**Experimental manipulation of selfish genetic elements links genes to microbial community function**

Steven D. Quistad[*1], Guilhem Doulcier[1] and Paul B. Rainey[*1,2]

[1]Laboratoire de Génétique de l'Evolution, Ecole Superieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI Paris Tech), CNRS UMR 8231, PSL Research University, 75231 Paris, France.

[2]Department of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön 24306, Germany.

[*]Corresponding authors: steven.quistad@gmail.com and rainey@evolbio.mpg.de

**Abstract**

Microbial communities underpin earth's biological and geochemical processes, but their complexity hampers understanding. Here we draw upon predictions from the theory concerning selfish genetic elements (SGEs), combined with approaches from experimental evolution, comparative metagenomics and biochemistry, and show how naturally occurring SGEs can be used to manipulate genes that underpin community function. Communities comprising hundreds of bacterial genera established from garden compost were propagated with bi-weekly transfer for one year in nitrogen-limited minimal medium with cellulose (paper) as sole carbon source. At each transfer, SGEs from one set of independent communities were collected, pooled and redistributed among this same set of communities (horizontal treatment). A control was included in which SGEs never moved beyond the community in which they were originally present (vertical treatment). SGEs along with genes of ecological relevance were rapidly amplified across horizontal communities and their dynamics tracked. Enrichment of genes implicated in nitrogen metabolism, and particularly ammonification, prompted biochemical assays that showed a significant functional difference between communities subject to horizontal versus vertical treatments. This simple experimental strategy offers a powerful new approach for unravelling dynamical processes underpinning microbial community function.

## Introduction

Microbial communities underpin all major biological and biogeochemical processes (Widder et al. 2016). Appropriate functioning of communities has direct implications for health of people (Knight et al. 2017; Gilbert et al. 2018), terrestrial (Tyson et al. 2004; Allison and Martiny 2008), marine (Fuhrman et al. 2015) and freshwater (Anantharaman et al. 2016) ecosystems and even global climate (Bardgett et al. 2008). While the species composition of many communities have been exhaustively documented (Venter et al. 2004; Quince et al. 2017), the link between genes and community function is poorly understood (Widder et al. 2016).

New strategies for investigation are required that are process-focused (Koskella et al. 2017). Ideally such approaches will provide knowledge on the complex interconnections between community members, their combined effects, including feedbacks that shape the evolution of community members (Hug and Co 2018). Recent advances draw upon new strategies for linking patterns of sequence diversity in metagenomic data sets to population genetic processes (Truong et al. 2017; Garud et al. 2019). Knowledge of such processes provides a link between variation and the likelihood that particular traits fix.

Equally desirable are approaches that provide a link between genes, their spatial and temporal dynamics, and community function, achieved using strategies that leave as far as possible the natural complexity of communities undisturbed. Here we show this can be achieved via an experiment in which we combine theory governing the behaviour of selfish genetic elements (Doolittle and Sapienza 1980; Orgel and Crick 1980; Bergstrom et al. 2000; Burt and Trivers 2006) (SGEs) with experimental evolution (Elena and Lenski 2003; Rainey et al. 2017), comparative metagenomics (Maltez Thomas et al. 2018) and functional assay.

## Results

### *Manipulating the fate of SGEs*

SGEs including bacteriophages and plasmids are abundant in microbial communities (Aminov 2011; Hall et al. 2017). Typically capable of autonomous replication — or able to exploit mobilisation genes provided in *trans* — SGEs can disseminate across diverse hosts with

50 movement often coincidental with genes of ecological significance to which, by chance, they
51 become linked. Long-term persistence requires occasional exposure to new hosts: frequent
52 exposure to new hosts is expected to drive rapid spread and evolution of SGEs and linked genes
53 (Burt and Trivers 2006).
54
55 Recognising that SGEs are naturally adept at mobilising not only themselves, but also
56 ecologically significant genes, and that their activity is fuelled by exposure to new hosts, we
57 reasoned that promoting the activity of SGEs would bring about changes in the abundance and
58 distribution of the genetic determinants of fitness, allow tracking of dynamics and even
59 measurement of impacts on community function.
60
61 A set of ten independent experimental communities were established from 1 g samples of kitchen
62 / garden compost and propagated with bi-weekly transfer in static nitrogen-limited minimal M9
63 medium with cellulose (paper) as the sole carbon source. After one month, each of the ten
64 communities was used to found two new communities, giving a total of 20 paired communities.
65 One of each pair was assigned to the "horizontal regime", while the remainder were assigned to
66 the "vertical regime" (Fig. 1). The terms *horizontal* (H) and *vertical* (V) indicate the imposed
67 transfer regime: the former is expected to enhance the evolutionary success of SGEs, the latter to
68 castrate.
69
70 From the moment of the initial split (time point $T_0$) and for the ensuing 48 weeks (24 transfers
71 ($T_{24}$)) 1 ml of each community, including the slurry of any remaining paper, was transferred to a
72 fresh mesocosm containing 19 ml of M9 medium plus fresh paper. This ensured serial passage
73 of microbes plus SGEs from each mesocosm. Half of these mesocosms (one of each pair),
74 labelled vertical communities (VCs), received no further treatment other than continued serial
75 passage on a bi-weekly basis.
76
77 The remaining ten horizontal communities (HCs) received — at the time of serial transfer — an
78 aliquot of a "SGE cocktail" that was a pooled sample of SGEs derived from each horizontal
79 mesocosm (Fig. 1). This was prepared by taking a sample from from each of the ten
80 communities, followed by centrifugation and passage of the supernatant through a 0.22 μm filter.

81 After pooling the filtrate from each of the ten horizontal treatments a 1ml aliquot of the mixture
82 was placed in each HC.
83
84 *Detection of SGE activity*
85 To determine whether the horizontal regime promoted movement of SGEs, total DNA was
86 extracted from each $T_0$ community and each of the 20 $T_1$ communities (two weeks after the split
87 of communities into vertical and horizontal treatments) and sequenced (Fig. 1 and
88 Supplementary Table 1). Analysis was straightforward given the paired experimental design.
89 For each set of $T_0$ and descendant HCs and VCs (e.g., $T_0$-1 and descendants $T_1$V-1 and $T_1$H-1)
90 DNA sequence data were interrogated to identify reads found solely in (and therefore unique to)
91 the horizontal mesocosm (e.g., $T_1$H-1). Such unique reads are likely to stem from SGEs, or be
92 associated with SGE activity, originating from an allopatric community.
93
94 Analysis of the sequences unique to each horizontal community resulted in identification of ~26
95 million unique reads from a total of ~152 million reads across all ten communities, with an
96 average of ~2.6 million unique reads per community (Supplementary Table 2). Assembly of
97 reads into contigs revealed an average of 3,352 contigs per HC greater than 1 kb in length. The
98 mean maximum contig size per assembly was 82 kb (Supplementary Table 2).
99
100 After extracting Open Reading Frames (ORFs), interrogation of the Conserved Domain Database
101 (CDD) (Marchler-Bauer et al. 2015) showed 1,279 ORFs predicted to encode phage-associated
102 proteins involved in capsid formation, baseplate assembly, and phage-induced lysis of bacterial
103 cells (Supplementary Table 3). An example of a contig ($T_1$H-1_35969) containing numerous
104 genes characteristic of phages is shown in Supplementary Fig. 1a.
105
106 The distribution and abundance of contig $T_1$H-1_35969 and 19 other representative phage-like
107 entities (Supplementary Table 3) across independent mesocosms was determined by mapping
108 total reads from each horizontal and vertical metagenome onto each of the 20 phage-like contigs
109 (Fig. 2). The predominance of these contigs in HCs is evidence of rapid amplification and
110 dissemination of genetic material. $T_1$H-1_35969, originally assembled from the unique set of
111 sequences from $T_1$H-1, is also present in both $T_1$H-2 and $T_1$V-2 and was therefore likely present

112 in the founding community ($T_0$-2). Within two weeks this contig has amplified and spread to
113 horizontal communities 1, 5 and 6 (Fig. 2). Contig T1H2_39307 is below the level of detection
114 in all vertical treatments, but two weeks after imposition of the horizontal treatment, this contig
115 was detected in seven of the independent HCs (Fig. 2, Supplementary Table 3). From the
116 community perspective, less than four of the 20 phage-like contigs were detected on average per
117 VC, whereas the mean number of contigs per HCs was 11 (Fig. 2).
118
119 In addition to phage-like contigs, assemblies of the unique sequences from HCs contained genes
120 predicted to encode traits of likely ecological significance (Supplementary Fig. 2). These
121 include genes implicated in cellulose degradation, defence against SGE invasion and nitrogen
122 metabolism (Supplementary Table 4) (Seed et al. 2013; Ramisetty and Santhosh 2016;
123 Bustamante and Iredell 2017). Particularly notable is enrichment of reads mapping to *β*-
124 glucosidase genes that define the rate-limiting step in cellulose degradation (Singhania et al.
125 2013). On occasion these ecological significant genes were linked to genes encoding features
126 found on SGEs (Supplementary Table 5), but often these contigs lacked such features
127 (Supplementary Fig. 1b). Nonetheless they were amplified and disseminated through the course
128 of the selection experiment with the same dynamic as expected of a mobile genetic element (see
129 below).
130
131 *Diversity and dynamics*
132 Given evidence indicative of the amplification and dissemination of DNA via the horizontal
133 treatment, combined with ability to detect such sequences, communities were allowed to evolve
134 for a total of 48 weeks with, at each transfer, the HCs experiencing exposure to a new cocktail of
135 SGEs. Continuance was in part motivated by the expectation that movement of DNA —
136 particularly that encoding genes of ecological relevance — mediated by SGEs would result in a
137 measurable change in the function of HCs relative to VCs. To complement existing DNA
138 sequence from $T_0$ and $T_1$, DNA was extracted from all HCs and VCs at time points $T_2$, $T_3$, $T_4$,
139 $T_{10}$, $T_{16}$, $T_{20}$ and $T_{24}$ yielding a total of 180 metagenomes (5.2 billion reads, each ~150 bp).
140
141 The extraordinary flux of DNA detected within two weeks of establishing the horizontal
142 treatment regime likely relies upon a minimum threshold of community-level diversity. Given

143 propagation of communities on minimal medium with a single recalcitrant (cellulose) carbon
144 source, we firstly interrogated the metagenomic data for sequences mapping to 16S rDNA to
145 determine whether diversity was maintained (Supplementary Table 6). Communities at $T_0$
146 harboured on average ~140 bacterial genera (SD $\pm$ 23). At $T_{24}$ the number of genera detected
147 increased in all communities to an average of ~200 (SD $\pm$ 32), indicating an increase in
148 abundance of rare types (Supplementary Table 6, Supplementary Fig. 3a). No significant
149 difference was observed between HCs and VCs (Supplementary Fig. 3a-c). Rank abundance
150 distributions from pooled HCs and VCs at T1 and T24 are shown in Fig. 3 and reveal the tail of
151 the $T_{24}$ distribution to be markedly flatter in both cases, consistent with an overall tendency for
152 rare types to increase during the course of the year long selection. Rank abundance distributions
153 for each $T_1$ and $T_{24}$ HC and VC, plus the rank order (and change therein) of the most common
154 genera are shown in Supplementary Figs 4a and 4b. Overall communities differed markedly in
155 genera composition. Among the most abundant genera, only two harbour species commonly
156 associated with cellulolytic ability (Wilson 2011) (*Cytophaga* and *Cellvibrio*), while others are
157 better known for roles in ammonification (Simon 2002) (*Azospirillum* and *Paenibacillus*
158 (nitrogen fixation); *Rhodanobacter* and *Pseudomonas* (dissimilatory nitrate reduction /
159 denitrification)) (Supplementary Figs 4a and 4b). Single celled eukaryotes were also present in
160 all mesocosms with several even maintaining populations of nematodes through the 48-week
161 selection period following the horizontal / vertical split.
162
163 The 180 metagenomic data sets (nine time points, 20 communities per time point
164 (Supplementary Table 1)) spread across 48 weeks allowed the abundance and distribution of
165 SGEs and associate genes to be determined by read mapping. The spatial and temporal
166 dynamics of two phage-like contigs and two contigs encoding *β*-glucosidases are shown in Fig.
167 4a-d.
168
169 **Ecologically significant genes**
170 To determine whether the horizontal transfer regime affected the abundance and distribution of
171 functional genes, the MG-RAST database was interrogated with sequence reads from all $T_0$ and
172 all $T_{24}$ HCs and VCs (Supplementary Table 7). The relative abundance of reads assigned to 13
173 of 28 functional categories at Subsystems Level 1 (Overbeek et al. 2005), including genes

174 involved in virulence, motility and nitrogen metabolism, changed significantly during the course
175 of the year (Supplementary Fig. 5a). HCs and VCs did not always change to the same extent, or
176 in same direction (Supplementary Figs 5a and 5b). No significant change occurred in the overall
177 category of carbohydrate metabolism (Supplementary Fig. 5b), but a significant increase in *β*-
178 glucosidase metabolism — the rate limiting step in cellulose degradation — was detected in both
179 HCs and VCs (Supplementary Fig. 5a)
180
181 The nitrogen-limited nature of the culture medium, combined with evidence of selection
182 favouring genes involved in nitrogen metabolism, led to closer focus on this essential resource.
183 The MG-RAST database was again interrogated, but this time at Subsystems Level 2 (Overbeek
184 et al. 2005), which provides information on functional categories within the broader category of
185 nitrogen metabolism (Fig. 5a). Significant changes were observed in the relative abundance of
186 reads mapping to genes involved in both ammonia assimilation and dissimilation
187 (ammonification). Such changes are consistent with the nature of the selective environment.
188 This prompted a query of the extent to which these changes may have been influenced by
189 movement of genes via SGEs.
190
191 A modified version of the bioinformatic pipeline developed to identify sequences unique to HCs
192 (Methods) was applied to the $T_{24}$ metagenomes. $T_{24}$ HC metagenomes were compared to their
193 paired $T_{24}$ VC metagenomes as well as to all metagenomes from VCs from earlier time points
194 (Supplementary Fig. 6a, Supplementary Table 8, Supplementary Methods Fig. 1). This showed a
195 greater number of unique sequences in the HCs, but also showed a fraction of unique sequences
196 present in VCs. The latter is to be expected given that $T_0$ communities must contain rare
197 sequences that through the course of the selection experiment became common — and is
198 consistent with data on changes in abundance of genera (Fig. 3). To see whether it was possible
199 to reduce this signal, genomic DNA samples from $T_0$ were sequenced on the HiSeq platform
200 resulting in additional 320 million reads per community. The effect of increasing depth of
201 sequence made minimal impact on detection of unique sequences (Supplementary Fig. 6b).
202 Nonetheless, the "deep" $T_0$ metagenomes were used in all subsequent analyses.
203

204 To determine the fraction of unique reads that mapped to genes involved in nitrogen metabolism,
205 the final sets of unique reads were obtained from HCs and VCs; they were assembled, ORFs
206 identified and functionally categorised using the CDD. HCs contained significantly more unique
207 ORFs predicted to be involved in nitrogen metabolism compared to VCs (Supplementary Fig. 7).
208 HCs were enriched in several functional classes of gene, especially those with predicted roles in
209 regulation and ammonification (Fig. 5b, Supplementary Table 9).
210
211 To link to function, we asked whether there was evidence of an effect of the horizontal treatment
212 on a measurable community property. To this end, and with focus on the process of
213 ammonification (production of ammonia through either fixation, or reduction of nitrate / nitrite)
214 the concentration of nitrate, nitrite, and ammonia were determined in the $T_{24}$ HCs and VCs at the
215 end of the two-week period immediately prior to serial passage. The concentration of ammonia
216 was significantly higher in HCs (Fig. 6a, Supplementary Table 10). To check the robustness of
217 this finding the concentration of ammonia was measure during the course of a two-week period.
218 Despite substantive variability in composition of the independent communities (Supplementary
219 Fig. 4a and 4b), a significantly greater concentration of ammonia was detected in HCs at each
220 one of nine sampling occasions throughout the two week period (Fig. 6b, Supplementary Table
221 11). A comparison between T0 communities and T24 VCs showed no change (Supplementary
222 Fig. 8) supporting the conclusion that movement of ecologically significant genes via SGEs was
223 responsible for the functional difference between horizontal and vertical communities.
224
225 **Discussion**
226 Microbial communities are dynamical systems defined by complex ecological interactions and
227 evolutionary feedbacks (Widder et al. 2016). The challenges associated with understanding this
228 complexity are immense. Here we have demonstrated the utility of a new approach in which we
229 have used naturally occurring SGEs to manipulate the abundance and distribution of genes of
230 ecological significance, shown that the dynamics of these genes can be followed, and established
231 that the magnitude of effect is sufficient to allow links to community function to be made.
232 Central to the strategy is a simple manipulation that promotes the evolutionary success of SGEs.
233 That this involves nothing more than frequent exposure of SGEs to new hosts, combined with
234 routine metagenomic analysis, means that it is readily transferable to a wide range of microbial

235 communities, from experimental constructs as here, through microbiomes of plants and animals,
236 to naturally occurring marine and terrestrial communities.
237
238 The study delivered numerous surprises. Among these, ability to maintain exceptionally diverse
239 communities through the course of a year-long selection experiment was particularly notable.
240 The transfer regime, involving passage of 1/20th of the two-week culture into fresh medium,
241 meant possibility for resource partitioning and cross-feeding, but to witness diversity in all
242 replicate mesocosms increase over the course of the experiment was unexpected (Rainey et al.
243 2000). Although barely touched in our analysis so far, cycles for replenishing all major nutrients
244 were established within each mesocosm, with that involving nitrogen being most evident. Each
245 community harboured a single dominant free-living nitrogen-fixing bacterium, but the identity
246 — even at the genus level — differed between mesocosms. Attempts to establish such complex
247 communities by choosing key players before the fact, would likely fail. That such diversity and
248 stability can arise from a single gram of compost speaks to an abundance of redundancy in these
249 natural systems (Rosenfeld 2002; Louca et al. 2018).
250
251 A second surprise was the speed and magnitude of the response of communities to the mixing
252 regime. Within just two weeks of implementation, ~17 % of DNA sequence reads within HCs
253 were derived from allopatric communities. That these reads could be assembled into contigs
254 containing genes of ecological relevance, including those involved in cellulose degradation,
255 emphasises the dynamism of processes driven by SGEs. Ability to so clearly detect movement
256 of this DNA via comparative metagenomic approaches was an additional, but critical, bonus.
257
258 The association established between genes involved in nitrogen metabolism and enhanced
259 activity of SGEs in HCs is correlative, but is nonetheless linked to horizontally transferred
260 sequences that map to genes with predicted roles in ammonification. The association is further
261 linked to data that demonstrate a significant effect of the experimental treatment on
262 concentrations of ammonia in HCs (Fig. 6), but no effect (compared to $T_0$ communities) in VCs
263 (Supplementary Fig. 8). Additionally, increased ammonia in the HCs is consistent with the
264 prediction that horizontal movement of DNA will increase the rate at which community function
265 improves.

266

267 The two processes that generate ammonia: nitrogen fixation and nitrate ammonification, both
268 require environments devoid of oxygen (or mostly so in the case of nitrate ammonification)
269 (Simon 2002; Lam and Kuypers 2011). It is possible that enhanced metabolic activity associated
270 with cellulose degradation causes lower oxygen conditions at the paper surface (where
271 community members grow as biofilms) compared to the VCs, leading to enhanced production of
272 ammonia. Intriguingly, the cattle industry promotes addition of ammonia to feed because of
273 beneficial effects on digestibility of plant matter in the rumen (Givens et al. 1988). This raises
274 an alternate possibility, which is that elevated ammonia levels are similarly beneficial to
275 digestion of cellulose in the experimental mesocosms, and reflect a more rapid response to
276 selection in the HCs.
277

278 The genomic era has done much to resolve controversy surrounding the importance of horizontal
279 gene transfer as a driver of evolutionary change (Goldenfeld and Woese 2007; Koonin 2009;
280 Boto 2010). Direct evidence of gene transfer events — many facilitated by SGEs — from one
281 organism to another are common (Ochman et al. 2000; Werren 2011), but the possibility that the
282 process assumes far greater evolutionary significance when operating within communities, has,
283 to the best of our knowledge, been little considered, let alone directly studied. That the vast
284 diversity of DNA sequence encompassed within complex microbial communities might exist in
285 fluid association with SGEs generating permutations of effects significantly beyond those
286 observed through study of individual SGEs and their hosts, is both interesting and plausible.
287

288 **Methods**

289 **Establishment of cellulose-degrading microbial communities**

290 Microbial communities were initially established by placing ten independent 1 g samples of fresh
291 compost into 20 mL of M9 minimal media supplemented with a 4 $cm^2$ piece of cellulose paper
292 (Whatman cellulose filter paper). All incubations were performed in 140 mL sterile glass bottles
293 with loosened screw caps allowing for gas exchange between the community and the
294 environment. Incubation was at room temperature without shaking. Compost was sampled from
295 the Square Theodore-Monod compost heap (Paris, France) in February 2016. During the primary
296 incubation period of 2 weeks the piece of cellulose was suspended by a wire from the screw cap

297 into the media allowing for paper colonization. After two weeks, the community was transferred
298 by moving the wire-suspended cellulose paper into 20 mL of fresh M9 minimal media that also
299 contained a new 4 $cm^2$ piece of cellulose paper in suspension. Two more weeks of incubation
300 provided the opportunity for colonization of the new piece of suspended cellulose paper thus
301 eliminating the need for a wire-suspended community for each transfer. All community transfers
302 were performed by vortexing each bottle at maximum speed until the cellulose fibres were
303 dissolved into a slurry.
304
305 **Horizontal and vertical transfer regimes**
306 Vertical and horizontal transfers were performed at the exact same time every two weeks. Before
307 each transfer aliquots of the community were taken for glycerol stocks, DNA extraction, and
308 phage storage. Glycerol stocks were created by mixing 500 µl of cellulose-microbial slurry with
309 500 µl of 80% glycerol and stored at -80 ºC. For DNA extractions 2 mL of cellulose-microbial
310 slurry was centrifuged at 13,000 g for 10 min and the pellet was stored at -80 ºC for later
311 processing. DNA was extracted from the founding $T_0$ communities (before the horizontal /
312 vertical split) as well as transfers 1, 2, 3, 4, 10, 16, 20, 24 ($T_1$, $T_2$, $T_3$, $T_4$, $T_{10}$, $T_{16}$, $T_{20}$, $T_{24}$) using
313 the soil DNA extraction kit (Norgen Biotek). DNA sequencing was performed using the MiSeq
314 and NextSeq platforms for all time points and $T_0$ communities with additional deep sequencing
315 of the $T_0$ communities using the HiSeq platform.
316
317 Transfers at two-week intervals were performed as described in Fig. 1. The "SGE-cocktail" was
318 prepared by mixing 1 mL of homogenized community (n=10) followed by centrifugation and
319 passage through a 0.22 µm filter to remove microbial cells. Recognising the possibility that ultra
320 micro-bacteria might escape sedimentation and pass through the filter, metagenomic sequence
321 reads were recruited to seven available reference genomes. A maximum of 0.018 % of reads
322 across all horizontal and vertical mesocosms at 48 weeks matched these reference genomes
323 (Supplementary Fig. 9).
324
325 **Bioinformatic analysis**
326 DNA sequences were demultiplexed using bcl2fastq, paired ends were joined using FLASh
327 (Magoc and Salzberg 2011) and initial fastq files were generated. Preprocessing of fastq files

328 was performed using PrinSeq (Schmieder and Edwards 2011) with a minimum length of 100
329 base pairs, minimum quality score of 25, and a maximum percentage of N's of 10%. All
330 metagenomes were uploaded to the MG-RAST metagenomic analysis server and are publicly
331 available (Supplementary Table 1). Gene category abundances (Supplementary Fig. 4) and genus
332 identifications (Supplementary Table 6) were determined using the MG-RAST pipeline (Wilke
333 et al. 2016). Sequence matches were determined using BLASTn (Altschul et al. 1990) with a
334 minimum e-value threshold of 1E-05, minimum alignment length of 100 base pairs, and a
335 minimum percentage identity of 90%. "Unique" Horizontal and Vertical sequences were
336 assembled using MEGAHIT (Li et al. 2015), ORFs with a minimum length of 100 amino acids
337 were extracted using getORF (Rice et al. 2000), compared to the Conserved Domain Database
338 (CDD) (Marchler-Bauer et al. 2015) using RPS-BLAST, and top domain hits were extracted
339 (Supplementary Table 4). Phage-like contigs were classified based on phage-specific protein
340 domains (Supplementary Table 3) and their distribution was determined using Fr-Hit (Niu et al.
341 2011). Sequences matches to phage-like contig were based on a minimum alignment length of
342 100 base pairs with a percentage identity of at least 90%. Nitrogen-related genes were based on
343 gene function classifications and descriptions with a total of 136 domains (Supplementary Table
344 9). To identify unique sequences a custom bioinformatic pipeline was developed
345 (Supplementary Methods Fig. 1). Each Horizontal metagenome was compared to the founding $T_0$
346 metagenome (before the Horizontal vs. Vertical split) and the paired Vertical metagenome using
347 BLASTn. Horizontal-$T_0$ and Horizontal-Vertical sequence matches were removed, thus leaving a
348 set of unique Horizontal sequences that could not be identified in neither the founding $T_0$-
349 metagenome, nor the paired Vertical-metagenome. Sequence matches were determined using a
350 minimum e-value threshold of 1E-05, minimum alignment length of 100 base pairs, and a
351 minimum percent identity of 90%. The unique Horizontal sequences were then assembled, Open
352 Reading Frames (ORFs) were extracted, and predicted protein domains were determined through
353 comparison to the Conserved Domain Database (CDD). To determine the impact of increasing
354 the sequencing depth of the $T_0$ communities on the final percentage of unique sequences, $T_0$
355 metagenomes were sequenced using the HiSeq platform and reads from each mesocosm split
356 arbitrarily into eight metagenomes each consisting of about ~40 million sequences. The unique
357 analysis pipeline was then applied using the first arbitrary $T_0$ metagenome and any matching
358 sequences were removed. The process was repeated with all eight $T_0$ arbitrarily defined

359 metagenomes thus revealing the extent to which additional sequencing affected ability to detect
360 rare sequences.
361
362 **Nitrogen assays**
363 Vertical and Horizontal communities from the one-year time point (24 transfers, $T_{24}$) and $T_0$
364 were re-established by placing 100 µl of glycerol stock into 20 mL of M9 minimal media
365 supplemented with a 4 $cm^2$ piece of cellulose paper and incubated for two weeks. 1 mL of the
366 cellulose-community slurry was transferred to 19 mL of M9 Minimal Media with a fresh piece of
367 cellulose and incubated for an additional two weeks. 1 mL of the community was then
368 transferred to a new bottle and 100 µl of surrounding media was sampled at various time points
369 during the two-week incubation period to determine ammonia/ammonium, nitrate, and nitrite
370 concentrations. The concentrations of all three nitrogen species were determined using
371 fluorometric assay kits according the manufacturers protocol (Ammonia Assay Kit, Sigma-
372 Alrdrich, and nitrate/nitrite fluorometric assay kit) (Fig. 6 and Supplementary Table 11).
373
374 **Data availability**
375 All metagenomes are publicly available here: https://www.mg-
376 rast.org/mgmain.html?mgpage=project&project=mgp18485. Supplementary Table 1 contains
377 metagenome descriptions and MG-RAST identifiers
378
379 All associated datasets are publicly available here: https://zenodo.org/record/3293312 and
380 Supplementary Table 12 contains file descriptions and associated metadata
381
382 **Supplementary files**
383 Supplementary Figure 1: contigs assembled from unique sequence reads.
384
385 Supplementary Figure 2: Ecologically significant genes.
386
387 Supplementary Figure 3: Changes in diversity.
388
389 Supplementary Figure 4: Changes in patterns of diversity seen in rank abundance plots
390
391 Supplementary Figure 5: Functional gene categories.
392
393 Supplementary Figure 6: Distinguishing sequences transferred by SGEs from rare sequences.

394
395 Supplementary Figure 7: Nitrogen-related genes are enriched in "unique" horizontal sequences.
396
397 Supplementary Figure 8: No change in function of communities subject to the vertical transfer
398 regime after 48 weeks.
399
400 Supplementary Figure 9: DNA sequence reads mapping to ultra-microbacteria reference
401 genomes.
402
403 Supplementary Methods Figure 1: Bioinformatic pipeline for identifying "unique" sequences.
404
405 Supplementary Table 1: MGRAST IDs and metadata associated with metagenomes.
406
407 Supplementary Table 2: Summary of unique sequences assembled from $T_1$ Horizontal
408 mesocosms.
409
410 Supplementary Table 3: Sequences matching to phage-like contigs in Fig. 2;
411 annotation of phage-like contigs depicted in Fig. 2; all $T_1$ unique contigs containing an
412 ORF associated with bacteriophages.
413
414 Supplementary Table 4: Number of unique $T_1$ horizontal sequences containing an
415 ORF with a domain of interest depicted in Supplementary Fig. 2.
416
417 Supplementary Table 5: Full annotation of the two phage-like and two beta-
418 glucosidase contigs depicted in Fig. 4.
419
420 Supplementary Table 6: Genus abundance of all sequenced time points for all
421 mesocosms based on the MG-RAST annotation pipeline.
422
423 Supplementary Table 7: The relative abundance of functional gene categories in $T_0$,
424 $T_{24}$ vertical, and $T_{24}$ horizontal metagenomes based on the MG-RAST annotation
425 pipeline.
426
427 Supplementary Table 8: Summary of unique sequences assembled from $T_{24}$ horizontal
428 and $T_{24}$ vertical metagenomes.
429
430 Supplementary Table 9: Number of nitrogen-related ORFs assembled from unique
431 T24 horizontal sequences depicted in Fig. 5b; number of nitrogen-related ORFs
432 assembled from unique T24 horizontal sequences.
433
434 Supplementary Table 10: Nitrate, nitrite, and ammonia / ammonium concentrations of
435 $T_{24}$ horizontal and $T_{24}$ vertical mesocosms at the end of the two-week incubation
436 period depicted in Figure 6a.
437

438 Supplementary Table 11: Time series of ammonia / ammonium concentrations of $T_{24}$
439 horizontal and $T_{24}$ vertical mesocosms during the two-week incubation period.
440
441 Supplementary Table 12: Description of all associated datasets publicly available at
442 https://zenodo.org/record/3293312 .
443
444 Supplementary Table 13: P-values of statistical tests performed in study.
445

446 **References**

447


448 Allison, S. D. and J. B. Martiny. 2008. Colloquium paper: resistance, resilience, and redundancy
449 in microbial communities. Proc Natl Acad Sci U S A 105 Suppl 1:11512-11519.
450 Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment
451 search tool. J Mol Biol 215:403-410.
452 Aminov, R. I. 2011. Horizontal gene exchange in environmental microbiota. Front Microbiol
453 2:158.
454 Anantharaman, K., C. T. Brown, L. A. Hug, I. Sharon, C. J. Castelle, A. J. Probst, B. C. Thomas,
455 A. Singh, M. J. Wilkins, U. Karaoz, E. L. Brodie, K. H. Williams, S. S. Hubbard, and J. F.
456 Banfield. 2016. Thousands of microbial genomes shed light on interconnected
457 biogeochemical processes in an aquifer system. Nat. Commun. 7:13219.
458 Bardgett, R. D., C. Freeman, and N. J. Ostle. 2008. Microbial contributions to climate change
459 through carbon cycle feedbacks. ISME J 2:805-814.
460 Bergstrom, C. T., M. Lipsitch, and B. R. Levin. 2000. Natural selection, infectious transfer and
461 the existence conditions for bacterial plasmids. Genetics 155:1505-1519.
462 Boto, L. 2010. Horizontal gene transfer in evolution: facts and challenges. Proc Biol Sci
463 277:819-827.
464 Burt, A. and R. L. Trivers. 2006. Genes in Conflict: the Biology of Selfish Genetic Elements.
465 Belknap Press, Harvard, Cambridge, MA.
466 Bustamante, P. and J. R. Iredell. 2017. Carriage of type II toxin-antitoxin systems by the growing
467 group of IncX plasmids. Plasmid 91:19-27.
468 Doolittle, W. F. and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome
469 evolution. Nature 284:601-603.
470 Elena, S. F. and R. E. Lenski. 2003. Evolution experiments with microorganisms: the dynamics
471 and genetic bases of adaptation. Nat Rev Genet 4:457-469.
472 Fuhrman, J. A., J. A. Cram, and D. M. Needham. 2015. Marine microbial community dynamics
473 and their ecological interpretation. Nat Rev Microbiol 13:133-146.
474 Garud, N. R., B. H. Good, O. Hallatschek, and K. S. Pollard. 2019. Evolutionary dynamics of
475 bacteria in the gut microbiome within and across hosts. PLoS Biol 17:e3000102.
476 Gilbert, J. A., M. J. Blaser, J. G. Caporaso, J. K. Jansson, S. V. Lynch, and R. Knight. 2018.
477 Current understanding of the human microbiome. Nat Med 24:392-400.
478 Givens, D. I., A. H. Adamson, and J. M. Cobby. 1988. The effect of ammoniation on the
479 nutritive value of wheat, barley and oat straws. II. Digestibility and energy value
480 measurements in vivo and their prediction from laboratory measurements. Anim Feed Sci
481 Tech 19:173-184.
482 Goldenfeld, N. and C. Woese. 2007. Biology's next revolution. Nature 445:369.

483 Hall, J. P. J., M. A. Brockhurst, and E. Harrison. 2017. Sampling the mobile gene pool:
484 innovation via horizontal gene transfer in bacteria. Phil Trans R Soc Lond B 372.
485 Hug, L. A. and R. Co. 2018. It takes a village: Microbial communities thrive through interactions
486 and metabolic handoffs. mSystems 3:e00152-00117.
487 Knight, R., C. Callewaert, C. Marotz, E. R. Hyde, J. W. Debelius, D. McDonald, and M. L.
488 Sogin. 2017. The microbiome and human biology. Annu Rev Genomics Hum Genet 18:65-
489 86.
490 Koonin, E. V. 2009. Darwinian evolution in the light of genomics. Nucleic Acids Res 37:1011-
491 1034.
492 Koskella, B., L. J. Hall, and C. J. E. Metcalf. 2017. The microbiome beyond the horizon of
493 ecological and evolutionary theory. Nat Ecol Evol 1:1606-1615.
494 Lam, P. and M. M. Kuypers. 2011. Microbial nitrogen cycling processes in oxygen minimum
495 zones. Ann Rev Mar Sci 3:317-345.
496 Li, D., C. M. Liu, R. Luo, K. Sadakane, and T. W. Lam. 2015. MEGAHIT: an ultra-fast single-
497 node solution for large and complex metagenomics assembly via succinct de Bruijn graph.
498 Bioinformatics 31:1674-1676.
499 Louca, S., M. F. Polz, F. Mazel, M. B. N. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann,
500 A. S. Hahn, D. S. Srivastava, S. A. Crowe, M. Doebeli, and L. W. Parfrey. 2018. Function
501 and functional redundancy in microbial systems. Nat Ecol Evol 2:936-943.
502 Magoc, T. and S. L. Salzberg. 2011. FLASH: fast length adjustment of short reads to improve
503 genome assemblies. Bioinformatics 27:2957-2963.
504 Maltez Thomas, A., F. Prata Lima, L. Maria Silva Moura, A. Maria da Silva, E. Dias-Neto, and
505 J. C. Setubal. 2018. Comparative Metagenomics. Methods Mol Biol 1704:243-260.
506 Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C.
507 Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song,
508 N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, and S. H. Bryant. 2015. CDD:
509 NCBI's conserved domain database. Nucleic Acids Res. 43:D222-226.
510 Niu, B., Z. Zhu, L. Fu, S. Wu, and W. Li. 2011. FR-HIT, a very fast program to recruit
511 metagenomic reads to homologous reference genomes. Bioinformatics 27:1704-1705.
512 Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of
513 bacterial innovation. Nature 405:299-304.
514 Orgel, L. E. and F. H. C. Crick. 1980. Selfish DNA: The ultimate parasite. Nature 284:604-607.
515 Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang, M. Cohoon, V. de Crecy-
516 Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E. D. Frank, S. Gerdes, E. M. Glass, A.
517 Goesmann, A. Hanson, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, L. Krause, M. Kubal, N.
518 Larsen, B. Linke, A. C. McHardy, F. Meyer, H. Neuweger, G. Olsen, R. Olson, A.
519 Osterman, V. Portnoy, G. D. Pusch, D. A. Rodionov, C. Ruckert, J. Steiner, R. Stevens, I.
520 Thiele, O. Vassieva, Y. Ye, O. Zagnitko, and V. Vonstein. 2005. The subsystems approach
521 to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids
522 Res. 33:5691-5702.
523 Quince, C., A. W. Walker, J. T. Simpson, N. J. Loman, and N. Segata. 2017. Shotgun
524 metagenomics, from sampling to analysis. Nat Biotechnol 35:833-844.
525 Rainey, P. B., A. Buckling, R. Kassen, and M. Travisano. 2000. The emergence and maintenance
526 of diversity: insights from experimental bacterial populations. Trends Ecol Evol 15:243-
527 247.

528 Rainey, P. B., P. Remigi, A. D. Farr, and P. A. Lind. 2017. Darwin was right: where now for
529 experimental evolution? Curr Opin Genet Dev 47:102-109.
530 Ramisetty, B. C. and R. S. Santhosh. 2016. Horizontal gene transfer of chromosomal Type II
531 toxin-antitoxin systems of Escherichia coli. FEMS Microbiol Lett 363.
532 Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European Molecular Biology Open
533 Software Suite. Trends Genet. 16:276-277.
534 Rosenfeld, J. S. 2002. Functional redundancy in ecology and conservation. Oikos 98:156-162.
535 Schmieder, R. and R. Edwards. 2011. Quality control and preprocessing of metagenomic
536 datasets. Bioinformatics 27:863-864.
537 Seed, K. D., D. W. Lazinski, S. B. Calderwood, and A. Camilli. 2013. A bacteriophage encodes
538 its own CRISPR/Cas adaptive response to evade host innate immunity. Nature 494:489-
539 491.
540 Simon, J. 2002. Enzymology and bioenergetics of respiratory nitrite ammonification. FEMS
541 Microbiol Rev 26:285-309.
542 Singhania, R. R., A. K. Patel, R. K. Sukumaran, C. Larroche, and A. Pandey. 2013. Role and
543 significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production.
544 Bioresour Technol 127:500-507.
545 Truong, D. T., A. Tett, E. Pasolli, C. Huttenhower, and N. Segata. 2017. Microbial strain-level
546 population structure and genetic diversity from metagenomes. Genome Res 27:626-638.
547 Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V.
548 Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield. 2004. Community structure and
549 metabolism through reconstruction of microbial genomes from the environment. Nature
550 428:37-43.
551 Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Y. Wu, I.
552 Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K.
553 Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch,
554 Y. H. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the
555 Sargasso Sea. Science 304:66-74.
556 Werren, J. H. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. Proc
557 Natl Acad Sci U S A 108 Suppl 2:10863-10870.
558 Widder, S., R. J. Allen, T. Pfeiffer, T. P. Curtis, C. Wiuf, W. T. Sloan, O. X. Cordero, S. P.
559 Brown, B. Momeni, W. Shou, H. Kettle, H. J. Flint, A. F. Haas, B. Laroche, J. U. Kreft, P.
560 B. Rainey, S. Freilich, S. Schuster, K. Milferstedt, J. R. van der Meer, T. Grobetakopf, J.
561 Huisman, A. Free, C. Picioreanu, C. Quince, I. Klapper, S. Labarthe, B. F. Smets, H.
562 Wang, F. Isaac Newton Institute, and O. S. Soyer. 2016. Challenges in microbial ecology:
563 building predictive understanding of community function and dynamics. ISME J. 10:2557-
564 2568.
565 Wilke, A., J. Bischof, W. Gerlach, E. Glass, T. Harrison, K. P. Keegan, T. Paczian, W. L.
566 Trimble, S. Bagchi, A. Grama, S. Chaterji, and F. Meyer. 2016. The MG-RAST
567 metagenomics database and portal in 2015. Nucleic Acids Res. 44:D590-594.
568 Wilson, D. B. 2011. Microbial diversity of cellulose hydrolysis. Curr Opin Microbiol 14:259-
569 263.
570
571
572
573 **Acknowledgements**

We thank members of the Rainey lab at ESPCI and the MPI for Evolutionary Biology for valuable discussion, Boris Shraiman and Daniel Fisher for fuelling interest in the intractability of microbial communities, Sven Kuenzel for DNA sequencing, and Rob Edwards and Jeremy Barr for comments on the manuscript. S.Q. acknowledges receipt of funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 747527.

**Author information**

Concept & experimental design PBR & SQ, experimentation SQ & PBR, bioinformatic analysis SQ and GD, writing PBR & SQ.

## Figures

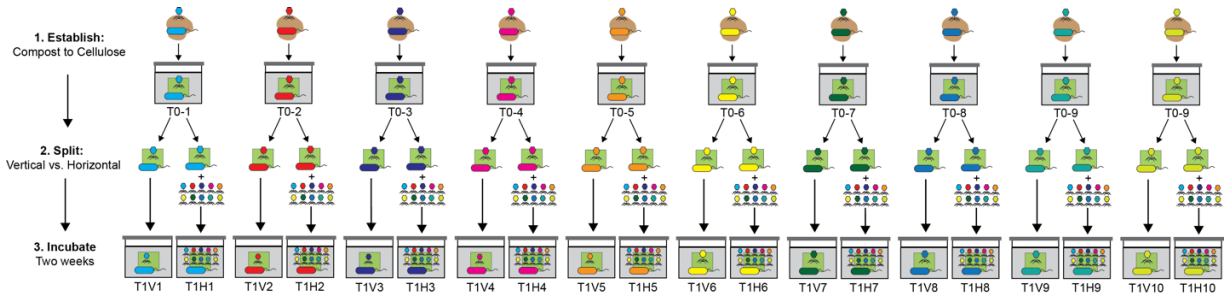


**Figure 1: Experimental setup.**

Cellulose-degrading microbial communities were established by placing ten independent 1 gram samples of fresh compost into 20 mL of M9 Minimal Medium contained within a 100 ml glass sample bottle (mesocosm) with a 4 $cm^2$ piece of paper as the sole carbon source. Following four weeks of incubation at room temperature and without shaking, the communities were homogenized and divided in two. Thereafter — and for the next 48 weeks — one of each pair was subject to a "vertical" transfer regime and the other a "horizontal" transfer regime ("vertical" and "horizontal" refer to the regime affecting fate of selfish genetic elements (SGEs)). For the vertical transfer regime 1 mL of homogenized community was transferred to 19 mL of fresh M9 Minimal Medium (plus paper). In this treatment movement of SGEs is confined to a single community. For the horizontal transfer regime, 1 mL of homogenized community was transferred using the same method as for the vertical regime, but additionally, each horizontal mesocosm received 1 mL of "SGE-cocktail". The "SGE-cocktail" was prepared by sampling 1 mL of homogenized community from each of the ten horizontal communities. These independent samples were then centrifuged and the supernatant passed through a 0.22 µm filter to remove microbial cells. The ten SGE-containing samples were then pooled to produce the "SGE-cocktail". Under the horizontal transfer regime, SGEs repeatedly encounter hosts from other mesocosms. Paired horizontal and vertical communities were incubated for two weeks and total DNA was extracted for metagenomic sequencing. At each transfer samples were collected for DNA sequencing and glycerol stocks of each community were created and stored at -80 °C.

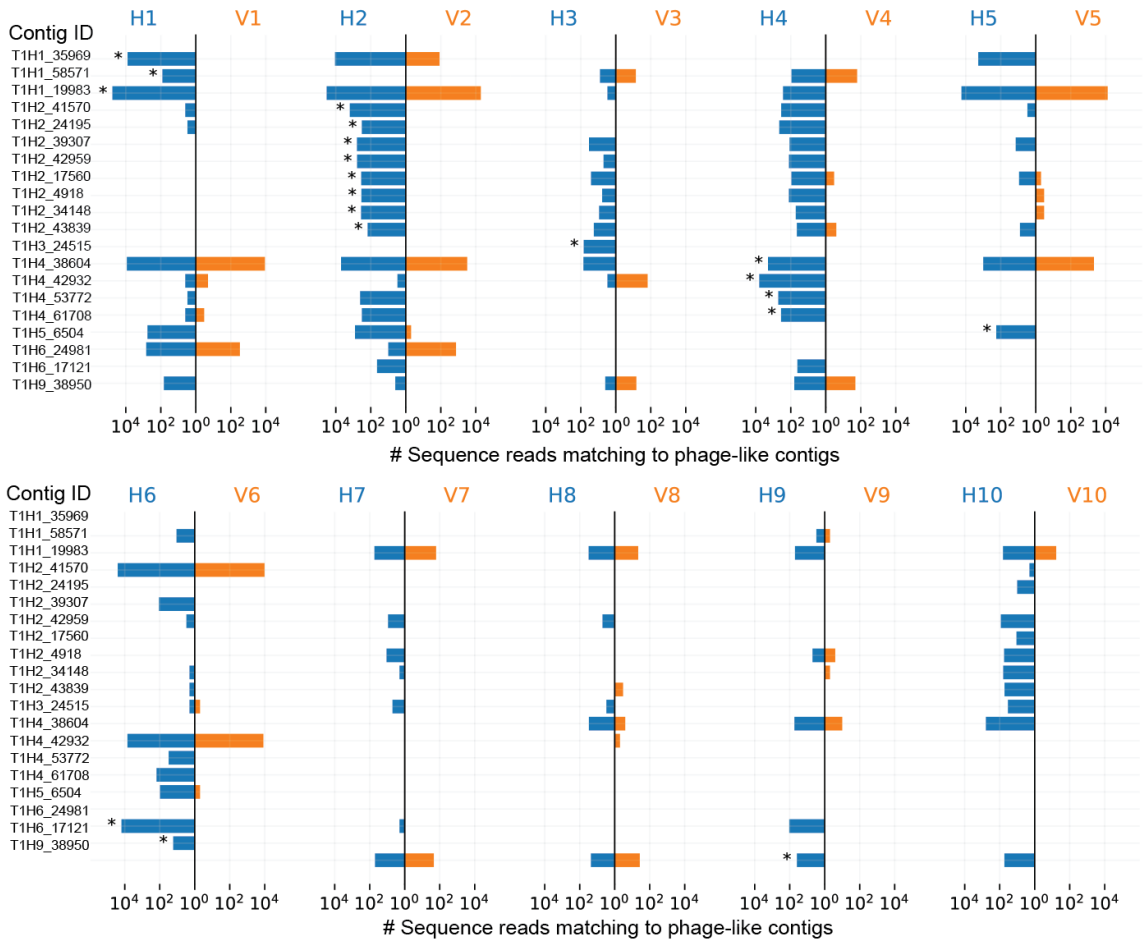


**Figure 2: Movement of phage-like elements between horizontal communities.**

Twenty phage-like contigs assembled from sequence reads unique to horizontal communities are listed on the y-axes. Each vertical line represents a single pair of mesocosms with communities subject to the horizontal and vertical regimes depicted in blue and orange, respectively. The x-axis of the butterfly plot is the number of reads mapping to each contig. Astrices denote the source community from which the contigs were assembled.

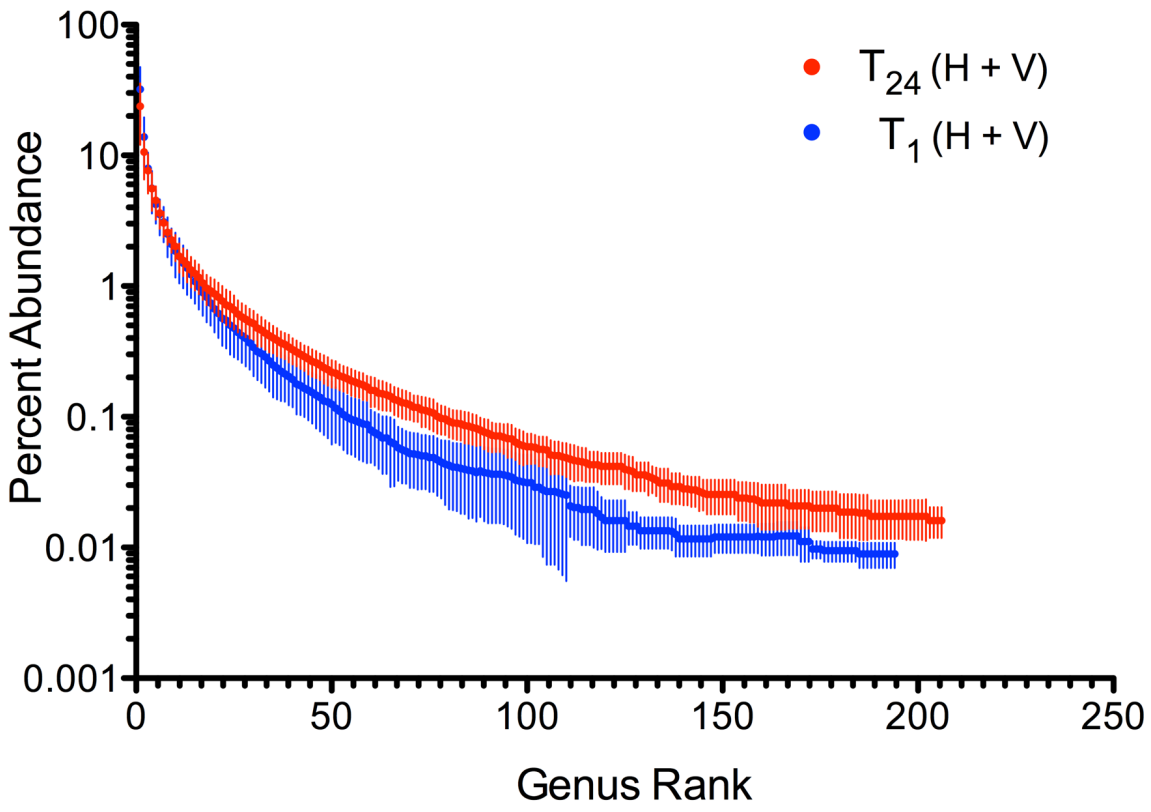


**Figure 3: Genus-level rank abundance curves.**

The blue line depicts the genus-level rank abundance curve at time point $T_1$ two weeks after the divide of founding communities into horizontal and vertical treatments. The red line is at time point $T_{24}$ (48 weeks). Because there were no differences between vertical and horizontal communities at $T_1$ or $T_{24}$ data from both regimes were combined. Data are means and standard deviations from 20 mesocosms.

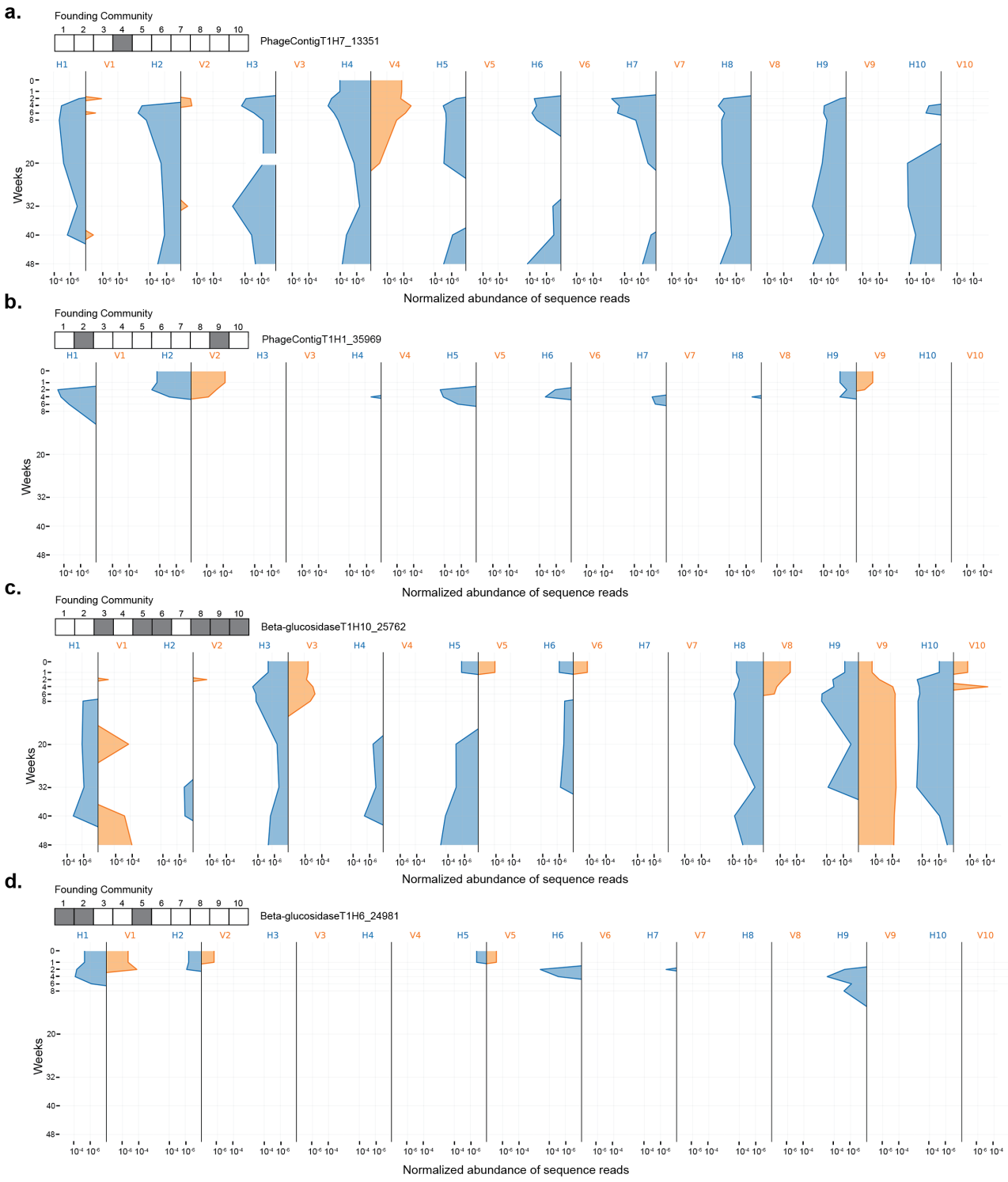


**Figure 4: Dynamics of contigs unique to horizontal communities.**

The dynamics of two-phage like elements (**a**, **b**) and two elements containing predicted beta glucosidases (**c**, **d**) identified from assembling unique reads from horizontal communities at time point $T_1$. Each mesocosm is represented as a vertical line and dynamics were tracked through the course of the 48 week experiment. The butterfly plot depicts normalised abundance of sequence reads mapping to each of the representative contigs from horizontal and vertical mesocosms (blue and orange, respectively). Grey / white boxes above each figure depict presence / absence of sequences from each contig in the founding $T_0$ community (before the vertical / horizontal split).

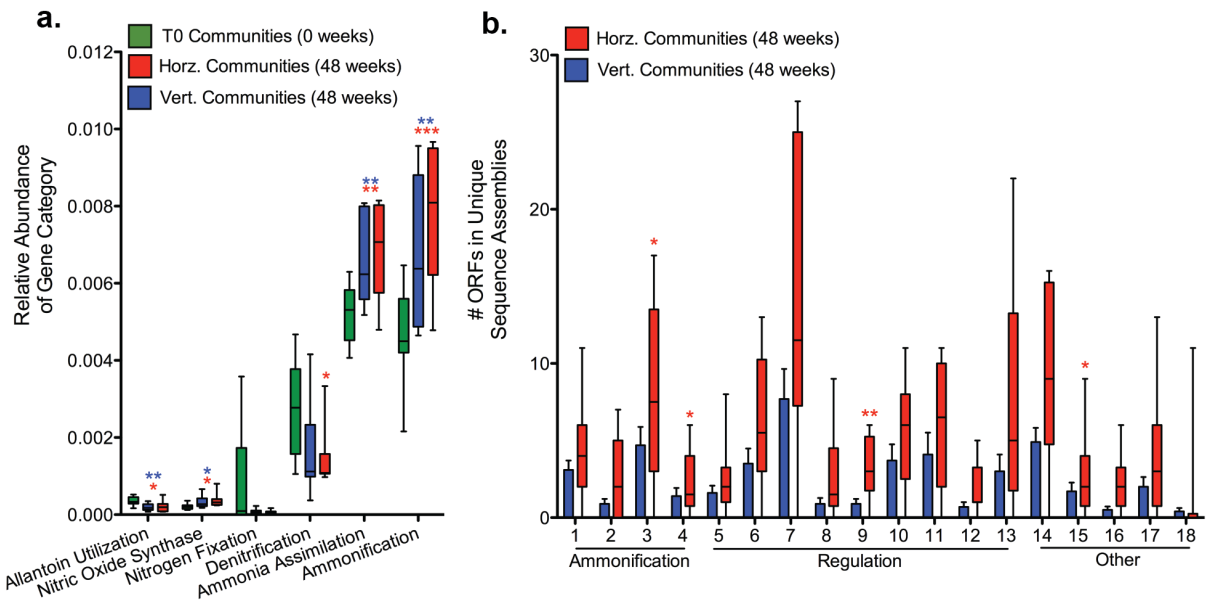


**Figure 5: Enrichment of nitrogen metabolism genes in horizontal communities.**

**a,** Relative abundance of nitrogen metabolism gene categories based on sequence annotation using MG-RAST. Data are shown as box and whisker plots depicting median, inter quartile range (box) and full data spread from 10 replicate communities; * = $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (unpaired *t*-test). **b,** Total number of Open Reading Frames (ORFs) predicted to encode specific domains associated with ammonification, regulation of nitrogen metabolism, and other associated nitrogen metabolism genes in the assemblies of unique vertical and unique horizontal sequences. Data are shown as box and whisker plots depicting median, inter quartile range (box) and full data spread from 10 replicate mesocosms; * = $P < 0.05$ (unpaired *t*-test). 1, NifS; 2, NifE; 3, NapA; 4, nitrilase; 5, NtrA; 6, NifR3; 7, NtrY; 8, NifL; 9, PtsN; 10, NtrB; 11, NtrC; 12, NAC; 13, FixI; 14, FixG; 15 & 16 nitroreductases; 17, arginase; 18, cytochrome D1. Full descriptions of domains are provided in Supplementary Table 9.

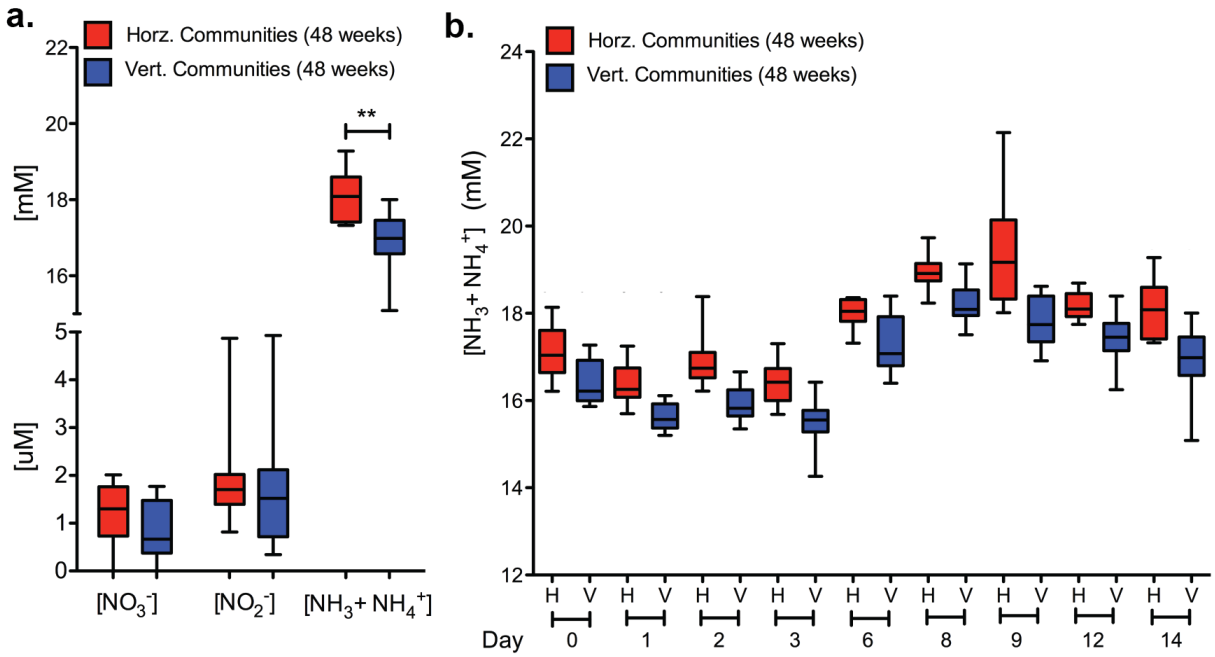


**Figure 6: Enhanced ammonia production in horizontal communities.**

**a,** Concentrations of nitrate, nitrite, and ammonia / ammonium from mesocosms subject to horizontal and vertical regimes at 48 weeks ($T_{24}$). Data are shown as box and whisker plots depicting median, inter quartile range (box) and full data spread from 10 replicate communities; $^{**}P < 0.01$ (unpaired *t*-test). **b,** Ammonia / ammonium concentrations measured at nine time points during the two-week incubation period for communities from horizontal and vertical regimes at 48 weeks. Data are shown as box and whisker plots depicting median, inter quartile range (box) and full data spread from 10 replicate communities (triplicate measures were obtained from each community on each occasion). Analysis of time series data using a mixed ANOVA design showed a significant effect of treatment (horizontal vs. vertical; $P = 1.96 \times 10^{-5}$) and a significant effect of time ($P = 2.7 \times 10^{-24}$, after Greenhouse & Geisser correction for sphericity). The interaction between treatment regime and time was not significant $P = 0.69$). Comparison between the vertical regime at 48 ($T_{24}$) and community function at the beginning of the experiment ($T_0$) showed no difference (see Supplementary Fig. 8).