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, Chemistry, Biology and Innovation (CBI) UMR8231, ESPCI Paris, CNRS, PSL Research University, 10 rue Vauquelin, 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

Keywords: Bacteriophages — Cellulose degradation — Diversity — Horizontal gene transfer — Ammonification

Abstract

Microbial communities underpin earth's biological and geochemical processes, but their complexity hampers understanding. Motivated by the challenge of diversity and the need to forge ways of capturing dynamical behaviour connecting genes to function, biologically independent experimental communities comprising hundreds of microbial genera were established from garden compost and propagated on nitrogen-limited minimal medium with cellulose (paper) as sole carbon source. After one year of bi-weekly transfer, communities retained hundreds of genera. To connect genes to function we used a simple experimental manipulation that involved periodic collection of selfish genetic elements (SGEs) from separate communities, followed by pooling and redistribution across communities. The treatment was predicted to promote amplification and dissemination of SGEs and thus horizontal gene transfer (HGT). Confirmation came from comparative metagenomics, which showed substantive movement of ecologically significant genes whose dynamic across space and time could be followed. Enrichment of genes implicated in nitrogen metabolism, and particularly ammonification, prompted biochemical assays that revealed a measurable impact on community function. Our simple experimental strategy offers a conceptually new approach for unravelling dynamical processes affecting microbial community function.

Introduction

Microbial communities underpin all major biological and biogeochemical processes [1]. Appropriate functioning of communities has direct implications for health of people [2, 3], terrestrial [4, 5], marine [6] and freshwater [7] ecosystems and even global climate [8]. While the species composition of many communities have been exhaustively documented [9, 10], the link between genes and community function is poorly understood [1].

New strategies for investigation are required that are process-focused [11]. 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 [12-15]. Recent advances draw upon new strategies for linking patterns of sequence diversity in metagenomic data sets to population genetic processes [16-19]. 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 — in a proof-of-principle experiment — that this can be achieved via an experimental strategy that combines theory governing the behaviour of selfish genetic elements [20-23] (SGEs) and expected effects on horizontal gene transfer (HGT), with approaches from experimental evolution [24, 25], comparative metagenomics [26] and functional assay.

Results

Diversity in experimental communities

Experimental communities were established from a diverse primary source: garden compost. Ten independent 1 g samples of compost were taken from ten different regions of a single 1 m^3 compost heap and each placed in one of ten 140 ml bottles (mesocosms) containing 20 ml nitrogen-limited minimal M9 medium plus cellulose (a 4 cm^2 piece of paper) as sole carbon source. Mesocosms were incubated on a laboratory bench without shaking and with lids left unsealed. After a two week period the initial piece of paper from each mesocosm was separately transferred to a new bottle containing fresh M9 medium plus a new piece of paper. Following a further two week incubation period we considered the communities to have adjusted to laboratory conditions. At this time point mesocosms were vortex-mixed and 1 ml of cellulosic slurry from each bottle was transferred to *two* new mesocosms containing 19 ml fresh M9 medium plus paper giving a total of 20 paired communities. This transfer, four weeks after initial establishment, was deemed time point zero (T_0). Thereafter and for the ensuing 48 weeks, communities were serially propagated every two weeks (T_1 to T_{24}) by transferring 1 ml of slurry to fresh bottles. One set of each pair of bottles was labelled "*vertical*" (vertical communities (VC)) and the second of each pair labelled "*horizontal*" (horizontal communities (HC)). Details of the treatments and their significance are elaborated below, but at this stage, with focus solely on the question of diversity in cellulose-based mesocosms, we acknowledge treatment effects by name only.

Diversity through 48 weeks of propagation was addressed by metagenomics: total DNA was extracted from each of the ten T_0, and twenty T_1 and T_{24} communities and the data interrogated for sequence reads mapping to 16S rDNA (Supplementary Table 1). Communities at T_0 harboured on average ~140 bacterial genera (SD $\pm$ 23). At T_{24} the number of genera detected increased in all communities to an average of ~200 (SD $\pm$ 32), indicating an increase in abundance of rare types (Supplementary Table 1, Supplementary Fig. 1a). Diversity levels were highly similar in both HCs and VCs (Supplementary Fig. 1a-c). Rank abundance distributions from pooled HCs and VCs at T1 and T24 are shown in Fig. 1 and reveal the tail of the T_{24} distribution to be markedly flatter in both cases, consistent with an overall tendency for rare types to have increased during the course of the year long selection. Rank abundance

70 distributions for each T_1 and T_{24} HC and VC, plus the rank order (and change therein) of the
71 most common genera are shown in Supplementary Figs 2a and 2b. Supplementary Fig. 3 is a
72 multidimensional scaling plot showing that communities at the start and end of the experiment
73 differed markedly in genera composition and there was no sign of convergence after 48 weeks.
74 Among the most abundant genera, only two harbour species commonly associated with
75 cellulolytic ability [27] (*Cytophaga* and *Cellvibrio*), while others are better known for roles in
76 ammonification [28] (*Azospirillum* and *Paenibacillus* (nitrogen fixation); *Rhodanobacter* and
77 *Pseudomonas* (dissimilatory nitrate reduction / denitrification)) (Supplementary Figs 2a and 2b).
78 Single celled eukaryotes were also present in all mesocosms with several unexpectedly
79 maintaining populations of nematodes through the 48-week selection period.
80
81 *Manipulating community-level process*
82 An important goal is development of strategies that link genes to community function by way of
83 process. An ideal experiment is one that alters a single ecological or evolutionary process and
84 compares the outcome to a set of replicates in which the process is unaffected. For example,
85 evolutionary biologists interested in the evolutionary consequences of sex have taken advantage
86 of yeast in which it is possible to genetically engineer sexual and asexual types that are otherwise
87 isogenic [29, 30]. Differences in evolutionary outcome between sexual and asexual types after a
88 period of selection under identical environmental conditions can thus be attributable to sex.
89
90 We reasoned that a similar manipulation might be possible in the context of microbial
91 communities. Recombination mediated by horizontal gene transfer (HGT) is loosely analogous
92 to sex, albeit to a form of sex that takes place at the level of communities. Moreover, like
93 genetic manipulations affecting sexual reproduction in yeast, the extent of HGT can be
94 controlled by manipulating the evolutionary fate of selfish genetic elements (SGEs) [20, 23].
95 Such elements, including phages, plasmids, transposons and integrative and conjugative
96 elements are prime vehicles for the movement of ecologically significant genes among diverse
97 bacteria [31-33]. We note that when SGEs acquire genes that aid host survival they are no
98 longer "selfish" and are more appropriately referred to as "mobile" genetic elements, however,
99 the theory we draw upon comes from thinking about elements that are costly to maintain and
100 thus we continue to use the more generic SGE label.

101

102 The critical manipulation is one that affects the likelihood that SGEs encounter new hosts.
103 Frequent exposure to new hosts drives dissemination and evolution of SGEs and linked genes
104 [23]. Conversely, in the absence of new hosts, and thus opportunity for infectious spread,
105 selection is powerless to prevent loss of transfer ability leading to degradation of SGEs and thus
106 reduction in HGT. That loss occurs is evidenced by the remnants of prophages that litter
107 bacterial genomes [34].

108

109 The *vertical* and *horizontal* treatment designations referred to above reflect manipulations that
110 are predicted to either restrict, or facilitate, respectively, HGT mediated by SGEs. The vertical
111 regime involved serial transfer of material from mesocosm to mesocosm. SGEs within these
112 communities were denied opportunity to encounter new hosts: transfer of SGEs was vertical and
113 the extent of HGT is expected to be limited (Fig. 2) (and eliminated altogether over longer
114 timescales).

115

116 Communities subject to the horizontal regime were transferred by serial dilution to fresh
117 mescosms as per the vertical regime, but received additionally — at the time of serial transfer —
118 an aliquot of "SGE cocktail". The SGE cocktail was a pooled sample of SGEs derived from
119 each horizontal mesocosm, which was then redistributed among each HC (Fig. 2). SGEs within
120 these communities were thus repeatedly (every two weeks) provided with opportunity to
121 encounter new hosts. For example, phages from HC 1 could infect hosts present in HCs 2-9 and
122 may additional mobilise ecologically significant genes acquired en route. The horizontal regime
123 thus breathes evolutionary life into SGEs and HGT is expected to be rampant.

124

125 *Detection of SGE activity*

126 To determine whether the horizontal regime promoted movement of SGEs, total DNA was
127 extracted from each T_0 community and each of the 20 T_1 communities (two weeks after the split
128 of communities into vertical and horizontal treatments) and sequenced (Fig. 2 and
129 Supplementary Table 2). Analysis was straightforward given the paired experimental design.
130 For each set of T_0 and descendant HCs and VCs (e.g., T_0-1 and descendants T_1V-1 and T_1H-1)
131 DNA sequence data were interrogated to identify reads found solely in (and therefore unique to)

132 the horizontal mesocosm (e.g., T_1H-1). Such unique reads are likely to stem from SGEs, or be
133 associated with SGE activity, originating from an allopatric community.
134
135 Analysis of the sequences unique to each horizontal community resulted in identification of ~26
136 million unique reads from a total of ~152 million reads across all ten communities, with an
137 average of ~2.6 million unique reads per community (Supplementary Table 3). Assembly of
138 reads into contigs revealed an average of 3,352 contigs per HC greater than 1 kb in length. The
139 mean maximum contig size per assembly was 82 kb (Supplementary Table 3).
140
141 After extracting Open Reading Frames (ORFs), interrogation of the Conserved Domain Database
142 (CDD) [35] showed 1,279 ORFs predicted to encode phage-associated proteins involved in
143 capsid formation, baseplate assembly, and phage-induced lysis of bacterial cells (Supplementary
144 Table 3). An example of a contig (T_1H-1_35969) containing numerous genes characteristic of
145 phages is shown in Supplementary Fig. 4a.
146
147 The distribution and abundance of contig T_1H-1_35969 and 19 other representative phage-like
148 entities (Supplementary Table 4) across independent mesocosms was determined by mapping
149 total reads from each horizontal and vertical metagenome onto each of the 20 phage-like contigs
150 (Fig. 3). The predominance of these contigs in HCs is evidence of rapid amplification and
151 dissemination of genetic material. T_1H-1_35969, originally assembled from the unique set of
152 sequences from T_1H-1, was also present in both T_1H-2 and T_1V-2 and was therefore likely
153 present in the founding community (T_0-2). Within two weeks this contig had amplified and
154 spread to horizontal communities 1, 5 and 6 (Fig. 3). Contig T1H2_39307 was below the level of
155 detection in all vertical treatments, but two weeks after imposition of the horizontal treatment,
156 this contig was detected in seven of the independent HCs (Fig. 3, Supplementary Table 4). From
157 the community perspective, less than four of the 20 phage-like contigs were detected on average
158 per VC, whereas the mean number of contigs per HCs was 11 (Fig. 3).
159
160 In addition to phage-like contigs, assemblies of the unique sequences from HCs contained genes
161 predicted to encode traits of likely ecological significance (Supplementary Fig. 5). These
162 include genes implicated in cellulose degradation, defence against SGE invasion, nitrogen

163 metabolism and siderophore biosynthesis (Supplementary Table 5) [36-38]. Particularly notable
164 was enrichment of reads mapping to *β*-glucosidase genes that define the rate-limiting step in
165 cellulose degradation [39]. On occasion these ecologically significant genes were linked to
166 genes encoding features found on SGEs (Supplementary Table 6), but often these contigs lacked
167 such features (Supplementary Fig. 4b). Nonetheless they were amplified and disseminated
168 through the course of the selection experiment with the same dynamic as expected of a mobile
169 genetic element (see below).
170
171 *Tracking the dynamics of SGE amplification and dissemination*
172 Given evidence indicative of the amplification and dissemination of DNA via the horizontal
173 treatment just two weeks after splitting T_0 communities into VCs and HCs, we complemented
174 existing DNA sequence from T_0 and T_1 and T_{24} with DNA extracted from all HCs and VCs at
175 time points T_2, T_3, T_4, T_{10}, T_{16}, and T_{20} yielding a total of 180 metagenomes (5.2 billion reads,
176 each ~150 bp). These 180 metagenomic data sets (nine time points, 20 communities per time
177 point (Supplementary Table 2)) spread across 48 weeks allowed the abundance and distribution
178 of SGEs and associate genes to be determined by read mapping to contigs assembled from T_0
179 data. The spatial and temporal dynamics of two phage-like contigs and two contigs encoding *β*-
180 glucosidases are shown in Fig. 4a-d.
181
182 Focussing on PhageContigT_1H_7_13351 (Fig 4a), this phage-like element was detected only in
183 community C4 at T_0 , but was likely present also in communities C1 and C2, although below the
184 limit of detection. By week 3, the phage-like element was present in all horizontal communities.
185 In vertical communities the element persisted in community C4 four for 22 weeks and transiently
186 in C1 and C2, but was present at 48 weeks in all but one horizontal community. Within
187 horizontal communities a pattern observed on four occasions saw the phage-like element go
188 extinct (or fall below the level of detection), but re-emerge and persist at latter time points.
189
190 *Ecologically significant genes*
191 To determine whether the horizontal transfer regime affected the abundance and distribution of
192 functional genes, the MG-RAST database was interrogated with sequence reads from all T_0 and
193 all T_{24} HCs and VCs (Supplementary Table 7). For the ensuing analysis we have avoided

194 standard statistical testing because horizontal microcosms, by virtue of the movement of SGEs
195 and linked genes among communities, cannot be treated as independent entities. Nonetheless,
196 trends are clearly evident from the data displayed in Fig. 5 (and corresponding Supplementary
197 Figs 6a and 6b), which include for each functional category — for each vertical, horizontal and
198 T_0 community — median value, interquartile range and full data range.
199
200 The relative abundance of reads assigned to 13 of 28 functional categories at Subsystems Level 1
201 [40], including genes involved in virulence, motility and nitrogen metabolism, changed during
202 the course of the year (Supplementary Fig. 6a). HCs and VCs did not always change to the same
203 extent, or in the same direction (Supplementary Figs 6a and 6b). No obvious change occurred in
204 the overall category of carbohydrate metabolism (Supplementary Fig. 6b), but an increase in β-
205 glucosidase metabolism — the rate limiting step in cellulose degradation — was detected in both
206 HCs and VCs (Supplementary Fig. 6a).
207
208 The nitrogen-limited nature of the culture medium, combined with evidence of selection
209 favouring genes involved in nitrogen metabolism, led to closer focus on this essential resource.
210 The MG-RAST database was again interrogated, but this time at Subsystems Level 2 [40], which
211 provides information on functional categories within the broader category of nitrogen
212 metabolism (Fig. 5a). Changes were observed in the relative abundance of reads mapping to
213 genes involved in both ammonia assimilation and dissimilation (ammonification). Such changes
214 are consistent with the nature of the selective environment. This prompted a query of the extent
215 to which these changes may have been influenced by movement of genes via SGEs.
216
217 A modified version of the bioinformatic pipeline developed to identify sequences unique to HCs
218 (Methods) was applied to the T_{24} metagenomes. T_{24} HC metagenomes were compared to their
219 paired T_{24} VC metagenomes as well as to all metagenomes from VCs from earlier time points
220 (Supplementary Fig. 7a, Supplementary Table 8, Supplementary Methods Fig. 1). This showed a
221 greater number of unique sequences in the HCs, but also showed a fraction of unique sequences
222 present in VCs. The latter is to be expected given that T_0 communities must contain rare
223 sequences that through the course of the selection experiment became common — and is
224 consistent with data on changes in abundance of genera (Fig. 1). To see whether it was possible

225 to reduce this signal, genomic DNA samples from T_0 were sequenced on the HiSeq platform
226 resulting in additional 320 million reads per community. The effect of increasing depth of
227 sequence made minimal impact on detection of unique sequences (Supplementary Fig. 7b).
228 Nonetheless, the "deep" T_0 metagenomes were used in all subsequent analyses.
229
230 To determine the fraction of unique reads that mapped to genes involved in nitrogen metabolism,
231 the final sets of unique reads were obtained from HCs and VCs; they were assembled, ORFs
232 identified and functionally categorised using the CDD. HCs contained more unique ORFs
233 predicted to be involved in nitrogen metabolism compared to VCs (Supplementary Fig. 8). HCs
234 were enriched in several functional classes of gene, especially those with predicted roles in
235 regulation and ammonification (Fig. 5b, Supplementary Table 9).
236
237 To link to function, we asked whether there was evidence of an effect of the horizontal treatment
238 on a measurable community property. To this end, and with focus on the process of
239 ammonification (production of ammonia through either fixation, or reduction of nitrate / nitrite)
240 the concentration of nitrate, nitrite, and ammonia were determined in the T_{24} HCs and VCs at the
241 end of the two-week period immediately prior to serial passage (Fig. 6). Bearing in mind the
242 non-independence of horizontal communities, two different, but appropriate statistical tests were
243 employed. The first, a two-sample Kolmogorov-Smirnov (K-S) test revealed no significant
244 difference among HCs and VCs for nitrate or nitrite production, but a highly significant
245 difference in the amount of ammonia produced. The second, a Linear Mixed Effect Model, with
246 "community" assigned as a random effect and regime (vertical / horizontal) as a fixed effect,
247 produced results that mirrored those of the K-S test. Details of the tests are provided in the
248 caption to Fig. 6.
249
250 To check the robustness of this finding, the concentration of ammonia was measured during the
251 course of a two-week period. Despite substantive variability in composition of the independent
252 communities (Supplementary Fig. 2a,b and Supplementary Fig. 3), a significantly greater
253 concentration of ammonia was detected in HCs at each of nine sampling occasions throughout
254 the two week period (Fig. 6b, Supplementary Table 11). A comparison between T0 communities
255 and T24 VCs showed no change (Supplementary Fig. 9) supporting the conclusion that

256 movement of ecologically significant genes via SGEs was responsible for the functional
257 difference between horizontal and vertical communities.

Discussion

Simplification is a powerful scientific approach. For example, detailed insight into the dynamic of adaptive evolution has come from the field of experimental evolution, where single (often asexual) genotypes are propagated under defined laboratory conditions with variation arising solely by *de novo* mutation and where selection is the primary arbiter of mutation success [24, 25]. This minimal experimental design avoids mixing (migration) between populations, sex and recombination, thus allowing observed evolutionary change to be attributed directly to specific mutations that can be connected causally to changes in fitness.

We sought to apply the same kind of simplifying experimental approach to the study of microbial communities, however doing so is not straightforward. Simply substituting a community of microbes in place of a single bacterial genotype followed by recapitulation of a standard serial transfer experiment, where the community is propagated on a simple carbon source, such as glucose, would yield little insight into processes relevant to natural communities and may even add confusion. In a serial transfer experiment, focal populations adapt genetically by natural selection. Substitution of communities for single populations does not result in communities adapting genetically by natural selection: individual bacterial lineages within the community will evolve — although the extent to which their evolution is adaptive is an open and fascinating question [41-43] — but it makes no sense to think about community-level adaptation, because communities (in a standard serial transfer experiment) are not units of selection. To observe adaptation at the community level, communities must participate in the process of evolution by natural selection. This requires establishment of an experimental regime in which communities are treated as evolving lineages subject to a birth-death process [25, 44-47].

If it makes no sense to assume that communities adapt in a manner analogous to single bacterial populations, then attention shifts away from focus on mutation and selection, and turns to possibilities that might stem from manipulation of evolutionary processes such as recombination, migration [15], or genetic drift. We chose to manipulate the former, seeing possibility to promote or retard HGT via a simple mixing regime predicted to affect the evolutionary fate of SGEs, which in turn, might connect genes to community function.

289

290 Before proceeding to discuss choice of experimental design, we note subtleties and
291 complications concerning processes driving eco-evolutionary dynamics, the correspondence
292 between these dynamics and diversity, and the connection between scale of the dynamic and
293 impact on community function. While much is unknown, it seems reasonable to presume a
294 connection, which suggests that communities composed by too few types may fail to capture
295 processes intrinsic to the system, particularly if that process relies on HGT.
296
297 Such considerations were instrumental in the decision to found experimental communities from a
298 highly diverse source (garden compost) and to propagate communities on cellulose as sole
299 carbon source. For a start, garden compost is rich in cellulose thus ensuring an *in vitro*
300 environment not too dissimilar from that encountered in nature, additionally, there exists, in
301 principle, the possibility that, in future work, component species might be isolated and cultured
302 on laboratory agar with glucose as sole carbon source (glucose being the utilisable metabolic
303 break-down product of cellulose). However, the primary consideration for use of cellulose was
304 that it is difficult to degrade [27], thus allowing the possibility that communities might be
305 cultured over prolonged periods under nutrient limitation without dominance by *r*-selected types.
306 In effect, while the means and mode of propagation involved serial transfer, communities likely
307 experienced culture conditions more reminiscent of those provided in chemostat culture.
308
309 Levels of diversity recorded from analysis of metagenomic data were remarkably high,
310 remaining stable — and even increasing — through the course of the 48 week period of
311 propagation (Fig. 1). This is unexpected given propagation on a single carbon source, where few
312 types should be stably maintained [48-50]. Were the community propagated on glucose instead
313 of cellulose, almost certainly diversity would have been minimal and dominated by a small
314 number of fast growing types. That diversity increased over time (Fig. 1) indicates that rare
315 genera that were initially below the level of detection at the start of the experiment, counter to
316 expectation, increased in frequency. The ecological reasons are not known, but point to
317 unanticipated ecological complexity and presumably abundant opportunity for cross feeding.
318 The high and stably maintained levels of diversity in cellulose-based mesocosms founded from

319 garden compost renders them ideally suited for the kind of experimental work that we have
320 initiated here.
321
322 Although noted only in passing, and from examination of the mapping of DNA sequence reads to
323 genes connected with nutrient cycling (data not shown), cycles for replenishing all major
324 nutrients appeared to have been established within each mesocosm, with that involving nitrogen
325 being most evident. Each community harboured a single dominant free-living nitrogen-fixing
326 bacterium, but the identity — even at the genus level — differed between mesocosms. Attempts
327 to establish such complex communities by choosing key players before the fact, would likely
328 fail. That such diversity and stability can arise from a single gram of compost speaks to an
329 abundance of redundancy in these natural systems [51, 52].
330
331 Given the experimental design with vertical and horizontal treatment regimes, it is natural to
332 query the relationship between treatment effects and patterns of diversity. At the level of number
333 of bacterial genera and shape of rank abundance curves, no difference was detected and in fact
334 communities showed remarkably similar overall trends. However, differences are apparent when
335 focussing on the rank of individual genera. These are evident in the butterfly plots in
336 Supplementary Fig. 2b and show difference both between independent communities but also
337 differences between vertical and horizontal treatments within communities. Because our
338 experimental design did not include technical replicates (a decision made based on the need to
339 balance experimental feasibility with intuition that we should maximise the number of
340 independent communities) no statistical analysis is possible. Nonetheless, notable are the
341 similarity of changes in genus rankings between vertical and horizontal treatments from the same
342 communities. Also notable is the fact that there was no evidence of convergence in diversity in
343 the horizontal treatment after 48 weeks (Supplementary Fig. 3). Although unsurprising, this
344 emphasises the distinctness of the communities despite being linked via movement of SGEs and
345 linked genes.
346
347 The pooling of material from each mesocosm that passed through a 0.2 μm filter to produce a
348 SGE-cocktail, followed by its mixing and regular distribution among HCs, was evidently
349 effective in both mobilising SGEs and promoting horizontal transfer of genes of ecological

350 significance. In fact the magnitude of the effect combined with capacity to detect it via
351 identification of DNA sequence reads unique to each horizontal community, exceeded
352 expectation. Within just two weeks of implementation, ~17 % of DNA sequence reads within
353 HCs were derived from allopatric communities. That these reads could be assembled into
354 contigs containing genes of ecological relevance, including those involved in cellulose
355 degradation, emphasises the dynamism of processes driven by SGEs.
356
357 One immediate consequence is utility of the experimental approach to detect — at the level of
358 communities — traits under selection. The suite of unique DNA sequences amplified and
359 disseminated across communities suggests that these genes matter. In the context of conditions
360 experienced by the compost-derived communities this makes sense. Community members are
361 expected to have experienced selection for ability to degrade cellulose: this is indicated by the
362 abundance of horizontally transferred genes predicted to be involved cellulose degradation
363 (Supplementary Fig. 5). Similarly, as mesocosms were nitrogen limited, selection was expected
364 to favour movement of genes involved in nitrogen metabolism. That this happened is indicated
365 by the prevalence of genes implicated in nitrogen metabolism among the set of DNA sequence
366 reads unique to HCs. Not appreciated at the outset was that communities, according to the
367 signature acquired from unique sequence reads, were iron limited. With hindsight this was to be
368 expected. That genes encoding CRISPR are also enriched is especially significant and suggests
369 an unappreciated dynamic between hosts, the defence systems maintained to protect bacterial
370 hosts against phages, and benefits that accrue from the horizontal dissemination of these systems
371 [53]. Application of vertical vs horizontal treatment regimes — which are readily imposed
372 across numerous experimental systems — thus stand to reveal, from the perspective of the
373 community, the nature of traits under selection.
374
375 Without further experimentation, the identity of the elements responsible for horizontal spread of
376 genes is unknown, and so too is the relationship between the genes amplified and disseminated
377 across mesocosms, and the vehicles for their dissemination. While some assembled contigs
378 show features typical of phages, many do not. This may be a consequence of working with
379 metagenomic data, but may also reflect the existence of elements whose nature and dynamic
380 remains to be discovered.

381

382 One of the larger assembled contigs shown in Supplementary Figure 4b is a case in point: at ~80
383 kb, hints as to its identity are expected, and yet while carrying genes predicted to encode
384 glycoside hydrolyses — implicated in cellulose degradation — the only suggestion of ability to
385 be both mobilised and disseminated is via linkage to a single integrase. Acquisition of long-read
386 sequence data may aid understanding, but it is also possible that the captured dynamic — the
387 rapid amplification and dissemination of ecologically significant genes — is attributable to
388 "fortunate accidents" mediated by, for example, lytic phages or other elements that happen on
389 rare occasion to capture genes of utility to hosts, with the feedback between genes that prove
390 useful to hosts and the amplifying effects of selection driving their dynamic. This points to the
391 possibility of processes that unfold at the level of diverse communities that remain to be
392 understood.
393

394 Irrespective of the nature of the disseminating vehicles, the experimental approach outlined
395 allows the dynamics of elements amplified and disseminated to be tracked. Examples are shown
396 in Figs 3 and 4. Although mechanistic explanations are unavailable, the dynamics of
397 PhageContigT1H7_13351 (Fig. 3A) raise numerous questions. Extinction within VC4, but
398 persistence in all HCs at 48 weeks (with the exception of H1) raises questions as to the causes.
399 Using recently developed Hi-C [16, 19, 54] and population genetic approaches [55] it ought to be
400 possible, in the future, to disentangle ecological from evolutionary effects. Nonetheless, that
401 persistence is evident in HCs, points to a significant role for SGEs.
402

403 The association established between genes involved in nitrogen metabolism and enhanced
404 activity of SGEs in HCs is correlative, but is nonetheless linked to horizontally transferred
405 sequences that map to genes with predicted roles in ammonification. The association is further
406 linked to data that demonstrate a significant effect of the experimental treatment on
407 concentrations of ammonia in HCs (Fig. 6), but no effect (compared to T_0 communities) in VCs
408 (Supplementary Fig. 9). Additionally, increased ammonia in the HCs is consistent with the
409 prediction that horizontal movement of DNA will increase the rate at which community function
410 improves.
411

412 The two processes that generate ammonia: nitrogen fixation and nitrate ammonification, both
413 require environments devoid of oxygen (or mostly so in the case of nitrate ammonification) [28,
414 56]. It is possible that enhanced metabolic activity associated with cellulose degradation causes
415 lower oxygen conditions at the paper surface (where community members grow as biofilms)
416 compared to the VCs, leading to enhanced production of ammonia. Intriguingly, the cattle
417 industry promotes addition of ammonia to feed because of beneficial effects on digestibility of
418 plant matter in the rumen [57]. This raises an alternate possibility, which is that elevated
419 ammonia levels are similarly beneficial to digestion of cellulose in the experimental mesocosms,
420 and reflect a more rapid response to selection in the HCs.
421
422 The genomic era has done much to resolve controversy surrounding the importance of horizontal
423 gene transfer as a driver of evolutionary change [58-60]. Direct evidence of HGT — many
424 facilitated by SGEs — from one organism to another are common [61, 62], but the possibility
425 that the process assumes far greater evolutionary significance when operating within
426 communities, has, to the best of our knowledge, been little considered, let alone directly studied.
427 That the vast diversity of DNA sequence encompassed within complex microbial communities
428 might exist in fluid association with SGEs generating permutations of effects significantly
429 beyond those observed through study of individual SGEs and their hosts, is both interesting and
430 plausible.
431
432 Microbial communities are dynamical systems defined by complex ecological interactions and
433 evolutionary feedbacks [1]. The challenges associated with understanding this complexity are
434 immense. Here we have demonstrated the utility of an approach in which we have employed a
435 simple experimental manipulation predicted (and shown) to fuel the evolutionary life of SGEs
436 and thus the process of HGT. In effect, mixing (or not) SGEs among communities promotes (or
437 limits) what might be considered community-level sex. The manipulation affects the abundance
438 and distribution of genes of ecological significance, whose individual dynamics can be followed.
439 Further, the effects are sufficient to connect — at a correlative level — genes to community
440 function. That this involves nothing more than frequent exposure of SGEs to new hosts,
441 combined with metagenomic analysis, means that the experimental strategy can be readily
442 transferred to a wide range of microbial communities, from experimental communities as here,

443 through microbiomes of plants and animals, to naturally occurring marine and terrestrial
444 communities.

445 **Methods**

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

447 Microbial communities were initially established by placing ten independent 1 g samples of fresh
448 compost into 20 mL of M9 minimal media (the nitrogen source was ammonium chloride (0.935
449 mM)) supplemented with a 4 cm^2 piece of cellulose paper (Whatman cellulose filter paper). All
450 incubations were performed in 140 mL sterile glass bottles with loosened screw caps allowing
451 for gas exchange between the community and the environment. Incubation was at room
452 temperature without shaking. Compost was sampled from the Square Theodore-Monod compost
453 heap (Paris, France) in February 2016. During the primary incubation period of two weeks the
454 piece of cellulose was suspended by a wire from the screw cap into the media allowing for paper
455 colonization. After two weeks, the community was transferred by moving the wire-suspended
456 cellulose paper into 20 mL of fresh M9 minimal media that also contained a new 4 cm^2 piece of
457 cellulose paper in suspension. Two more weeks of incubation provided the opportunity for
458 colonization of the new piece of suspended cellulose paper thus eliminating the need for a wire-
459 suspended community for each transfer. All community transfers were performed by vortexing
460 each bottle at maximum speed until the cellulose fibres were disaggregated into a slurry.

461

462 **Horizontal and vertical transfer regimes**

463 Vertical and horizontal transfers were performed at the exact same time every two weeks. Before
464 each transfer aliquots of the community were taken for glycerol stocks, DNA extraction, and
465 phage storage. Glycerol stocks were created by mixing 500 µl of cellulose-microbial slurry with
466 500 µl of 80% glycerol and stored at -80 ºC. For DNA extractions 2 mL of cellulose-microbial
467 slurry was centrifuged at 13,000 g for 10 min and the pellet was stored at -80 ºC for later
468 processing. DNA was extracted from the founding T_0 communities (before the horizontal /
469 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
470 the soil DNA extraction kit (Norgen Biotek). DNA sequencing was performed using the MiSeq
471 and NextSeq platforms for all time points and T_0 communities with additional deep sequencing
472 of the T_0 communities using the HiSeq platform.

473

474 Transfers at two-week intervals were performed as described in Fig. 2. The "SGE-cocktail" was
475 prepared by mixing 1 mL of homogenized community (n=10) followed by centrifugation and

476 passage through a 0.2 µm filter to remove microbial cells. Recognising the possibility that ultra
477 micro-bacteria might escape sedimentation and pass through the filter, metagenomic sequence
478 reads were recruited to seven available reference genomes. A maximum of 0.018 % of reads
479 across all horizontal and vertical mesocosms at 48 weeks matched these reference genomes
480 (Supplementary Fig. 10).
481
482 **Bioinformatic analysis**
483 DNA sequences were demultiplexed using bcl2fastq, paired ends were joined using FLASh [63]
484 and initial fastq files were generated. Preprocessing of fastq files was performed using PrinSeq
485 [64] with a minimum length of 100 base pairs, minimum quality score of 25, and a maximum
486 percentage of N's of 10%. All metagenomes were uploaded to the MG-RAST metagenomic
487 analysis server and are publicly available (Supplementary Table 1). Gene category abundances
488 (Supplementary Fig. 5) and genus identifications (Supplementary Table 6) were determined
489 using the MG-RAST pipeline [65]. Sequence matches were determined using BLASTn [66] with
490 a minimum e-value threshold of 1E-05, minimum alignment length of 100 base pairs, and a
491 minimum percentage identity of 90%. "Unique" Horizontal and Vertical sequences were
492 assembled using MEGAHIT [67], ORFs with a minimum length of 100 amino acids were
493 extracted using getORF [68], compared to the Conserved Domain Database (CDD) [35] using
494 RPS-BLAST, and top domain hits were extracted (Supplementary Table 4). Phage-like contigs
495 were classified based on phage-specific protein domains (Supplementary Table 3) and their
496 distribution was determined using Fr-Hit [69]. Sequences matches to phage-like contig were
497 based on a minimum alignment length of 100 base pairs with a percentage identity of at least
498 90%. Nitrogen-related genes were based on gene function classifications and descriptions with a
499 total of 136 domains (Supplementary Table 9). To identify unique sequences a custom
500 bioinformatic pipeline was developed (Supplementary Methods Fig. 1). Each Horizontal
501 metagenome was compared to the founding T_0 metagenome (before the Horizontal vs. Vertical
502 split) and the paired Vertical metagenome using BLASTn. Horizontal-T_0 and Horizontal-Vertical
503 sequence matches were removed, thus leaving a set of unique Horizontal sequences that could
504 not be identified in neither the founding T_0-metagenome, nor the paired Vertical-metagenome.
505 Sequence matches were determined using a minimum e-value threshold of 1E-05, minimum
506 alignment length of 100 base pairs, and a minimum percent identity of 90%. The unique

507 horizontal sequences were then assembled, Open Reading Frames (ORFs) were extracted, and
508 predicted protein domains were determined through comparison to the Conserved Domain
509 Database (CDD). To determine the impact of increasing the sequencing depth of the T_0
510 communities on the final percentage of unique sequences, T_0 metagenomes were sequenced
511 using the HiSeq platform and reads from each mesocosm split arbitrarily into eight metagenomes
512 each consisting of about ~40 million sequences. The unique analysis pipeline was then applied
513 using the first arbitrary T_0 metagenome and any matching sequences were removed. The process
514 was repeated with all eight T_0 arbitrarily defined metagenomes thus revealing the extent to which
515 additional sequencing affected ability to detect rare sequences.
516
517 **Nitrogen assays**
518 Vertical and Horizontal communities from the one-year time point (24 transfers, T_{24}) and T_0
519 were re-established by placing 100 µl of glycerol stock into 20 mL of M9 minimal media
520 supplemented with a 4 cm^2 piece of cellulose paper and incubated for two weeks. 1 mL of the
521 cellulose-community slurry was transferred to 19 mL of M9 Minimal Media with a fresh piece of
522 cellulose and incubated for an additional two weeks. 1 mL of the community was then
523 transferred to a new bottle and 100 µl of surrounding media was sampled at various time points
524 during the two-week incubation period to determine ammonia/ammonium, nitrate, and nitrite
525 concentrations. The concentrations of all three nitrogen species were determined using
526 fluorometric assay kits according the manufacturers protocol (Ammonia Assay Kit, Sigma-
527 Alrdrich, and nitrate/nitrite fluorometric assay kit) (Fig. 6 and Supplementary Table 10).
528
529 **Data availability**
530 All metagenomes are publicly available here: https://www.mg-
531 rast.org/mgmain.html?mgpage=project&project=mgp18485. Supplementary Table 1 contains
532 metagenome descriptions and MG-RAST identifiers
533
534 All associated datasets are publicly available here: https://zenodo.org/record/3293312 and
535 Supplementary Table 12 contains file descriptions and associated metadata
536
537 **Supplementary files**

538 Supplementary Figure 1: Changes in diversity.
539
540 Supplementary Figure 2: Changes in patterns of diversity seen in rank abundance plots.
541
542 Supplementary Figure 3: Among community diversity.
543
544 Supplementary Figure 4: Contigs assembled from unique sequence reads.
545
546 Supplementary Figure 5: Ecologically significant genes.
547
548 Supplementary Figure 6: Functional gene categories.
549
550 Supplementary Figure 7: Distinguishing sequences transferred by SGEs from rare sequences.
551
552 Supplementary Figure 8: Nitrogen-related genes are enriched in "unique" horizontal sequences.
553
554 Supplementary Figure 9: No change in function of communities subject to the vertical transfer
555 regime after 48 weeks.
556
557 Supplementary Figure 10: DNA sequence reads mapping to ultra-microbacteria reference
558 genomes.
559
560 Supplementary Methods Figure 1: Bioinformatic pipeline for identifying "unique" sequences.
561
562 Supplementary Table 1: Genus abundance of all sequenced time points for all
563 mesocosms based on the MG-RAST annotation pipeline.
564
565 Supplementary Table 2: MGRAST IDs and metadata associated with metagenomes.
566
567 Supplementary Table 3: Summary of unique sequences assembled from T_1 Horizontal
568 mesocosms.
569
570 Supplementary Table 4: Sequences matching to phage-like contigs in Fig. 2;
571 annotation of phage-like contigs depicted in Fig. 2; all T_1 unique contigs containing an
572 ORF associated with bacteriophages.
573
574 Supplementary Table 5: Number of unique T_1 horizontal sequences containing an
575 ORF with a domain of interest depicted in Supplementary Fig. 2.
576
577 Supplementary Table 6: Full annotation of the two phage-like and two beta-
578 glucosidase contigs depicted in Fig. 4.
579
580 Supplementary Table 7: The relative abundance of functional gene categories in T_0,
581 T_{24} vertical, and T_{24} horizontal metagenomes based on the MG-RAST annotation
582 pipeline.
583

584 Supplementary Table 8: Summary of unique sequences assembled from T_{24} horizontal
585 and T_{24} vertical metagenomes.
586
587 Supplementary Table 9: Number of nitrogen-related ORFs assembled from unique
588 T24 horizontal sequences depicted in Fig. 5b; number of nitrogen-related ORFs
589 assembled from unique T24 horizontal sequences.
590
591 Supplementary Table 10: Nitrate, nitrite, and ammonia / ammonium concentrations of
592 T_{24} horizontal and T_{24} vertical mesocosms at the end of the two-week incubation
593 period depicted in Figure 6a.
594
595 Supplementary Table 11: Time series of ammonia / ammonium concentrations of T_{24}
596 horizontal and T_{24} vertical mesocosms during the two-week incubation period.
597
598 Supplementary Table 12: Description of all associated datasets publicly available at
599 https://zenodo.org/record/3293312.

References

[1] Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., Cordero, O. X., Brown, S. P., Momeni, B., Shou, W., et al. 2016 Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J.* **10**, 2557-2568. (DOI:10.1038/ismej.2016.45).

[2] Knight, R., Callewaert, C., Marotz, C., Hyde, E. R., Debelius, J. W., McDonald, D. & Sogin, M. L. 2017 The microbiome and human biology. *Annu Rev Genomics Hum Genet* **18**, 65-86. (DOI:10.1146/annurev-genom-083115-022438).

[3] Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V. & Knight, R. 2018 Current understanding of the human microbiome. *Nat Med* **24**, 392-400. (DOI:10.1038/nm.4517).

[4] Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S. & Banfield, J. F. 2004 Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43. (DOI:10.1038/nature02340).

[5] Allison, S. D. & Martiny, J. B. 2008 Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A* **105 Suppl 1**, 11512-11519. (DOI:10.1073/pnas.0801925105).

[6] Fuhrman, J. A., Cram, J. A. & Needham, D. M. 2015 Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* **13**, 133-146. (DOI:10.1038/nrmicro3417).

[7] Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., Thomas, B. C., Singh, A., Wilkins, M. J., Karaoz, U., et al. 2016 Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, 13219. (DOI:ARTN 13219 10.1038/ncomms13219).

[8] Bardgett, R. D., Freeman, C. & Ostle, N. J. 2008 Microbial contributions to climate change through carbon cycle feedbacks. *ISME J* **2**, 805-814. (DOI:10.1038/ismej.2008.58).

[9] Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D. Y., Paulsen, I., Nelson, K. E., Nelson, W., et al. 2004 Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74. (DOI:10.1126/science.1093857).

[10] Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. 2017 Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* **35**, 833-844. (DOI:10.1038/nbt.3935).

[11] Koskella, B., Hall, L. J. & Metcalf, C. J. E. 2017 The microbiome beyond the horizon of ecological and evolutionary theory. *Nat Ecol Evol* **1**, 1606-1615. (DOI:10.1038/s41559-017-0340-2).

[12] Hansen, S. K., Rainey, P. B., Haagensen, J. A. & Molin, S. 2007 Evolution of species interactions in a biofilm community. *Nature* **445**, 533-536.

[13] Lawrence, D., Fiegna, F., Behrends, V., Bundy, J. G., Phillimore, A. B., Bell, T. & Barraclough, T. G. 2012 Species interactions alter evolutionary responses to a novel environment. *PLoS Biol* **10**, e1001330. (DOI:10.1371/journal.pbio.1001330).

[14] Hug, L. A. & Co, R. 2018 It takes a village: Microbial communities thrive through interactions and metabolic handoffs. *mSystems* **3**, e00152-00117. (DOI:ARTN e00152-17 10.1128/mSystems.00152-17).

645 [15] Robinson, C. D., Klein, H. S., Murphy, K. D., Parthasarathy, R., Guillemin, K. &
646 Bohannan, B. J. M. 2018 Experimental bacterial adaptation to the zebrafish gut reveals a
647 primary role for immigration. *PLoS Biol.* **16**. (DOI:ARTN e2006893
648 10.1371/journal.pbio.2006893).
649 [16] Marbouty, M., Baudry, L., Cournac, A. & Koszul, R. 2017 Scaffolding bacterial genomes
650 and probing host-virus interactions in gut microbiome by proximity ligation (chromosome
651 capture) assay. *Sci. Adv.* **3**, e1602105. (DOI:10.1126/sciadv.1602105).
652 [17] Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. 2017 Microbial strain-
653 level population structure and genetic diversity from metagenomes. *Genome Res* **27**, 626-
654 638. (DOI:10.1101/gr.216242.116).
655 [18] Garud, N. R., Good, B. H., Hallatschek, O. & Pollard, K. S. 2019 Evolutionary dynamics of
656 bacteria in the gut microbiome within and across hosts. *PLoS Biol* **17**, e3000102.
657 (DOI:10.1371/journal.pbio.3000102).
658 [19] Yaffe, E. & Relman, D. A. 2019 Tracking microbial evolution in the human gut using Hi-C.
659 *BioRxiv*, http://dx.doi.org/10.1101/594903.
660 [20] Doolittle, W. F. & Sapienza, C. 1980 Selfish genes, the phenotype paradigm and genome
661 evolution. *Nature* **284**, 601-603.
662 [21] Orgel, L. E. & Crick, F. H. C. 1980 Selfish DNA: The ultimate parasite. *Nature* **284**, 604-
663 607.
664 [22] Bergstrom, C. T., Lipsitch, M. & Levin, B. R. 2000 Natural selection, infectious transfer
665 and the existence conditions for bacterial plasmids. *Genetics* **155**, 1505-1519.
666 [23] Burt, A. & Trivers, R. L. 2006 Genes in Conflict: the Biology of Selfish Genetic Elements.
667 Harvard, Cambridge, MA., Belknap Press.
668 [24] Elena, S. F. & Lenski, R. E. 2003 Evolution experiments with microorganisms: the
669 dynamics and genetic bases of adaptation. *Nat Rev Genet* **4**, 457-469.
670 [25] Rainey, P. B., Remigi, P., Farr, A. D. & Lind, P. A. 2017 Darwin was right: where now for
671 experimental evolution? *Curr Opin Genet Dev* **47**, 102-109.
672 (DOI:10.1016/j.gde.2017.09.003).
673 [26] Maltez Thomas, A., Prata Lima, F., Maria Silva Moura, L., Maria da Silva, A., Dias-Neto,
674 E. & Setubal, J. C. 2018 Comparative Metagenomics. *Methods Mol Biol* **1704**, 243-260.
675 (DOI:10.1007/978-1-4939-7463-4_8).
676 [27] Wilson, D. B. 2011 Microbial diversity of cellulose hydrolysis. *Curr Opin Microbiol* **14**,
677 259-263. (DOI:10.1016/j.mib.2011.04.004).
678 [28] Simon, J. 2002 Enzymology and bioenergetics of respiratory nitrite ammonification. *FEMS*
679 *Microbiol Rev* **26**, 285-309. (DOI:10.1111/j.1574-6976.2002.tb00616.x).
680 [29] Goddard, M. R., Godfray, H. C. J. & Burt, A. 2005 Sex increases the efficacy of natural
681 selection in experimental yeast populations. *Nature* **434**, 636-640.
682 (DOI:10.1038/nature03405).
683 [30] McDonald, M. J., Rice, D. P. & Desai, M. M. 2016 Sex speeds adaptation by altering the
684 dynamics of molecular evolution. *Nature* **531**, 233-+. (DOI:10.1038/nature17143).
685 [31] Aminov, R. I. 2011 Horizontal gene exchange in environmental microbiota. *Front*
686 *Microbiol* **2**, 158. (DOI:10.3389/fmicb.2011.00158).
687 [32] Colombi, E., Straub, C., Kunzel, S., Templeton, M. D., McCann, H. C. & Rainey, P. B.
688 2017 Evolution of copper resistance in the kiwifruit pathogen Pseudomonas syringae pv.
689 actinidiae through acquisition of integrative conjugative elements and plasmids. *Environ*
690 *Microbiol* **19**, 819-832. (DOI:10.1111/1462-2920.13662).

691 [33] Hall, J. P. J., Brockhurst, M. A. & Harrison, E. 2017 Sampling the mobile gene pool:
692 innovation via horizontal gene transfer in bacteria. *Phil Trans R Soc Lond B* **372**.
693 (DOI:10.1098/rstb.2016.0424).
694 [34] Casjens, S. 2003 Prophages and bacterial genomics: what have we learned so far? *Mol.*
695 *Microbiol.* **49**, 277-300. (DOI:10.1046/j.1365-2958.2003.03580.x).
696 [35] Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y.,
697 Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., et al. 2015 CDD: NCBI's conserved domain
698 database. *Nucleic Acids Res.* **43**, D222-226. (DOI:10.1093/nar/gku1221).
699 [36] Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. 2013 A bacteriophage
700 encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**,
701 489-491. (DOI:10.1038/nature11927).
702 [37] Ramisetty, B. C. & Santhosh, R. S. 2016 Horizontal gene transfer of chromosomal Type II
703 toxin-antitoxin systems of Escherichia coli. *FEMS Microbiol Lett* **363**.
704 (DOI:10.1093/femsle/fnv238).
705 [38] Bustamante, P. & Iredell, J. R. 2017 Carriage of type II toxin-antitoxin systems by the
706 growing group of IncX plasmids. *Plasmid* **91**, 19-27. (DOI:10.1016/j.plasmid.2017.02.006).
707 [39] Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C. & Pandey, A. 2013 Role and
708 significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production.
709 *Bioresour Technol* **127**, 500-507. (DOI:10.1016/j.biortech.2012.09.012).
710 [40] Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de
711 Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., et al. 2005 The subsystems approach to
712 genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.*
713 **33**, 5691-5702. (DOI:10.1093/nar/gki866).
714 [41] Chu, H. Y., Sprouffske, K. & Wagner, A. 2018 Assessing the benefits of horizontal gene
715 transfer by laboratory evolution and genome sequencing. *BMC Evol. Biol.* **18**, 54.
716 (DOI:10.1186/s12862-018-1164-7).
717 [42] Frazão, N., Sousa, A., Lässig, M. & Gordo, I. 2019 Horizontal gene transfer overrides
718 mutation in Escherichia coli colonizing the mammalian gut. *Proc Natl Acad Sci USA*,
719 201906958. (DOI:10.1073/pnas.1906958116).
720 [43] Zhao, S. J., Lieberman, T. D., Poyet, M., Kauffman, K. M., Gibbons, S. M., Groussin, M.,
721 Xavier, R. J. & Alm, E. J. 2019 Adaptive evolution within gut microbiomes of healthy
722 people. *Cell Host & Microbe* **25**, 656-667. (DOI:10.1016/j.chom.2019.03.007).
723 [44] Wilson, D. S. & Sober, E. 1989 Reviving the superorganism. *J Theor Bio.* **136**, 337-356.
724 [45] Swenson, W., Wilson, D. S. & Elias, R. 2000 Artificial ecosystem selection. *Proc Natl Acad*
725 *Sci U S A* **97**, 9110-9114. (DOI:10.1073/pnas.150237597 150237597).
726 [46] Xie, L., Yuan, A. E. & Shou, W. Y. 2019 Simulations reveal challenges to artificial
727 community selection and possible strategies for success. *PLoS Biol.* **17**. (DOI:ARTN
728 e3000295 10.1371/journal.pbio.3000295).
729 [47] Black, A. J., Bourrat, P. & Rainey, P. B. 2020 Ecological scaffolding and the evolution of
730 individuality. *Nat Ecol Evol* **In press**.
731 [48] Gause, G. F. 1934 *The Struggle for Existence*. Baltimore, Williams & Wilkins.
732 [49] Rosenzweig, R. F., Sharp, R. R., Treves, D. S. & Adams, J. 1994 Microbial evolution in a
733 simple unstructured environment: Genetic differentiation in *Escherichia coli*. *Genetics* **137**,
734 903-917.

735 [50] Rainey, P. B., Buckling, A., Kassen, R. & Travisano, M. 2000 The emergence and
736 maintenance of diversity: insights from experimental bacterial populations. *Trends Ecol Evol*
737 **15**, 243-247.
738 [51] Rosenfeld, J. S. 2002 Functional redundancy in ecology and conservation. *Oikos* **98**, 156-
739 162.
740 [52] Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I.,
741 Ackermann, M., Hahn, A. S., Srivastava, D. S., Crowe, S. A., et al. 2018 Function and
742 functional redundancy in microbial systems. *Nat Ecol Evol* **2**, 936-943.
743 (DOI:10.1038/s41559-018-0519-1).
744 [53] Landsberger, M., Gandon, S., Meaden, S., Rollie, C., Chevallereau, A., Buckling, A.,
745 Westra, E. R. & van Houte, S. 2018 Anti-CRISPR phages cooperate to overcome CRISPR-
746 Cas immunity. *Cell* **174**, 908-916. (DOI:10.1016/j.cell.2018.05.058).
747 [54] Marbouty, M., Cournac, A., Flot, J. F., Marie-Nelly, H., Mozziconacci, J. & Koszul, R.
748 2014 Metagenomic chromosome conformation capture (meta3C) unveils the diversity of
749 chromosome organization in microorganisms. *eLife* **3**, e03318. (DOI:10.7554/eLife.03318).
750 [55] Garud, N. R., Good, B. H., Hallatschek, O. & Pollard, K. S. 2019 Evolutionary dynamics of
751 bacteria in the gut microbiome within and across hosts. *PLoS Biol.* **17**, e3000102.
752 (DOI:10.1371/journal.pbio.3000102).
753 [56] Lam, P. & Kuypers, M. M. 2011 Microbial nitrogen cycling processes in oxygen minimum
754 zones. *Ann Rev Mar Sci* **3**, 317-345. (DOI:10.1146/annurev-marine-120709-142814).
755 [57] Givens, D. I., Adamson, A. H. & Cobby, J. M. 1988 The effect of ammoniation on the
756 nutritive value of wheat, barley and oat straws. II. Digestibility and energy value
757 measurements in vivo and their prediction from laboratory measurements. *Anim Feed Sci*
758 *Tech* **19**, 173-184.
759 [58] Goldenfeld, N. & Woese, C. 2007 Biology's next revolution. *Nature* **445**, 369.
760 (DOI:10.1038/445369a).
761 [59] Koonin, E. V. 2009 Darwinian evolution in the light of genomics. *Nucleic Acids Res* **37**,
762 1011-1034. (DOI:10.1093/nar/gkp089).
763 [60] Boto, L. 2010 Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci* **277**,
764 819-827. (DOI:10.1098/rspb.2009.1679).
765 [61] Ochman, H., Lawrence, J. G. & Groisman, E. A. 2000 Lateral gene transfer and the nature
766 of bacterial innovation. *Nature* **405**, 299-304.
767 [62] Werren, J. H. 2011 Selfish genetic elements, genetic conflict, and evolutionary innovation.
768 *Proc Natl Acad Sci U S A* **108 Suppl 2**, 10863-10870. (DOI:10.1073/pnas.1102343108).
769 [63] Magoc, T. & Salzberg, S. L. 2011 FLASH: fast length adjustment of short reads to improve
770 genome assemblies. *Bioinformatics* **27**, 2957-2963. (DOI:10.1093/bioinformatics/btr507).
771 [64] Schmieder, R. & Edwards, R. 2011 Quality control and preprocessing of metagenomic
772 datasets. *Bioinformatics* **27**, 863-864. (DOI:10.1093/bioinformatics/btr026).
773 [65] Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P., Paczian, T.,
774 Trimble, W. L., Bagchi, S., Grama, A., et al. 2016 The MG-RAST metagenomics database
775 and portal in 2015. *Nucleic Acids Res.* **44**, D590-594. (DOI:10.1093/nar/gkv1322).
776 [66] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990 Basic local
777 alignment search tool. *J Mol Biol* **215**, 403-410. (DOI:10.1016/S0022-2836(05)80360-2).
778 [67] Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. 2015 MEGAHIT: an ultra-fast
779 single-node solution for large and complex metagenomics assembly via succinct de Bruijn
780 graph. *Bioinformatics* **31**, 1674-1676. (DOI:10.1093/bioinformatics/btv033).

781 [68] Rice, P., Longden, I. & Bleasby, A. 2000 EMBOSS: the European Molecular Biology Open
782 Software Suite. *Trends Genet.* **16**, 276-277.
783 [69] Niu, B., Zhu, Z., Fu, L., Wu, S. & Li, W. 2011 FR-HIT, a very fast program to recruit
784 metagenomic reads to homologous reference genomes. *Bioinformatics* **27**, 1704-1705.
785 (DOI:10.1093/bioinformatics/btr252).

786 **Figure Captions**

787

788 **Figure 1: Genus-level rank abundance curves.**

789

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

795

796

797 **Figure 2: Experimental protocol for manipulation SGEs and thus the extent of HGT**

798

799 Cellulose-degrading microbial communities were established by placing ten independent 1 gram
800 samples of fresh compost (taken from separate locations in a single compost heap) into minimal
801 M9 medium contained within a glass mesocosm with paper as the sole carbon source. Following
802 four weeks of incubation at room temperature communities were homogenized and divided in
803 two. Thereafter — and for the next 48 weeks — one of each pair was subject to a "vertical" and
804 the other a "horizontal" transfer regime. The vertical regime involved serial transfer of a sample
805 from each community every two weeks. The horizontal regime involved the same serial transfer
806 protocol, but additionally, at the time of transfer a sample of supernatant was collected from each
807 of the ten horizontal communities and passed through a 0.2 μm filter producing a "SGE-
808 cocktail". This sample of ten SGE-cocktails was pooled (indicated by the 10 different coloured
809 phage cartoons) and redistributed across all horizontal microcosms. Note: DNA was not
810 extracted from the SGE-cocktail prior to mixing (or subsequently).

811

812

813 **Figure 3: Movement of phage-like elements between horizontal communities**

814

815 Twenty phage-like contigs assembled from sequence reads unique to horizontal communities are
816 listed on the y-axes. Each vertical line represents a single pair of mesocosms with communities

817 subject to the horizontal and vertical regimes depicted in blue and orange, respectively. The x-
818 axis of the butterfly plot is the number of reads mapping to each contig. Astrices denote the
819 source community from which the contigs were assembled.
820
821
822 **Figure 4: Dynamics of contigs unique to horizontal communities**
823
824 The dynamics of two-phage like elements (**a**, **b**) and two elements containing predicted beta
825 glucosidases (**c**, **d**) identified from assembling unique reads from horizontal communities at time
826 point T_1. Each mesocosm is represented as a vertical line and dynamics were tracked through
827 the course of the 48 week experiment. The butterfly plot depicts normalised abundance of
828 sequence reads mapping to each of the representative contigs from horizontal and vertical
829 mesocosms (blue and orange, respectively). Grey / white boxes above each figure depict
830 presence / absence of sequences from each contig in the founding T_0 community (before the
831 vertical / horizontal split).
832
833
834 **Figure 5: Enrichment of nitrogen metabolism genes in horizontal communities**
835
836 **a,** Relative abundance of nitrogen metabolism gene categories based on sequence annotation
837 using MG-RAST. Data are shown as box and whisker plots depicting median, inter quartile
838 range (box) and full data spread from 10 replicate communities. **b,** Total number of Open
839 Reading Frames (ORFs) predicted to encode specific domains associated with ammonification,
840 regulation of nitrogen metabolism, and other associated nitrogen metabolism genes in the
841 assemblies of unique vertical and unique horizontal sequences. Data are shown as box and
842 whisker plots depicting median, inter quartile range (box) and full data spread from 10 replicate
843 mesocosms. 1, NifS; 2, NifE; 3, NapA; 4, nitrilase; 5, NtrA; 6, NifR3; 7, NtrY; 8, NifL; 9, PtsN;
844 10, NtrB; 11, NtrC; 12, NAC; 13, FixI; 14, FixG; 15 & 16 nitroreductases; 17, arginase; 18,
845 cytochrome D1. Full descriptions of domains are provided in Supplementary Table 9.
846
847

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

849

850 **a,** Concentrations of nitrate, nitrite, and ammonia / ammonium from mesocosms subject to
851 horizontal and vertical regimes at 48 weeks (T_{24}). Data are shown as box and whisker plots
852 depicting median, inter quartile range (box) and full data spread from 10 replicate communities.
853 To test for differences in nitrate, nitrite and ammonia / ammonium concentrations at time T_{24} for
854 the ten communities in vertical and horizontal regimes (3 independent measures per community)
855 were compared using a two-sample Kolmogorov-Smirnov test (implemented in python 3.7.2,
856 package scipy 1.3.1). Nitrate and nitrite concentration distributions are not significantly different
857 in the horizontal compared to vertical regime (both $P > 0.1$). However, the distribution of
858 ammonia / ammonium concentrations were significantly different (KS-statistic = 0.6, N = 30 per
859 sample, $P < 0.001$). Analysis was repeated using a Linear Mixed Effect Model. Community
860 identity (1 - 10) was assigned as a random effect and regime (horizontal / vertical) as a fixed
861 effect (implemented in python 3.7.2, package stats model 0.10.2, restricted maximum likelihood
862 fit). Results are in agreement with the KS-test. Nitrate and nitrite concentrations were not
863 significantly affected by regime (for both Wald-test results, $P > 0.3$). However, the
864 concentration of ammonia / ammonium was significantly higher in the horizontal regime (mean
865 effect = 1.207, SD = 0.151, 95% CI: [0.912, 1.503], N = 60, Wald-test, $P < 0.001$). **b,** Ammonia
866 / ammonium concentrations measured at nine time points during the two-week incubation period
867 for communities from horizontal and vertical regimes at 48 weeks. Data are shown as box and
868 whisker plots depicting median, inter quartile range (box) and full data spread from 10 replicate
869 communities (triplicate measures were obtained from each community on each occasion).
870 Analysis of time series data using a mixed ANOVA design showed a significant effect of
871 treatment (horizontal vs. vertical; $P = 1.96 \times 10^{-5}$) and a significant effect of time ($P = 2.7 \times 10^{-}$
872 ^{24}, after Greenhouse & Geisser correction for sphericity). The interaction between treatment
873 regime and time was not significant $P = 0.69$). Comparison between the vertical regime at 48
874 (T_{24}) and community function at the beginning of the experiment (T_0) showed no difference (see
875 Supplementary Fig. 9).

Acknowledgements

We thank members of the Rainey lab at ESPCI and the MPI for Evolutionary Biology, and especially Dave Rogers and Andy Farr, 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

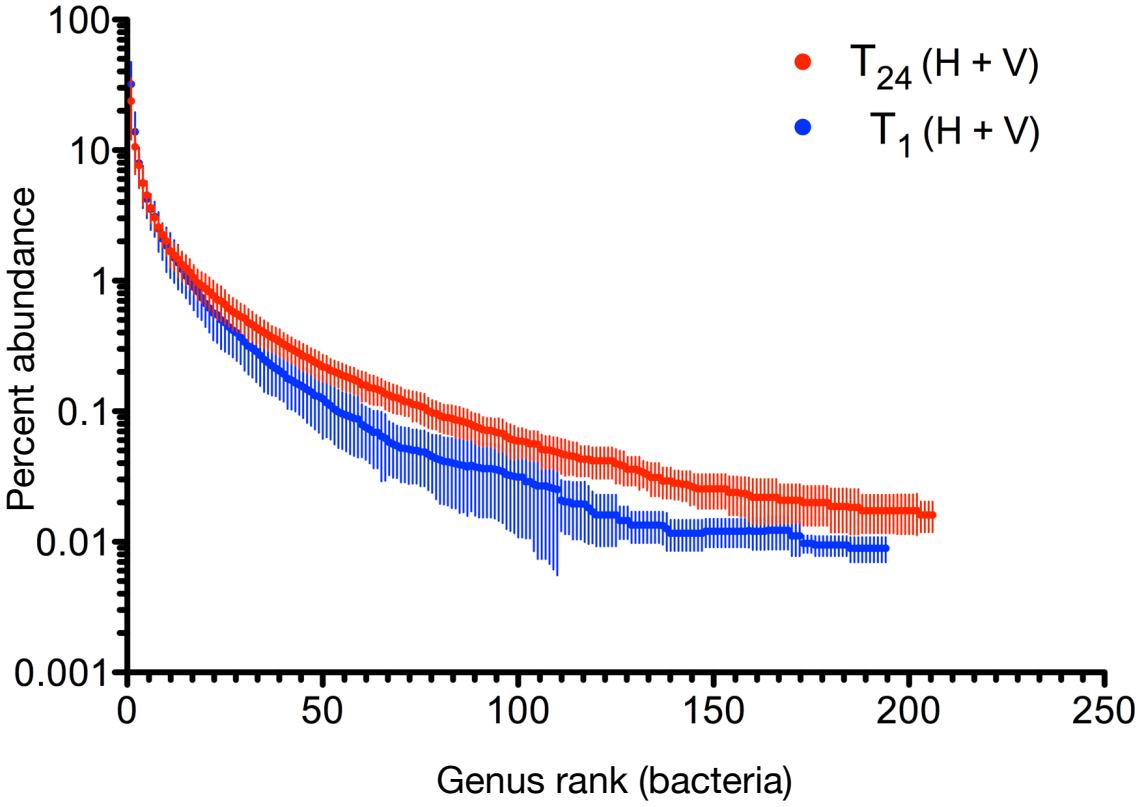

Figure 1: Genus-level rank abundance curves

The blue line depicts the genus-level rank abundance curve for bacteria 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.

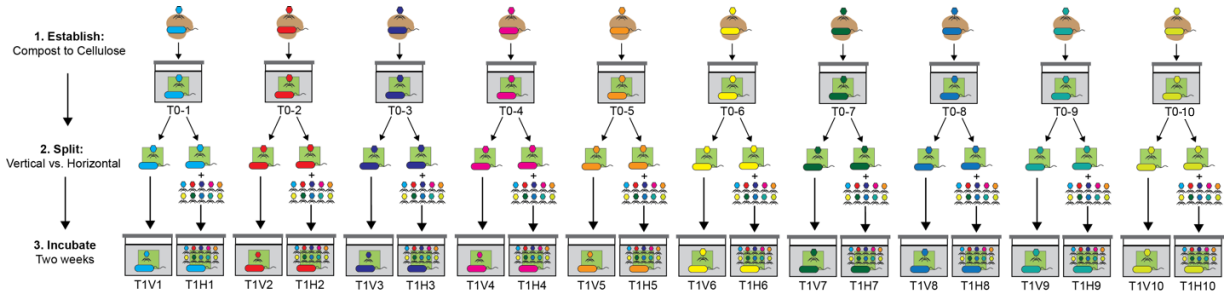

Figure 2: Experimental protocol for manipulation SGEs and thus the extent of HGT

Cellulose-degrading microbial communities were established by placing ten independent 1 gram samples of fresh compost (taken from separate locations in a single compost heap) into minimal M9 medium contained within a glass mesocosm with paper as the sole carbon source. Following four weeks of incubation at room temperature communities were homogenized and divided in two. Thereafter — and for the next 48 weeks — one of each pair was subject to a "vertical" and the other a "horizontal" transfer regime. The vertical regime involved serial transfer of a sample from each community every two weeks. The horizontal regime involved the same serial transfer protocol, but additionally, at the time of transfer a sample of supernatant was collected from each of the ten horizontal communities and passed through a 0.2 µm filter producing a "SGE-cocktail". This sample of ten SGE-cocktails was pooled (indicated by the 10 different coloured phage cartoons) and redistributed across all horizontal microcosms. Note: DNA was *not* extracted from the SGE-cocktail prior to mixing (or subsequently).

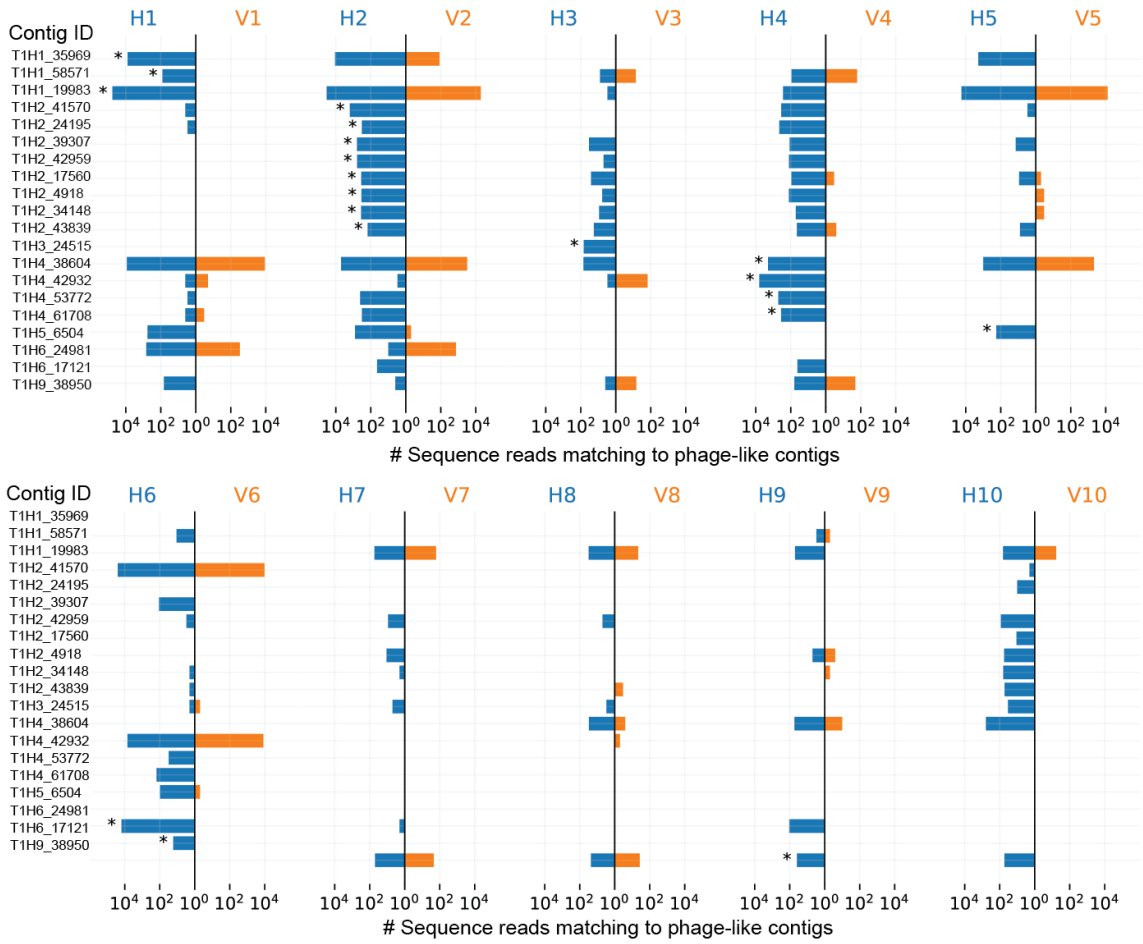

Figure 3: Movement of phage-like elements between horizontal communities

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.

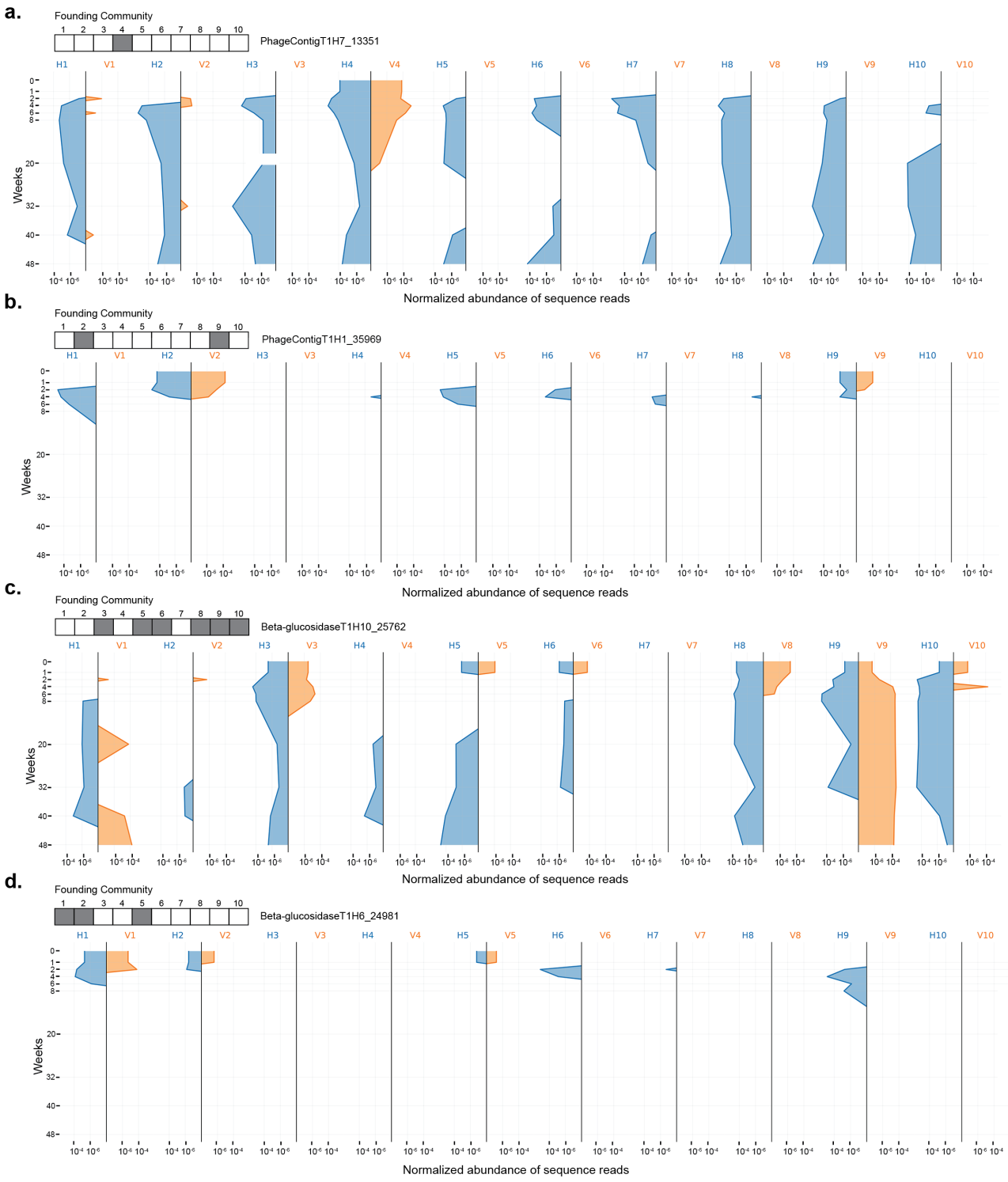

Figure 4: Dynamics of contigs unique to horizontal communities

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).

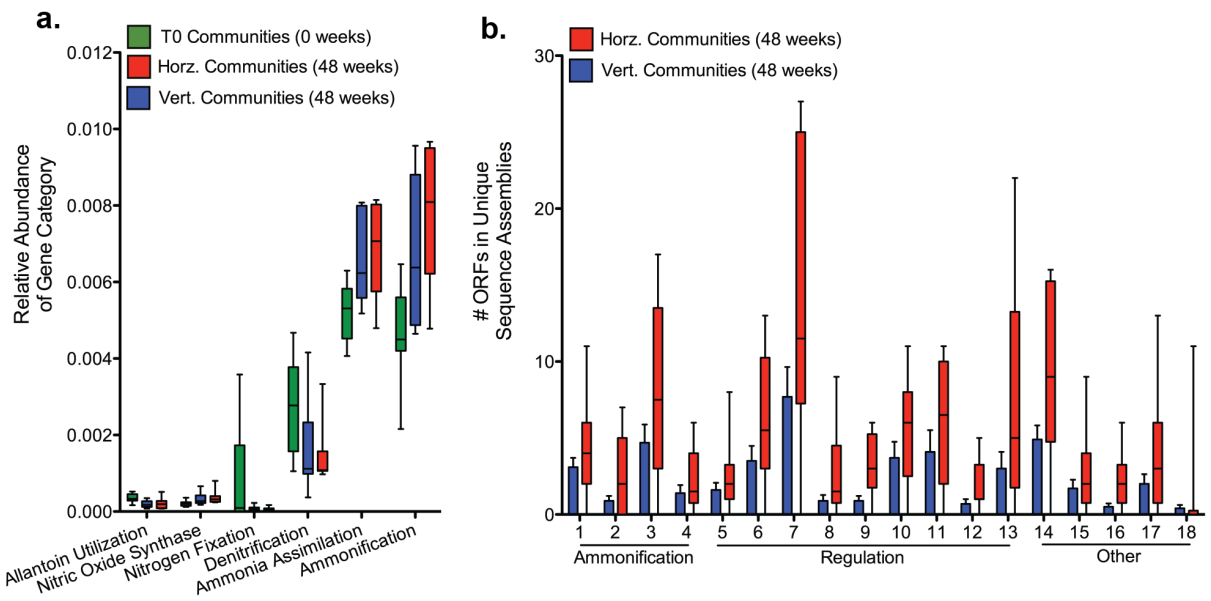

Figure 5: Enrichment of nitrogen metabolism genes in horizontal communities

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. **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. 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.

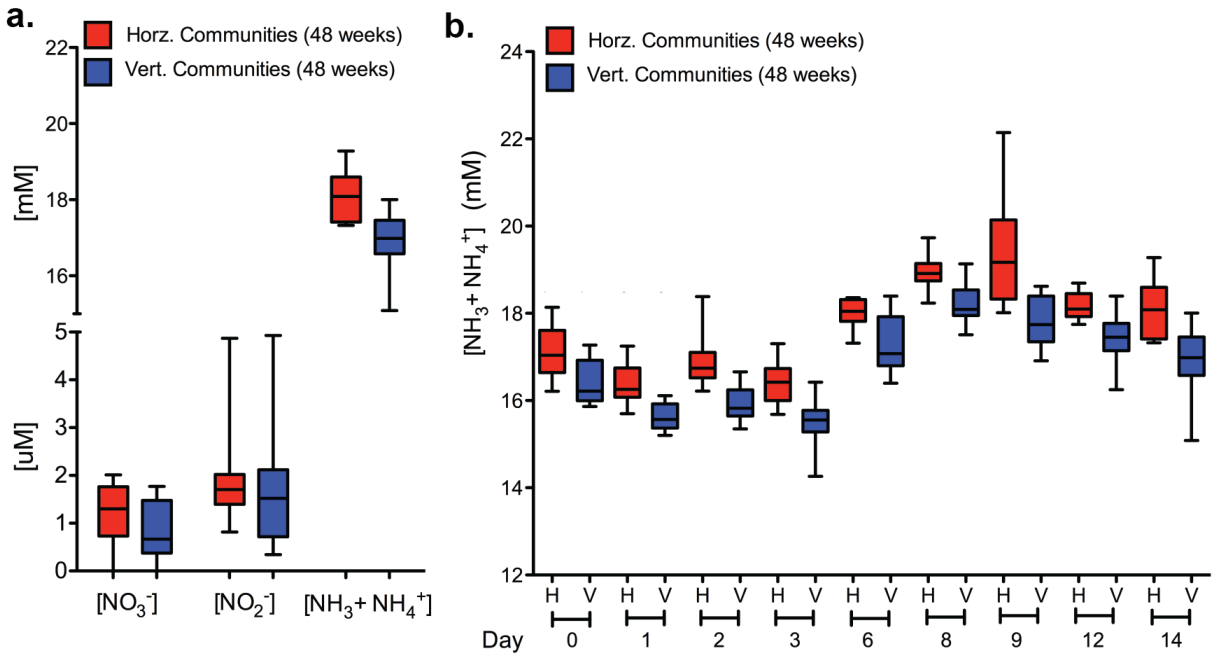

Figure 6: Enhanced ammonia production in horizontal communities.

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. To test for differences in nitrate, nitrite and ammonia / ammonium concentrations at time T_{24} for the ten communities in vertical and horizontal regimes (3 independent measures per community) were compared using a two-sample Kolmogorov-Smirnov test (implemented in python 3.7.2, package scipy 1.3.1). Nitrate and nitrite concentration distributions are not significantly different in the horizontal compared to vertical regime (both $P > 0.1$). However, the distribution of ammonia / ammonium concentrations were significantly different (KS-statistic = 0.6, N = 30 per sample, $P < 0.001$). Analysis was repeated using a Linear Mixed Effect Model. Community identity (1 - 10) was assigned as a random effect and regime (horizontal / vertical) as a fixed effect (implemented in python 3.7.2, package stats model 0.10.2, restricted maximum likelihood fit). Results are in agreement with the KS-test. Nitrate and nitrite concentrations were not significantly affected by regime (for both Wald-test results, $P > 0.3$). However, the concentration of ammonia / ammonium was significantly higher in the horizontal regime (mean effect = 1.207, SD = 0.151, 95% CI: [0.912, 1.503], N = 60, Wald-test, $P < 0.001$). **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. 9).