://doi.org/10.1099/ijs.0.000198.

91. Romano I, Lama L, Nicolaus B, Poli A, Gambacorta A, Giordano A. 2006. Halomonas alkaliphila sp. nov., a novel halotolerant alkaliphilic bacterium isolated from a salt pool in Campania (Italy). J Gen Appl Microbiol 52:339–348. https://doi.org/10.2323/jgam.52.339.

92. Mormile MR, Romine MF, Garcia MT, Ventosa A, Bailey TJ, Peyton BM. 1999. Halomonas campisalis sp. nov., a denitrifying, moderately haloalkaliphilic bacterium. Syst Appl Microbiol 22:551–558. https://doi.org/10.1016/S0723-2020(99)80008-3.

93. Xu L, Sun J-Q, Wang L-J, Liu X-Z, Ji Y-Y, Shao Z-Q, Wu X-L. 2017. Aliidiomarina soli sp. nov., isolated from saline-alkaline soil. Int J Syst Evol Microbiol 67:724–728. https://doi.org/10.1099/ijsem.0.001709.

94. Chiu H-H, Rogozin DY, Huang S-P, Degermendzhy AG, Shieh WY, Tang S-L. 2014. Aliidiomarina shirensis sp. nov., a halophilic bacterium isolated from Shira Lake in Khakasia, southern Siberia, and a proposal to transfer Idiomarina maris to the genus Aliidiomarina. Int J Syst Evol Microbiol 64:1334–1339. https://doi.org/10.1099/ijs.0.057851-0.

95. Srinivas TNR, Nupur A, Kumar P. 2012. Aliidiomarina haloalkalitolerans sp. nov., a marine bacterium isolated from coastal surface seawater. Antonie Van Leeuwenhoek 101:761–768. https://doi.org/10.1007/s10482-011-9690-y.

96. Cheng R, Wang X, Zhu H, Yan B, Shutes B, Xu Y, Fu B, Wen H. 2020. Isolation and characterization of a salt-tolerant denitrifying bacterium Alishewanella sp. F2 from seawall muddy water. Sci Rep 10:10002. https://doi.org/10.1038/s41598-020-66989-5.

97. Kamble K. 2015. Determination of potential of halophilic Bacillus and Alishewanella species for decolorization of acid blue dye. Int J Pur Appl Biosci 3:224–230.

98. Cumbers J, Rothschild LJ. 2014. Salt tolerance and polyphyly in the cyanobacterium Chroococcidiopsis (Pleurocapsales). J Phycol 50:472–482. https://doi.org/10.1111/jpy.12169.

99. Billi D, Friedmann EI, Hofer KG, Caiola MG, Ocampo-Friedmann R. 2000. Ionizing-radiation resistance in the desiccation-tolerant cyanobacterium Chroococcidiopsis. Appl Environ Microbiol 66:1489–1492. https://doi.org/10.1128/aem.66.4.1489-1492.2000.

100. Banerjee M, Debkumari S. 2005. Comparative studies on growth and phosphatase activity of endolithic cyanobacterial isolates of Chroococcidiopsis from hot and cold deserts. J Microbiol Biotechnol 15:125–130.

101. Shi W, Takano T, Liu S. 2012. Isolation and characterization of novel bacterial taxa from extreme alkali-saline soil. World J Microbiol Biotechnol 28:2147–2157. https://doi.org/10.1007/s11274-012-1020-7.

102. Connon SA, Lester ED, Shafaat HS, Obenhuber DC, Ponce A. 2007. Bacterial diversity in hyperarid Atacama Desert soils. J Geophys Res 112:311. https://doi.org/10.1029/2006JG000311.

103. Ferreira AC, Nobre MF, Moore E, Rainey FA, Battista JR, da Costa MS. 1999. Characterization and radiation resistance of new isolates of Rubrobacter radiotolerans and Rubrobacter xylanophilus. Extremophiles 3:235–238. https://doi.org/10.1007/s007920050121.

104. Kurth D, Belfiore C, Gorriti MF, Cortez N, Farias ME, Albarracín VH. 2015. Genomic and proteomic evidences unravel the UV-resistome of the polyextremophile Acinetobacter sp. Ver3. Front Microbiol 6:328. https://doi.org/10.3389/fmicb.2015.00328.

105. Diagne N, Djighaly PI, Ngom M, Pesce C, Champion A, Svistoonoff S, Hocher V, Tisa LS. 2020. Advances in Frankia genome studies and molecular aspects of tolerance to environmental stresses, p 381–389. In Sharma V, Salwan R, Al-Ani LKT (ed), Molecular aspects of plant beneficial microbes in agriculture. Academic Press, New York, NY.

106. Cheng B, Meng Y, Cui Y, Li C, Tao F, Yin H, Yang C, Xu P. 2016. Alkaline response of a halotolerant alkaliphilic halomonas strain and functional diversity of its Na+(K+)/H+ antiporters. J Biol Chem 291:26056–26065. https://doi.org/10.1074/jbc.M116.751016.

107. Shameer S. 2016. Haloalkaliphilic Bacillus species from solar salterns: an ideal prokaryote for bioprospecting studies. Ann Microbiol 66:1315–1327. https://doi.org/10.1007/s13213-016-1221-7.

108. Lingojwar D, Jadhav R, Gawai K. 2013. Alkaliphile-specific motif analysis of Stenotrophomonas species DL18 F_1F_0-ATP synthase c-subunit isolated from Indian alkaline Soda Lake, Lonar. Curr Sci 104:1216–1218.

109. Sajjad W, Qadir S, Ahmad M, Rafiq M, Hasan F, Tehan R, McPhail KL, Shah AA. 2018. Ectoine: a compatible solute in radio-halophilic Stenotrophomonas sp. WMA-LM19 strain to prevent ultraviolet-induced protein damage. J Appl Microbiol 125:457–467. https://doi.org/10.1111/jam.13903.

110. Warren-Rhodes KA, Lee KC, Archer SDJ, Cabrol N, Ng-Boyle L, Wettergreen D, Zacny K, Pointing SB, NASA Life in the Atacama Project Team. 2019. Subsurface microbial habitats in an extreme desert Mars-analog environment. Front Microbiol 10:69. https://doi.org/10.3389/fmicb.2019.00069.

111. Sand-Jensen K, Jespersen TS. 2012. Tolerance of the widespread cyanobacterium Nostoc commune to extreme temperature variations (−269 to 105°C), pH and salt stress. Oecologia 169:331–339. https://doi.org/10.1007/s00442-011-2200-0.

112. Pietrasiak N, Regus JU, Johansen JR, Lam D, Sachs JL, Santiago LS. 2013. Biological soil crust community types differ in key ecological functions. Soil Biol Biochem 65:168–171. https://doi.org/10.1016/j.soilbio.2013.05.011.

113. Brewer TE, Fierer N. 2018. Tales from the tomb: the microbial ecology of exposed rock surfaces. Environ Microbiol 20:958–970. https://doi.org/10.1111/1462-2920.14024.

114. Choe Y-H, Kim M, Woo J, Lee MJ, Lee JI, Lee EJ, Lee YK. 2018. Comparing rock-inhabiting microbial communities in different rock types from a high arctic polar desert. FEMS Microbiol Ecol 94:fiy070. https://doi.org/10.1093/femsec/fiy070.

115. Rosado T, Dias L, Lança M, Nogueira C, Santos R, Martins MR, Candeias A, Mirão J, Caldeira AT. 2020. Assessment of microbiota present on a Portuguese historical stone convent using high-throughput sequencing approaches. Microbiologyopen 9:1067–1084. https://doi.org/10.1002/mbo3.1030.

116. Vavourakis CD, Ghai R, Rodriguez-Valera F, Sorokin DY, Tringe SG, Hugenholtz P, Muyzer G. 2016. Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. Front Microbiol 7:211. https://doi.org/10.3389/fmicb.2016.00211.

117. Joshi AA, Kanekar PP, Kelkar AS, Shouche YS, Vani AA, Borgave SB, Sarnaik SS. 2008. Cultivable bacterial diversity of alkaline Lonar lake, India. Microb Ecol 55:163–172. https://doi.org/10.1007/s00248-007-9264-8.

118. Portland Cement Association. 2019. How cement is made. Portland Cement Association, Skokie, IL.

119. Marstorp H, Guan X, Gong P. 2000. Relationship between dsDNA, chloroform labile C and ergosterol in soils of different organic matter contents and pH. Soil Biol Biochem 32:879–882. https://doi.org/10.1016/S0038-0717(99)00210-2.

120. Shade A, Caporaso JG, Handelsman J, Knight R, Fierer N. 2013. A meta-analysis of changes in bacterial and archaeal communities with time. ISME J 7:1493–1506. https://doi.org/10.1038/ismej.2013.54.

121. Bowers RM, McCubbin IB, Hallar AG, Fierer N. 2012. Seasonal variability in airborne bacterial communities at a high-elevation site. Atmos Environ 50:41–49. https://doi.org/10.1016/j.atmosenv.2012.01.005.

122. Smit E, Leeflang P, Gommans S, van den Broek J, van Mil S, Wernars K. 2001. Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. Appl Environ Microbiol 67:2284–2291. https://doi.org/10.1128/AEM.67.5.2284-2291.2001.

123. Wardle DA. 1998. Controls of temporal variability of the soil microbial biomass: a global-scale synthesis. Soil Biol Biochem 30:1627–1637. https://doi.org/10.1016/S0038-0717(97)00201-0.

124. Illumina, Inc. 2013. 16S metagenomic sequencing library preparation. Illumina, Inc., San Diego, CA.

125. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.

126. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90. https://doi.org/10.1186/s40168-018-0470-z.

127. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. https://doi.org/10.1128/AEM.03006-05.

128. van der Loo MPJ. 2014. The stringdist package for approximate string matching. R J 6:111–122.

129. R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

130. Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C. 2016. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. BMC Bioinformatics 17:125. https://doi.org/10.1186/s12859-016-0976-y.

131. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Soft 67. https://doi.org/10.18637/jss.v067.i01.

132. Kuznetsova A, Brockhoff P, Christensen R. 2017. lmerTest Package: tests in linear mixed effects models. J Stat Softw 82:1–26. https://doi.org/10.18637/jss.v082.i13.

133. McLaren M. 2019. speedyseq: faster implementations of common phyloseq functions.

134. Paulson JN, Stine OC, Bravo HC, Pop M. 2013. Differential abundance analysis for microbial marker-gene surveys. Nat Methods 10:1200–1202. https://doi.org/10.1038/nmeth.2658.

135. Paulson JN, Talukder H, Bravo HC. 2017. Longitudinal differential abundance analysis of microbial marker-gene surveys using smoothing splines. bioRxiv.

136. Wood SN. 2017. Generalized additive models: an introduction with R, 2nd ed. CRC Press, Boca Raton, FL. https://doi.org/10.1201/9781315370279.

137. Biota Technology. 2019. SourceTracker2.

138. Parada AE, Needham DM, Fuhrman JA. 2016. 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. https://doi.org/10.1111/1462-2920.13023.

139. Pinto AJ, Raskin L. 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS One 7:e43093. https://doi.org/10.1371/journal.pone.0043093.

140. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2:e00191-16. https://doi.org/10.1128/mSystems.00191-16.

141. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. 2018. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods 15:796–798. https://doi.org/10.1038/s41592-018-0141-9.

142. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag. New York, NY.

143. Foster ZSL, Sharpton TJ, Grünwald NJ. 2017. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. PLoS Comput Biol 13:e1005404. https://doi.org/10.1371/journal.pcbi.1005404.