Nutrient supplementation experiments on saltern microbial communities support utilization of DNA as a source of phosphorus

Running Title: DNA as a microbial community's phosphorus source

Zhengshuang Hua[1], Matthew Ouellette[2,$], Andrea M. Makkay[2], R. Thane Papke[2,*], Olga Zhaxybayeva[1,3*]

[1] Department of Biological Sciences, Dartmouth College, Hanover, NH, USA

[2] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

[3] Department of Computer Science, Dartmouth College, Hanover, NH, USA

[$] Current affiliation: The Forsyth Institute, Cambridge, Massachusetts, USA and Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston, Massachusetts, USA

[*]Corresponding authors:

O.Z., olga.zhaxybayeva@dartmouth.edu;

R.T.P., thane@uconn.edu.

Keywords: extracellular DNA, dissolved DNA, hypersaline, haloarchaea, Halobacteria, DNA uptake, community diversity

Abstract

All environments including hypersaline ones harbor measurable concentrations of dissolved extracellular DNA (eDNA) that can be utilized by microbes as a nutrient. However, it remains poorly understood which eDNA components are used, and who in a community utilizes it. For this study, we incubated a saltern microbial community with combinations of carbon, nitrogen, phosphorus, and DNA, and tracked the community response in each microcosm treatment via 16S rRNA and *rpoB* gene sequencing. We show that microbial communities could use DNA only as a phosphorus source, and provision of other sources of carbon and nitrogen was needed to exhibit a substantial growth. The taxonomic composition of eDNA in the water column changed with the availability of inorganic phosphorus or supplied DNA, hinting at preferential uptake of eDNA from specific organismal sources. Especially favored for growth was eDNA from the most abundant taxa, suggesting some haloarchaea prefer eDNA from closely related taxa. Additionally, microcosms' composition shifted substantially depending on the provided nutrient combinations. These shifts allowed us to predict supplemented nutrients from microbial composition with high accuracy, suggesting that nutrient availability in an environment could be assessed from a taxonomic survey of its microbial community.

Introduction

Dissolved extracellular DNA (eDNA) in the environment plays important roles in global cycles of carbon (C), nitrogen (N) and phosphorus (P). It is found at measurable concentrations in every analyzed habitat [1–4] and is estimated to amount to gigatons globally [5]. Presence of eDNA in environments is predominantly influenced by release of DNA via microbial cell lysis from viral infections [6, 7], but also results from DNA freed up from spontaneous and programmed cell death, autolysis, and direct DNA secretion [8–11]. Subsequently, eDNA becomes available for resident organisms to utilize, and in some environments eDNA turnover rates can be as short as a few hours [12]. However, seasonal variability and type of habitat can increase the residence time of eDNA to months and years [13, 14], despite the presence and high activity of secreted extracellular nucleases [3]. Adsorption to minerals and particles can make eDNA unavailable for microorganisms [15], but factors like nutrient obtainability, and the presence of salt could also contribute to the stability, availability and utilization of the available eDNA [13, 16].

Many microorganisms secrete enzymes that degrade eDNA extracellularly [17]. The subsequent uptake of the DNA components could be used to support growth by being used as a C, N or P source [18, 19]. Many archaea and bacteria can also take up double-stranded eDNA as a high molecular weight molecule via a process of natural transformation, and use it either to repair damaged chromosomes or to increase genetic variation [15, 20]. However, since natural transformation has low efficiency rates and only one DNA strand is typically used in recombination, most of the eDNA transported into the cell may be consumed for other purposes, such as an energy source or as input building blocks (e.g., nucleotides) in metabolic pathways [21–25].

57 Hypersaline environments are reported to have the highest concentrations of eDNA [3]. These
58 typically nutrient-poor habitats have little to no primary production, especially near the saturation level of
59 NaCl, but support a high density of cells, usually dominated by heterotrophic, aerobic, obligate halophilic
60 archaea, informally known as Haloarchaea. A nutritionally competent model haloarchaeon *Haloferax*
61 *volcanii* can grow on eDNA, primarily as a P source [26], suggesting that in hypersaline habitats eDNA
62 plays an important role in the phosphorus cycle and also possibly in carbon and nitrogen cycles. Notably,
63 *Hfx. volcanii* discriminates different eDNA sources for growth, as cultures did not grow on purified *E.*
64 *coli* or herring sperm DNA, but grew on *Hfx. volcanii* (i.e., its own) or unmethylated *E. coli* DNA [23].
65 These observations raise a possibility of organismal preferences for selective eDNA utilization (perhaps
66 based on DNA methylation), a largely unexplored phenomenon. To better understand how eDNA
67 influences growth of microbial communities in hypersaline environments, and to investigate eDNA
68 utilization biases associated with organismal source of eDNA by different haloarchaea and hypersaline-
69 adapted bacteria, we conducted microcosm experiments on natural near-saturated hypersaline waters
70 collected from the Isla Cristina solar saltern in southern Spain that were amended by sources of C, N,
71 inorganic phosphorus (P_i) and DNA. We show that eDNA, which was either available in the water column
72 or provided as a supplement, was utilized by the microbial community as a source of phosphorus. Via
73 sequencing of *rpoB* and 16S rRNA genes from DNA collected from both the cells and water column
74 before and after the experiments, we observe that composition of both microbes in the community and
75 eDNA in the water column changes depending on nutrients, and infer that at least some of these shifts are
76 due to the ability to use eDNA and to prefer eDNA of specific taxonomic origin or taxonomic relatedness.

77 We also found that the shifts in taxonomic composition of a saltern microbial community and eDNA in

78 the water column are substantial enough to be predictive of the provided nutrients.

79

80 **Materials and Methods**

81 *Quantification of nutrients in Isla Cristina water samples*

82 The Isla Cristina water samples were analyzed for total organic carbon (TOC), total dissolved nitrogen

83 (TN), total phosphorus (TP) and inorganic phosphorus (P_i). The organic phosphorus in each sample (P_o)

84 was calculated as the difference between TP and P_i. See **Supplementary Methods** for details.

85

86 *Nutrient supplementation of samples, DNA purification,* rpoB *and 16S rRNA genes amplification, and*

87 *DNA sequencing*

88 Isla Cristina water samples were aliquoted into 50mL conical tubes, with each tube receiving 30mL of

89 water sample. These samples were supplemented with various combinations of C (0.5% w/v glucose), N

90 (5 mM NH_4Cl), and P_i (1 mM KH_2PO_4) sources, as well as with DNA (~6ng/µL of either *Hfx. volcanii*

91 DS2 or *E. coli dam*$^-$/*dcm*$^-$). Each combination of nutrient supplementation was performed in duplicate.

92 The microcosms were incubated, DNA was extracted, and *rpoB* and 16S rRNA genes were PCR-

93 amplified as described in **Supplementary Methods**. The purified amplicon products were used to

94 construct 69 *rpoB*-based and 28 16S-rRNA-based libraries. The libraries were sequenced using Illumina

95 MiSeq technology, as detailed in **Supplementary Methods**, collectively producing 3 567 797 and 2 277

96 899 paired-end raw reads for *rpoB* and 16S-rRNA genes, respectively.

97

98 *Sequence quality control, assembly, of contigs into the operational taxonomic units and assignment of*
99 *taxonomic affiliations*
100 The raw reads were trimmed from poor quality regions and assembled into contigs as described in
101 **Supplementary Methods**. This procedure resulted in the 1 174 289 *rpoB* and 1 320 673 16S *rRNA*
102 assembled contigs, which were clustered into operational taxonomic units (OTUs) using QIIME v1.9.1
103 [27] and assigned taxonomic affiliations using the Ribosomal Database Project classifier v2.2 [28], as
104 detailed in **Supplementary Methods**.
105
106 *Quantification of similarities and differences in taxonomic composition across and within samples*
107 For each sample, relative abundance of its OTUs, overall taxonomic composition and the OTU diversity
108 were calculated as described in **Supplementary Methods**. Additionally, the samples were clustered via
109 Principal Coordinates Analysis (PCoA), using beta-diversity matrices as inputs. The clusters were tested
110 for similarity and differences in the OTU composition (see **Supplementary Methods** for details).
111
112 *Examination of the impact of provided nutrients on community composition using linear regression*
113 Linear regression modeling was conducted with provided nutrients (C, N, P_i and P_o) as explanatory
114 variables and number of OTUs in each sample as dependent variables as described in **Supplementary**
115 **Methods**.
116
117 *Quantification of changes in relative abundances of the abundant haloarchaeal OTUs (ahOTUs) in*

118 *response to provided phosphorus*

119 Changes in ahOTU abundance in microcosms and eDNA pools in response to addition of DNA and P_i

120 were visualized as heatmaps mapped to a cladogram of relationships among ahOTUs (see

121 **Supplementary Methods** for details). Additionally, changes in relative abundance of an ahOTU in

122 microcosms and eDNA pools between two experimental treatments were summarized using two metrics,

123 $D = (ICW_a - ICW_b) - (ICC_a - ICC_b)$ and $ln(Odds\ ratio) = \ln(\frac{ICW_a/ICW_b}{ICC_a/ICC_b})$, where a and b denote two

124 compared treatments and ICC_i and ICW_i represent relative abundances of an ahOTU in communities and

125 eDNA pools of a treatment i, respectively.

126

127 *Prediction of provided nutrients from taxonomic composition of a sample*

128 Using the relative abundance of ahOTUs in a sample and categorization based on combination of

129 nutrients provided in the experiments, a random forest classifier [29] was trained and cross validated as

130 described in **Supplementary Methods**. The performance of classifier trained on the real data was

131 compared to that of a classifier trained on the re-shuffled samples.

132

133 *Data availability*

134 Raw sequencing data is available in GenBank under BioProject ID PRJNA646965. **Supplementary**

135 **Data**, which includes a script used for raw data quality control, assembled contigs, database used for

136 *rpoB*-based OTU taxonomy assignment, tables of OTU composition per sample, and nucleotide

137 sequences chosen to represent each OTU, are available in FigShare repository at

138 https://doi.org/10.6084/m9.figshare.12609068.

139

140 **Results**

141 **Isla Cristina's microbial communities grow on eDNA as a phosphorus source**

142 Supplementation of microbial communities from the Isla Cristina saltern (Huelva, Spain) with various
143 combinations of C, N, P_i, and DNA from either *Hfx. volcanii* DS2 (denoted as H) or *E. coli dam*$^-$/*dcm*$^-$
144 (denoted as E) revealed that unless both C and N were provided, no notable growth was observed (**Fig. 1a**
145 and **1b**). Supplementation with both C and N (+C+N treatment) yielded an average net OD_{600} increase of
146 0.414 (**Fig. 1a**) and 0.506 (**Fig. 1c**) in the two sets of conducted experiments. Providing P_i in addition to C
147 and N (+C+N+P_i treatment) resulted in the largest growth (**Fig. 1a** and **Fig. 1c**). Interestingly, samples
148 that were supplemented with DNA instead of P_i (+C+N+H or +C+N+E treatments) also exhibited a
149 comparable amount of growth, although net OD_{600} increase was ~12.9% lower than in the +C+N+P_i
150 treatment (One-way ANOVA test; P-value 0.0036; **Fig. 1c**). Nevertheless, the net OD_{600} increases in the
151 +C+N+H and +C+N+E treatments were 50% and 54% higher, respectively, than in the +C+N treatment
152 (One-way ANOVA test; P-values 0.0023 and 0.0018, respectively; **Fig. 1c**). This suggests that besides P_i,
153 the saltern community utilizes eDNA as a source of phosphorus.

154 The observed patterns of microbial community growth under different treatments are consistent with
155 the consumed amounts of C, N, and P_i, as measured by comparing the total organic carbon (TOC), total
156 nitrogen (TN), and total phosphorus (TP) in the filtered untreated water samples and in the treated filtered
157 water samples after the communities reached the stationary phase. The pre-incubation communities

158 contained small quantities of TOC, TN, and TP (**Supplementary Table S1**). In treatments where
159 communities exhibited no notable growth (i.e., C was not provided), there were also no notable changes
160 in TOC, TP and TN (**Fig. 1d**). In the microcosms with some growth (C was provided but N was not), a
161 31-36% decrease in TN was observed (**Fig. 1d** and **Supplementary Table S1**), suggesting that growth
162 was sustained until naturally occurring utilizable nitrogen resources were depleted. The fastest growing
163 microcosms, i.e. in which both C and N were supplied, exhibited the largest decreases in TN, TP and TOC
164 (**Fig. 1d**).
165 To further investigate the effectiveness of DNA as a source of phosphorus, we examined the
166 contributions of inorganic (P_i) and organic (P_o) phosphorus to the TP in the microcosms after they reached
167 stationary phase (**Fig. 1e** and **Supplementary Table S2**). We found that in the microcosms supplemented
168 with C and N but not provided with P_i (i.e., +C+N, +C+N+H or +C+N+E treatments), the remaining TP
169 was dominated by P_o (~91%, ~70% and ~74%, respectively). Since the microcosms under the +C+N,
170 +C+N+H or +C+N+E treatments exhibited substantial growth, we conjecture that while DNA (i.e., P_o)
171 was used by the communities as a source of phosphorus, P_i is generally preferred when available.
172
173 **The taxonomic composition of the microbial communities and eDNA pool varies across treatments**
174 To better understand how the different nutrient availability affects both organismal composition of a
175 community and organismal sources of DNA in the eDNA pool, we sequenced 16S rRNA and *rpoB* genes
176 from the DNA extracted from the Isla Cristina living Cells (ICC) and the Isla Cristina Water column
177 (ICW) before and after the above-described microcosm experiments.

178 Analysis of the 16S rRNA OTUs revealed that, as expected for a saltern environment, the pre-
179 incubation community (ICC sample "X" on **Fig. 2a**) is numerically dominated by archaea from class
180 Halobacteria (relative abundance of 92.9%) and by bacteria from the genus *Salinibacter* (relative
181 abundance of 6.6%). The remaining 0.5% of OTUs belong to either other archaeal classes or other
182 bacterial genera (see **Supplementary Data**). Halobacteria and *Salinibacter* also dominate all post-
183 treatment communities, with their relative abundances in 92-100% and 0.1%-7.3% range, respectively
184 (**Fig. 2a**).
185 The DNA in water column is also dominated by eDNA from Halobacteria and *Salinibacter* OTUs
186 (**Fig. 2a**). However, DNA originating from bacteria on average represents a higher fraction in the eDNA
187 pools than in the microbial communities, with the largest fraction (26%) found in the pre-incubation water
188 column (ICW sample "X" on **Fig. 2a)**. Additionally, some eDNA comes from OTUs that were either not
189 observed in any of the ICC samples (**Fig. 2b**) or were found in the ICC samples in low abundances (see
190 **Supplementary Data**). These eDNA fragments may have originated from living microorganisms
191 elsewhere and were transported to the saltern environment. Alternatively, they may represent resistant-to-
192 consumption eDNA that has slowly accumulated from the rare members in the pre-incubation community
193 (see **Supplementary Results** for specific examples.)
194 Overall, OTUs present at >1% abundance constitute the greatest majority of OTUs in every sample
195 (**Fig. 2c**). Unfortunately, many haloarchaea are known to carry multiple non-identical copies of 16S rRNA
196 gene [30–32], with some copies exhibiting within-genome identity as low as 91.3% (**Supplementary**
197 **Table S5**). Since this intragenomic heterogeneity is likely to affect accuracy of the 16S rRNA-based

198 relative abundance estimates, we also amplified the *rpoB* gene, which is known to exist in haloarchaeal
199 genomes in a single copy, and was shown to resolve genera assignments better than the 16S rRNA gene
200 [33–35]. Because the 16S rRNA analysis indicates that the microbial communities are dominated by
201 Halobacteria and do not contain any OTUs from other archaeal classes at > 1% abundance, we amplified
202 the *rpoB* gene using the Halobacteria-specific primers. Of the 5 905 *rpoB*-based OTUs detected across all
203 samples, 5 851 (99.1%) are assigned to class Halobacteria, demonstrating that the primers designed to
204 amplify this class of Archaea have high specificity. Relative abundances of Halobacterial genera observed
205 in the *rpoB*-based and 16S rRNA-based analyses are comparable, with few discrepancies detailed in
206 **Supplementary Results**.
207 Thirty-four of the *rpoB*-based OTUs have > 1% relative abundance in at least one ICC sample.
208 Thirty two of the 34 OTUs maintain > 1% relative abundance in at least one ICW sample. Additional 29
209 halobacterial OTUs have > 1% relative abundance in at least one ICW sample, and found in low
210 abundance in at least one ICC sample. For the remainder of the analyses we designated these 34 + 29 = 63
211 OTUs as *abundant halobacterial OTUs* (ahOTUs). Notably, an individual sample contains only a fraction
212 of the 63 ahOTUs: 16 ± 1.8 of ahOTUs per ICC sample and 18 ± 2.8 ahOTUs per ICW sample. Because
213 in each sample the cumulative relative abundance of ahOTUs is 67-78% and 80 - 95% for ICC and ICW
214 samples, respectively (**Fig. 2d**), there are non-negligible differences in the ahOTU composition and their
215 relative abundances across samples. Below, we examine in detail how these differences relate to nutrient
216 availability.
217

218 **Carbon, DNA and inorganic phosphorus alter taxonomic composition of microbial communities**

219 **and eDNA pools**

220 Based on the relative abundances of both the *rpoB*-based and the 16S rRNA-based OTUs, the samples

221 aggregated into three distinct clusters **(Fig. 3a** and **Supplementary Fig. S1a)**. Differences in OTU

222 composition of the three clusters are significant (PERMANOVA: R = 0.74, P = 0.001 [*rpoB*] and R =

223 0.72, P = 0.001 [16S rRNA]; ANOSIM: R = 0.87, P = 0.001 [*rpoB*] and R = 0.82, P = 0.001 [16S rRNA]).

224 Cluster 1 consists of the cellular communities' samples that come from the pre-treatment and all

225 microcosms that either exhibited no notable or some growth (**Fig. 1b**), hereafter referred as "slow-

226 growing communities". Cluster 2 consists of the water column samples that correspond to the cellular

227 communities of Cluster 1. Cluster 3 consists of the cellular communities' and water column samples from

228 all experiments that produced fast-growing microcosms, that is, the microcosms that were supplemented

229 simultaneously with C and N. Such clear taxonomic separation of fast- and slow-growing microcosms

230 suggests substantial shifts in community composition in response to nutrient supplementation. Clusters 1,

231 2 and 3 also emerge even if OTU abundance is ignored and only OTU richness is considered

232 (**Supplementary Fig. S2**), suggesting that the observed clustering is not driven solely by a few abundant

233 community members, but also influenced by rare microorganisms and their DNA in the water column.

234 Presence and abundance of specific bacterial and archaeal OTUs, and especially those with > 1%

235 abundance, is behind the observed shifts in taxonomic composition among the three clusters (**Fig. 2**).

236 Among Halobacteria, the relative abundances of 48 of the 63 ahOTUs differ significantly among the three

237 clusters (the Mann-Whitney *U* test; see **Supplementary Table S6** for P values; **Fig. 4a**). Cluster 1

238 samples are dominated mostly by Haloferacaceae ahOTUs (**Fig. 4b** and **4c**). In contrast, eDNA from
239 Halorubraceae ahOTUs (**Fig. 4b** and **4d**) dominate Cluster 2 samples. Finally, in Cluster 3 samples,
240 Haloferacaceae and Halorubraceae ahOTUs become much less abundant and are replaced by
241 Haloarculaceae and Halobacteriaceae ahOTUs (**Fig. 4b** and **4e**).

242 Within the three clusters, there are visibly distinct sub-clustering of samples that imply more
243 nuanced differences in taxonomic composition of the samples. For the Cluster 1 communities and their
244 water columns (Cluster 2), the pre-treatment sample, samples without notable growth (no C added) and
245 some growth (C added) clearly form separate clusters, with further subdivision due to provision of *Hfx.*
246 *volcanii* DNA (**Fig. 3b-c** and **Supplementary Fig. S1b-c**). These groups arise primarily due to the
247 differences in relative abundance of Halorubraceae, Haloarculaceae and Haloferacaceae OTUs (**Fig. 3e**).
248 Within fast-growing microcosms (Cluster 3), sub-clustering is less pronounced but reflects availability
249 and type of P sources provided in treatments (**Fig. 3d** and **Supplementary Fig. S1d**). Taxonomically, the
250 observed groups are due to differences in relative abundance of Haloferacaceae, Halorubraceae,
251 Haloarculaceae and Halobacteriaceae OTUs (**Fig. 3e**).

252 Taken together, supplementation of nutrients significantly influences composition of both the
253 microbial community and the eDNA pool. Linear regression modeling with nutrients as categorical
254 variables revealed that provision of C is the primary driver of microbial community composition shifts
255 ($R_{adj}^2 = 0.45$; $P = 0.002$; *t* value: 4.5), suggesting unequal abilities of various taxa to compete for added
256 glucose. Provision of N alone does not affect community composition, suggesting that community
257 members have equal abilities in its assimilation. Provision of P_i and DNA significantly alters the

258 composition of eDNA in the water column (R_{adj}^2 = 0.327; P = 0.001; *t* value: 3.6), raising a possibility that
259 there are organismal preferences for phosphate and specific "taxonomic affiliation" of DNA as a P source.
260

261 **Shifts in eDNA OTU composition depending on available sources of phosphorus suggest**
262 **preferential eDNA uptake**

263 It is surprising that the DNA content of the initial community and the slow-growing microcosms (Cluster
264 1) are quite different from the eDNA pools of their water columns (Cluster 2). This distinction, albeit less
265 apparent, also exists in the fast-growing microcosms (Cluster 3; **Fig. 3d**). Assuming that all cells in a
266 microcosm have equal probability of birth and death, it would be expected to see similar OTU
267 compositions of a microbial community and of the eDNA from the corresponding water column.
268 Interestingly, in the slow-growing communities, provision of *Hfx. volcanii* DNA (+H and +C+H
269 treatments) significantly increases the eDNA OTU richness (the Mann-Whitney *U* test, P = 1.6×10^{-5};
270 **Supplementary Fig. S3**). In the fast-growing communities, the increase eDNA OTU richness, albeit not
271 significant, is caused by the addition of P_i (+C+N+P_i treatment; **Supplementary Fig. S3**). Therefore, we
272 hypothesize that the observed differences in OTU compositions of cellular communities and their water
273 columns are due to biased eDNA uptake from certain taxonomic groups. Furthermore, we propose that the
274 availability of *Hfx. volcanii* DNA and inorganic phosphorus in slow- and fast-growing communities,
275 respectively, results in the reduced uptake of available eDNA as a P source, leading to eDNA
276 accumulation in the environment and causing the observed changes in eDNA OTU composition. And in
277 reverse, the observed limited OTU diversity when P sources are scarce may be due to depletion of eDNA

278 from specific taxa through taxonomically-biased eDNA uptake.

279

280 **Preferences for eDNA of specific taxonomic origins by the saltern microbial community**

281 Complexity and variability of microcosms across experimental treatments (**Fig. 3e**), as well as impacts of
282 sequencing depth and sequencing biases, make it difficult to establish exact eDNA preferences of specific
283 OTUs from our data. However, by comparing relative abundances of specific OTUs in microcosms with
284 and without additional P sources, we can indirectly infer some taxonomic biases of eDNA uptake in this
285 saltern microbial community as a whole. Specifically, we will continue to assume that all cells in a
286 microcosm have equal probability of birth and death. As a result, OTUs that do not substantially change
287 their relative abundance in treatments with and without an extra P source are not expected to show
288 substantial changes in the relative abundance of their eDNA in the water column. Deviations from these
289 patterns would suggest an eDNA uptake bias, and we developed and applied two metrics to quantify such
290 differences (see **Materials and Methods** for details).

291 From comparisons of slow-growing microcosms with and without added *Hfx. volcanii* DNA, we
292 found that when *Hfx. volcanii* DNA is not provided, eDNA from five Halorubraceae and Haloarculaceae
293 ahOTUs is depleted (see highlighted taxa in **Supplementary Fig. S4** and **Supplementary Table S7**) and
294 eDNA of 305 additional OTUs become undetectable in the water columns (**Supplementary Table S8**). Of
295 these 305 OTUs, 98 (32%) belong to Halorubraceae and Haloarculaceae families. These findings suggest
296 preferential consumption of Halorubraceae and Haloarculaceae eDNA in slow-growing microcosms. In
297 fast-growing microcosms, eDNA from Haloarculaceae, Halorubraceae and Halobacteriaceae ahOTUs are

298 depleted in water columns without provided P source (either *Hfx. volcanii* DNA, *E. coli* DNA or P_i; see
299 highlighted taxa in **Supplementary Fig. S5** and **Supplementary Table S9**). In comparisons between
300 +C+N and +C+N+P_i treatments, eDNA from 158 OTUs become depleted in the water column without P_i,
301 and 62 of the 158 OTUs (39%) belong to Haloarculaceae, Halorubraceae and Halobacteriaceae families
302 (**Supplementary Table S10**). These observations suggest preferential consumption of eDNA from these
303 three families in fast-growing microcosms. Interestingly, the most abundant of the depleted eDNA
304 belongs to ahOTUs from families that represent a substantial fraction of the microcosms (Halorubraceae
305 for the slow-growing and Haloarculaceae for the fast-growing communities; **Fig. 3e**), indicating that
306 eDNA most closely related to the most abundant OTUs in the community is preferentially consumed.
307 Intriguingly, several ahOTUs become depleted in the +C+N+P_i water columns when compared to the
308 other three treatments with fast growth (+C+N, +C+N+H, and +C+N+E) (**Supplementary Fig. S5** and
309 **Supplementary Table S9**), suggesting that some eDNA may be more preferred by some organisms than
310 inorganic phosphorus.

311 We also hypothesize that the saltern community can also utilize bacterial eDNA, which can be
312 exemplified by Chitinophagaceae DNA (**Supplementary Fig. S6**). Chitinophagaceae OTUs are not
313 detected in the pre-incubation cellular community sample, but their DNA is abundant in the pre-
314 incubation water column. Under all experimental treatments, the Chitinophagaceae OTUs continue to be
315 undetected in the microbial communities, suggesting they are at best extremely rare community members.
316 However, their DNA is either absent or found at very low abundance in the treatments' water columns,
317 suggesting that Chitinophagaceae eDNA in the initial, pre-incubation water column was consumed by the

318 microcosms' communities.

319 Provision of large amounts of *Hfx. volcanii* DNA allowed us to identify a taxon that likely prefers to
320 utilize *Hfx. volcanii* DNA as a P source. The ahOTU *denovo3413*, which is assigned to *Haloarculaceae*
321 family and has 91.4% of nucleotide identity to *Halapricum* sp. strain SY-39 (MF443862.1), has < 1%
322 relative abundance in the slow-growing communities without provided *Hfx. volcanii* DNA. However,
323 when *Hfx. volcanii* DNA is supplied, *denovo3413* becomes significantly more abundant both in microbial
324 communities (the Mann-Whitney *U* test, $P = 7.4\times10^{-5}$; **Supplementary Table S7**) and the corresponding
325 water column samples (the Mann-Whitney *U* test, $P = 1.1\times10^{-5}$; **Supplementary Table S7**).

326

327 **Provided nutrients can be predicted from taxonomic affiliations of microbial community members**
328 **and of water column DNA**

329 The shifts in community composition in response to availability of C, N, and P suggest that nutrients
330 provided in the experiments may be predictable from the OTU composition associated with DNA in cells
331 and dissolved in water columns. We designed a random forest classifier that takes ahOTU composition as
332 an input and predicts if a sample falls into one of the five treatment groups: "No C was provided (no C)",
333 "C but no N were provided (+C)", "Only C and N were provided (no P)", "C, N and P_i were provided
334 ($+P_i$)", or "C, N and DNA were provided (+H/+E)". We found that a random forest classifier achieves
335 95.5% accuracy in a 20-fold cross-validation and 94.6% accuracy in out-of-bag (OOB) error estimate.
336 These predictions have significantly higher accuracies than those obtained in the classifier trained on the
337 randomly shuffled samples (Student's *t* test, $P < 2.2*10^{-16}$ for cross validation approach [**Fig. 5a**] and

338 Kolmogorov-Smirnov test, $P < 2.2*10^{-16}$ for OOB approach [**Fig. 5b**]). Predictions remain robust even for
339 smaller training data sets (**Fig. 5c**). Overall, 63 ahOTUs can predict the available nutrients with the mean
340 prediction accuracy of 62%. The predictions are strongly influenced by 24 ahOTUs where each one
341 contributes ≥1% of the prediction accuracy (**Fig. 5d**) and their removal decreases the overall mean
342 classifier accuracy by 49%. Consistent with this observation, 23 of the 24 ahOTUs have significantly
343 different abundances in samples from the five groups (**Fig. 5e**; one-way ANOVA test, all 23 P values <
344 0.01).
345

346 **Discussion**

347 Our microcosm experiments with saltern-derived microbial communities revealed that carbon, nitrogen
348 and phosphorus are limiting resources, since their simultaneous provision resulted in the largest growth of
349 cells. Having a readily accessible carbon source is most crucial for this microbial community, as no
350 notable growth was observed without supplementation of carbon. Nitrogen, on the other hand, appears to
351 be more easily assimilated from the naturally available sources, as microcosms with no supplemented
352 nitrogen exhibited growth until the environmental nitrogen was depleted. Interestingly, the communities
353 supplied with carbon and nitrogen grew well even if not supplied with inorganic phosphorus, suggesting
354 that they were able to utilize either provided DNA or naturally occurring eDNA as a source of
355 phosphorus. Nutrient usage of eDNA is not unique to the saltern microorganisms, as other microbes are
356 known to use DNA as a nutrient by enzymatic cleavage of eDNA (either extracellularly or intracellularly)
357 into deoxyribose, inorganic orthophosphate, and nucleobases [14, 17]. Yet, there is also a possibility that
358 the observed growth in our microcosms was instead due to utilization of internal cellular DNA as a source

359 of phosphorus. For example, feeding experiments on a model haloarchaeon *Hfx. volcanii* with the same
360 types of nutrients demonstrated that in the absence of supplemented phosphate *Hfx. volcanii* utilized its
361 chromosomal DNA, rather than eDNA, as a P source [23, 26], which is possible because *Hfx. volcanii* is
362 known to maintain multiple copies of its chromosome and plasmids [36]. Such polyploidy has been
363 shown for all tested Haloarchaea [37, 38], and is likely widespread in the Euryarchaeota phylum [39].
364 Hence, many microorganisms populating the communities in our experiments could be polyploid and
365 consequently could use the extra copies of their chromosomes as a P source.

366 Given the presumed utilization of eDNA as a P source, it is puzzling why so much of the available
367 eDNA has remained unused. One possibility is that the community reached the stationary phase because it
368 ran out of other nutrients first. For instance, nitrogen was used up in the fast-growing communities
369 (**Supplementary Table S1**). Alternatively, the microorganisms may not have been able to utilize all
370 available eDNA. For instance, aquatic bacteria are hypothesized to specialize on different sizes of DNA
371 fragments [19], and naturally competent *Haemophilus influenzae* and *Neisseria gonorrhoeae* use species-
372 specific DNA uptake recognition sequences and therefore would not import each other's DNA [40–42].
373 DNA methylation patterns are important for discriminating different sources of DNA during the uptake
374 process [23, 43]. As a result, some eDNA might not be recognized as a P source. Consistent with these
375 observations, the changes in eDNA composition of the water columns associated with our experimental
376 microcosms showed that organismal origin of eDNA indeed matters for growth on eDNA, with some
377 bacterial and archaeal eDNA being clearly preferred. Especially preferred is eDNA from the most
378 dominant members of the microcosms, suggesting some haloarchaea prefer eDNA that is most closely

379 related to their own. This preference may be based on DNA methylation patterns [23]. Future studies are
380 needed to better understand the specific reasons behind the observed biases in eDNA utilization.
381 The observed patterns of preferential eDNA consumption and differential growth under various
382 nutrient availability regimes are associated with substantial shifts in the taxonomic composition and
383 overall diversity of underlying microcosm communities, suggesting that in salterns the microbial
384 community assembly is determined by the available resources. Such resource-based assembly also has
385 been demonstrated in synthetic bacterioplankton communities [44], and therefore may represent a general
386 principle for a microbial community assembly. The changes in a microbial community structure in
387 response to environmental challenges are highly dynamic [45], and are likely due to a combination of
388 differential abilities of lineages to survive under severe nutrient limitation [46, 47], to compete for
389 assimilation of limited nutrients [48], to compete for uptake of abundant nutrients [49], to inhibit growth
390 of their competitors [50], and to create mutualistic relationships via metabolic exchange and cross-feeding
391 [45, 51]. Although relative importance of these factors in our saltern community remains to be
392 established, differential abundance of Haloferaceae, Halorubraceae and Halobacteriaceae families in
393 response to availability of carbon and phosphorus indicate that competition for the available nutrients
394 plays a major role in shaping saltern communities. Moreover, after addition of limiting nutrients, the
395 microcosms exhibited increased biomass production and reduction in OTU richness. Such loss of species
396 diversity in response to increase in nutrient availability has been also demonstrated in bacterial biofilms
397 [52], soil microbial communities [53], phytoplankton [54] and plants [55], and was interpreted as a
398 support for the niche dimensionality theory [56]. The theory postulates that biodiversity is maintained

399 under limiting resources because competition for them can be achieved via different trade-offs, which in
400 turn creates multiple niches that can be occupied by similar but not identical species or lineages.
401 Removing nutrient limitation eliminates trade-offs and leads to a decreased diversity. Yet, presence of
402 several closely related and abundant OTUs from *Haloarcula* and *Halomicroarcula* genera in the fast-
403 growing microcosms cannot be explained by competition under the niche dimensionality framework.
404 Therefore, other processes, such as mutual exchange of metabolic by-products, may be at play to promote
405 their coexistence [51]. Future metagenomic surveys of metabolic genes within these microcosms will help
406 to test this hypothesis.

407 Changing microbial communities can in turn have a great impact on the environment they inhabit
408 [57, 58] . By applying a machine learning algorithm to OTU abundance and nutrient treatments data from
409 our microcosm experiments, we found that the nutrient availability could be predicted with high accuracy
410 from the DNA composition of a microbial community and water column. Notably, to achieve the high
411 accuracy it is sufficient to have data from only few abundant OTUs. Our findings suggest that nutrient
412 availability in a natural environment (or its past and future changes) could be predicted from a taxonomic
413 survey of a microbial community, if appropriate training data from laboratory experiments is available.
414 This could become a useful tool for advancing our understanding of ecological processes happening in
415 particular environments and for practical applications, like bioremediation projects or treatments of
416 microbiome-impacted diseases.
417

Acknowledgements

We thank Drs. Antonio Ventosa, Paulina Corral, and Cristina Sánchez-Porro (University of Sevilla), who helped with saltern sampling. We also thank Dr. Jialiang Kuang (University of Oklahoma) for discussions about random forest predictions, and Dr. Shaopeng Li (East China Normal University) for discussions about community assembly and niche dimensionality. The work was supported through a NASA Exobiology award NNX15AM09G to R.T.P. and O.Z., and by the Simons Foundation Investigator in Mathematical Modeling of Living Systems award 327936 to O.Z.

Author contributions

Z.S.H., O.Z., and R.T.P. conceived the study. M.O. and A.M. performed the microcosm experiments, DNA extraction and sequencing. Z.S.H. performed the metagenomic and statistical analyses. Z.S.H., O.Z., R.T.P. and M.O. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Karl DM, Bailiff MD. The measurement and distribution of dissolved nucleic acids in aquatic environments. *Limnol Oceanogr* 1989; **34**: 543–558.

438 2. Paul JH, Cazares LH, David AW, DeFlaun MF, Jeffrey WH. The distribution of dissolved DNA in an
439 oligotrophic and a eutrophic river of Southwest Florida. *Hydrobiologia* 1991; **218**: 53–63.
440 3. Corinaldesi C, Beolchini F, Dell'anno A. Damage and degradation rates of extracellular DNA in
441 marine sediments: implications for the preservation of gene sequences. *Mol Ecol* 2008; **17**: 3939–
442 3951.
443 4. Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. Relic DNA is abundant in soil
444 and obscures estimates of soil microbial diversity. *Nat Microbiol* 2016; **2**: 1–6.
445 5. Dell'Anno A, Danovaro R. Extracellular DNA plays a key role in deep-sea ecosystem functioning.
446 *Science* 2005; **309**: 2179–2179.
447 6. Wilhelm SW, Suttle CA. Viruses and nutrient cycles in the sea: viruses play critical roles in the
448 structure and function of aquatic food webs. *BioScience* 1999; **49**: 781–788.
449 7. Fenchel T. The microbial loop – 25 years later. *J Exp Mar Biol Ecol* 2008; **366**: 99–103.
450 8. Azam F, Smith DC, Steward GF, Hagström Å. Bacteria-organic matter coupling and its significance
451 for oceanic carbon cycling. *Microb Ecol* 1994; **28**: 167–179.
452 9. Orellana MV, Pang WL, Durand PM, Whitehead K, Baliga NS. A role for programmed cell death in
453 the microbial loop. *PLoS One* 2013; **8**: e62595.
454 10. Steinmoen H, Knutsen E, Havarstein LS. Induction of natural competence in *Streptococcus*
455 *pneumoniae* triggers lysis and DNA release from a subfraction of the cell population. *Proc Natl Acad*
456 *Sci USA* 2002; **99**: 7681–7686.
457 11. Draghi JA, Turner PE. DNA secretion and gene-level selection in bacteria. *Microbiology* 2006; **152**:

458 2683–2688.

459 12. Paul JH, Jeffrey WH, DeFlaun MF. Dynamics of extracellular DNA in the marine environment. *Appl*
460 *Environ Microbiol* 1987; **53**: 170–179.

461 13. Salter I. Seasonal variability in the persistence of dissolved environmental DNA (eDNA) in a marine
462 system: The role of microbial nutrient limitation. *PLoS One* 2018; **13**: e0192409.

463 14. Nielsen KM, Johnsen PJ, Bensasson D, Daffonchio D. Release and persistence of extracellular DNA
464 in the environment. *Environ Biosafety Res* 2007; **6**: 37–53.

465 15. Lorenz MG, Wackernagel W. Bacterial gene transfer by natural genetic transformation in the
466 environment. *Microbiol Mol Biol Rev* 1994; **58**: 563–602.

467 16. Tehei M, Franzetti B, Maurel M-C, Vergne J, Hountondji C, Zaccai G. The search for traces of life:
468 the protective effect of salt on biological macromolecules. *Extremophiles* 2002; **6**: 427–430.

469 17. Mulcahy H, Charron–Mazenod L, Lewenza S. *Pseudomonas aeruginosa* produces an extracellular
470 deoxyribonuclease that is required for utilization of DNA as a nutrient source. *Environ Microbiol*
471 2010; **12**: 1621–1629.

472 18. Pinchuk GE, Ammons C, Culley DE, Li S-MW, McLean JS, Romine MF, et al. Utilization of DNA
473 as a sole source of phosphorus, carbon, and energy by *Shewanella* spp.: ecological and physiological
474 implications for dissimilatory metal reduction. *Appl Environ Microbiol* 2008; **74**: 1198–1208.

475 19. Lennon JT. Diversity and metabolism of marine Bacteria cultivated on dissolved DNA. *Appl Environ*
476 *Microbiol* 2007; **73**: 2799–2805.

477 20. Chen I, Dubnau D. DNA uptake during bacterial transformation. *Nat Rev Microbiol* 2004; **2**: 241–

249.

21. Finkel SE, Kolter R. DNA as a nutrient: novel role for bacterial competence gene homologs. *J Bacteriol* 2001; **183**: 6288–6293.

22. Vorkapic D, Pressler K, Schild S. Multifaceted roles of extracellular DNA in bacterial physiology. *Curr Genet* 2016; **62**: 71–79.

23. Chimileski S, Dolas K, Naor A, Gophna U, Papke RT. Extracellular DNA metabolism in *Haloferax volcanii*. *Front Microbiol* 2014; **5**.

24. Redfield RJ. Do bacteria have sex? *Nat Rev Genet* 2001; **2**: 634–639.

25. Palchevskiy V, Finkel SE. *Escherichia coli* competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *J Bacteriol* 2006; **188**: 3902–3910.

26. Zerulla K, Chimileski S, Näther D, Gophna U, Papke RT, Soppa J. DNA as a phosphate storage polymer and the alternative advantages of polyploidy for growth or survival. *PLoS One* 2014; **9**: e94819.

27. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010; **7**: 335–336.

28. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 2009; **37**: D141–D145.

29. Breiman L: Random forests. *Mach Learn* 2001; **45**: pp 5–32.

30. Mylvaganam S, Dennis PP. Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaebacterium *Haloarcula marismortui*. *Genetics* 1992; **130**: 399–410.

498 31. Vreeland RH, Straight S, Krammes J, Dougherty K, Rosenzweig WD, Kamekura M. *Halosimplex*
499 *carlsbadense* gen. nov., sp. nov., a unique halophilic archaeon, with three 16S rRNA genes, that grows
500 only in defined medium with glycerol and acetate or pyruvate. *Extremophiles* 2002; **6**: 445–452.
501 32. Boucher Y, Douady CJ, Sharma AK, Kamekura M, Doolittle WF. Intragenomic heterogeneity and
502 intergenomic recombination among haloarchaeal rRNA Genes. *J Bacteriol* 2004; **186**: 3980–3990.
503 33. Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and *rpoB*
504 genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* 2007; **73**: 278–
505 288.
506 34. Ki J-S, Zhang W, Qian P-Y. Discovery of marine *Bacillus* species by 16S rRNA and *rpoB* comparisons
507 and their usefulness for species identification. *J Microbiol Methods* 2009; **77**: 48–57.
508 35. Vos M, Quince C, Pijl AS, Hollander M de, Kowalchuk GA. A Comparison of *rpoB* and 16S rRNA
509 as markers in pyrosequencing studies of bacterial diversity. *PLoS One* 2012; **7**: e30600.
510 36. Soppa J. Evolutionary advantages of polyploidy in halophilic archaea. *Biochem Soc Trans* 2013; **41**:
511 339–343.
512 37. Breuert S, Allers T, Spohn G, Soppa J. Regulated polyploidy in halophilic Archaea. *PLoS One* 2006;
513 **1**: e92.
514 38. Ludt K, Soppa J. Polyploidy in halophilic archaea: regulation, evolutionary advantages, and gene
515 conversion. *Biochem Soc Trans* 2019; **47**: 933–944.
516 39. Soppa J. Ploidy and gene conversion in Archaea. *Biochem Soc Trans* 2011; **39**: 150–154.
517 40. Mathis LS, Scocca JJ. On the role of pili in transformation of *Neisseria gonorrhoeae*. *Microbiology*

518 1984; **130**: 3165–3173.

519 41. Redfield RJ, Findlay WA, Bossé J, Kroll JS, Cameron AD, Nash JH. Evolution of competence and
520 DNA uptake specificity in the Pasteurellaceae. *BMC Evol Biol* 2006; **6**: 82.

521 42. Hamilton HL, Dillard JP. Natural transformation of *Neisseria gonorrhoeae*: from DNA donation to
522 homologous recombination. *Mol Microbiol* 2006; **59**: 376–385.

523 43. Beauchamp JM, Leveque RM, Dawid S, DiRita VJ. Methylation-dependent DNA discrimination in
524 natural transformation of *Campylobacter jejuni*. *Proc Natl Acad Sci USA* 2017; **114**: E8053–E8061.

525 44. Fu H, Uchimiya M, Gore J, Moran MA. Ecological drivers of bacterial community assembly in
526 synthetic phycospheres. *Proc Natl Acad Sci USA* 2020; **117**: 3656–3662.

527 45. Zuñiga C, Li C-T, Yu G, Al-Bassam MM, Li T, Jiang L, et al. Environmental stimuli drive a transition
528 from cooperation to competition in synthetic phototrophic communities. *Nat Microbiol* 2019; **4**:
529 2184–2191.

530 46. Bergkessel M, Basta DW, Newman DK. The physiology of growth arrest: uniting molecular and
531 environmental microbiology. *Nat Rev Microbiol* 2016; **14**: 549–562.

532 47. Basta DW, Angeles-Albores D, Spero MA, Ciemniecki JA, Newman DK. Heat-shock proteases
533 promote survival of *Pseudomonas aeruginosa* during growth arrest. *Proc Natl Acad Sci USA* 2020;
534 **117**: 4358–4367.

535 48. Ni B, Colin R, Link H, Endres RG, Sourjik V. Growth-rate dependent resource investment in bacterial
536 motile behavior quantitatively follows potential benefit of chemotaxis. *Proc Natl Acad Sci USA* 2020;
537 **117**: 595–601.

538 49. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, et al. Recurring patterns in
539 bacterioplankton dynamics during coastal spring algae blooms. *eLife* 2016; **5**: e11888.
540 50. Ratzke C, Gore J. Modifying and reacting to the environmental pH can drive bacterial interactions.
541 *PLoS Biol* 2018; **16**: e2004248.
542 51. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity
543 in microbial community assembly. *Science* 2018; **361**: 469–474.
544 52. Mitri S, Clarke E, Foster KR. Resource limitation drives spatial organization in microbial groups.
545 *ISME J* 2016; **10**: 1471–1482.
546 53. Wang C, Liu D, Bai E. Decreasing soil microbial diversity is associated with decreasing microbial
547 biomass under nitrogen addition. *Soil Biol Biochem* 2018; **120**: 126–133.
548 54. Interlandi SJ, Kilham SS. Limiting resources and the regulation of diversity in phytoplankton
549 communities. *Ecology* 2001; **82**: 1270–1282.
550 55. Harpole WS, Sullivan LL, Lind EM, Firn J, Adler PB, Borer ET, et al. Addition of multiple limiting
551 resources reduces grassland diversity. *Nature* 2016; **537**: 93–96.
552 56. Hutchinson GE. Concluding Remarks. *Cold Spring Harb Symp Quant Biol.* Yale University, New
553 Haven, CT 1957, pp 415–427.
554 57. Baker BJ, Banfield JF. Microbial communities in acid mine drainage. *FEMS Microbiol Ecol* 2003; **44**:
555 139–152.
556 58. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut
557 microbiome with increased capacity for energy harvest. *Nature* 2006; **444**: 1027–1031.
558

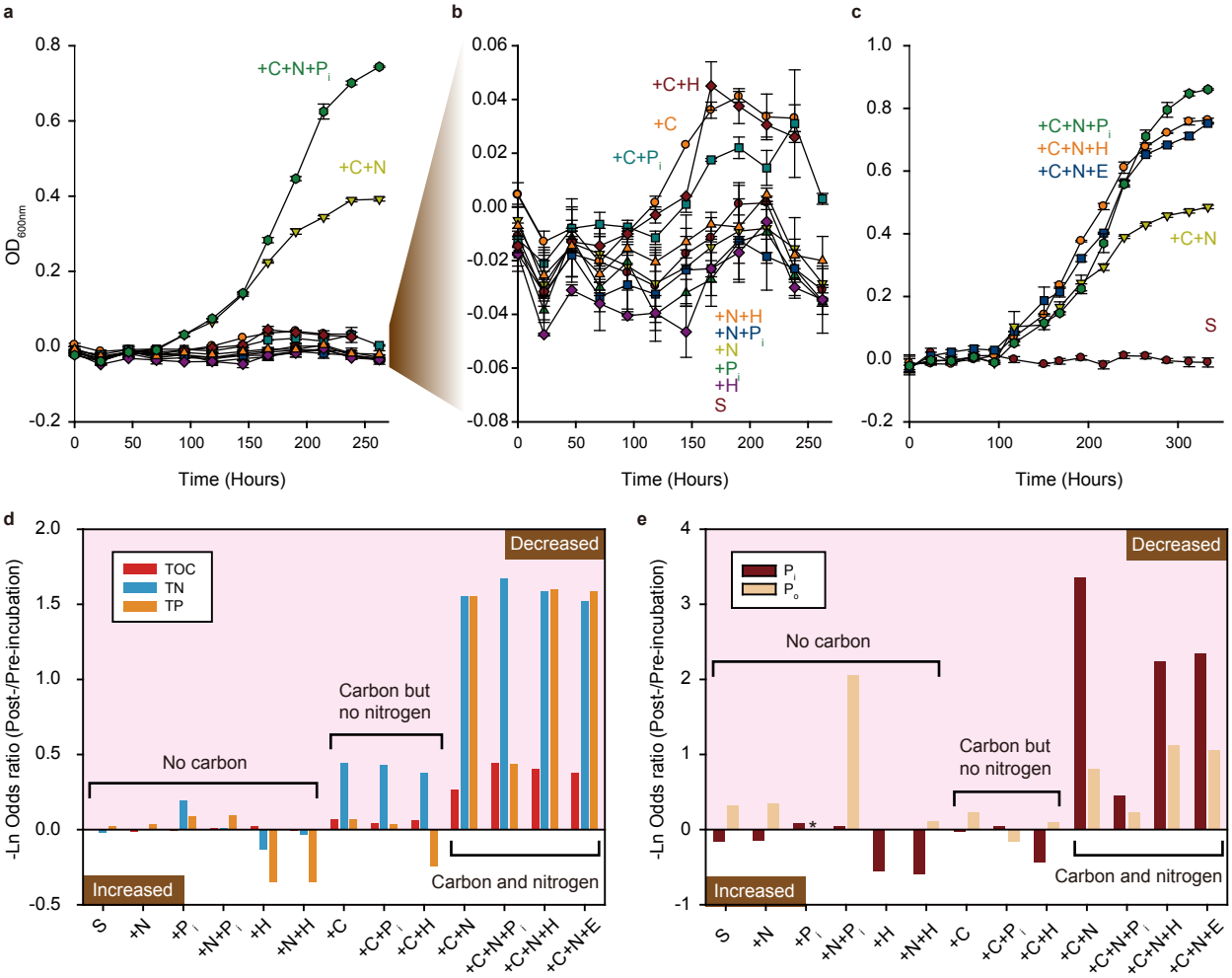

Figure 1. Microbial growth and associated changes in available nutrients in the experimental microcosms. (a-c) Growth curves under experimental treatments. Each line represents a separate experimental treatment. For each treatment, the added nutrients are denoted as "+" followed by a nutrient symbol of carbon (C), nitrogen (N), phosphorus (P_i), *Hfx. volcanii* DS2 DNA (H) and *E. coli dam-/dcm-* DNA (E). Starvation growth curve (S) serves as a control. Error bars represent the standard error of the mean. **Panel b** shows growth curves of the slow-growing microcosms from **panel a** plotted on a different scale. **(d, e)** Change in carbon, nitrogen and phosphorus after the microbial growth has stopped. The plotted changes are mean values from two experimental replicates. The experimental treatments (X axis) are grouped into three categories based on whether carbon and/or nitrogen were provided. Nutrient abbreviations: TOC, total organic carbon; TN, total nitrogen; TP, total phosphorus; P_i, inorganic phosphorus. Organic phosphorus (P_o) was computed as by subtracting P_i from TP. For "+P_i" treatment, the calculated value of P_o is replaced with an asterisk, due to errors associated with measurements of TP and P_i for that treatment. Actual values for the measured nutrient concentrations and the associated calculations are shown in **Supplementary Tables S1** and **S2**.

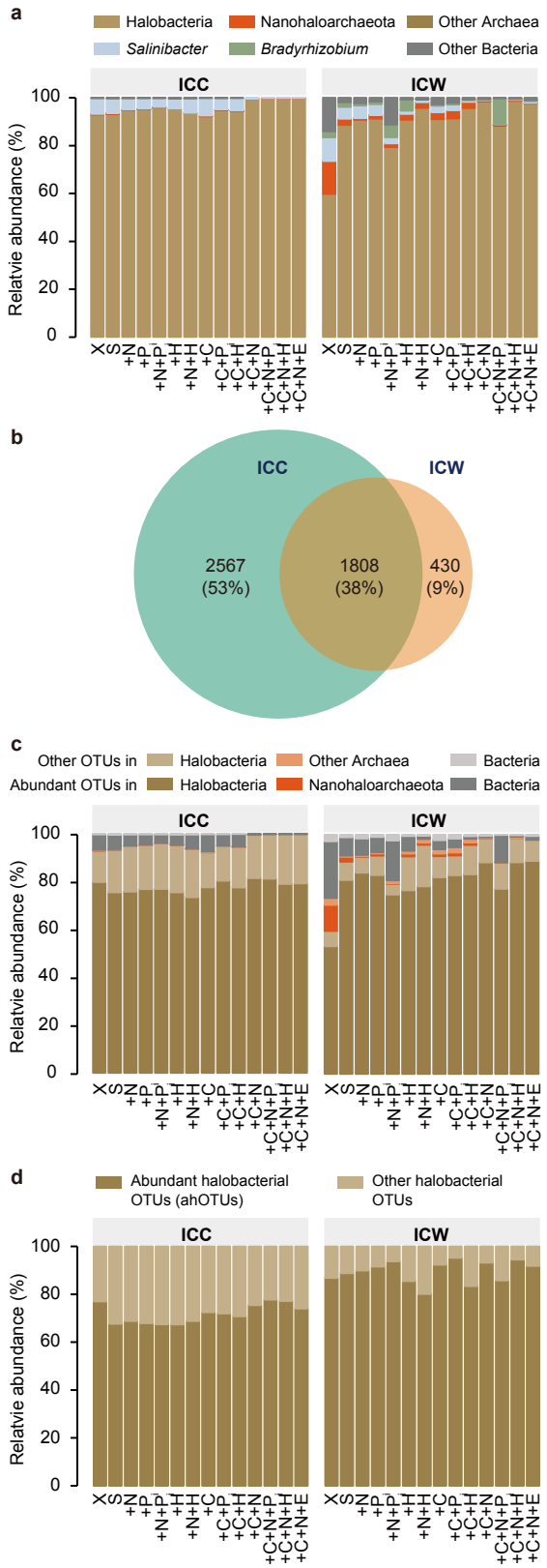

Figure 2. OTU composition of the microbial communities (ICC) and of eDNA in the associated water columns (ICW). (a) Relative abundance of 16S rRNA-based OTUs in the pre-incubation community (X), after starvation (S), and after nutrient-addition experiments. For experimental treatment abbreviations see **Figure 1** legend. Only a few selected taxonomic groups are highlighted, while other OTUs are pooled into "Other Archaea" and "Other Bacteria" categories. **(b)** Overlap of 16S rRNA-based OTUs in ICC and ICW across all samples combined. **(c)** Relative abundances of 16S rRNA-based OTUs that constitute ≥ 1% of at least one sample (designated as "abundant OTUs") in comparison to the OTUs with < 1% abundance (denoted as "Other OTUs"). **(d)** Relative abundances of *rpoB*-based OTUs from class Halobacteria that constitute ≥ 1% of at least one sample (designated as "ahOTUs") in comparison to the Halobacterial OTUs with < 1% abundance (denoted as "Other Halobacterial OTUs").

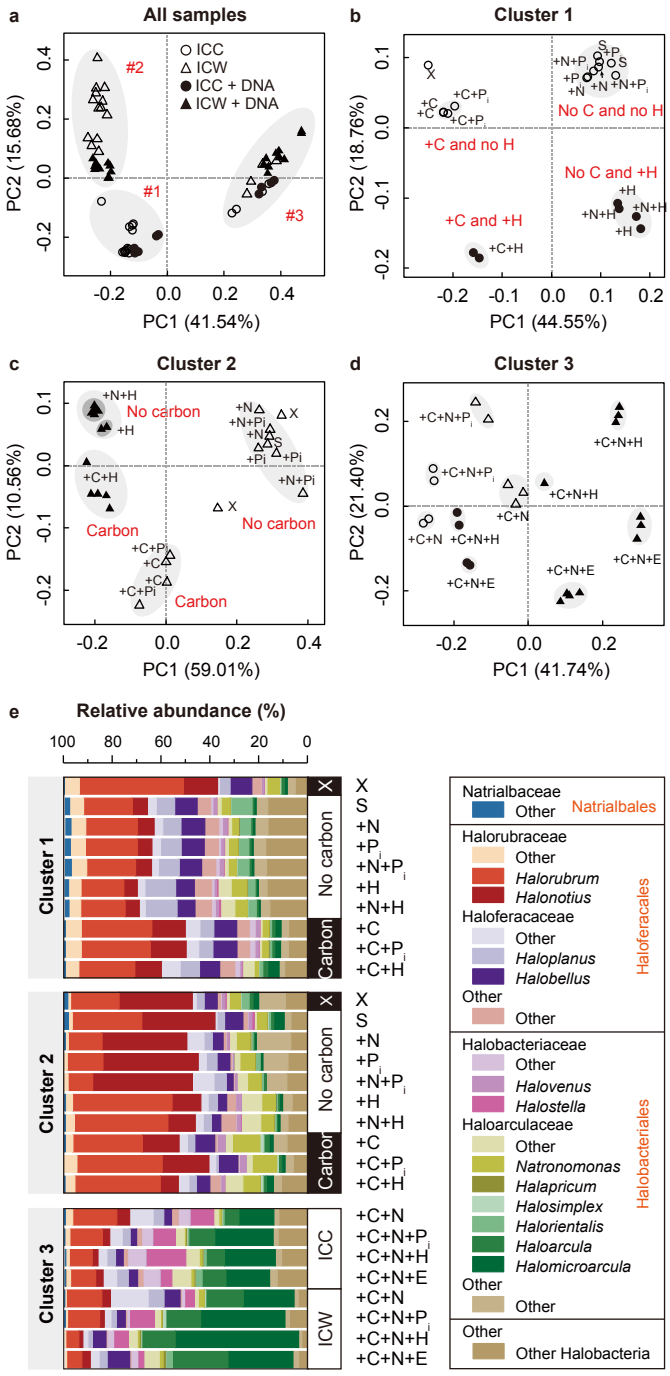

Figure 3. Comparison of the rpoB-based OTU compositions of the analyzed samples. (a) Principal Coordinate Analysis (PCoA) of all samples using the pairwise Bray-Curtis dissimilarity as a distance metric. The distances were calculated using all OTUs in a sample. Circles denote microbial community samples, while triangles refer to the water column samples. Within these, samples for treatments with added DNA are shown using filled symbols, while all remaining samples are shown using open symbols. For the PCoA analyses carried out using 16S rRNA-based OTUs, see **Supplementary Figure S2**. **(b, c** and **d)** PCoA of the samples within each of the three clusters. Same distance measure and same symbol notations as in panel a. **(e)** Taxonomic composition of samples within each of the three clusters. See **Supplementary Table S4** for the actual relative abundance values.

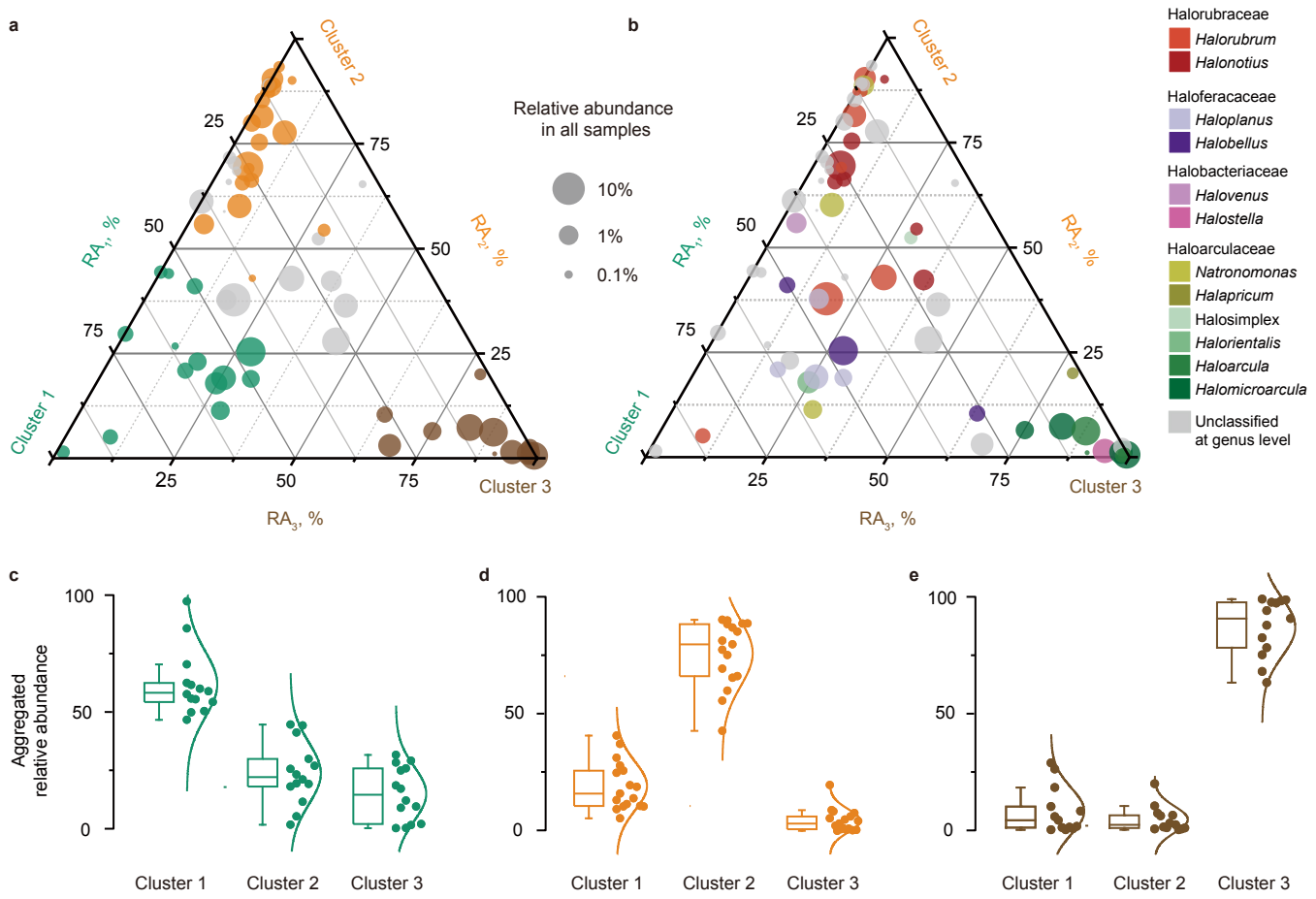

Figure 4. Relative abundance of ahOTUs across three clusters. (**a** and **b**) Relative abundance of ahOTUs in three clusters summarized in a ternary plot. Each vertex represents one of the three clusters. Each circle represents an ahOTU. The position of a circle is determined by relative abundance (RA) of the ahOTU in three clusters, and the circle size is proportional to the ahOTU's average relative abundance across all 69 samples. In **panel a**, ahOTUs with significantly higher abundance in Cluster 1, 2 or 3 than in the other two clusters are colored in green, orange and brown, respectively, while OTUs without significant difference in abundance are shown in grey. In **panel b**, the circles are colored according to the taxonomic assignment of the ahOTUs. (**c, d** and **e**) Aggregated relative abundances (aRA) of ahOTUs with significantly higher abundance in one cluster across each of the three clusters. Each point represents an ahOTU. For each OTU its aRA in a cluster *i* (*i* = 1,2,3) is defined as $aRA_i = RA_i / (\sum_{(j=1)}^{3} (RA_j))$, where RA_j is the relative abundance of the ahOTU in the cluster *i*. The distribution of aRAs within a cluster is summarized by a probability density function and a box-and-whisker plot, with whiskers extending to 1.5 of the interquartile range.

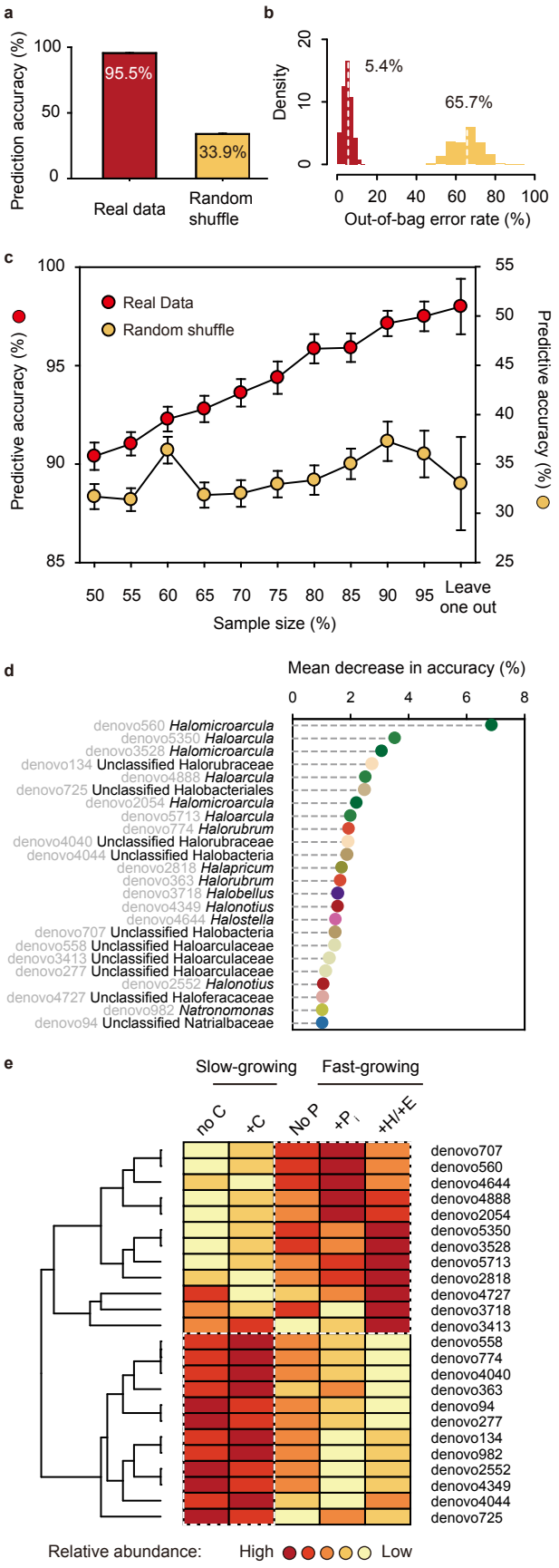

Figure 5. Performance of a random forest classifier in predicting available nutrients from the OTU composition of a microbial community. (**a**) Prediction accuracy in the 20-fold cross-validation of the classifier across 1000 training/validation datasets constructed from real and randomly reshuffled data. (**b**) Distribution of the out-of-bag (OOB) errors across 1000 training/validation datasets constructed from real and randomly reshuffled data. (**c**) Changes in prediction accuracy with the dataset size, which is shown as percent of the 69 samples used and as the leave-one-out strategy. (**d**) The decrease in prediction accuracy of 24 ahOTUs with at least 1% contribution to the predicted accuracy. The OTUs are colored according to notations shown in **Figure 3**. (**e**) The relative abundance of the top 24 ahOTUs in five categories of treatments. For each OTU, the relative abundance values were transformed into heatmap colors. The OTUs are clustered based on their relative abundance across five categories (as described in **Methods**). Abbreviations of five treatment categories: "+C", C but no N were provided; "no C", No C was provided; "+H/+E", C, N and DNA was provided; "+Pi", C, N and Pi were provided; "No P", only C and N were provided.