Relic DNA contributes minimally to estimates of microbial diversity

JT Lennon[1*], SA Placella[2], ME Muscarella[1,3]

[1]Indiana University, Department of Biology

[2]Michigan State University, W.K. Kellogg Biological Station

[3]University of Illinois, Department of Plant Biology

* corresponding author: lennonj@indiana.edu

Extracellular or "relic" DNA is one of the largest pools of nucleic acids in the biosphere [1,2]. Relic DNA can influence a number of important ecological and evolutionary processes, but it may also bias estimates of microbial abundance and diversity, which has implications for understanding environmental, engineered, and host-associated ecosystems. We developed theoretical predictions from a simulation model based on sampling a species abundance distribution to assess how and when relic DNA biases estimates of microbial diversity. We then tested our predictions by quantifying relic DNA and assessing its contribution to bacterial diversity using 16S rRNA gene sequences collected from soil, sediment, water, and mammalian gut environments. On average, relic DNA made up 33 % of the total bacterial DNA pool, but exceeded 80 % in some samples. Despite its abundance, relic DNA had a minimal effect on estimates of taxonomic and phylogenetic diversity, which is consistent with model simulations where the processes regulating the compositional turnover of relic and intact DNA are at steady state. When such conditions are met, relic DNA should not bias cultivation-independent estimates of microbial diversity.

When microorganisms die, their DNA leaks into the surrounding environment. The fate of this relic DNA has important implications for evolutionary and ecological processes. For example, relic DNA can be taken up and incorporated into the genomes of some microorganisms via transformation, thereby serving as a reservoir of genetic information that can confer new traits and fitness benefits to distantly related taxa [3]. In addition, relic DNA is a high-quality resource containing carbon, nitrogen, and phosphorus that is consumed by a diverse array of bacteria with consequences for microbial community structure and ecosystem processes [1,4].

34 Relic DNA also has the potential to bias cultivation-independent estimates of diversity, which are widely used for addressing questions concerning the assembly, biogeography, and
36 functioning of microbial communities. Microbial DNA extracted from environmental and host-associated samples is not solely derived from metabolically active organisms [5]. A large portion
38 of the individuals in a microbial community is dormant or dead [6,7]. Although nucleic acids can be temporarily retained in non-viable cells, DNA is ultimately released into the environment when
40 individuals die from autolysis, senescence, viral infection, or predation [8]. Together, these sources of mortality can create large pools of relic DNA [2,9]. For example, there is an estimated 0.45
42 petagrams of relic DNA in global ocean sediments, which is 70-fold greater than the amount of DNA contained in intact cells from the same environments [1]. If included in cultivation-
44 independent approaches, relic DNA could distort our understanding of the ecological and evolutionary processes that regulate the distribution, abundance, and function of microbial taxa.
46 The accumulation and turnover of relic DNA varies among ecosystems in ways that may bias estimates of microbial diversity. In some environments, DNA is rapidly broken down
48 outside the cell via hydrolysis, oxidation, and UV-mediated damage [2]. These degradation processes are reflected by the size distribution of relic DNA, which tends to be skewed towards
50 small fragments ranging between 100 and 500 base pairs [10,11]. In addition, the residence time of relic DNA is extremely short in some ecosystems. In freshwater and marine habitats,
52 extracellular DNA turns over in less than one day [12,13] while plasmid DNA begins to degrade in just minutes [14]. In other environments, the degradation of relic DNA can be much slower. For
54 example, DNA can bind with inorganic and organic substances in soils and sediments depending on cation availability and pH [9]. The resulting adsorption can increase the persistence of relic
56 DNA by protecting it from hydrolytic enzymes [15,16]. In addition, relic DNA is contained in

biofilms [8], and appears to be enriched in the gut mucosa of hosts with high-fat diets [17]. As a result, even in systems with high DNase activity, large pools of relic DNA can accumulate [18].

To date, the documented effects of relic DNA on estimates of diversity are idiosyncratic. Even in samples with relatively large relic DNA pools, bias can be non-existent or substantial, and lead to the overestimation or underestimation of diversity [19,20]. To delve further into this apparent unpredictability, we developed a theoretical framework to explore how relic DNA affects estimates of microbial diversity. While some variability in the effect of relic DNA may reflect taxonomic- or ecosystems-specific features, we hypothesized that the magnitude and direction of bias can be explained by the richness and evenness of sequences contained in the relic DNA pool. We used a simulation model to identify conditions where sampling effects lead to biased estimates of microbial diversity based on the size and species abundance distribution (SAD) of the relic DNA pool. We then tested our model predictions by quantifying the contribution of relic DNA on the abundance and diversity of bacterial communities in four distinct ecosystem types, including soils, sediments, surface waters, and mammalian guts.

RESULTS

We developed a model that represents the dynamics of relic DNA. The size of the relic DNA pool is affected by inputs associated with the death rates of microorganisms and by losses associated with the degradation of relic DNA. When death rates exceed degradation rates, the size of the relic DNA pool increases (Fig. 1a). However, a large relic DNA pool is not sufficient to create bias in diversity estimation. Rather, the direction and magnitude of bias reflects the simultaneous sampling of taxa from the species abundance distributions (SAD) associated with the relic and intact pools. Using sampling-based simulations (see Methods), we found that there

80 was no bias in estimating the number of microbial taxa (richness) when the SAD of the intact community was similar to the SAD of the relic pool (Fig. 1b-c). Such conditions are satisfied
82 when the death rates and degradation rates are proportional among taxa. Owing to sampling effects, our simulations reveal that bias can arise when the SADs of the relic and intact pools are
84 different. For example, our simulations predict that richness is overestimated when diversity (richness or evenness) is elevated in the relic pool, which reflects the increased probability of
86 sampling many unique taxa (Fig. 1b-c). Alternatively, richness can be underestimated when the diversity (richness or evenness) of the relic DNA pool is lower than the diversity of sequences in
88 the intact pool (Fig. 1b-c), which reflects an increased probability of sampling a few disproportionately abundant (i.e., dominant) taxa. Such conditions should arise when processes
90 (death and degradation) influencing the turnover of relic DNA are decoupled for different taxa in the intact and relic pools (Fig. 1a). The conditions leading to the over- or under-estimation of
92 richness also translate into biased measurements of beta-diversity (between-samples) due to compositional changes in the relative abundances of sequences (Fig. S1). Thus, we predict that
94 properties of the species abundance distribution – in particular, richness and evenness – determine whether or not relic DNA biases estimates of microbial diversity.
96

Variation in relic DNA pool size — We evaluated our model predictions by collecting samples
98 from different ecosystem types (soil, sediment, water, and mammalian guts) that are hypothesized to vary in relic DNA turnover rates [10,16]. Using quantitative PCR with 16S rRNA
100 primers, we quantified intact DNA after treating an aliquot of sample with DNase to remove relic DNA. In addition, we quantified total DNA (intact + relic) from a control aliquot of sample that
102 was not treated with DNase (see Methods and Fig. S2). Relic DNA accounted for a substantial

but variable portion of the total bacterial DNA. The proportion of relic DNA was normally distributed (mean ± sd = 0.33 ± 0.218) and ranged from 0 to 0.83 across 34 samples obtained from different ecosystem types (Fig. 2). These results are in line with the amount of relic DNA that has been reported in other studies [19,20,24]. Even though host and environmental features associated with these habitats are thought to influence relic DNA dynamics, there was only a marginally significant effect of ecosystem type on the proportion of relic DNA (one-way ANOVA, $F_{3, 30} = 2.43$, $P = 0.08$) [2,19]. Nevertheless, the large average pool size and range in relic DNA provided us with the opportunity to test our model predictions regarding the magnitude and direction of bias that relic DNA should have on estimates of microbial diversity.

Magnitude and direction of bias on microbial diversity — We sequenced 16S rRNA genes from the intact and total DNA pools to determine whether and to what degree relic DNA biased the taxonomic and phylogenetic alpha-diversity of bacterial communities. Despite accounting for a substantial portion of the total DNA (Fig. 2), relic DNA had no effect on estimates of richness, evenness, or phylogenetic diversity (PD) (Table S1). We expressed each of these alpha-diversity metrics as a ratio (total / intact), where values > 1 represent overestimation bias and values < 1 represent underestimation bias. Relic DNA had no effect on the diversity ratios based on the observation that the 95 % confidence intervals overlapped with 1.0 (Fig. 3). Furthermore, the 95 % confidence intervals of the diversity ratios overlapped across ecosystem types, indicating that the contribution of relic DNA to all measures of diversity was low irrespective of the microbial habitat sampled. Finally, using simple linear regression, we tested our model predictions that the magnitude of bias has the potential to increase with increasing relic DNA pool size (see Fig. 1b-c). For all diversity ratios (richness, evenness, and PD), the slopes were not different from zero

126 ($P > 0.65$) and the intercepts were not different from 1.0 ($P > 0.18$) (Fig. S3). These findings are consistent with simulations predicting that in order for bias to arise, there must be differences in
128 the diversity of the intact and relic DNA pools (Fig. 1b-c).

130 **Contribution of relic DNA to community composition** — We estimated the effects of relic DNA on the compositional diversity for the intact and total DNA pools. Principal Coordinates
132 Analysis (PCoA) revealed that bacterial composition was strongly affected by ecosystem type (Fig. S4), which was supported by permutational multivariate analysis of variance
134 (PERMANOVA) using both taxonomic ($R^2 = 0.35$, $F_{3,50} = 9.33$, $P = 0.001$) and phylogenetic ($R^2 = 0.65$, $F_{3,50} = 31.8$, $P = 0.001$) distance matrices. However, multiple lines of evidence suggest
136 that relic DNA had no appreciable effect on the observed bacterial composition. First, the intact and total DNA pools were significantly and highly correlated with one another when we
138 performed a Mantel test using Bray-Curtis distances, which reflect taxonomic dissimilarity ($P = 0.001$, $r = 0.959$), and UniFrac distances, which capture phylogenetic dissimilarity ($P = 0.001$, r
140 $= 0.996$). Second, we tested for the effect of relic DNA on beta-diversity by calculating centroid distance ratios (see Methods). If the centroid distance ratio was > 1, we concluded that relic
142 DNA inflated beta-diversity; if the distance ratio was < 1, we concluded that relic DNA homogenized beta-diversity. Relic DNA had no effect on the centroid distance ratios based on
144 the observation that the 95 % confidence intervals overlapped with 1.0 (Fig. 4). Furthermore, the 95 % confidence intervals for the centroid distance ratios overlapped across ecosystem types,
146 indicating that the contribution of relic DNA to beta-diversity was unaffected by different taxa or environmental conditions in the distinct microbial habitats. Last, we used simple linear
148 regression to test our model predictions that bias in estimates of community composition would

increase with increasing relic DNA pool size (see Fig. S1). The proportion of relic DNA in a
150 sample had no effect on the slopes for centroid distance ratios regardless of whether they were calculated using taxonomic ($P = 0.48$) or phylogenetic ($P = 0.75$) distance matrices. Similarly,
152 the intercepts for these regression relationships were not significantly different from 1 ($P > 0.3$) (Fig. S5). These empirical findings are consistent with simulations predicting that in order for
154 bias to arise, there must be differences in the diversity between the intact and relic DNA pools (Fig. S1).

156

DISCUSSION

158 Rarely, if ever, are biological communities completely censused. As a result, estimates of diversity are often based on incomplete sampling, which introduces uncertainty and potential
160 bias. For microbiologists using cultivation-independent approaches, diversity estimation is further complicated by the accumulation and persistence of relic DNA, which may have an
162 abundance distribution that is distinct from that of the sequences contained in viable cells. In such cases, our simulation models predict that microbial diversity can be overestimated or
164 underestimated depending on the size, richness, and evenness of the relic DNA pool (Fig. 1b-c). We tested our model predictions by analyzing the intact and total DNA in a range of ecosystems
166 with relic DNA pool sizes that could bias estimates of bacterial diversity. We found that relic DNA had minimal to no effect on commonly estimated measures of taxonomic and phylogenetic
168 diversity, suggesting that the compositional turnover of relic and intact DNA may approach steady state conditions in a range of ecosystems despite different environmental conditions and
170 divergent microbial assemblages (Fig. S4). In the following sections, we discuss scenarios where relic DNA should and should not lead to bias in the estimation of microbial diversity.

172 Relic DNA should have no effect on estimates of microbial diversity when the composition of sequences in the intact DNA pool and the relic DNA pool are at steady state. 174 Such conditions will be met when the taxon-specific production of relic DNA (via death) equals the taxon-specific loss of relic DNA (via degradation), regardless of whether the absolute rates of 176 these processes are fast or slow (Fig. 1). When these assumptions are met, the diversity of the relic and intact DNA pools should be identical. Therefore, sampling from the relic DNA pool 178 should not bias estimates of diversity in the intact DNA pool, even if relic DNA makes up a large portion of the total DNA pool. Despite sampling a range of habitats with attributes that 180 reportedly affect the turnover of relic DNA, our data suggest that measures of taxonomic and phylogenetic alpha- and beta-diversity were unbiased by relic DNA even when it accounted for > 182 80 % of the total DNA (Figs. S3 and S5). Similar findings are reported elsewhere. For example, there was no bias in the estimate of bacterial richness for a soil sample that contained nearly 80 184 % relic DNA [20]. Similarly, more than 90 % of the DNA recovered in lung tissues was found to be sensitive to DNase, yet the diversity of sequenced bacteria in the enzymatically treated and 186 untreated samples was the same [19].

Nevertheless, our model predicts that bias can arise when there are different species 188 abundance distributions (SADs) for intact and relic DNA pools (Fig. 1b-c, Fig. 2S). These predictions are consistent with some reports demonstrating that relic DNA can inflate estimates 190 of bacterial and fungal diversity [20,31]. While most studies have emphasized the potential to overestimate diversity, our model predicts that relic DNA pools that are low in richness or 192 dominated by only a few taxa can also lead to the underestimation of richness, which has been observed in some samples [20]. Such findings suggest that the compositional turnover of intact and 194 relic DNA pools can become decoupled. There are a number of scenarios where deviations from

steady-state conditions may arise. First, the DNA associated with certain taxa may be
196 preferentially preserved if it is contained in soil aggregates or biofilms that physically protect relic DNA from enzymes or other conditions that would otherwise lead to more rapid
198 degradation. Second, the immigration of taxa that are poorly adapted to local conditions may die and enrich the relic DNA pool with sequences that are dissimilar to sequences found in the local
200 community. A similar effect could arise when dead bacteria are transported across ecosystem boundaries, a phenomenon that occurs, for example, when marine snow is exported from surface
202 waters to marine sediments [32]. Furthermore, any abiotic or biotic perturbation that removes a substantial amount of living biomass could lead to divergence in the composition of sequences in
204 the intact and relic DNA pool. For example, virulent phage can dramatically reduce the abundance of bacterial prey and in the process release large quantities of DNA into the
206 environment, which could bias estimates of microbial diversity [31]. Last, although not fully explored here, shifts in rank abundance of taxa that are independent of richness or evenness
208 could results in biased estimates of microbial diversity.

Our understanding of the microbial biosphere has been transformed by the development
210 and application of molecular-based cultivation-independent techniques. The ability to rapidly obtain millions of gene sequences from a range of environments has yielded valuable insight into
212 the processes that regulate community assembly and function [25,26], and has also paved the way for the discovery of new metabolisms [27], tests for unifying patterns of biodiversity [28], and an
214 updated tree of life [29]. There are limitations, however, associated with culture-independent techniques, which include inefficient nucleic-acid extraction methods [29] and "universal" primers
216 that over-represent some taxonomic groups while overlooking others [30]. Sequencing of relic DNA is another important concern, which can potentially lead to the overestimation or

218 underestimation of microbial diversity. However, this bias requires the decoupling of processes that regulate the compositional turnover of the relic and intact DNA pools. While some recent
220 evidence suggests this can arise [20], our results from diverse ecosystems suggest that relic DNA contributes minimally to the characterization of microbial community structure.

222

METHODS

224

Relic DNA Model — We used sampling-based simulations to model the mixing of intact and
226 relic DNA pools. For each simulation, we sampled a regional species pool consisting of 10,000 taxa with a lognormal abundance distribution [21]. The intact community consisted of 1,000,000
228 individuals sampled from this regional pool. We then combined this intact community with a relic community at proportions from 0.01 to 0.96. In the first set of simulations, we altered the
230 evenness of the regional pool from which the relic community was sampled. We altered the evenness by changing the scale parameter of the lognormal distribution. To decrease the
232 evenness of the relic DNA pool, we increased the scale parameter from 0.98 to 2. To increase the evenness of the relic DNA pool, we decreased the parameter from 0.98 to 0.1. In the second set
234 of simulations, we altered the richness of the regional pool from which the relic community was sampled. We altered richness by increasing the magnitude of the regional pool while maintaining
236 the order of species and the scale parameter of the lognormal distribution. To increase the richness of the relic DNA pool, we increased the size of the regional pool from 10,000 to 50,000
238 taxa. To decrease the richness of the relic DNA pool, we decreased the size of the regional pool from 10,000 to 5,000 taxa. After mixing the intact and relic communities for both simulations,
240 we rarified the total community to 10,000 observations and calculated richness and Bray-Curtis distances to estimate compositional differences between the intact and total DNA pools. To

242 estimate contribution of relic DNA to diversity, we calculate ratios (total DNA / intact DNA) for richness and Bray-Curtis distances. All simulations and estimations were performed in the R
244 statistic environment (v 3.3.2)[46] using the vegan package as well as custom functions.

246 **Sample collection and DNA pools—** We collected samples from a range of environmental and host-associated ecosystems. First, we sampled sediments and surface water from lakes near the
248 Michigan State University, W.K. Kellogg Biological Station (KBS) in Hickory Corners, Michigan, USA. Soils were sampled from the main sites and surrounding areas at the KBS Long-
250 Term Ecological Research site [33]. We also collected fresh feces as representative gut microbiome samples from cows, dogs, horses, rabbits, and humans. In each of these ecosystem types, we
252 obtained samples from 6 - 8 independent sites. After transporting samples to the laboratory, we added DNase to an aliquot of a sample to remove relic DNA. Our method was based on
254 procedures that have previously been used to quantify relic DNA in marine sediments [24], host tissue [19], and drinking-water biofilms [23]. The DNase-based method has proven to be effective not
256 only at removing extracellular DNA, but also DNA that is contained inside of dead cells [23]. Furthermore, it has been shown that DNase does not affect the integrity of living cells [22,23]. Thus,
258 we assumed that the remaining "intact DNA" from the enzymatically treated aliquot was contained in potentially viable cells. We used another aliquot without DNase from the same
260 sample as a control, which we refer to as "total DNA". For aqueous samples, we filtered 250 mL samples onto 47 mm 0.2 μm Supor filter (Pall Corporation) at 10 mm Hg of vacuum pressure,
262 cut the filter in half and randomly assigned one half to a DNase treatment and used the other half as the control. For non-aquatic samples (gut, sediment, and soils), 0.25 g aliquots were
264 transferred into two 2 mL microcentrifuge tubes. We then added 440 mL of a DNase buffer

(382.5 μL molecular-grade H_2O, 5 μL of 1 M $MgCl_2$, 2.5 μL of bovine serum albumin BSA [10
266 mg/mL], and 50 μL of 0.5 M Tris HCl [pH = 7.5]) to the microcentrifuge tubes and randomly
assigned one to a DNase treatment and the other to a DNase control. Prior to analyzing samples,
268 we conducted an experiment to determine the optimal DNase concentration for effective
hydrolysis (see Fig. S6). For samples receiving the DNase treatment, we added 40 μL of 10
270 U/μL DNase (Roche, Indianapolis, IN) and 20 μL of ultrapure H_2O to the 2 mL microcentrifuge
containing buffer while the control received 60 μL of ultrapure H_2O. For each sample, we
272 measured pH using a micro pH probe (Orion 9110DJWP, Thermo Scientific) and adjusted to
7.3 - 7.7, which is in the optimum range for DNase. We then incubated the microcentrifuge tubes
274 horizontally at 37 ºC for 60 min to ensure proper mixing and enzymatic activity. To stop the
DNase reaction, we added 25 μL of 0.5 M EDTA to each tube and incubated at 75 ºC for 10 min.
276 The contents of the microcentrifuge tubes were then transferred to 15 mL Falcon tubes
containing 0.1 mm glass beads. We added 1 mL cetyltrimethylammonium bromide (1X, CTAB)
278 buffer to each Falcon tube to precipitate proteins, lipids, and polysaccharides. We then initiated
DNA extraction by adding 1 mL of phenol:chloroform:isoamylalcohol (25:24:1) to the 15 mL
280 Falcon tube. After vortexing for 10 min, we centrifuged the samples at 7,000 x *g* for 10 min. We
transferred the top aqueous layer to a new tube, added an equal volume of
282 chloroform:isoamylalcohol (24:1), and centrifuged at 7,000 x *g* for 5 min. We transferred 400 μL
of the top aqueous layer to a new sterile 1.5 mL tube and isolated DNA using the MO BIO
284 PowerLyzer PowerSoil DNA Isolation Kit. The resulting purified DNA was stored at -80 °C.

286 **Contribution of relic DNA to bacterial abundance** — We used 16S rRNA gene copy numbers
generated from quantitative PCR (qPCR) assays [34] to estimate the proportion of relic DNA in a
288 sample as 1 - (intact DNA / total DNA). The 30 μL qPCR reactions contained 1 μL of DNA

template, 0.5 μL of each primer (10 μmol/L), 14.5 μL of nuclease-free H_2O, and 13.5 μL of 5
290 Prime 2.5x Real MasterMix SYBR ROX (5 Prime, Inc. Gaithersburg, MD, USA). We amplified
a 200 base-pair fragment of the 16S rRNA gene with Eub 338 (forward) and Eub518 (reverse)
292 primers [35]. PCR assays were performed with an Eppendorf Mastercycler Realplex^2 system using
previously reported thermal cycle conditions [35]. We generated qPCR standards from bacterial
294 genomic DNA (*Micrococcus* sp.) using the TOPO TA Cloning Kit (Invitrogen). We extracted
plasmids from transformed cells [36], and used the M13 forward and reverse primers to generate
296 PCR products. The PCR products were quantified and used to generate a standard curve
capturing a range of $10^2 – 10^7$ gene copies per μL. The coefficients of determination (r^2) for our
298 assays ranged from 0.96 and 0.99, while amplification efficiencies fell between 93 and 99 %.
Based on melting curve analyses, we found no evidence for primer dimers or non-specific
300 amplification.

302 **Contribution of relic DNA to bacterial diversity**

Community sequencing — We estimated the contribution of relic DNA to bacterial diversity
304 using high-throughput sequencing of the 16S rRNA gene. Specifically, we amplified the V4
region of the 16S rRNA gene from the intact and total DNA pools of each sample using
306 barcoded primers (515F and 806R) designed to work with the Illumina MiSeq platform [37]. We
cleaned the sequence libraries using the AMPure XP purification kit, quantified the resulting
308 products using the QuantIt PicoGreen kit (Invitrogen), and pooled libraries at equal molar ratios
(final concentration: 20 ng per library). We then sequenced the pooled libraries with the Illumina
310 MiSeq platform using paired end reads (Illumina Reagent Kit v2, 500 reaction kit) at the Indiana
University Center for Genomics and Bioinformatics Sequencing Facility. Paired-end raw 16S
312 rRNA sequence reads were assembled into contigs using the Needleman algorithm [38]. We

obtained a total of 12,916,632 16S rRNA sequences from 42 samples representing 21 sites. After
314 quality trimming with a moving average quality score (window 50 bp, minimum quality score
35), we aligned the sequences to the Silva Database (version 123) using the Needleman
316 algorithm. Chimeric sequences were removed using the UCHIME algorithm [39]. After this
filtering, there was an average (± SEM) of 222,701 ± 9,560 sequences per site. We created
318 operational taxonomic units (OTUs) by first splitting the sequences based on taxonomic class
(using the RDP taxonomy) and then binning sequences into OTUs based on 97% sequence
320 similarity. Our depth of sequencing led to a high degree of coverage across samples (minimum
Good's Coverage = 0.98). For phylogenetic analysis, we picked representative sequences for
322 each OTU by using the most abundant unique sequence. We used FastTree [40] to generate a
phylogenetic tree from the representative sequences using the generalized time-reversible model
324 of nucleotide evolution. We calculated phylogenetic distances using weighted UniFrac distances
[41]. All initial sequence processing was completed using the software package mothur (version
326 1.38.1) [42].

328 *Alpha Diversity* — We estimated the effects of relic DNA on richness, evenness, and
phylogenetic diversity for the intact and total DNA pools with a sample. To estimate the number
330 of OTUs (i.e., richness), we used a resampling approach that subsampled each sample to an
equal number of sequences per sample and summed the number of OTUs that were represented
332 [43]. Briefly, we subsampled to 30,000 observations, resampled 999 additional times, and then
calculated the average richness estimates (± SEM) for each sample. To estimate the equitability
334 in abundance among taxa in a sample (evenness), we used the same resampling approach and
calculated average evenness estimates (± SEM) using Simpson's Evenness index [44]. To test
336 whether relic DNA affected the phylogenetic diversity within a sample, we subsampled

communities to 30,000 observations and then calculated Faith's *D* statistic, which sums the
338 branch lengths for each species found in a sample from the root to the tip of the phylogenetic tree [45]. All estimations were performed in the R statistic environment (v 3.3.2) [46] using the vegan,
340 ape, ade4, picante, and plyr packages, along with custom functions.

342 *Beta Diversity* — We estimated the effects of relic DNA on between-sample (i.e., beta) diversity by comparing the taxonomic and phylogenetic diversity of bacterial communities in the intact
344 and total DNA pools. First, we conducted a Principal Coordinates Analysis (PCoA) on $\log_{10}$-transformed relative abundance data to visualize the effects of relic DNA removal (via DNase
346 treatment) on bacterial community composition within and among ecosystem types. The PCoA was performed with Bray-Curtis and UniFrac distances to assess taxonomic and phylogenetic
348 effects, respectively. In addition, we used PERMANOVA to test for differences in taxonomic and phylogenetic composition based on ecosystem type for the total DNA pool. Second, we
350 conducted a Mantel test to assess the correlation between the community resemblance matrices (either Bray-Curtis or UniFrac) represented by the intact and total DNA pools. Last, we tested
352 whether relic DNA altered beta-diversity within an ecosystem type by comparing centroid distances. To calculate this metric of sample dispersion, we determined the centroid from a
354 PCoA with either Bray-Curtis or UniFrac distances for the total DNA pool for all sites within a given ecosystem type. We then measured the Euclidean distances between the centroid and all
356 samples (total and intact) to determine the centroid distances.

358 **ACKNOWLEDGMENTS**

We acknowledge constructive feedback from KD Webster, RZ Moger-Reischer, EK Hall, NI
360 Wisnoski, and WR Shoemaker. We thank S. Kuenzel and B.K. Lehmkuhl for technical support.

This work was supported by the USDA National Institute of Food and Agriculture Grant 2011-67019-3022 (JTL), the National Science Foundation Dimensions of Biodiversity Grant 1442246 (JTL), and US Army Research Office Grant W911NF-14-1-0411 (JTL). All code and data used in this study can be found in a public GitHub repository (https://www.github.com/LennonLab/relicDNA) and the NCBI SRA.

368 **REFERENCES**

1 Dell'Anno, A. & Danovaro, R. Extracellular DNA plays a key role in deep-sea ecosystem
370 functioning. *Science* **309**, 2179-2179 (2005).

2 Torti, A., Lever, M. A. & Jorgensen, B. B. Origin, dynamics, and implications of
372 extracellular DNA pools in marine sediments. *Mar. Genomics* **24**, 185-196, doi:10.1016/j.margen.2015.08.007 (2015).

374 3 Redfield, R. J. Genes for breakfast - the have-your-cake-and-eat-it-too of bacterial transformation. *J. Hered.* **84**, 400-404 (1993).

376 4 Lennon, J. T. Diversity and metabolism of marine bacteria cultivated on dissolved DNA. *Appl. Environ. Microbiol.* **73**, 2799-2805 (2007).

378 5 Klein, D. A. Microbial communities in nature: a postgenomic perspective. *Microbe* **2**, 591-595 (2007).

380 6 Luna, G. M., Manini, E. & Danovaro, R. Large fraction of dead and inactive bacteria in coastal marine sediments: Comparison of protocols for determination and ecological
382 significance. *Appl. Environ. Microbiol.* **68**, 3509-3513 (2002).

7 Lennon, J. T. & Jones, S. E. Microbial seed banks: the ecological and evolutionary
384 implications of dormancy. *Nat. Rev. Microbiol.* **9**, 119-130 (2011).

8 Vorkapic, D., Pressler, K. & Schild, S. Multifaceted roles of extracellular DNA in
386 bacterial physiology. *Curr. Genet.* **62**, 71-79 (2016).

9 Levy-Booth, D. J. *et al.* Cycling of extracellular DNA in the soil environment. *Soil Biol.*
388 *Biochem.* **39**, 2977-2991 (2007).

10 Deflaun, M. F., Paul, J. H. & Jeffrey, W. H. Distribution and molecular-weight of dissolved DNA in subtropical estuarine and oceanic environments. *Mar. Ecol. Prog. Ser.* **38**, 65-73 (1987).

11 Beebee, T. J. C. Identification and analysis of nucleci acids in natural freshwaters. *Sci. Total. Env.* **135**, 123-129 (1993).

12 Paul, J. H., Jeffrey, W. H., David, A. W., Deflaun, M. F. & Cazares, L. H. Turnover of extracellular DNA in eutrophic and oligotrophic freshwater environments of southwest Florida. *Appl. Environ. Microbiol.* **55**, 1823-1828 (1989).

13 Dell'Anno, A. & Corinaldesi, C. Degradation and turnover of extracellular DNA in marine sediments: Ecological and methodological considerations. *Appl. Environ. Microbiol.* **70**, 4384-4386 (2004).

14 Deflaun, M. F. & Paul, J. H. Detection of exogenous gene-sequences in dissolved DNA from aquatic environments. *Microb. Ecol.* **18**, 21-28 (1989).

15 Romanowski, G., Lorenz, M. G. & Wackernagel, W. Adsorption of plasmid DNA to mineral surfaces and protection against DNase-1. *Appl. Environ. Microbiol.* **57**, 1057-1061 (1991).

16 Morrissey, E. M. *et al.* Dynamics of extracellular DNA decomposition and bacterial community composition in soil. *Soil Biol. Biochem.* **86**, 42-49 (2015).

17 Sun J, Wang P. & Shi, Y. High fat diet decrease diversity of extracellular DNA of mice gut microbiota. *FASEB J.* **30**, ib204 (2016).

18 Corinaldesi, C., Barucca, M., Luna, G. M. & Dell'Anno, A. Preservation, origin and genetic imprint of extracellular DNA in permanently anoxic deep-sea sediments. *Mol. Ecol.* **20**, 642-654 (2011).

412 19 Pezzulo, A. A. *et al.* Abundant DNase I-sensitive bacterial DNA in healthy porcine lungs and its implications for the lung microbiome. *Appl. Environ. Microbiol.* **79**, 5936-5941
414 (2013).

20 Carini, P. *et al.* Relic DNA is abundant in soil and obscures estimates of soil microbial
416 diversity. *Nat. Microbiol.* **2**, 16242 (2017).

21 Shoemaker, W. R., Locey, K. J. & Lennon, J. T. A macroecological theory of microbial
418 biodiversity. *Nat. Ecol. Evol.* **1** (2017).

22 Dell'Anno, A., Bompadre, S. & Danovaro, R. Quantification, base composition, and fate
420 of extracellular DNA in marine sediments. *Limnol. Oceanogr.* **47**, 899-905 (2002).

23 Villarreal, J. V., Jungfer, C., Obst, U. & Schwartz, T. DNase I and proteinase K eliminate
422 DNA from injured or dead bacteria but not from living bacteria in microbial reference systems and natural drinking water biofilms for subsequent molecular biology analyses.
424 *J. Microbiol. Methods* **94**, 161-169 (2013).

24 Corinaldesi, C., Danovaro, R. & Dell'Anno, A. Simultaneous recovery of extracellular
426 and intracellular DNA suitable for molecular studies from marine sediments. *Appl. Environ. Microbiol.* **71**, 46-50 (2005).

428 25 Lindstrom, E. S. & Langenheder, S. Local and regional factors influencing bacterial community assembly. *Environ. Microbiol. Rep.* **4**, 1-9 (2012).

430 26 Aanderud, Z. T., Jones, S. E., Fierer, N. & Lennon, J. T. Resuscitation of the rare boisphere contributes to pulses of ecosystem activity. *Front. Microbiol.* **6**, (2015).

432 27 Beja, O. *et al.* Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* **289**, 1902-1906 (2000).

434 28 Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5970-5975 (2016).

436 29 Feinstein, L. M., Sul, W. J. & Blackwood, C. B. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl. Environ. Microbiol.* **75**, 5428-
438 5433 (2009).

30 Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for
440 classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).

442 31 Corinaldesi, C., Tangherlini, M., Luna, G. M. & Dell'Anno, A. Extracellular DNA can preserve the genetic signatures of present and past viral infection events in deep
444 hypersaline anoxic basins. *Proc. R. Soc. Lond. Biol.* **281,** 20133299 (2014).

32 Azam, F. & Long, R. A. Oceanography - Sea snow microcosms. *Nature* **414**, 497-498,
446 (2001).

33 Hamilton, S. K., Doll, J. E. & Robertson, G. P. *The ecology of agricultural landscapes:*
448 *Long-term research on the path to sustainability*. (Oxford University Press, 2015).

34 Lau, J. A. & Lennon, J. T. Evolutionary ecology of plant-microbe interactions: soil
450 microbial structure alters selection on plant traits. *New Phytol.* **192**, 215-224 (2011).

35 Fierer, N., Jackson, J. A., Vilgalys, R. & Jackson, R. B. Assessment of soil microbial
452 community structure by use of taxon-specific quantitative PCR assays. *Appl. Environ. Microbiol.* **71**, 4117-4120 (2005).

454 36 Sambrook, J. & Russell, D. W. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory Press, 2001).

456 37 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621-1624 (2012).

458 38 Needleman S.B. & C.D., W. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443-453 (1970).

460 39 Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-2200 (2011).

462 40 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2-approximately maximum-ikelihood trees for large alignments. *PLOS ONE* **5**, (2010).

464 41 Lozupone, C. A. & Knight, R. Species divergence and the measurement of microbial diversity. *FEMS Microbiol. Rev.* **32**, 557-578 (2008).

466 42 Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ.*
468 *Microbiol.* **75**, 7537-7541 (2009).

43 Muscarella, M. E., Jones, S. E. & Lennon, J. T. Species sorting along a subsidy gradient
470 alters community stability. *Ecology* (2016).

44 Smith, B. & Wilson, J. B. A consumer's guide to evenness indices. *Oikos* **76**, 70-82
472 (1996).

45 Faith, D. P. Conservation evaluation and phylogenetic diersity. *Biol. Conserv.* **61**, 1-10
474 (1992).

46 R Core Development Team R: A language and environment for statistical computing,
476 reference index version 2.8.1. R Foundationfor Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://R-project.org. (2009).

478

FIGURE CAPTIONS

480

Figure 1 | A model of relic DNA dynamics. **a** The amount of relic DNA in a microbial
482 environment is determined by inputs associated with the mortality of viable individuals with intact DNA and by losses associated with the degradation of relic DNA. If the diversity of
484 sequences contained in the relic DNA pool is sufficiently different from that in the intact DNA pool, then relic DNA may bias estimates of microbial biodiversity (as indicated by different
486 colored boxes) when sampling from the total (intact + relic) DNA pool. We developed a sampling-based simulation model that created communities populated with individuals with
488 intact DNA sampled from a lognormal species abundance distribution. We then mixed the intact community with relic individuals so that the relic contribution to total community ranged from
490 0.01 to 0.96. The relic community was also populated with individuals sampled from a lognormal species distribution, but which varied in (**b**) evenness (*E*) or (**c**) richness (*S*). To
492 quantify bias in diversity, we calculated "richness ratios" which reflect the number of species in the total DNA pool (intact + relic) divided by the number of species in the intact DNA pool.
494 When values richness ratios = 1, relic DNA has no effect on diversity; when richness ratios > 1, relic DNA leads to overestimates of diversity; richness ratios <1, relic DNA lead to
496 underestimation of diversity.

498 **Figure 2 | Proportion of bacterial relic DNA in different ecosystem types.** We quantified the amount of intact DNA in a sample after removing relic DNA with a DNase treatment. We then
500 estimated the proportion of relic DNA as 1 - (intact DNA / total DNA), where the total DNA concentration was quantified without DNase treatment. Relic DNA constituted an appreciable

502 portion of the total DNA pool, but was not affected by the ecosystem type from which the sample was collected (gut, soil, sediment, and water). Data are represented as means ± 95%
504 confidence intervals.

506 **Figure 3 | Bias of relic DNA on within-sample bacterial diversity in different ecosystem types. a-c** We tested for the effects of bias caused by relic DNA by calculating diversity ratios
508 for (**a**) richness, (**b**) evenness, and (**c**) phylogenetic diversity. The ratios reflect the diversity of the total DNA pool (intact + relic) divided by the diversity of the intact DNA pool. Relic DNA
510 did not bias any measures of diversity in any of the ecosystem types. Data are represented as means ± 95% confidence intervals. Richness was calculated as the number of operational
512 taxonomic units (97% sequence similarity of the 16S rRNA gene), evenness was calculated using Simpson's evenness index, and phylogenetic diversity was calculated using Faith's *D* index.
514

Figure 4 | Bias of relic DNA on the among-sample bacterial diversity in different ecosystem types. a-b We tested for the effects of bias caused by relic DNA by calculating a beta-diversity ratio based on centroid distances. Centroid distances were estimated after performing Principle Coordinates Analyses (PCoA) using taxonomic (**a**) and phylogenetic (**b**) distance metrics (Bray-Curtis and UniFrac, respectively). The centroid distance ratio was calculated on each sample within an ecosystem type and reflects the composition of the total DNA pool (intact + relic) relative to the intact DNA pool. Relic DNA had no effect on beta-diversity for any of the ecosystem types sampled. Data are represented as means ± 95% confidence intervals.

516 **Fig. 1.**

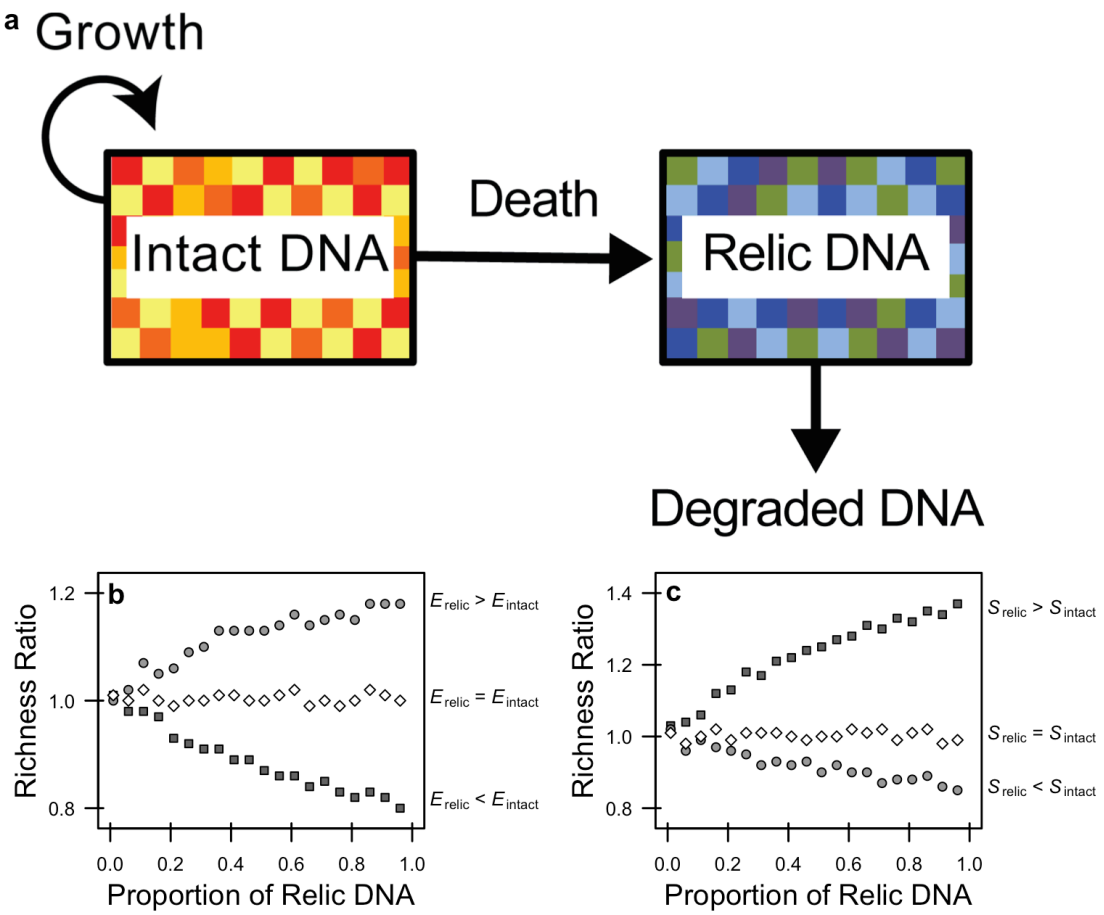

Figure 1 | A model of relic DNA dynamics. **a** The amount of relic DNA in a microbial environment is determined by inputs associated with the mortality of viable individuals with intact DNA and by losses associated with the degradation of relic DNA. If the diversity of sequences contained in the relic DNA pool is sufficiently different from that in the intact DNA pool, then relic DNA may bias estimates of microbial biodiversity (as indicated by different colored boxes) when sampling from the total (intact + relic) DNA pool. We developed a sampling-based simulation model that created communities populated with individuals with intact DNA sampled from a lognormal species abundance distribution. We then mixed the intact community with relic individuals so that the relic contribution to total community ranged from 0.01 to 0.96. The relic community was also populated with individuals sampled from a lognormal species distribution, but which varied in (**b**) evenness (*E*) or (**c**) richness (*S*). To quantify bias in diversity, we calculated "richness ratios" which reflect the number of species in the total DNA pool (intact + relic) divided by the number of species in the intact DNA pool. When values richness ratios = 1, relic DNA has no effect on diversity; when richness ratios > 1, relic DNA leads to overestimates of diversity; richness ratios <1, relic DNA lead to underestimation of diversity.

518 **Fig. 2.**

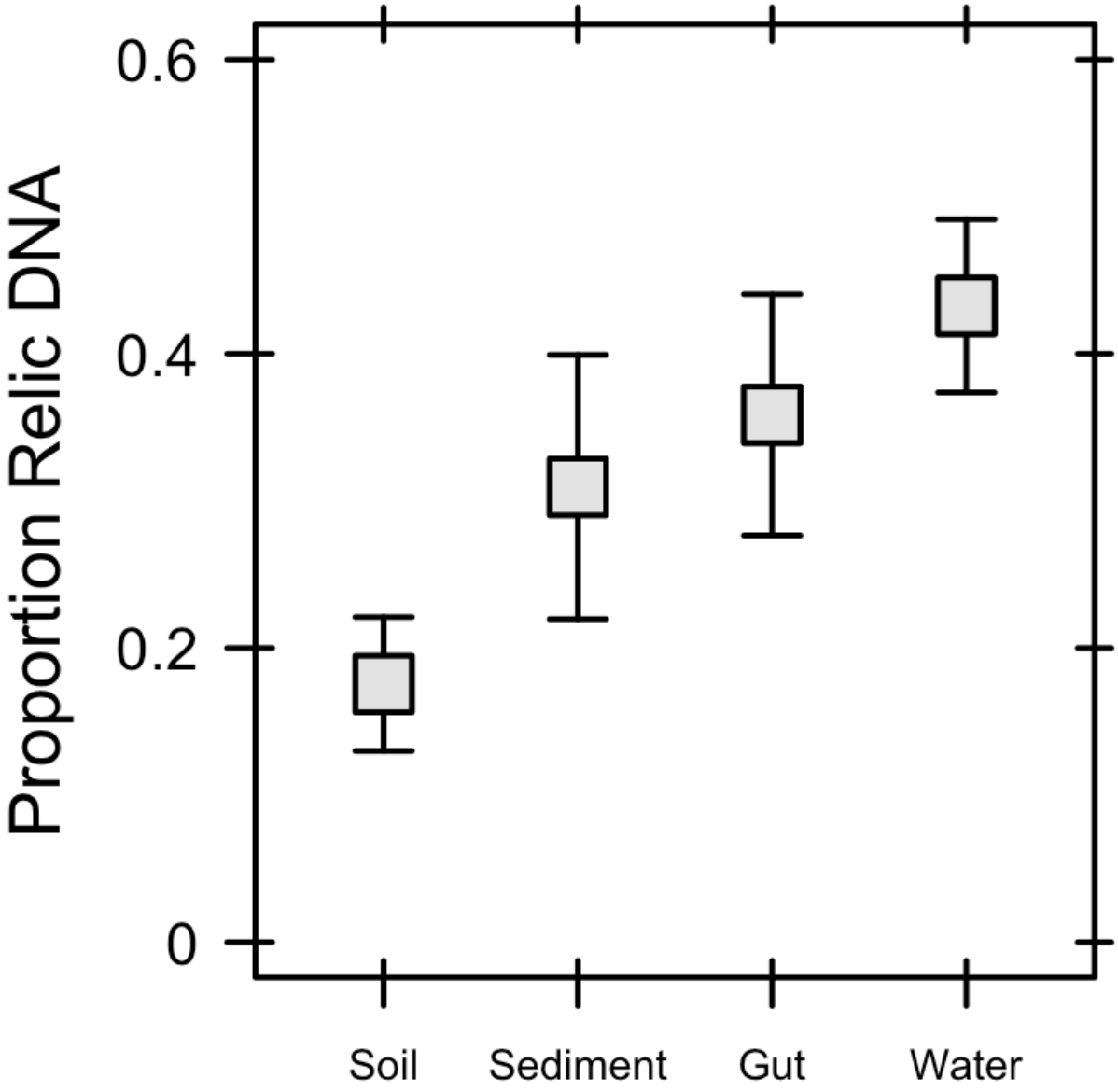

Figure 2 | Proportion of bacterial relic DNA in different ecosystem types. We quantified the amount of intact DNA in a sample after removing relic DNA with a DNase treatment. We then estimated the proportion of relic DNA as 1 - (intact DNA / total DNA), where the total DNA concentration was quantified without DNase treatment. Relic DNA constituted an appreciable portion of the total DNA pool, but was not affected by the ecosystem type from which the sample was collected (gut, soil, sediment, and water). Data are represented as means ± 95% confidence intervals.

520 **Fig. 3.**

522

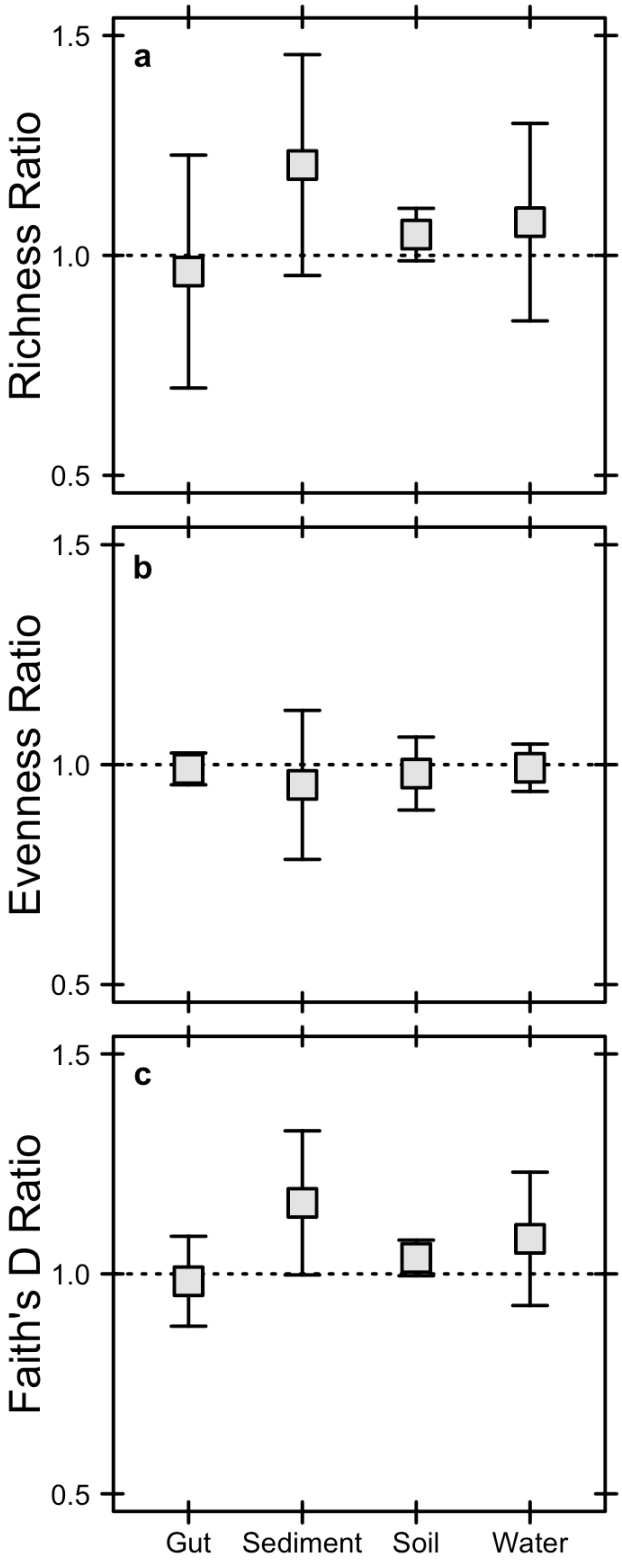

Figure 3 | Bias of relic DNA on within-sample bacterial diversity in different ecosystem types. a-c We tested for the effects of bias caused by relic DNA by calculating diversity ratios for (**a**) richness, (**b**) evenness, and (**c**) phylogenetic diversity. The ratios reflect the diversity of the total DNA pool (intact + relic) divided by the diversity of the intact DNA pool. Relic DNA did not bias any measures of diversity in any of the ecosystem types. Data are represented as means ± 95% confidence intervals. Richness was calculated as the number of operational taxonomic units (97% sequence similarity of the 16S rRNA gene), evenness was calculated using Simpson's evenness index, and phylogenetic diversity was calculated using Faith's *D* index.

524 **Fig. 4.**

526

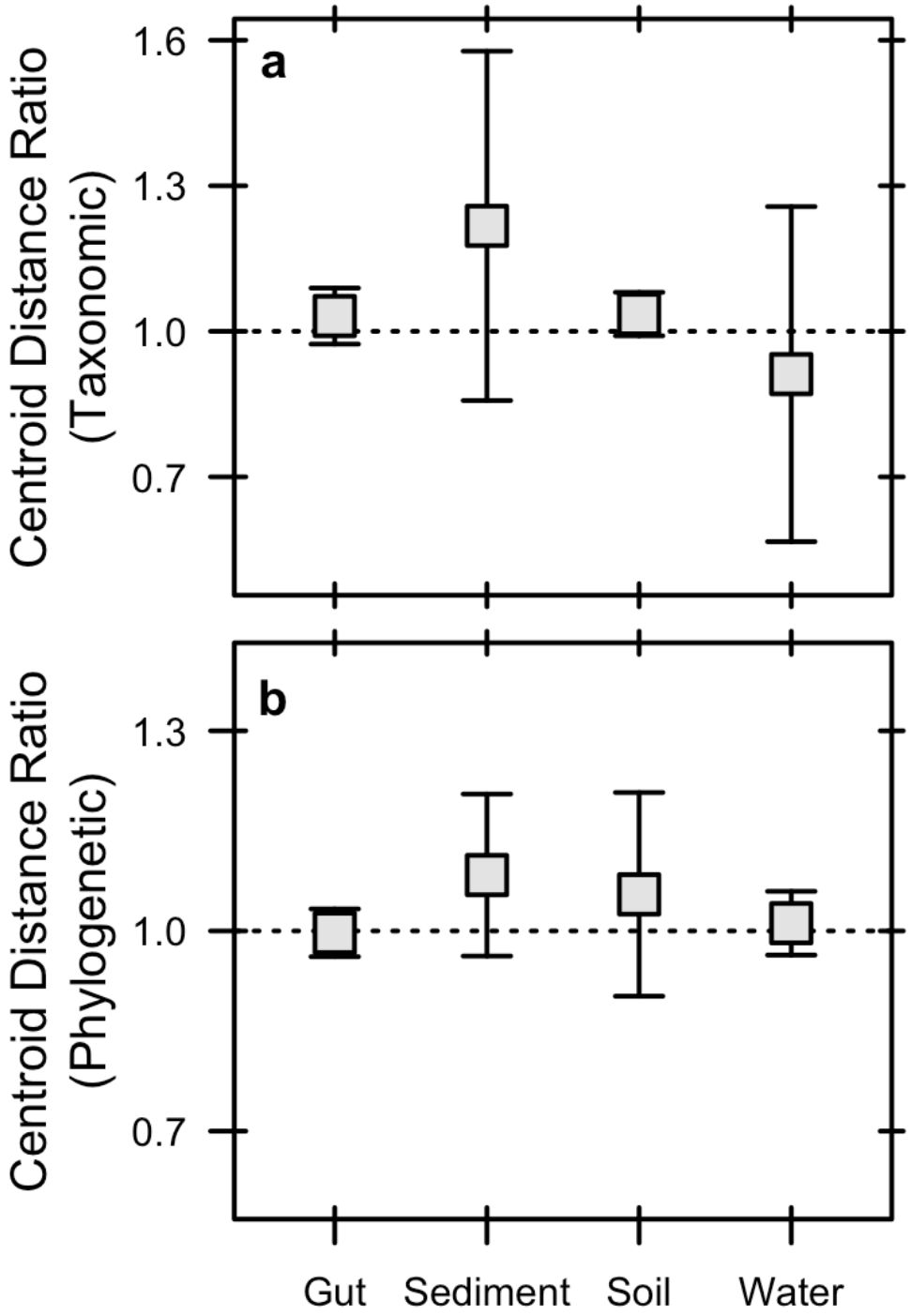

Figure 4 | Bias of relic DNA on the among-sample bacterial diversity in different ecosystem types. a-b We tested for the effects of bias caused by relic DNA by calculating a beta-diversity ratio based on centroid distances. Centroid distances were estimated after performing Principle Coordinates Analyses (PCoA) using taxonomic (**a**) and phylogenetic (**b**) distance metrics (Bray-Curtis and UniFrac, respectively). The centroid distance ratio was calculated on each sample within an ecosystem type and reflects the composition of the total DNA pool (intact + relic) relative to the intact DNA pool. Relic DNA had no effect on beta-diversity for any of the ecosystem types sampled. Data are represented as means ± 95% confidence intervals.