



LAWRENCE  
LIVERMORE  
NATIONAL  
LABORATORY

# Codon bias, nucleotide selection, and genome size predict in situ bacterial growth rate and transcription in rewetted soil

P. F. Chuckran, K. Y. Estera-Molina, A. M. Nicolas, E. T. Sieradzki, P. Dijkstra, M. K. Firestone, J. Pett-Ridge, S. J. Blazewicz

July 12, 2024

Proceedings of the National Academy of Sciences

## **Disclaimer**

---

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

1 **Codon bias, nucleotide selection, and genome size predict *in situ* bacterial growth rate and**  
2 **transcription in rewetted soil**

3 Peter F. Chuckran<sup>1\*</sup>, Katerina Estera-Molina<sup>1,2</sup>, Alexa M. Nicolas<sup>1,3,4</sup>, Ella T. Sieradzki<sup>1,5</sup>, Paul  
4 Dijkstra<sup>6</sup>, Mary K. Firestone<sup>1</sup>, Jennifer Pett-Ridge<sup>2,7,8</sup>, Steven J. Blazewicz<sup>2</sup>

5

6 <sup>1</sup>. Department of Environmental Science, Policy, and Management, University of California,  
7 Berkeley, CA, USA

8 <sup>2</sup>. Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore,  
9 CA, USA

10 <sup>3</sup>. Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

11 <sup>4</sup>. Department of Biological Sciences, Columbia University, NY, NY, USA

12 <sup>5</sup>. Laboratoire Ampère, École Centrale de Lyon, Lyon, France

13 <sup>6</sup>. Center for Ecosystem Science and Society (ECOSS) and Department of Biological Sciences,  
14 Northern Arizona University, Flagstaff, AZ, USA

15 <sup>7</sup>. Life & Environmental Sciences Department, University of California Merced, Merced, CA,  
16 USA

17 <sup>8</sup>. Innovative Genomics Institute, University of California Berkeley, Berkeley, CA, USA

18 \* Corresponding author

19

## 20 ABSTRACT

21 In soils, the first rain after a prolonged dry period represents a major pulse event impacting soil  
22 microbial community function, yet we lack a full understanding of the genomic traits associated  
23 with the microbial response to rewetting. Genomic traits such as codon usage bias and genome  
24 size have been linked to bacterial growth in soils—however, often through measurements in  
25 culture. Here, we used metagenome-assembled genomes with  $^{18}\text{O}$ -water stable isotope probing  
26 and metatranscriptomics to track genomic traits associated with growth and transcription of soil  
27 microorganisms over one week following rewetting of a grassland soil. We found that codon bias  
28 in ribosomal protein genes was the strongest predictor of growth rate. We also found higher  
29 growth rates in bacteria with smaller genomes, suggesting that reduced genome size enables a  
30 faster response to pulses in soil bacteria. Faster transcriptional upregulation of ribosomal protein  
31 genes was associated with high codon bias and increased nucleotide skew. We found several of  
32 these relationships existed within phyla, indicating that these associations between genomic traits  
33 and activity could be generalized characteristics of soil bacteria. Finally, we used publicly  
34 available metagenomes to assess the distribution of codon bias across a pH gradient and found  
35 that microbial communities in higher pH soils—which are often more water-limited and pulse-  
36 driven—have higher codon usage bias in their ribosomal protein genes. Together, these results  
37 provide evidence that genomic characteristics affect soil microbial activity during rewetting and  
38 pose a potential fitness advantage for soil bacteria where water and nutrient availability are  
39 episodic.

40 SIGNIFICANCE STATEMENT:

41 The first rain event after a dry season represents a period of high activity in soils, during which a  
42 large portion of annual carbon turnover occurs. Despite the importance of this phenomenon, we  
43 lack an understanding of the features which dictate the microbial response to rewetting. We  
44 found that genomic traits such as codon usage, genome size, and nucleotide frequency, were  
45 predictive of activity. This work not only provides insight into the rewetting response, but helps  
46 inform our understanding of the trade-offs embedded in genomic traits and their broad scale  
47 distribution in soils.

## 48 INTRODUCTION

49 Soil microbial communities are often exposed to pulse events (1, 2)—discrete, sudden  
50 changes in the environment (3). Identifying biological traits that predict response of organisms to  
51 pulse events is a central challenge in determining how these events impact ecosystem function  
52 (3). Rewetting of dry soil is a common and ecologically important pulse event where microbial  
53 traits associated with growth and activity may play an outsized role in biogeochemical cycling,  
54 community structure, and function. In ecosystems characterized by dry and wet seasons, the first  
55 rain event following the dry season results in a flux of CO<sub>2</sub> (known as the Birch Effect) (4),  
56 which can account for a large portion of annual soil C loss (5). During these events, rapid  
57 stimulation of microbial activity and growth is driven by both the release of water stress as well  
58 as an influx of bioavailable carbon compounds sourced from osmolytes (6–8), microbial  
59 necromass (9, 10), slaking of microaggregates (11), and increased connectivity in the soil matrix  
60 (12, 13). These microbial responses to rewetting therefore have important consequences for  
61 nutrient cycling and ecosystem function. However, the microbial traits that underly the growth  
62 and activity during these events remain unknown.

63 Trait-based frameworks have long-been used in ecology to help understand complex  
64 community patterns (14, 15), and this approach has gained considerable attention in microbial  
65 ecology (16, 17), where advances in sequencing technology have expanded the ability to probe  
66 microbial attributes. Analysis of bacterial traits enables two key insights into soils. First, since  
67 traits are the product of evolutionary forces, understanding the distribution of traits progresses  
68 fundamental questions concerning the evolution and selection for certain attributes (18). Second,  
69 considering the complexity of microbial community composition, traits serve as valuable metrics  
70 for assessing community dynamics across scales (19, 20) and with ecosystem function (17).

71 Traits associated with growth rate are of particular interest, as growth rate is often central in life-  
72 strategy frameworks (17, 21, 22), functional metrics such as carbon use efficiency (23, 24), and  
73 understanding changing community dynamics.

74 Trait-based analyses using sequence data generally rely on two broad sets of metrics that we  
75 operationally define here: functional potential (i.e. the description of specific functions) and  
76 genomic traits (characteristics of a genome that do not describe taxonomy, anatomy or a specific  
77 function, e.g., GC-%, genome size, codon usage, etc.). Functional potential is often evaluated  
78 based on the presence or absence of functional gene sequences, the completeness of their  
79 pathway, and their expression in different environmental conditions. During growth, expression  
80 of specific genes varies considerably depending on function (25) and environmental constraints  
81 (26). Several functional genes have been associated with growth in soil microbial communities  
82 (27), but the gene-growth relationship is often tied to a specific set of conditions (e.g., favoring  
83 specific nutrient acquisition strategies or metabolic pathways) and may not be predictive outside  
84 of that environment. Outside of ribosomal protein genes, which are reliably upregulated during  
85 growth, there are relatively few genes that have been universally tied to growth. The number of  
86 rRNA gene copies in a genome positively correlates with maximum growth rate (28, 29)—  
87 making rRNA gene copies a useful trait dimension used in microbial ecology trait-based  
88 frameworks (21, 22, 30). However, even the link between rRNA levels and growth/activity is  
89 inconsistent (31), highlighting the limitations of using functional genes and gene transcription as  
90 predictive traits of growth.

91 In contrast, genomic traits (such as genome size, nucleotide frequency, and codon usage)  
92 may provide comprehensive metrics that are predictive of activity and growth (29, 32), despite  
93 not explicitly describing any particular function. For example, transcription and growth rate may

94 be strongly influenced by the frequency with which certain codons occur in a gene sequence.  
95 Generally, this occurs when codon frequency aligns with the tRNA pool (33). Multiple codons  
96 can encode for the same amino acid, and these “synonymous codons” have a greater affinity for  
97 tRNA with corresponding anticodons. When the synonymous codons of a gene sequence occur  
98 in a similar frequency to the pool of tRNA anticodons, this can increase the rate of both  
99 transcription and translation—and the degree of this alignment is referred to as codon  
100 optimization (33). For translation, codon optimization impacts the rate of elongation, protein  
101 folding, initiation, and termination (34–36). For transcription, codon optimization often predicts  
102 mRNA abundance since more optimized codons generally increases mRNA stability (37, 38)  
103 and is related to higher levels of transcription (39, 40). Codon bias—the degree of redundancy of  
104 codon usage in the genetic code for a particular gene or genome—correlates with the degree of  
105 codon optimization (41). High levels of codon bias in ribosomal proteins is associated with rapid  
106 growth in bacteria (29, 42) and has increasingly been used to predict growth rate (32, 43).

107 Outside of an intrinsic relationship between GC content and codon usage (where very high or  
108 low GC content tend to naturally have high levels of codon bias) (33), there appears to be no  
109 clear relationship between doubling time and GC content in bacteria (44). However, the cost of  
110 nucleotide synthesis does influence transcription, and rapidly transcribed genes often use less  
111 energy expensive (‘cheaper’) nucleotides at synonymous sites to reduce the cost of transcription  
112 (45). Using metabolic models of *E. coli*, Chen et al. 2014 (45) estimated that guanine requires  
113 4.6 more ATP to produce than cytosine, and adenine requires 7.7 more ATP to produce than  
114 uracil. Therefore, genes with cytosine and thymine at synonymous sites tend to be transcribed  
115 faster than genes with guanine and adenine (45). At nonsynonymous sites (where a nucleotide  
116 substitution changes the encoded amino acid), there is an inverse relationship between nucleotide

117 cost and the cost of their encoded amino acid sites (45)—such that a higher frequency of more  
118 expensive nucleotides at nonsynonymous sites is associated with higher levels of expression.

119 Genome size is also thought to influence growth rate; however, the relationship is less clear  
120 than codon usage or rRNA copy number. In oligotrophic marine environments, extremely small  
121 genomes are thought to arise in response to nutrient limitation to curb the cost of reproduction  
122 (termed genomic streamlining; (46)), but may have slower maximum growth rates than  
123 copiotrophs with larger genomes (21). Although streamlined genomes have been documented in  
124 soils (47–49), soil bacterial genomes tend to be large relative to other ecosystems (50)—  
125 potentially because of the increased metabolic diversity required to utilize complex substrates  
126 (51). It has been hypothesized that these large genomes might come at the expense of growth rate  
127 (52), as the increased energy required for reproduction in large genomes may slow growth.  
128 However, the evidence regarding the relationship between genome size and growth rate in soil  
129 bacteria is inconclusive (53, 54), necessitating further investigation.

130 In soils, a major hurdle in assessing how genomic traits relate to growth rate has been our  
131 inability to effectively measure these traits *in situ*. Much of the work linking these traits to  
132 growth has been conducted using soil isolates where growth rates are assessed in pure culture  
133 experiments; however, traits associated with media-based growth are not reliable predictors for  
134 microbial growth in a natural environment (54). Recent method and technical advances have  
135 created new opportunities to track microbial growth *in situ*. In quantitative stable isotope probing  
136 (qSIP), the incorporation of added stable isotopes into DNA tracks the growth rate of microbes  
137 (55). This approach has enabled valuable insights into metagenomic features and metagenome-  
138 assembled genomes (MAGs) associated with growth (10, 27, 56, 57), and shows promise for  
139 linking measurable microbial traits and growth rates *in situ*. In this study, we combined

140 metagenomic, metatranscriptomic, and qSIP data to assess how genomic traits correspond with  
141 activity and growth during the rewetting of seasonally dry soil in a Mediterranean grassland.  
142 Specifically, we investigated how traits such as genome size, nucleotide selection, codon usage,  
143 and ribosomal protein nucleotide and codon frequency relate to growth and transcription. We  
144 hypothesized that the bacterial response to wet-up would be associated the following set of  
145 genomic traits: 1) codon bias in ribosomal protein genes would be a strong predictor of both  
146 transcription and growth rate, as based on the well-established relationship between codon bias  
147 and growth (32); 2) fast-growing and transcribing bacteria would use biosynthetically cheaper  
148 nucleotides (i.e. C and T/U, as opposed to G and A) at synonymous substitution sites, and; 3)  
149 genome size would have little impact on growth, as previous studies have shown factors such as  
150 genome and cell size do not strongly relate to growth in soil bacteria (53, 54). Through a  
151 combined multi-omics and qSIP approach, we aim to test these hypotheses and gain a better  
152 understanding of the mechanisms which affect this fundamental response in soils.

## 153 METHODS

154 The field study was located at Hopland Extension and Research Center (HERC),  
155 Hopland, California, USA (39° 00' 14.6" N, 123° 05' 09.1" W), which resides on the ancestral  
156 home of the Shóqowa and Hopland people. The region features a Mediterranean climate of  
157 warm, dry summers and cool, wet winters. *Avena barbata* (wild oat grass) dominated the studied  
158 field site. This study consisted of 16 plots 3.24 m<sup>2</sup>, with rainout shelters constructed around each  
159 plot, either allowing full or 50% mean annual precipitation for the two preceding years. The soils  
160 are of the Squawrock-Witherell complex, with pH of 7.3, total C content of 15.1 mg/g and total  
161 N of 1.5 mg/g. A full description of the field experimental set-up can be found in Fossum et al.  
162 2022 (58).

163 Soils were collected in September 2018 at the end of the dry season and 25 days before  
164 the first rainfall of the wet season. At the time of collection, the soil gravimetric water content  
165 was approximately 3%. Topsoil samples (0-15 cm deep, roughly 4700 cm<sup>3</sup>) were taken from  
166 eight plots (four full and four reduced precipitation). Samples were transported to Lawrence  
167 Livermore National Laboratory where soil from each field plot was separately homogenized and  
168 sieved (2 mm) to remove large rocks and roots.

169

170 *Wet-up experiment*

171 Details of the soil H<sub>2</sub><sup>18</sup>O labeling can be found in Nicolas et al. 2023 and Sieradzki et al. 2022  
172 (10, 59). In brief, homogenized soil from each plot was separated into 11 microcosms containing  
173 5 g of soil each, for 88 microcosms in total. Soils were brought to 22% gravimetric water content  
174 by adding 1 ml of either natural abundance water or 98 atom % H<sub>2</sub><sup>18</sup>O, and then incubated in the  
175 dark within sealed 500 ml mason jars. Four replicates of each sample type (<sup>18</sup>O labeled and

176 unlabeled) from each plot were destructively harvested at 3, 24, 48, 72, and 168 h post wet-up—  
177 in addition to a set of dry soil controls (i.e. 0 h). Samples were immediately frozen in liquid-N<sub>2</sub>  
178 and stored in a freezer at -80 °C.

179

### 180 *Metagenomic qSIP*

181 A full description of the metagenome DNA extraction, sequencing, assembly, binning,  
182 and qSIP calculations is provided in Sieradzki et al., 2022. Briefly, DNA from three plots per  
183 treatment and timepoint was extracted in triplicate using a phenol chloroform extraction protocol  
184 adapted from Barnard et al. (2015) (60) and then pooled. 5 µg DNA samples were then spun at  
185 20 °C for 108 hours (176,284 RCF<sub>avg</sub>) on a Beckman Coulter Optima XE-90 ultracentrifuge in a  
186 cesium chloride solution (density of 1.730 g mL<sup>-1</sup>) to create a density gradient following a  
187 technique previously described in Blazewicz et al. 2020 (61). The contents of the ultracentrifuge  
188 tube were then separated into 36 fractions using Lawrence Livermore National Laboratory's  
189 high-throughput SIP pipeline (57), each of which was assessed for density and DNA  
190 concentration. Fractions were combined into five groups along the density gradient, purified and  
191 concentrated, and then sequenced on an Illumina Novaseq S4 2 x 150 bp platform at Novogene  
192 (Sacramento, CA). Adapters were trimmed and reads were QC filtered using bbdduk (62) and  
193 Sickle (63). QC filtered reads were assembled into contigs using MEGAHIT (v1.2.9; (64)),  
194 binned with MaxBin 2.0 (65) and MetaBAT2 (66), and refined with MetaWrap (67).

195 We combined this set of MAGs with another set generated from a study at a nearby site  
196 at HREC (27) and dereplicated the combined set of MAGs using dRep, which also assesses  
197 MAG completeness (68). Open reading frames were identified using Prodigal (69) and annotated  
198 with KEGG Orthologs (70) using DRAM (71).

199 Reads from each SIP fraction were then mapped against the combined set of MAGs using  
200 BMap (62). The mapping of reads from each fraction was used to calculate atom fraction  
201 excess (AFE) for each MAG (described in ref. 56), which quantitates the amount of isotopic  
202 label incorporated into a genome. Since reproducing genomes incorporate the  $^{18}\text{O}$  isotope from  
203 the added  $^{18}\text{O}$ -water into their DNA, AFE values can be used as an index of growth (55).

204

### 205 *RNA Extraction and sequencing*

206 RNA was extracted from four soil samples from each time-point and precipitation  
207 treatment using the RNeasy PowerSoil Total RNA kit (Qiagen) according to manufacturer  
208 instructions. Extracted RNA was treated with RNase-free DNase (Qiagen) and stored at  $-80^{\circ}\text{C}$ .  
209 RNA concentration was determined using a Qubit fluorometer (Invitrogen) and quality was  
210 assessed using a Nanodrop One Spectrophotometer (Thermo Fisher Scientific Inc). rRNA  
211 depletion and sequencing were performed at the Joint Genome Institute (JGI; Berkeley,  
212 California, USA). From 100 ng of RNA, rRNA was depleted using three QIAseq FastSelect kits  
213 (Qiagen): 5S/16S/23S, rRNA Plant, and rRNA Yeast. One heavily degraded sample was  
214 discarded. Then, paired-end 2 x 151 bp libraries were sequenced for the remaining 47 samples on  
215 an Illumina NovaSeq platform. Generated RNA sequences were prepared for metatranscriptomic  
216 analyses using the JGI Integrated Microbial Genomes (IMG) pipeline v.5.1.5 (72) and can be  
217 found under the GOLD project ID Gp0612223. IMG assemblies are not included in this analysis,  
218 and a more detailed description of sequencing conditions, initial sequence QC, and assembly  
219 details can be found in Chuckran et al. 2024 (73)

220

### 221 *Metatranscriptome analysis*

222 Raw reads were downloaded from the JGI genome portal and were QC filtered using  
223 bbdduk (62) and Sickle (63). BBmap was then used to map QC filtered reads to a concatenated  
224 reference MAG file (minid=0.95). Counts of reads mapping to DRAM-annotated genes were  
225 identified using featureCounts (74). Expression and normalization counts of mapped transcripts  
226 for each annotated gene (excluding rRNA genes) were generated using DESeq2 (75) using  
227 default parameters. Differential expression was calculated as the log<sub>2</sub>-fold change compared to  
228 the gene expression of the dry soil control group.

229

### 230 *Genomic traits*

231 Genomic traits were calculated using custom scripts written in Python (v 3.8.2)—using  
232 the packages *pandas* (76) and NumPy (77); they can be found at  
233 [https://github.com/PChuckran/Wet\\_up\\_traits](https://github.com/PChuckran/Wet_up_traits). To capture genome size for each MAG, we  
234 corrected the measured genome size with the completeness of the bin—for medium to high  
235 quality MAGs (> 70% complete, <10 % contamination) we divided the total assembled base  
236 pairs by genome completeness as a fraction. For each gene in each MAG we calculated the  
237 effective number of codons (ENC') as described by Novembre 2002 (78)—which uses  
238 background nucleotide frequencies to assess levels of codon bias. Lower ENC' values in a gene  
239 represent a fewer number of unique codons used in that gene, and therefore higher codon bias.

240 To determine the relative bias of ribosomal protein genes, we calculated  $\Delta ENC'$  (29):

$$241 \quad \Delta ENC' = \frac{ENC'_{all} - ENC'_{ribo}}{ENC'_{all}}$$

242 where  $ENC'_{all}$  is the effective number of codons for the whole genome, and  $ENC'_{ribo}$  is the  
243 effective number of codons of the ribosomal protein genes. These  $\Delta ENC'$  values represent the

244 degree of bias in ribosomal protein genes relative to the rest of the genome, where higher  $\Delta\text{ENC}$ '  
245 values represent greater relative codon bias.

246 Nucleotide frequencies and skews were determined at synonymous and nonsynonymous  
247 substitution sites of ribosomal protein genes according to nucleotide degeneracy detailed in Chen  
248 et al. 2016 (45). Synonymous site nucleotide frequencies were calculated from nucleotide  
249 frequencies at fourfold degenerate sites (sites where any substitution at that location results in the  
250 same amino acid), and nonsynonymous frequencies from sites where any substitution would  
251 change the encoded amino acid. GC and AT skew were calculated as:

$$252 \quad GC \text{ skew} = \frac{G - C}{G + C}$$

253 And:

$$254 \quad AT \text{ skew} = \frac{A - T}{A + T}$$

255 The cost of amino acid synthesis—measured as the total number of phosphate bonds used for  
256 synthesis—was derived from Akashi & Gojobori 2002 (79).

257 The number of 16S rRNA gene copies has been shown to be associated with growth rate  
258 (29); however, we have not included them in this analysis. Ribosomal RNA genes are often  
259 missing in MAGs assembled from short-reads because of the highly conserved nature of the 16S  
260 rRNA gene (80). Although there are pipelines which address this issue (81), reducing our MAG  
261 dataset to only those with recovered rRNA genes would limit our analysis. The other traits we  
262 focused on are accessible from short-read metagenome assemblies, and are therefore easier to  
263 compare across samples or bins. Future metagenomic SIP work could potentially use long-read  
264 sequencing to more accurately capture patterns in 16S rRNA gene copy traits.

265

266 *Analysis and model selection*

267 All statistical analyses were conducted in R version 4.2.1 (82) using the tidyverse  
268 package (83) and visualized with ggplot2 (84). The effect of the precipitation treatments  
269 preceding wet-up were not considered for this analysis. MAGs were grouped by transcriptional  
270 response according to when each MAG was most transcriptionally active, as measured by the  
271 total percentage of expressed genes which were upregulated:

$$272 \frac{\text{\# of upregulated genes}}{\text{total number of expressed genes}} * 100$$

273 We divided these responses into 4 groups: early responders (most active 3 h post wet-up), middle  
274 responders (most active 24, 48, 72 h), late responders (168 h), and sensitive (down-regulated post  
275 wet-up; Fig. S1). Differences in genomic traits between transcriptional response groups were  
276 determined using an analysis of variance (ANOVA). Tukey's HSD was used for pairwise  
277 comparison between groups. Relationships between genomic traits and AFE values were  
278 determined using multiple linear regression. Since we had several related measures of codon  
279 bias, to reduce the number of redundant variables, we ran an initial set of model comparisons to  
280 assess which codon bias metric best predicted growth rate. Model comparison was conducted  
281 using Akaike information criterion (AIC) (85)—with a threshold of  $-\Delta 4$ , indicating an  
282 improvement of a model with the addition of a parameter.

283

#### 284 *Database analysis*

285 We accessed metagenomes collected by the National Ecological Observation Network  
286 (NEON) in September of 2024 from the JGI Integrated Microbial Genomes and Microbiomes  
287 (IMG/M) database (86). For 148 metagenomes, we downloaded the sequences for bacteria-  
288 specific ribosomal protein genes according to the KEGG Orthology database, with the following

289 KO numbers: KO:K02887, KO:K02888, KO:K02899, KO:K02902, KO:K02911, KO:K02913,  
290 KO:K02914, KO:K02916, KO:K02959, KO:K02963, KO:K02990.

291 For each ribosomal protein gene, we then estimated the Effective Number of Codons  
292 (ENC) as defined by Wright 1990 (87). This estimate does not require the background nucleotide  
293 frequencies for a genome and therefore allows us to estimate codon usage for individual genes in  
294 a metagenome without the need for binning. Scaffold read-depth was used to generate a  
295 weighted-average ENC value for each KO number in a metagenome. These metagenome-level  
296 ENC estimations were then paired with plot-level soil pH data collected (88).

297

## 298 RESULTS AND DISCUSSION

299           The rewetting of dry soil represents a reoccurring and impactful pulse event in the soil  
300 microbial environment. The response of microbial taxa to this change may be in part determined  
301 by traits; however, the influence of the genomic traits commonly associated with growth and  
302 transcription on the response to wet-up has yet to be determined. In this study, we re-wet  
303 seasonally dry soils in a laboratory incubation to create a time-series of the genomic and  
304 transcriptomic composition of soils in response to rewetting. To link traits associated with  
305 growth and activity during rewetting, we assessed the relationship between genomic traits of  
306 metagenome assembled genomes (MAGs) and the affiliated populations' *in situ* growth rates  
307 (measured via MAG targeted quantitative stable isotope probing; qSIP), and transcriptional  
308 responses (as measured through metatranscriptomics).

309

310 *Traits associated with growth*

311           We assessed the relationship between growth rate and several genome-level and  
312 ribosomal protein gene-level traits. This included: codon usage bias, GC content, amino acid  
313 synthesis cost (~P bonds used in synthesis), average amino acid C:N, AT and GC-skew, and  
314 genome size. We used atom fraction excess (AFE) values as the response variable, with isotopic  
315 enrichment being used as index of growth. Functional genes were not considered in our analysis  
316 for two reasons: 1) previous work has examined how functional genes influence the response to  
317 rewetting (59) and for this analysis we were specifically interested in how genomic traits would  
318 relate to growth; and 2) a comparative analysis of functional genes vs traits would be difficult to  
319 do with the number of MAGs (i.e. the number of gene/trait combinations risk model-overfitting).

320           The best model (indicated through AIC and  $R^2$  values) predicted growth using the codon  
321 bias in ribosomal protein genes (measured as the effective number of codons;  $ENC'_{\text{ribo}}$ ),  
322 ribosomal protein GC content, and genome size ( $R^2 = 0.45$ , Table S1a&b). Amongst these traits,  
323  $ENC'_{\text{ribo}}$  predicted growth better than estimated genome size, or ribosomal GC content (Fig. 1,  
324 Table S1A), supporting our first hypothesis.  $ENC'_{\text{ribo}}$  also predicted growth better than the codon  
325 bias of the entire genome ( $ENC'$ , Table S1A)—demonstrating that ribosomal protein codon bias  
326 can be predictive of growth not only in pure cultures, as previously established (29), but also in  
327 complex soil microbial communities responding to a natural pulse phenomenon. This result is  
328 not only important for identifying traits associated with growth, but also in verifying commonly  
329 used metrics for assessing growth. Although qSIP is commonly used in soil microbial ecology  
330 for measuring growth rate (55, 61, 89–91), and ribosomal codon bias is used in algorithms to  
331 predict growth (32, 92), there has yet to be a comparison of these two metrics in a soil dataset.  
332 The alignment of the two approaches is a promising indicator of the potential for qSIP and codon  
333 usage to capture growth and growth potential respectively. However, this work examines a pulse  
334 response, and more work must be done to understand how traits relate to growth in less dynamic  
335 conditions—as we could potentially expect to see more growth with lower codon bias under  
336 relatively static environmental conditions. Methods such as labeling with  $^{18}\text{O}$  water-vapor (93),  
337 which does not require the direct addition of water to the soil, has been shown as an effective  
338 method to track growth under static conditions (94), and is an exciting future avenue for  
339 understanding trait-growth relationships outside of acute perturbations.

340           Ribosomal GC content was negatively correlated with growth rate (Table S1, Fig. 1C)—  
341 provided that codon bias was high (i.e.  $ENC'_{\text{ribo}}$  was low, Fig. 1C, Fig. S2). Although what  
342 drives this relationship is unclear, these results suggest that a lower GC content poses an

343 advantage for rapid growth in response to wet-up. We hypothesize that given the greater  
344 metabolic (ATP) cost of the GC base pair compared to the AT base pair (45), higher GC content  
345 could slow growth rate. GC content is also associated with metabolic strategy, where lower GC  
346 content is typically associated with greater preference for glycolytic carbon sources, such as  
347 sugars and compounds with a higher carbon to nitrogen ratio (C:N) (95). It's possible that the  
348 sudden increase in bioavailable carbohydrates upon rewetting (96) gives an advantage to lower  
349 GC bacteria. This trend is particularly notable considering that genomic GC-content is generally  
350 predictive of codon bias (33), where more extreme GC values will naturally result in higher  
351 levels of codon bias (an example of this relationship with the MAGs used in this study is shown  
352 in Fig. S3). In this way, the emergence of neutral GC content and high codon bias requires that  
353 each of these traits be selected for independently in taxa with these traits. Further, though we  
354 found a similar trend with genome-GC content, the relationship between growth and ribosomal  
355 GC content was stronger (Table S1), further emphasizing the importance of ribosomal protein  
356 gene expression for growth.

357         Refuting our initial hypothesis that genome size would not impact growth, we found that  
358 smaller genomes (< 3 Mbp) were associated with higher growth rates at later time points,  
359 particularly at 168 h post wet-up (Fig. 1B)—where a 75% reduction in genome size resulted in  
360 ~200% increase in growth rate. Smaller genomes may reduce the cost of replication, a key factor  
361 in responding to sudden changes in environmental conditions. As a corollary, large genomes that  
362 enable metabolic flexibility may come at the expense of growth rate (45). This may also explain  
363 why smaller genomes are more prevalent in arid and carbon poor soils (97–99), since the ability  
364 to quickly respond to the momentary availability of water and resources may be fundamental to  
365 survival and persistence.

366 Our results contrast with observations from marine systems, where free-living bacteria  
367 with small streamlined genomes (1-2 Mbp) tend to have slower growth rates (46). This biome-  
368 scale inconsistency between growth rate and genome size may reflect fundamentally different  
369 life-strategies that both leverage simplified genomes. In marine environments, reduced genomic  
370 complexity and a high surface area to volume ratio occur in response to nutrient limitation—  
371 reducing the total cost of replication while increasing the likelihood of capturing dissolved  
372 nutrients (46). However, these traits in marine environments are typically not associated with  
373 faster growth rates (21). In contrast, for soil communities responding to a rapid pulse of  
374 resources or change in environmental conditions, the lower costs associated with replicating  
375 smaller genomes may allow for faster growth in response to flux in nutrient availability. The  
376 reduced cost of smaller genomes might serve both strategies well, and the relationship between  
377 genome size and growth rate could be environment specific.

378

### 379 *Traits associated with transcription*

380 To assess the influence of genomic traits on transcription, we separated our dataset into 4  
381 temporal categories according to peak transcriptional responses—based on the proportion of  
382 expressed genes that were significantly upregulated at each timepoint (see methods, Fig. S1).  
383 High codon bias, in addition to being a predictor of growth, has also been shown to increase gene  
384 expression (39, 40). We found that the codon bias of ribosomal protein genes was significantly  
385 different between transcriptional response group (ANOVA;  $p < 0.01$ ), with a pairwise  
386 comparison (Tukeys HSD,  $p < 0.05$ ) indicating the exception being between sensitive (down-  
387 regulated post wet-up) and middle or late-responding taxa. Codon bias corresponded with  
388 transcription timing, where early responding taxa had a significantly higher level of codon bias

389 than all other groups, followed by middle-responding taxa, and then late responding and  
390 sensitive taxa (Fig 2A). Regulation of ribosomal protein genes corresponded to transcriptional  
391 response group (Fig. S4), and although this is not a particularly surprising result considering  
392 response categories were curated, it does demonstrate a relationship between codon bias and  
393 transcriptional regulation of individual sets of genes. Although the importance of codon bias for  
394 transcription has been shown for individual taxa (39, 40), this result elucidates that codon usage  
395 may be an important feature in dictating the response of bacteria living in complex soil  
396 communities. Further, this result demonstrates that codon usage is not only important for the  
397 growth response of bacteria during rewetting, but also plays a role in the timing of transcriptional  
398 responses.

399 Nucleotide selection, in addition to its inherent relationship with codon usage, can  
400 influence transcription due to the cost of nucleotide synthesis. Each nucleotide requires a  
401 different amount of energy to synthesize, where  $A > T$ ,  $G > C$ , and  $G+C > A+T$  (45). At  
402 synonymous sites, where nucleotide substitutions do not change the encoded amino acid,  
403 nucleotides which are cheaper to synthesize have been shown to be associated with faster rates of  
404 transcription (45). We found that early responding organisms displayed a lower AT ( $A-T/A+T$ )  
405 and GC skew ( $G-C/G+C$ ) at synonymous sites for ribosomal protein genes (Fig. 2 b&c). Our  
406 findings align with results from Chen et al. (2016) (45), which demonstrated that the lower cost  
407 of U and C at synonymous sites was associated with increased gene transcription. No such effect  
408 was observed at nonsynonymous sites (Fig. 2 b&c), showing that cheaper nucleotide selection is  
409 maintained only when peptide sequence is not impacted. This distinction between synonymous  
410 and non-synonymous sites extends to all transcriptional response categories evidencing the cost-  
411 saving encoding of lower GC-and AT-skew only when amino acid sequence is not compromised

412 (Fig. 2 b&c). Notably, GC skew and AT skew of synonymous substitutions on ribosomal protein  
413 genes were weakly negatively correlated to growth rate at certain time points (Fig. S5A). We did  
414 find a relationship between nonsynonymous nucleotide skew and growth (Fig. S5 C&D), which  
415 we discuss in the Supplemental Results and Discussion. These correlations are weakly  
416 significant in the context of other traits, but suggest that these cost-saving traits may be  
417 associated with fast-growing taxa. These results underscore the disparity between traits affecting  
418 transcription and those having a more pronounced influence on growth.

419

#### 420 *The relationship between growth and transcription*

421 Transcription, although necessary for growth, has been argued to be a poor indicator of  
422 activity (31). Translational efficiency may additionally influence growth rate, and while we  
423 cannot directly measure the contribution of higher transcription vs translational efficiency on  
424 growth in these data, we can assess the relationship between transcriptional response and growth.  
425 We did not find a strong relationship between the overall expression of ribosomal protein genes  
426 and growth rate, as measured by AFE (Fig. 3a), nor did we find a significant relationship  
427 between transcriptional response group and growth rate (Fig. 3b). These results demonstrate that  
428 although transcription may be necessary for growth, levels of RNA may not serve as a reliable  
429 quantitative proxy for growth rate at the community scale, and corroborates prior work indicating  
430 rRNA serves as a poor metric for activity (31).

431 We next explored whether transcriptional response patterns corresponded with growth  
432 rate. We found that the transcriptionally most active early were more likely to be growing early,  
433 and taxa that were most transcriptionally active later were more likely to be growing later (Fig.  
434 3C). This connection between transcriptional activity and growth does suggest that even if

435 transcription may not be a good predictor or proxy for growth, the timing of transcription  
436 generally follows the timing of growth after rewetting.

437

438 *Phylogenetic conservation of trait relationships*

439 Genomic traits such as genome size (100), codon usage, and nucleotide selection are  
440 phylogenetically conserved—raising the question as to the relative impact of these traits within  
441 the context of taxonomy. For example, early growing taxa were dominated by Proteobacteria  
442 (37-57% of growing taxa) (Fig. 4A), which often had higher levels of ribosomal codon bias (Fig.  
443 4B, Fig. S7A). This would suggest that the predictive power of genomic traits are  
444 phylogenetically conserved. The phylogenetic distribution of these traits may explain results  
445 from other rewetting studies, which have found a high abundance of Proteobacteria in response  
446 to wet-up (61, 101).

447 However, the observed relationships between traits and activity could also exist within a  
448 taxonomic group—meaning that the influence of genomic traits on the rewetting response would  
449 be more generalized features of soil bacteria. The number of MAGs recovered in this study does  
450 not allow for a comprehensive quantitative analysis of the contribution of phylogeny vs traits to  
451 activity; however, we can assess whether patterns between traits and growth and transcription  
452 emerge within a phylogenetic group. We chose to analyze the relationships between growth,  
453 transcription, and traits among the two most abundant and prevalent phyla at HREC,  
454 Proteobacteria and Actinobacteria (Fig. 4A). Consistent with total community observations,  
455 higher AFE values tended to correlate with higher levels of codon bias (Fig. 4C). We also found  
456 that transcriptional response categories followed a similar trend to community level observations,  
457 with early transcriptional response being associated with higher levels of codon bias (Fig. 4D).

458 Although we cannot properly assess the influence of other genomic traits in a complex model  
459 due to sample size constraints, we did find some notable trends when separated by phylum:  
460 genome size in Proteobacteria was significantly negatively related to growth at 48 and 168 h  
461 (Fig. S7B), and nucleotide skew followed community-level trends with respect to transcriptional  
462 response category (Fig. S7 C&D). These results suggest that relationships between growth or  
463 transcription with genomic traits are not strictly reliant on phylum-level taxonomy and may  
464 represent fundamental trade-offs for soil bacteria.

465 In contrast, we did not find a relationship between ribosomal GC-content and growth on  
466 the phylum level. Rather, we found this relationship in two fast-growing families of  
467 Proteobacteria: *Sphingomonadaceae* and *Burkholderiaceae* (Fig. S7E). This indicates that low GC-  
468 content is not a trait universally related to growth rate in soil bacterial taxa, but that this is a  
469 phylogenetically conserved relationship. Previous studies observed a high representation of  
470 *Sphingomonadaceae* in growing bacteria post rewetting (101, 102) and these genomic trait  
471 relationships might be related to a unique life strategy posing a fitness advantage—perhaps due  
472 to substrate preference, which has been shown to be related to GC content (95).

473

#### 474 *Broad-scale distribution of codon bias*

475 Our experimental results demonstrate a set of genomic traits important for soil bacteria  
476 experiencing transient water availability—enabling quicker responses to sudden pulse events.  
477 Finding that the codon bias of ribosomal protein genes was the trait most associated with growth,  
478 we were interested in the broad-scale distribution of this trait. The distribution of genome size  
479 and GC content in soils has been the subject of several studies (54, 97, 99, 103, 104), but few  
480 studies have looked at the distribution of codon usage across scales. From the JGI IMG database,

481 we selected 148 metagenomes across 18 sites from the National Ecological Observation Network  
482 (NEON), calculated the codon bias of bacteria-specific ribosomal protein genes (using the  
483 estimate ENC from Wright 1990), and paired these data with measured soil pH. We focused on  
484 soil pH because it is reflective of the water balance in soils, where greater evapotranspiration  
485 relative to precipitation is linked to higher pH (105) and lower carbon availability (106). This  
486 makes soil pH an effective metric for capturing patterns in water pulse dynamics across a large  
487 scale. In low pH soils where water is more persistent and carbon is more available, we tend to  
488 find larger genomes (97) with greater metabolic diversity (99), which is thought to come at the  
489 expense of rapid growth rate (52). We hypothesized that, based on our experimental results,  
490 codon optimization follows this pattern, and demonstrate this trade-off between steady-growth  
491 and low pH and greater capacity for rapid-response at high pH.

492       Accordingly, we found higher levels of ribosomal protein codon bias with higher pH.  
493 This was true when codon bias was calculated combining all bacteria-specific ribosomal protein  
494 genes ( $R^2=0.65$ ,  $p < 0.01$ ), but also for many of these genes individually. Notably, for all  
495 ubiquitous bacteria-specific ribosomal protein genes—i.e. genes found in every metagenome  
496 samples—we found higher codon bias with increasing pH (Fig. 5). These results follow a similar  
497 trend to previous findings showing genome size in soil tends to decrease at higher pH (97, 99).  
498 Together, these relationships complement our experimental SIP wet-up results by indicating that  
499 microbes in higher pH soils—where water is more limiting—may show a greater capacity for  
500 rapid response to resource pulse events. This finding highlights the widespread nature of this  
501 tradeoff.

502

503 *Conclusion*

504 Rapid environmental change such as rewetting may play an important role in the selection of  
505 traits in soil bacteria. Using a combination of metagenome-assembled genomes, transcriptomics,  
506 and <sup>18</sup>O-water qSIP, we evaluated genomic traits associated with transcriptional activity and  
507 growth of soil microorganisms in the period immediately following soil rewetting. Growth was  
508 not proportionally related to levels of transcription; however, timing of transcription and growth  
509 were generally coupled. We found that transcription was associated with nucleotide synthesis  
510 cost as well as codon usage bias, and that faster growth was associated with high codon usage  
511 bias, lower GC content, and smaller genome size. This work highlights the importance of  
512 genomic traits to short-term responses in systems characterized by pulses of water and nutrient  
513 availability, enhancing our understanding of metrics for assessing the functional potential of soil  
514 bacteria and offering important perspective on the broad scale ecological distribution of genomic  
515 traits in soil.  
516

517 Data Availability:

518 Metatranscriptome data are located on the NCBI database, IMG/M database, and JGI genome  
519 portal. Sample names, links to sequence data, and descriptions can be found in the data release  
520 (<https://doi.org/10.1128/mra.00322-24>) (73). MAGs can be accessed from the NCBI database  
521 under the BioProject PRJNA856348 and PRJNA718849.

522

523

524 Acknowledgements:

525 We would like to thank Mengting Maggie Yuan, Erin Nuccio, Aaron Chew, Anne Kakouridis,  
526 Kate Zhalnina, Nameer Baker, Rachel Hestrin, Caleb Herman, Javier Ceja-Navarro, Christina  
527 Ramon, Alex Greenlon, Ryan Gini, Ilexis Chu-Jacoby, Melissa Lafler, Keith Morrison, Amrita  
528 Bhattacharyya, Heejung Cho, Angela Hodge, Albert Molina, Don Herman, and Xiao Bin Max Li  
529 for their help with the experimental set-up and sample processing. We thank Alex Greenlon,  
530 Jillian Banfield, and Rohan Sachdeva for their assistance with MAG creation and bioinformatic  
531 expertise, and Nora Borp and Carl Roybal for intellectual contributions and manuscript proofing.  
532 This research was supported by the U.S. Department of Energy (DOE), Office of Biological and  
533 Environmental Research (BER), Genomic Science Program Lawrence Livermore National  
534 Laboratory (LLNL) ‘Microbes Persist’ Soil Microbiome Scientific Focus Area SCW1632, and a  
535 subaward to UC Berkeley. Field plots and precipitation management were initially generated via  
536 DOE BER awards DE-SC0020163 and DE-SC0016247 (to MKF) and awards SCW1589, and  
537 SCW1421 (to JPR). Sequencing was conducted at the Joint Genome Institute (JGI) via a  
538 Department of Energy Biological and Environmental Support Science award #508594 (to JPR).  
539 Research Work at Lawrence Livermore National Laboratory was conducted under the auspices of  
540 the U.S. DOE under contract DE-AC52-07NA27344.

541

## 542 REFERENCES:

- 543 1. A. Shade, *et al.*, Fundamentals of microbial community resistance and resilience.  
544 *Frontiers in Microbiology* **3**, 417 (2012).
- 545 2. S. L. Collins, *et al.*, Pulse dynamics and microbial processes in aridland ecosystems.  
546 *Journal of Ecology* **96**, 413–420 (2008).
- 547 3. A. Jentsch, P. White, A theory of pulse dynamics and disturbance in ecology. *Ecology*  
548 **100**, e02734 (2019).
- 549 4. H. F. Birch, The effect of soil drying on humus decomposition and nitrogen availability.  
550 *Plant and Soil* **10**, 9–31 (1958).
- 551 5. P. Jarvis, *et al.*, Drying and wetting of Mediterranean soils stimulates decomposition and  
552 carbon dioxide emission: the “Birch effect.” *Tree Physiology* **27**, 929–940 (2007).
- 553 6. N. Fierer, J. P. Schimel, A Proposed Mechanism for the Pulse in Carbon Dioxide  
554 Production Commonly Observed Following the Rapid Rewetting of a Dry Soil. *Soil Science*  
555 *Society of America Journal* **67**, 798 (2003).
- 556 7. E. W. Slessarev, *et al.*, Cellular and extracellular C contributions to respiration after  
557 wetting dry soil. *Biogeochemistry* **147**, 307–324 (2020).
- 558 8. C. R. Warren, Response of osmolytes in soil to drying and rewetting. *Soil Biology and*  
559 *Biochemistry* **70**, 22–32 (2014).
- 560 9. S. J. Blazewicz, E. Schwartz, M. K. Firestone, Growth and death of bacteria and fungi  
561 underlie rainfall-induced carbon dioxide pulses from seasonally dried soil. *Ecology* **95**, 1162–  
562 1172 (2014).
- 563 10. A. M. Nicolas, *et al.*, A subset of viruses thrives following microbial resuscitation during  
564 rewetting of a seasonally dry California grassland soil. *Nat Commun* **14**, 5835 (2023).
- 565 11. P. M. Homyak, *et al.*, Effects of altered dry season length and plant inputs on soluble soil  
566 carbon. *Ecology* **99**, 2348–2362 (2018).
- 567 12. F. E. Moyano, S. Manzoni, C. Chenu, Responses of soil heterotrophic respiration to  
568 moisture availability: An exploration of processes and models. *Soil Biology and Biochemistry* **59**,  
569 72–85 (2013).
- 570 13. A. P. Smith, *et al.*, Shifts in pore connectivity from precipitation versus groundwater  
571 rewetting increases soil carbon loss after drought. *Nature Communications* **8** (2017).
- 572 14. J. P. Grime, Evidence for the Existence of Three Primary Strategies in Plants and Its  
573 Relevance to Ecological and Evolutionary Theory. *The American Naturalist* **111**, 1169–1194  
574 (1977).
- 575 15. C. Violle, *et al.*, Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).
- 576 16. M. Westoby, *et al.*, Trait dimensions in bacteria and archaea compared to vascular  
577 plants. *Ecology Letters* **24**, 1487–1504 (2021).
- 578 17. A. A. Malik, *et al.*, Defining trait-based microbial strategies with consequences for soil  
579 carbon cycling under climate change. *ISME Journal* **14**, 1–9 (2020).
- 580 18. J. B. H. Martiny, S. E. Jones, J. T. Lennon, A. C. Martiny, Microbiomes in light of traits: A  
581 phylogenetic perspective. *Science* **350** (2015).

- 582 19. A. Barberán, A. Fernández-Guerra, B. J. M. Bohannon, E. O. Casamayor, Exploration of  
583 community traits as ecological markers in microbial metagenomes. *Molecular Ecology* **21**,  
584 1909–1917 (2012).
- 585 20. J. Wan, T. W. Crowther, Uniting the scales of microbial biogeochemistry with trait-based  
586 modelling. *Functional Ecology* **36**, 1457–1472 (2022).
- 587 21. F. M. Lauro, *et al.*, The genomic basis of trophic strategy in marine bacteria. *Proceedings*  
588 *of the National Academy of Sciences of the United States of America* **106**, 15527–33 (2009).
- 589 22. N. Fierer, Embracing the unknown: Disentangling the complexities of the soil  
590 microbiome. *Nature Reviews Microbiology* **15**, 579–590 (2017).
- 591 23. R. L. Sinsabaugh, S. Manzoni, D. L. Moorhead, A. Richter, Carbon use efficiency of  
592 microbial communities: Stoichiometry, methodology and modelling. *Ecology Letters* **16**, 930–  
593 939 (2013).
- 594 24. D. Gao, E. Bai, D. Wasner, F. Hagedorn, Global prediction of soil microbial growth rates  
595 and carbon use efficiency based on the metabolic theory of ecology. *Soil Biology and*  
596 *Biochemistry* **190**, 109315 (2024).
- 597 25. S. Klumpp, T. Hwa, Bacterial growth: global effects on gene expression, growth feedback  
598 and proteome partition. *Current Opinion in Biotechnology* **28**, 96–102 (2014).
- 599 26. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, Interdependence of Cell  
600 Growth and Gene Expression: Origins and Consequences. *Science* **330**, 1099–1102 (2010).
- 601 27. A. Greenlon, *et al.*, Quantitative Stable-Isotope Probing (qSIP) with Metagenomics Links  
602 Microbial Physiology and Activity to Soil Moisture in Mediterranean-Climate Grassland  
603 Ecosystems. *mSystems* **7**, e00417-22 (2022).
- 604 28. J. A. Klappenbach, J. M. Dunbar, T. M. Schmidt, rRNA operon copy number reflects  
605 ecological strategies of bacteria. *Applied and Environmental Microbiology* **66**, 1328–1333  
606 (2000).
- 607 29. S. Vieira-Silva, E. P. C. Rocha, The Systemic Imprint of Growth and Its Uses in Ecological  
608 (Meta)Genomics. *PLoS Genetics* **6**, e1000808 (2010).
- 609 30. S. Krause, *et al.*, Trait-based approaches for understanding microbial biodiversity and  
610 ecosystem functioning. *Frontiers in Microbiology* **5**, 251 (2014).
- 611 31. S. J. Blazewicz, R. L. Barnard, R. A. Daly, M. K. Firestone, Evaluating rRNA as an indicator  
612 of microbial activity in environmental communities: Limitations and uses. *ISME Journal* **7**, 2061–  
613 2068 (2013).
- 614 32. J. L. Weissman, S. Hou, J. A. Fuhrman, Estimating maximal microbial growth rates from  
615 cultures, metagenomes, and single cells via codon usage patterns. *Proceedings of the National*  
616 *Academy of Sciences* **118**, e2016810118 (2021).
- 617 33. J. B. Plotkin, G. Kudla, Synonymous but not the same: The causes and consequences of  
618 codon bias. *Nature Reviews Genetics* **12**, 32–42 (2011).
- 619 34. Y. Liu, Q. Yang, F. Zhao, Synonymous but Not Silent: The Codon Usage Code for Gene  
620 Expression and Protein Folding. *Annual Review of Biochemistry* **90**, 375–401 (2021).
- 621 35. C. H. Yu, *et al.*, Codon Usage Influences the Local Rate of Translation Elongation to  
622 Regulate Co-translational Protein Folding. *Molecular Cell* **59**, 744–754 (2015).
- 623 36. M. Zhou, *et al.*, Non-optimal codon usage affects expression, structure and function of  
624 clock protein FRQ. *Nature* **2013** 495:7439 **495**, 111–115 (2013).

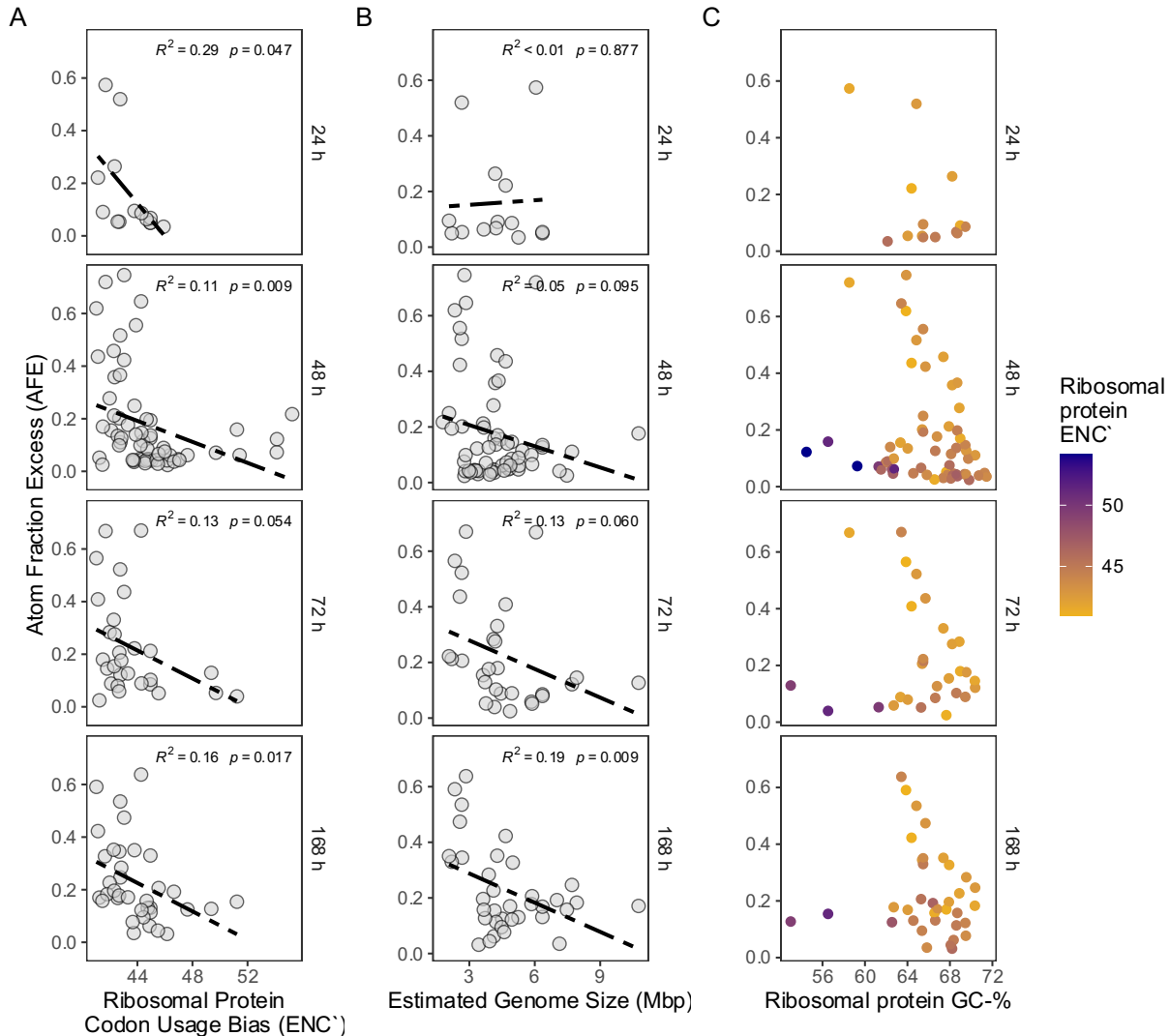
- 625 37. C. Dressaire, *et al.*, Role of mRNA Stability during Bacterial Adaptation. *PLOS ONE* **8**,  
626 e59059 (2013).
- 627 38. V. Presnyak, *et al.*, Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**,  
628 1111–1124 (2015).
- 629 39. Z. R. Newman, J. M. Young, N. T. Ingolia, G. M. Barton, Differences in codon bias and GC  
630 content contribute to the balanced expression of TLR7 and TLR9. *Proceedings of the National*  
631 *Academy of Sciences of the United States of America* **113**, E1362–E1371 (2016).
- 632 40. Z. Zhou, *et al.*, Codon usage is an important determinant of gene expression levels  
633 largely through its effects on transcription. *Proceedings of the National Academy of Sciences of*  
634 *the United States of America* **113**, E6117–E6125 (2016).
- 635 41. S. Bahiri-Elitzur, T. Tuller, Codon-based indices for modeling gene expression and  
636 transcript evolution. *Computational and Structural Biotechnology Journal* **19**, 2646–2663  
637 (2021).
- 638 42. D. L. Kirchman, Growth rates of microbes in the oceans. *Annual Review of Marine*  
639 *Science* **8**, 285–309 (2016).
- 640 43. A. M. Long, S. Hou, • J Cesar, I. -Espinoza, J. A. Fuhrman, Benchmarking microbial growth  
641 rate predictions from metagenomes. *The ISME Journal* 1–13 (2020).  
642 <https://doi.org/10.1038/s41396-020-00773-1>.
- 643 44. L. Aliperti, *et al.*, r/K selection of GC content in prokaryotes. *Environmental Microbiology*  
644 **n/a** (2023).
- 645 45. W.-H. Chen, G. Lu, P. Bork, S. Hu, M. J. Lercher, Energy efficiency trade-offs drive  
646 nucleotide usage in transcribed regions. *Nature Communications* **7**, 11334 (2016).
- 647 46. S. J. Giovannoni, J. Cameron Thrash, B. Temperton, Implications of streamlining theory  
648 for microbial ecology. *The ISME Journal* **8**, 1553–1565 (2014).
- 649 47. T. E. Brewer, K. M. Handley, P. Carini, J. A. Gilbert, N. Fierer, Genome reduction in an  
650 abundant and ubiquitous soil bacterium ‘Candidatus Udaeobacter copiosus.’ *Nature*  
651 *Microbiology* **2**, 16198 (2017).
- 652 48. E. P. Starr, *et al.*, Stable isotope informed genome-resolved metagenomics reveals that  
653 Saccharibacteria utilize microbially-processed plant-derived carbon. *Microbiome* **6**, 122 (2018).
- 654 49. A. M. Nicolas, *et al.*, Soil Candidate Phyla Radiation Bacteria Encode Components of  
655 Aerobic Metabolism and Co-occur with Nanoarchaea in the Rare Biosphere of Rhizosphere  
656 Grassland Communities. *mSystems* **6**, 10.1128/msystems.01205-20 (2021).
- 657 50. M. Cobo-Simón, J. Tamames, Relating genomic characteristics to environmental  
658 preferences and ubiquity in different microbial taxa. *BMC Genomics* **18**, 499 (2017).
- 659 51. A. Barberán, *et al.*, Why are some microbes more ubiquitous than others? Predicting the  
660 habitat breadth of soil bacteria. *Ecology Letters* **17**, 794–802 (2014).
- 661 52. K. T. Konstantinidis, J. M. Tiedje, Trends between gene content and genome size in  
662 prokaryotic species with larger genomes. **10**, 3160–3165 (2004).
- 663 53. M. Westoby, *et al.*, Cell size, genome size, and maximum growth rate are near-  
664 independent dimensions of ecological variation across bacteria and archaea. *Ecology and*  
665 *Evolution* **11**, 3956–3976 (2021).
- 666 54. J. Li, *et al.*, Predictive genomic traits for bacterial growth in culture versus actual growth  
667 in soil. *The ISME Journal* 1 (2019). <https://doi.org/10.1038/s41396-019-0422-z>.

- 668 55. B. A. Hungate, *et al.*, Quantitative microbial ecology through stable isotope probing.  
669 *Applied and Environmental Microbiology* **81**, 7570–7581 (2015).
- 670 56. D. Vyshenska, *et al.*, A standardized quantitative analysis strategy for stable isotope  
671 probing metagenomics. *mSystems* **0**, e01280-22 (2023).
- 672 57. E. E. Nuccio, *et al.*, HT-SIP: a semi-automated stable isotope probing pipeline identifies  
673 cross-kingdom interactions in the hyphosphere of arbuscular mycorrhizal fungi. *Microbiome* **10**,  
674 199 (2022).
- 675 58. C. Fossum, *et al.*, Belowground allocation and dynamics of recently fixed plant carbon in  
676 a California annual grassland. *Soil Biology and Biochemistry* **165**, 108519 (2022).
- 677 59. E. T. Sieradzki, *et al.*, Functional succession of actively growing soil microorganisms  
678 during rewetting is shaped by precipitation history. *bioRxiv* 2022.06.28.498032 (2022).  
679 <https://doi.org/10.1101/2022.06.28.498032>.
- 680 60. R. L. Barnard, C. A. Osborne, M. K. Firestone, Changing precipitation pattern alters soil  
681 microbial community response to wet-up under a Mediterranean-type climate. *ISME Journal* **9**,  
682 946–957 (2015).
- 683 61. S. J. Blazewicz, *et al.*, Taxon-specific microbial growth and mortality patterns reveal  
684 distinct temporal population responses to rewetting in a California grassland soil. *ISME Journal*  
685 **14**, 1520–1532 (2020).
- 686 62. B. Bushnell, BBTools Software Package. (2014). Available at:  
687 <https://sourceforge.net/projects/bbmap/>.
- 688 63. Joshi NA, Fass JN., Sickle: A sliding-window, adaptive, quality-based trimming tool for  
689 FastQ files (Version 1.33). (2023). Deposited 10 May 2023.
- 690 64. D. Li, C. M. Liu, R. Luo, K. Sadakane, T. W. Lam, MEGAHIT: An ultra-fast single-node  
691 solution for large and complex metagenomics assembly via succinct de Bruijn graph.  
692 *Bioinformatics* **31**, 1674–1676 (2015).
- 693 65. Y.-W. Wu, B. A. Simmons, S. W. Singer, MaxBin 2.0: an automated binning algorithm to  
694 recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
- 695 66. D. D. Kang, *et al.*, MetaBAT 2: an adaptive binning algorithm for robust and efficient  
696 genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
- 697 67. G. V. Uritskiy, J. DiRuggiero, J. Taylor, MetaWRAP—a flexible pipeline for genome-  
698 resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
- 699 68. M. R. Olm, C. T. Brown, B. Brooks, J. F. Banfield, dRep: a tool for fast and accurate  
700 genomic comparisons that enables improved genome recovery from metagenomes through de-  
701 replication. *ISME J* **11**, 2864–2868 (2017).
- 702 69. D. Hyatt, *et al.*, Prodigal: Prokaryotic gene recognition and translation initiation site  
703 identification. *BMC Bioinformatics* **11**, 1–11 (2010).
- 704 70. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives  
705 on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**, D353–D361 (2017).
- 706 71. M. Shaffer, *et al.*, DRAM for distilling microbial metabolism to automate the curation of  
707 microbiome function. *Nucleic Acids Research* **48**, 8883–8900 (2020).
- 708 72. I.-M. A. Chen, *et al.*, IMG/M v.5.0: an integrated data management and comparative  
709 analysis system for microbial genomes and microbiomes. *Nucleic Acids Research* **47**, D666–  
710 D677 (2019).

- 711 73. P. F. Chuckran, *et al.*, Metatranscriptomes of California grassland soil microbial  
712 communities in response to rewetting. *Microbiology Resource Announcements* **0**, e00322-24  
713 (2024).
- 714 74. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for  
715 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
- 716 75. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for  
717 RNA-seq data with DESeq2. *Genome Biology* **15** (2014).
- 718 76. W. McKinney, pandas: a Foundational Python Library for Data Analysis and Statistics.  
719 *Python for High Performance and Scientific Computing* 1–9 (2011).
- 720 77. C. R. Harris, *et al.*, Array programming with NumPy. *Nature* **585**, 357–362 (2020).
- 721 78. J. A. Novembre, Accounting for Background Nucleotide Composition When Measuring  
722 Codon Usage Bias. *Molecular Biology and Evolution* **19**, 1390–1394 (2002).
- 723 79. H. Akashi, T. Gojobori, Metabolic efficiency and amino acid composition in the  
724 proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of  
725 Sciences of the United States of America* **99**, 3695–3700 (2002).
- 726 80. C. S. Miller, B. J. Baker, B. C. Thomas, S. W. Singer, J. F. Banfield, EMIRGE: reconstruction  
727 of full-length ribosomal genes from microbial community short read sequencing data. *Genome  
728 Biology* **12**, R44 (2011).
- 729 81. W. Song, S. Zhang, T. Thomas, MarkerMAG: linking metagenome-assembled genomes  
730 (MAGs) with 16S rRNA marker genes using paired-end short reads. *Bioinformatics* **38**, 3684–  
731 3688 (2022).
- 732 82. R. C. Team, R: A language and environment for statistical computing. *R Foundation for  
733 Statistical Computing. Vienna, Austria.* (2018).
- 734 83. H. Wickham, *et al.*, Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686  
735 (2019).
- 736 84. H. Wickham, “Elegant Graphics for Data Analysis” in *Elegant Graphics for Data Analysis*,  
737 (2016), pp. 3–10.
- 738 85. H. Akaike, A new look at the statistical model identification. *IEEE Transactions on  
739 Automatic Control* **19**, 716–723 (1974).
- 740 86. National Ecological Observatory Network (NEON), Soil microbe metagenome sequences  
741 (DP1.10107.001). (2021).
- 742 87. F. Wright, The ‘effective number of codons’ used in a gene. *Gene* **87**, 23–29 (1990).
- 743 88. National Ecological Observatory Network (NEON), Soil physical and chemical properties,  
744 periodic (DP1.10086.001). National Ecological Observatory Network (NEON). Deposited 2024.
- 745 89. M. M. Foley, *et al.*, Growth rate as a link between microbial diversity and soil  
746 biogeochemistry. *Nat Ecol Evol* 1–9 (2024). <https://doi.org/10.1038/s41559-024-02520-7>.
- 747 90. B. J. Koch, *et al.*, Estimating taxon-specific population dynamics in diverse microbial  
748 communities. *Ecosphere* **9**, e02090 (2018).
- 749 91. E. M. Morrissey, *et al.*, Phylogenetic organization of bacterial activity. *The ISME Journal*  
750 **1028**, 2336–2340 (2016).
- 751 92. J. L. Weissman, M. Peras, T. P. Barnum, J. A. Fuhrman, Benchmarking Community-Wide  
752 Estimates of Growth Potential from Metagenomes Using Codon Usage Statistics. *mSystems* **7**,  
753 e00745-22 (2022).

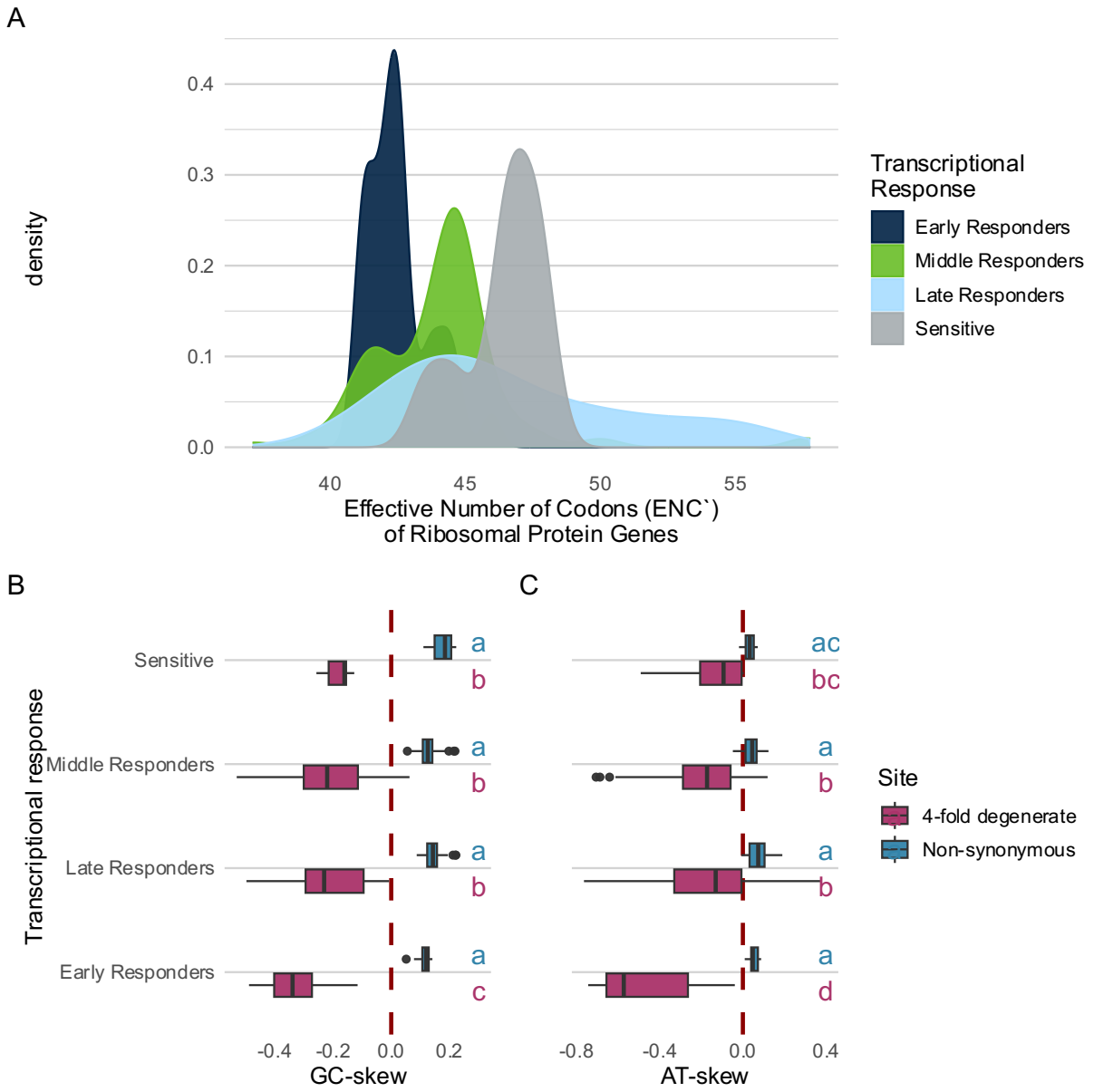
- 754 93. A. Canarini, *et al.*, Quantifying microbial growth and carbon use efficiency in dry soil  
755 environments via  $^{18}\text{O}$  water vapor equilibration. *Global Change Biology* **26**, 5333–5341 (2020).
- 756 94. D. Metze, *et al.*, Microbial growth under drought is confined to distinct taxa and  
757 modified by potential future climate conditions. *Nat Commun* **14**, 5895 (2023).
- 758 95. M. Gralka, S. Pollak, O. X. Cordero, Genome content predicts the carbon catabolic  
759 preferences of heterotrophic bacteria. *Nat Microbiol* 1–10 (2023).  
760 <https://doi.org/10.1038/s41564-023-01458-z>.
- 761 96. R. L. Barnard, S. J. Blazewicz, M. K. Firestone, Citation Classic Rewetting of soil: Revisiting  
762 the origin of soil  $\text{CO}_2$  emissions. *Soil Biology and Biochemistry* **147**, 107819 (2020).
- 763 97. P. F. Chuckran, *et al.*, Edaphic controls on genome size and GC content of bacteria in soil  
764 microbial communities. *Soil Biology and Biochemistry* **178**, 108935 (2023).
- 765 98. H. Liu, *et al.*, Warmer and drier ecosystems select for smaller bacterial genomes in  
766 global soils. *iMeta* **2**, e70 (2023).
- 767 99. G. Piton, *et al.*, Life history strategies of soil bacterial communities across global  
768 terrestrial biomes. *Nat Microbiol* 1–10 (2023). <https://doi.org/10.1038/s41564-023-01465-0>.
- 769 100. C. A. Martinez-Gutierrez, F. O. Aylward, Genome size distributions in bacteria and  
770 archaea are strongly linked to evolutionary history at broad phylogenetic scales. *PLOS Genetics*  
771 **18**, e1010220 (2022).
- 772 101. Z. T. Aanderud, J. T. Lennon, Validation of Heavy-Water Stable Isotope Probing for the  
773 Characterization of Rapidly Responding Soil Bacteria. *Applied and Environmental Microbiology*  
774 **77**, 4589–4596 (2011).
- 775 102. S. A. Placella, E. L. Brodie, M. K. Firestone, Rainfall-induced carbon dioxide pulses result  
776 from sequential resuscitation of phylogenetically clustered microbial groups. *Proceedings of the*  
777 *National Academy of Sciences of the United States of America* **109**, 10931–10936 (2012).
- 778 103. P. F. Chuckran, B. A. Hungate, E. Schwartz, P. Dijkstra, Variation in genomic traits of  
779 microbial communities among ecosystems. *FEMS Microbes* **2**, xtab020 (2022).
- 780 104. H. Liu, *et al.*, Warmer and drier ecosystems select for smaller bacterial genomes in  
781 global soils. *iMeta* **n/a**, e70.
- 782 105. E. W. Slessarev, *et al.*, Water balance creates a threshold in soil pH at the global scale.  
783 *Nature* **540**, 567–569 (2016).
- 784 106. A. A. Malik, *et al.*, Land use driven change in soil pH affects microbial carbon cycling  
785 processes. *Nature Communications* **9** (2018).
- 786
- 787

788 *Figure 1. Effect of a 1 week H<sub>2</sub><sup>18</sup>O wet-up experiment of annual grassland soils from Hopland,*  
 789 *CA on bacterial growth and genomic traits. (A) Relationships between taxon-specific growth*  
 790 *(atom fraction excess (AFE) of <sup>18</sup>O) and effective number of codons in ribosomal protein genes*  
 791 *(ENC') and (B) genome size (Mbp, derived from metagenome assembled genomes (MAGs)), for*  
 792 *samples collected at multiple post wet-up timepoints, and (C) versus GC content of ribosomal*  
 793 *protein genes post wet-up (the time where the most MAGs were represented), with color*  
 794 *indicating ENC values. Lower ENC values indicate a higher level of codon bias.*



795  
796

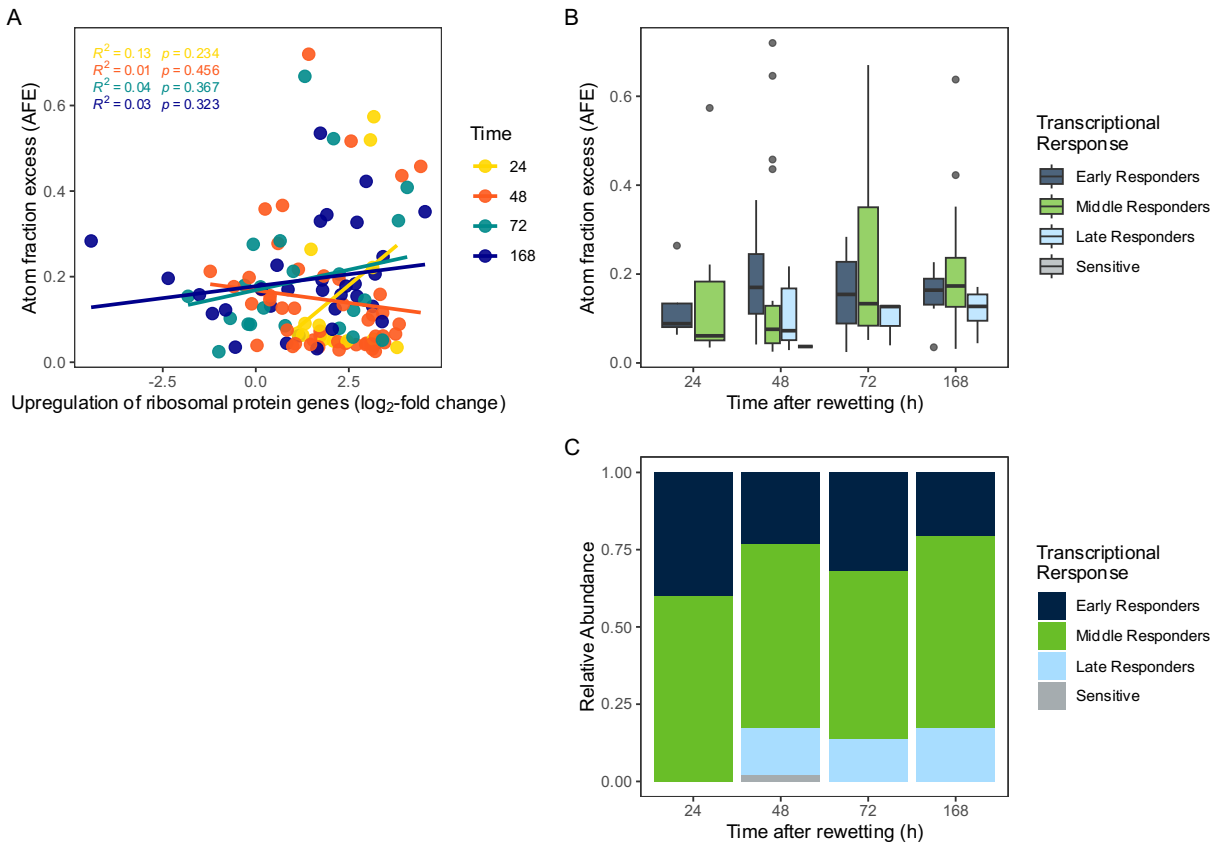
797 *Figure 2: Measures of ribosomal protein response to soil wet-up in a grassland soil. Codon*  
 798 *usage bias of ribosomal protein genes colored by transcriptional response type (A). The*  
 799 *synonymous and non-synonymous GC (B) and AT (C) skew by transcriptional response, with*  
 800 *color indicating nucleotide site-type. Four-fold degenerate sites describe a type of synonymous*  
 801 *substitution site where all possible nucleotide substitutions would result in the same amino acid.*  
 802



803

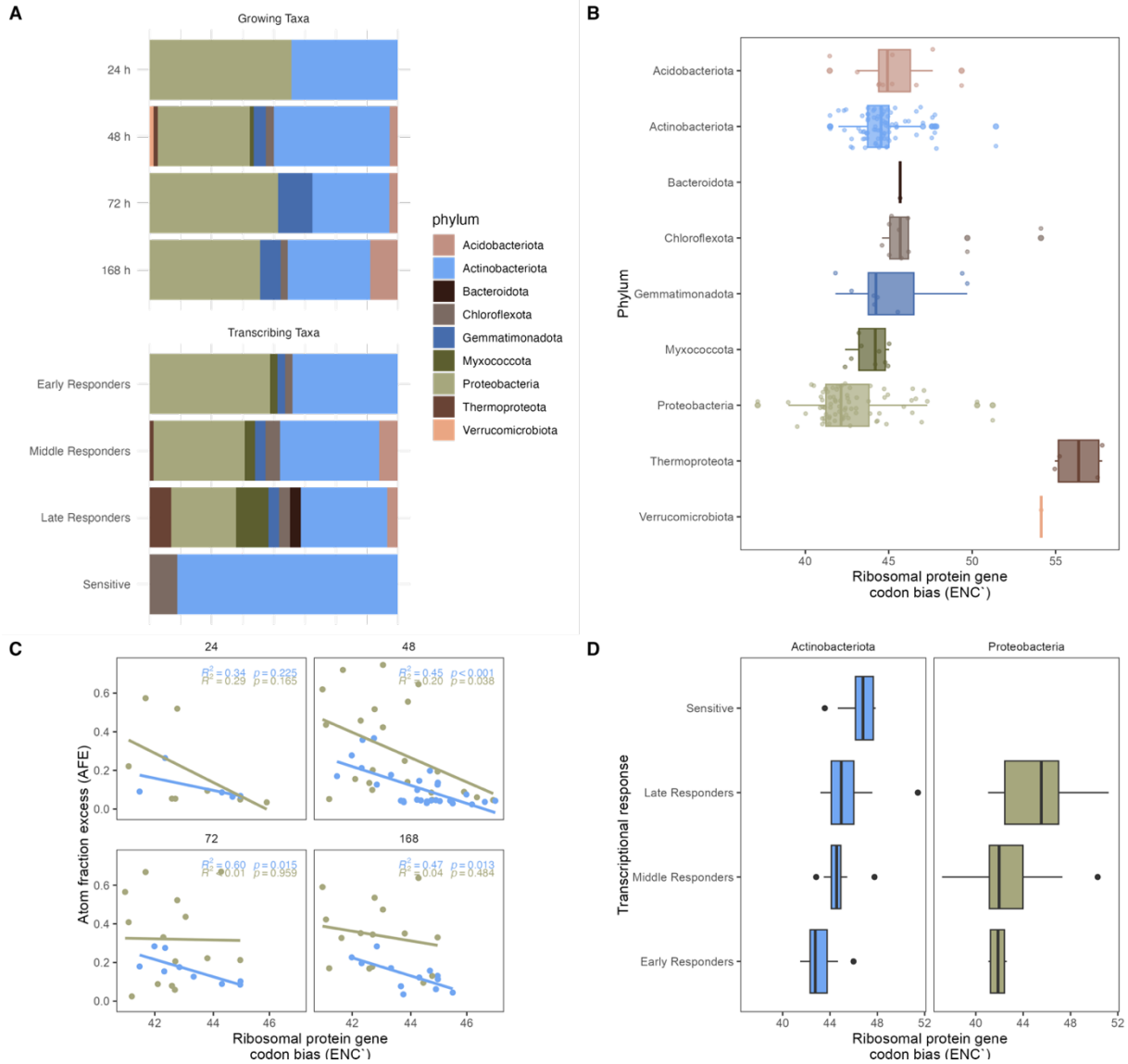
804

805 *Figure 3:*  
 806 *The relationship between growth (as indicated through Atom Fraction Excess, AFE) and*  
 807 *transcription after rewetting for metagenome assembled genomes (MAGs). The level of*  
 808 *upregulation of ribosomal protein genes in each MAG versus growth (AFE), with color*  
 809 *indicating time (A). AFE of each transcriptional response group over time (B), and the*  
 810 *proportion of growing MAGs that could be assigned to a transcriptional response group at each*  
 811 *time point (C).*  
 812



813  
814

815 *Figure 4:*  
 816 *Taxonomic breakdown of metagenome assembled genomes (MAGs) by phylum growing at each*  
 817 *time point and by transcriptional response category (A). Ribosomal protein gene codon usage*  
 818 *bias ( $ENC_{\text{ribo}}$ ) for all growing or transcribing MAGs separated by phylum (B). Atom fraction*  
 819 *excess, and index of growth, versus ribosomal protein codon usage bias ( $ENC'$ ), separated by*  
 820 *the two most represented phylum, Actinobacteria and Proteobacteria (C). Ribosomal protein*  
 821 *codon usage bias ( $ENC'$ ) by transcriptional response category and phylum (D).*  
 822

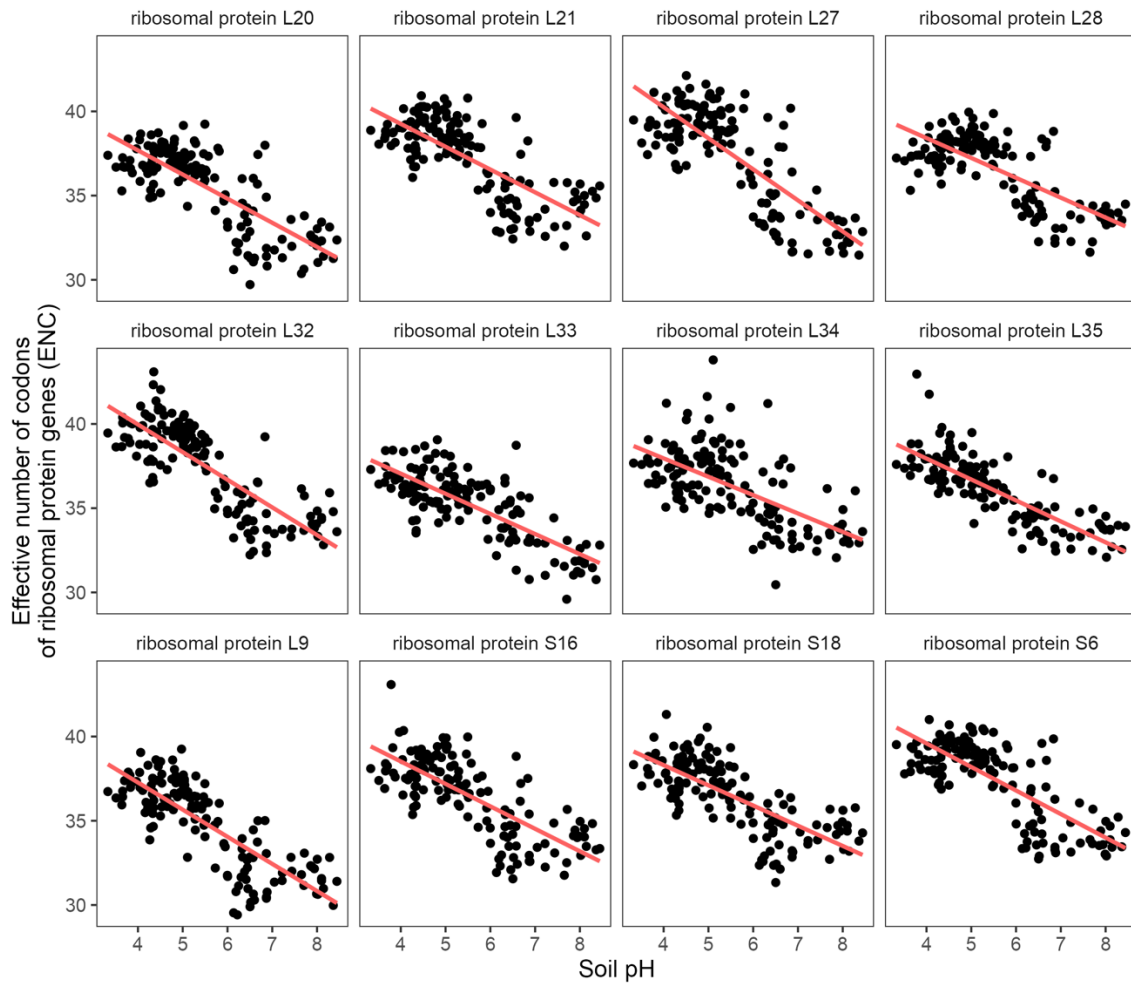


823

824

825 Figure 5:

826 *The relationship between codon usage bias (ENC) of ubiquitous bacteria-specific ribosomal*  
827 *protein genes and pH from metagenomes collected from the National Ecological Observation*  
828 *Network (NEON).*



829

830