bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A glyphosate-based herbicide selects for genetic changes while retaining within-species diversity in a freshwater bacterioplankton community

Emma Derrick[1], Naíla Barbosa da Costa[2,3], Rowan D. H. Barrett[4,5], B. Jesse Shapiro[1,3,5,6,7*]

1. Department of Microbiology and Immunology, McGill University, Montreal, Canada
2. Institut national de la recherche scientifique - Centre Eau Terre Environnement (INRS-ETE) 490 Couronne St, Québec, Canada
3. Groupe de Recherche Interuniversitaire en Limnologie (GRIL), Montreal, Canada
4. Department of Biology, McGill University, Montreal, Canada
5. Quebec Centre for Biodiversity Science, Montreal, Canada
6. McGill Genome Centre, McGill University, Montreal, Canada
7. McGill Centre for Microbiome Research, Montreal Canada

* correspondence: jesse.shapiro@mcgill.ca

1 **Abstract**

2 Bacterial populations evolve rapidly in the lab when faced with experimentally-applied selective
3 pressures. Yet how bacteria evolve in nature, in more complex multi-species communities, is both
4 challenging to study and essential to our understanding of ecosystem responses to rapid
5 anthropogenic change. It has been theorized that selection purges within-species diversity in
6 genome-wide selective sweeps, but the prevalence of such sweeps in response to known selective
7 pressures in nature remains unclear. To track bacterial evolution in a semi-natural context, we
8 applied Roundup, a glyphosate-based herbicide (GBH) as a selective pressure to 1000 L ponds
9 containing bacterioplankton communities from a pristine lake. Using metagenomic analyses, we
10 found that GBH treatment substantially affected community diversity, reducing species richness
11 twofold, but did not consistently purge within-species genetic diversity over the four weeks of the
12 experiment. We identified several functional categories of genes targeted by GBH selection across
13 11 different species of bacteria. There was no evidence for selection on the enzyme targeted by
14 glyphosate, which interferes with amino acid synthesis; however genes involved more broadly in
15 amino acid transport and metabolism were more likely to experience changes in allele frequency,
16 particularly in inferred GBH-sensitive species. Together, these results show how environmental
17 change can rapidly affect bacterial community structure while leaving within-species diversity
18 largely intact. Even without evident genome-wide selective sweeps, we identify consistent genetic
19 targets of selection, pointing to alternative mechanisms of GBH resistance in nature, and
20 suggesting a role for soft or gene-specific selective sweeps in adaptation.

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

21 **Introduction**

22 Microbial communities and the populations within them form the foundation of all ecosystems on
23 Earth (1). Many studies have focused on how natural microbial communities change on an
24 ecological level, with species changing in abundance in response to perturbations from
25 environmental stressors (2–6). It is becoming increasingly evident that microbial populations
26 within communities can also evolve on the same time scales as these ecological changes (7–9).
27 Rapid evolutionary change, for example of pathogens within infected patients, can have
28 consequences for virulence and disease persistence (10–13). More generally, within-species
29 diversity is important for the functioning and stability of bacterial communities (14–16). For
30 example, evolutionary diversification of a focal species can impact relative abundances of other
31 species within the community (17). Conversely, higher community diversity may promote or
32 constrain the diversification of species within a community (18). In the human gut, community
33 diversity promotes within-species diversity over time scales of a few months ("diversity begets
34 diversity"), until niches are filled (8). In other more diverse communities such as freshwater,
35 sediments, and soil microbiomes, the "diversity begets diversity" effect is negligible, presumably
36 because most niches are already filled (19). Similarly, natural compost communities challenged
37 with copper stress showed independent evolutionary and ecological changes, with no detectable
38 interaction between the two (7). In all cases, ecological and evolutionary changes within microbial
39 communities, along with the interactions in some cases, are expected to affect ecosystems
40 functions.

41 What types of evolutionary changes can occur within a community? Under the stable ecotype
42 model, selection on adaptive mutations can result in a genome-wide selective sweep, in which a
43 genome with an adaptive allele expands clonally (with relatively little recombination) and purges
44 genetic diversity from the population (20–22). Alternatively, if recombination is high, an adaptive
45 allele can be exchanged by horizontal gene transfer and spread through the population in a gene-
46 specific sweep, purging diversity in a region of the genome while maintaining genome-wide
47 diversity (23, 24). Or, if an adaptive allele is originally present in the population on multiple
48 genetic backgrounds, then, during a sweep, multiple strains carrying the adaptive allele may
49 increase in frequency, resulting in a soft sweep (25). Both soft and gene-specific selective sweeps
50 allow specific genes targeted by selection to adapt without purging diversity genome-wide. They

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

51 differ in that soft sweeps occur when recurrent adaptive mutations occur on different genomic
52 backgrounds (lineages or strains) while gene-specific sweeps require high rates of recombination
53 relative to selection.

54 In nature, there has been little evidence for pervasive genome-wide selective sweeps. Even in
55 simplified laboratory evolution experiments where genome-wide selective sweeps are theorized to
56 be more likely, diversity is often maintained in the population by diverse genetic targets of
57 selection and clonal interference (26, 27). In natural environments such as the human gut, high
58 rates of recombination allow for the exchange of genes within and between species (28, 29),
59 potentially promoting gene-specific rather than genome-wide sweeps (30). Studies of natural
60 populations of bacteria in freshwater lakes have identified evolutionary patterns consistent with
61 sweeps, but it remains unclear if these were driven by selection or genetic drift. Bendall et al. (31)
62 observed a gradual loss in diversity over nine years in one population of bacteria, consistent with
63 a genome-wide sweep, as well as other populations with low diversity in small genomic regions,
64 consistent with past gene-specific sweeps. More recently, Rower et al. (32) quantified changes in
65 species abundance and diversity throughout a 20-year time series and identified an association
66 with seasonal changes, as well as one possible soft sweep in *Nanopelagicus*. While these studies
67 suggest that selective sweeps are occurring in natural bacterioplankton populations, neither could
68 attribute sweeps to a known selective pressure.

69 Agrochemical pollution is an important selective pressure relevant to soil and aquatic microbial
70 communities. There is growing concern over agrochemicals entering freshwater from runoff and
71 leaching from agricultural land (33). Glyphosate-based herbicides (GBHs) are the most commonly
72 used herbicides worldwide, and while the active ingredient glyphosate is thought to strongly bind
73 to soil (34), it may also run off into rivers, streams, and lakes. Regulations in Canada limit the
74 concentration of glyphosate permitted for chronic (< 800 μg/L) and acute (< 27000 μg/L) aquatic
75 contamination (34). However, these guidelines are based on toxicity to eukaryotes, and ignore
76 potential effects on bacteria. Previous studies have shown GBH alters the composition of bacterial
77 communities in water (2) and honeybees (5). Further, GBH can cross-select for antibiotic
78 resistance genes in soil (35) and aquatic (36) communities.

79 GBHs inhibit plant growth by interfering with the shikimate pathway and preventing downstream
80 synthesis of essential aromatic amino acids (37). Glyphosate prevents the conversion of shikimate-

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

81 3-phosphate (S3P) and phosphoenolpyruvate (PEP) into 5-enolpyruvylshikimate 3-phosphate
82 (EPSP) by competitively binding EPSP synthase (EPSPS) (38). In addition to plants, the shikimate
83 pathway is also used by bacteria and some fungi. Bacteria with different EPSPS alleles encoding
84 specific amino acid changes vary in their resistance to glyphosate, and can be classified as
85 glyphosate resistant or sensitive based on their EPSPS allele (39). Bacteria can also be resistant to
86 GBH through other mechanisms, such as exporting glyphosate out of the cell with efflux pumps
87 (40) or by degrading it (41).

88 Here, we investigated how aquatic bacteria evolve in a semi-natural community faced with the
89 GBH Roundup as an experimentally applied selective pressure. We exposed replicate mesocosms
90 to two pulses of Roundup and sequenced metagenomes at five time points over eight weeks. To
91 study evolutionary responses to selection imposed by GBH, we focused on 11 bacterial species
92 (metagenome-assembled genomes; MAGs) present in multiple ponds after four weeks and tracked
93 their genetic diversity in control and GBH-treated ponds. While we found some evidence for GBH-
94 driven genome-wide selective sweeps in progress, these remained incomplete on the short
95 timescale of our study and tended not to be repeatable across replicate ponds. We hypothesized
96 that species predicted to be GBH-sensitive based on their EPSPS allele would be under stronger
97 selection than GBH-resistant species. Although predicted sensitive and resistant species had no
98 evident differences in genome-wide diversity after GBH treatment, they have distinct genetic
99 targets of selection. Particularly in GBH-sensitive species, GBH selected for single nucleotide
100 variants in genes involved in amino acid transport and metabolism, and for reduced copy number
101 of genes involved in transcription and translation. Together, our results show how a known
102 environmental stressor affects community diversity without consistently changing genome-wide
103 within-species diversity beyond a few specific targets of selection.

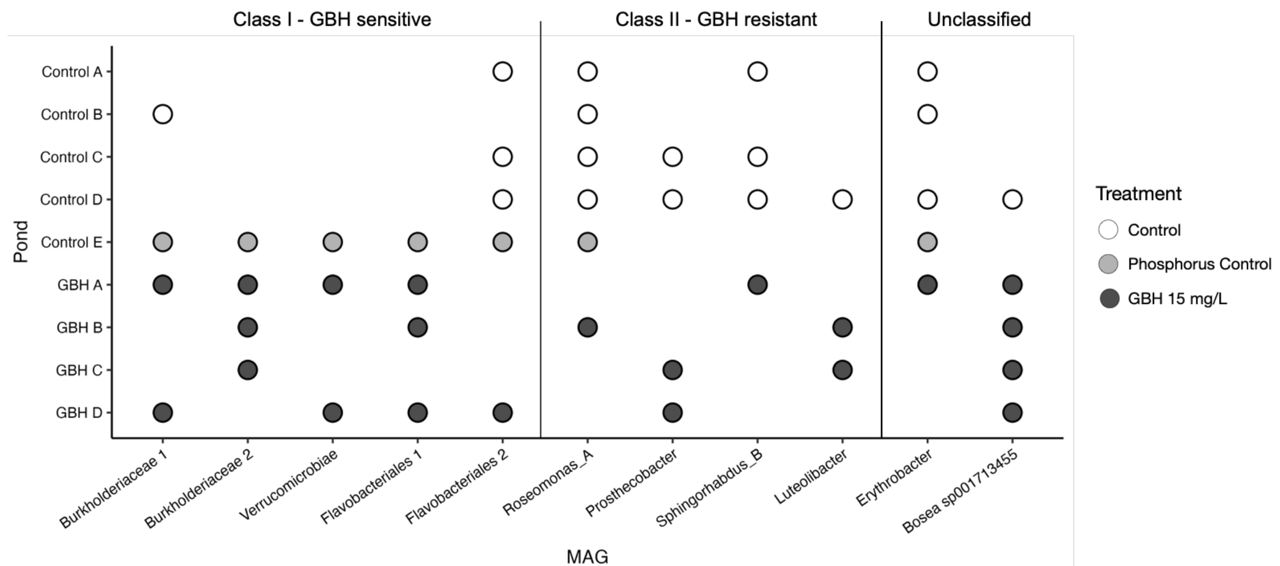
104 Results

105 Quantifying within-species diversity across experimental treatments

106 In this experiment we sampled nine 1000 L ponds filled with pristine lake water over eight weeks
107 to study how natural populations of aquatic bacteria evolve after exposure to Roundup, a
108 commonly used GBH. Treated ponds received two pulses of GBH (at 15 mg/L and 40 mg/L)
109 whereas controls received only phosphorus (as a control for the high phosphorus content of

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

110 glyphosate) or no treatment. For each of the nine ponds, we co-assembled metagenomic reads from
111 five time points throughout the 8-week experiment and binned the resulting contigs into
112 metagenome-assembled genomes (MAGs), yielding a database of 315 non-redundant MAGs,
113 which we define as distinct species (Methods). To quantify genetic diversity within species, we
114 competitively mapped reads from each time point to our non-redundant MAG database and
115 quantified the number and frequency of single nucleotide variants (SNVs) in each MAG in each
116 experimental pond. We identified 11 MAGs that were present at a minimum of 4x average depth
117 of coverage in at least one control and one GBH pond in the two samples collected at one and four
118 weeks after the 15 mg/L GBH pulse (**Figure 1**). These 11 MAGs had an average size of 3.04 Mbp,
119 completeness of 87.19%, and redundancy of 2.31% (**Table 1**). For some MAGs, SNV detection
120 rates were correlated with MAG coverage (**Figure S1**). To resolve this bias, we subsampled reads
121 such that each MAG had an equal coverage across ponds (Methods). The vast majority of these
122 MAGs (10/11) were at low or undetectable relative abundance at time point 1, then increased at
123 time point 2 in both control and treatment ponds (**Figure S2**). This suggests that this subset of
124 species are well-adapted to the pond environment.



125 **Figure 1. MAGs present in GBH treatment and control ponds after 15 mg/L GBH pulse.**
126 MAGs with a minimum of 4x average sequencing depth in both control and GBH ponds at time
127 point 2 (merged samples from day 7 and 28). Ponds are coloured by the treatment received. MAGs
128 are grouped by their predicted sensitivity to GBH based on their EPSPS allele.

129

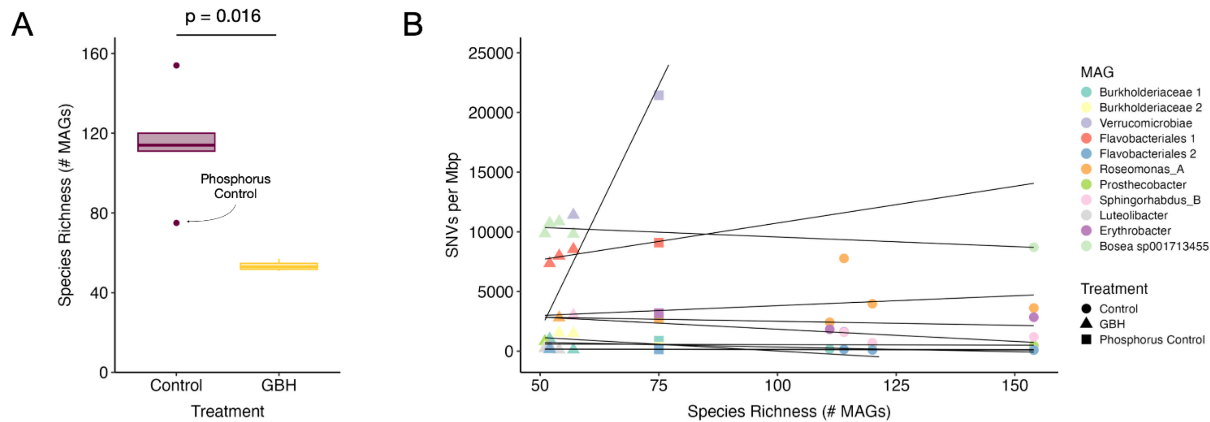
bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

130 **Table 1. Summary of MAGs analyzed.** Predicted sensitivity and resistance to glyphosate are
131 indicated by (S) and (R) respectively.

Species (MAG)	EPSPS Class	Completeness (%)	Redundancy (%)	Size (Mbp)	Contigs
Burkholderiaceae 1	I (S)	100.00	1.41	2.90	150
Burkholderiaceae 2	I (S)	90.14	2.82	3.11	104
Verrucomicrobiae	I (S)	97.18	1.41	2.61	8
Flavobacteriales 1	I (S)	94.37	1.41	2.96	139
Flavobacteriales 2	I (S)	77.46	1.41	2.71	358
Roseomonas_A	II (R)	78.87	5.63	4.23	400
Prosthecobacter	II (R)	88.73	0.00	4.05	321
Sphingorhabdus_B	II (R)	98.59	4.23	2.64	207
Luteolibacter	II (R)	78.87	1.41	3.84	737
Erythrobacter	Unclassified	81.69	4.23	2.46	195
Bosea sp001713455	Unclassified	73.24	1.41	1.95	501

132 If GBH inhibits the growth or viability of some bacterial species, we would expect GBH-treated
133 ponds to have fewer detectable species (MAGs) present compared to control ponds. We estimated
134 the species richness of each pond at time point 2 (after the 15 mg/L GBH pulse) by determining
135 the number of MAGs in our database that were present. Consistent with expectation and with an
136 earlier study (2), we found that GBH ponds had significantly lower species richness, approximately
137 two-fold lower than control ponds (Wilcoxon Rank Sum test, $U = 0$, $p = 0.016$) (**Figure 2A**). The
138 phosphorus control pond had lower richness than other controls, but still higher than any GBH-
139 treated pond (**Figure 2A**). For most species, there was no evident relationship between within-
140 species diversity and community diversity, with the exception of one *Verrucomicrobiae* MAG
141 which showed a positive "diversity begets diversity" relationship (**Figure 2B**). This suggests that,
142 on the time scale of our experiment, GBH has a significant effect on community-level diversity,
143 but this effect is generally independent of within-species diversity.

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



144

145 **Figure 2. Species richness differs between GBH and control ponds independently of within-**
146 **species diversity.** The number of species (MAGs) present in each pond at time point 2. A species
147 was considered present in a pond if the mean coverage was at least 1x and the breadth was > 50%.
148 (A) Box plot of the number of species (MAGs) detected in each pond. (B) The relationship between
149 within-species diversity (SNVs/Mbp) and species richness. A linear trendline was plotted for each
150 MAG.

151

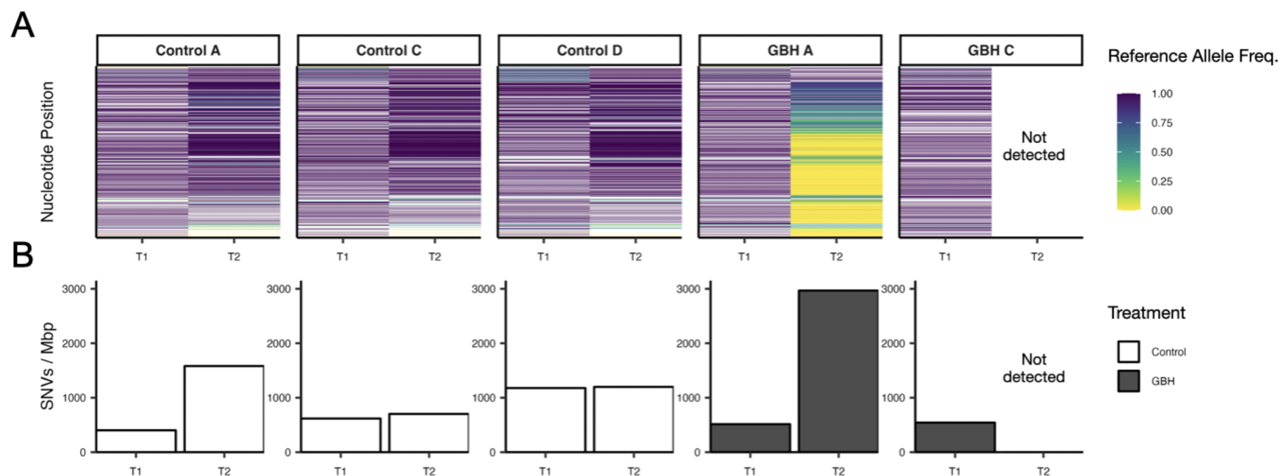
152 **Genome-wide selective sweeps are rare, incomplete, and not associated with EPSPS class**

153 To study within-species diversity changes in greater detail, we compared SNV frequencies within
154 each MAG between GBH and control ponds. If GBH drives genome-wide selective sweeps, this
155 should result in genome-wide purges of genetic diversity in GBH-treated ponds compared to
156 control ponds. One MAG, S*phingorhabdus_B*, classified as GBH-resistant, was the only species
157 recovered in ponds both before and after the GBH pulse, allowing us to track genetic diversity
158 over time. At time point 1, before any pond received GBH, the diversity in the *Sphingorhabdus_B*
159 population varied somewhat between the five ponds (402 to 1198 SNVs/Mbp) (**Figure 3A & 3B,**
160 **Table S1**). At time point 2, the diversity in control ponds increased slightly, but much more
161 dramatically in GBH-treated pond A, reaching nearly 3,000 SNVs/MBp (**Figure 3A & 3B**). This
162 increased diversity over time is not expected under a genome-wide selective sweep. Alternatively,
163 such a pattern could be explained by a soft selective sweep in which beneficial mutations rise to
164 fixation in multiple different genome backgrounds – for example, selecting for a rare strain with
165 an adaptive allele to increase in frequency. Consistent with a soft sweep, this GBH pond had over
166 5000 fixed (reference allele frequency = 0) or nearly fixed substitutions (yellow in **Figure 3A**,
167 **Table S1**) alongside the nearly 3000 polymorphic sites. In the other GBH replicate,

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

168 S*phingorhabdus_B* was not recovered at time point 2, suggesting strong selection drove this
169 species to extinction or below the limit of detection. *Sphingorhabdus_B* in the GBH A pond
170 experienced approximately 4x more gene copy number changes over time than control ponds
171 (**Figure S3, Table S2**). Further, 91% of genes with a copy number change in GBH A from T1 to
172 T2 were also identified as having a copy number change between GBH A and control ponds at T2,
173 compared to an average of 35% overlap among controls. This suggests that GBH is selecting for
174 gene copy number changes in *Sphingorhabdus_B* over time. Together, these results suggest that
175 natural selection imposed by GBH drives changes in genetic diversity over time, consistent with a
176 soft or gene-specific sweep, or an incomplete genome-wide sweep that might have gone to
177 completion given more time.

178



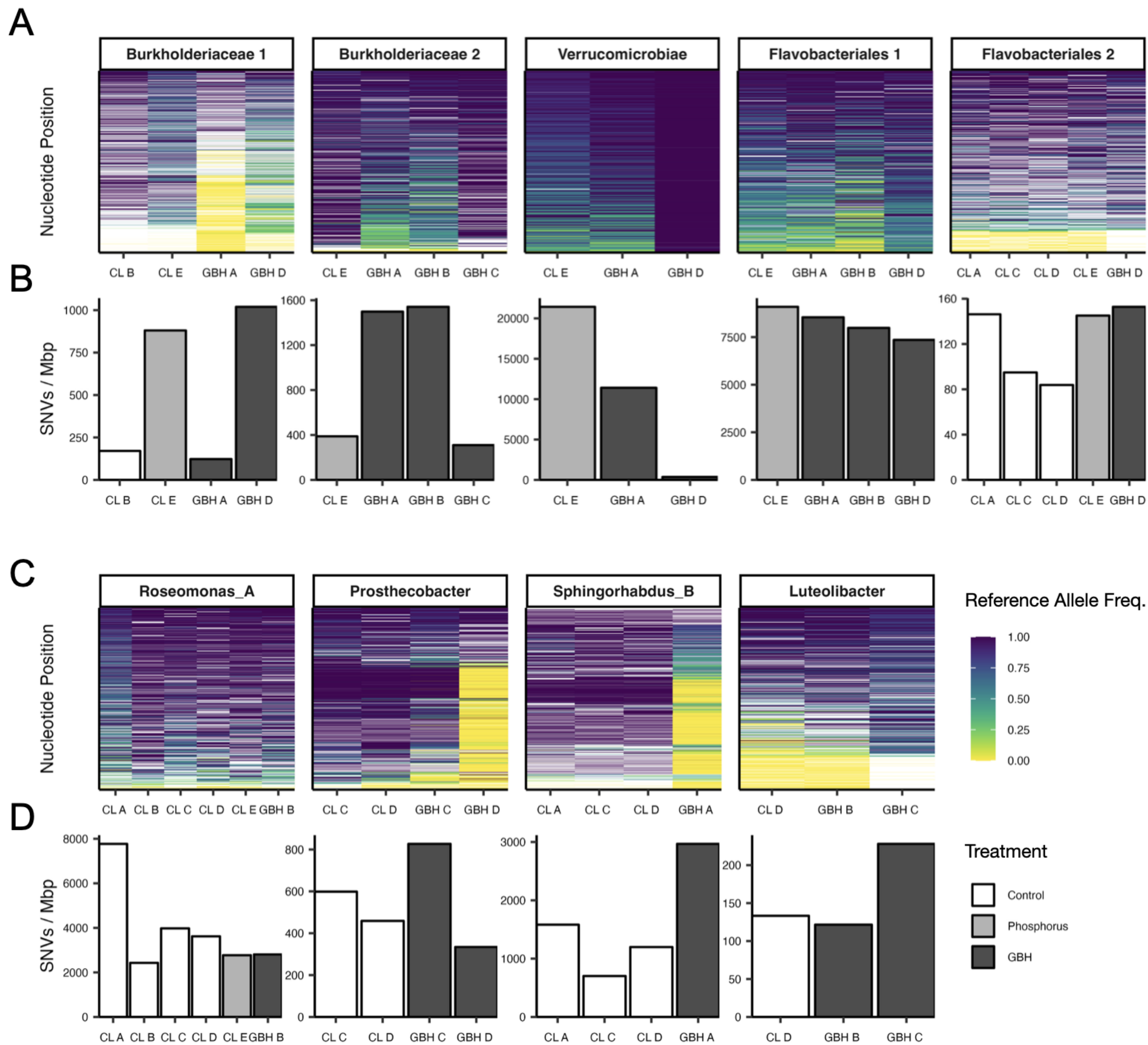
179 **Figure 3. Temporal diversity changes in S*phingorhabdus_B* before and after a GBH pulse.**
180 (A) Each row in the heat map represents a nucleotide position which is polymorphic or fixed on a
181 non-reference allele in at least one pond. The rows are coloured based on the reference allele
182 frequency. Reference allele frequency at each genome position is calculated as the proportion of
183 reads mapping to that site in the reference MAG that match the reference allele. Rows are ordered
184 by the mean reference allele frequency across all ponds. (B) Total number of polymorphic sites in
185 the MAG population divided by the MAG genome size (bp) x 10^6. T1 (time point 1) includes the
186 sample taken before any GBH pulse was added on day 0. T2 (time point 2) includes two samples
187 taken after a 15 mg/L GBH pulse on days 7 and 28.

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

188 Although we lacked time series data for the other ten MAGs (which were too rare at T1 to call
189 SNVs), we were still able to compare within-species diversity in control vs. GBH ponds at T2. Of
190 these MAGs, none showed unequivocal evidence for a complete and repeatable GBH-driven
191 genome-wide selective sweep (**Figure 4, Figure S3**). While in some cases there were differences
192 in diversity between ponds, all populations contained measurable diversity in both control and
193 GBH treatment ponds. In one example of a potential genome-wide selective sweep, the
194 *Verrucomicrobiae* population had lower diversity in both GBH ponds compared to a control
195 (**Figure 4A & 4B**). The *Verrucomicrobiae* population in the control pond contained 21,438
196 SNVs/Mbp, which was reduced by almost half in one GBH pond (11,410 SNVs/Mbp) and by over
197 50-fold in another (373 SNVs/Mbp; **Figure 4B, Table S1**). This is consistent with a genome-wide
198 selective sweep in progress, favouring the reference allele. In the *Burkholderiaceae* 1 population,
199 diversity was reduced in one replicate GBH pond but not the other (**Figure 4B**), suggesting a
200 genome-wide sweep that was not repeatable, or was more rapid in one replicate. Both GBH ponds
201 had more fixed non-reference alleles or low reference allele frequency SNVs compared to control
202 ponds, suggesting a replacement of one dominant strain by another (**Figure 4A**). Similarly, a
203 potential genome-wide sweep in the *Prosthecobacter* population only reduced diversity in one
204 GBH pond compared to controls, resulting in the fixation of non-reference alleles (**Figure 4C &
205 4D).** The other GBH pond shared some of these fixed substitutions, but remained more diverse
206 than either control pond (**Figure 4C & 4D)**.

207 Overall, the evidence for genome-wide selective sweeps was equivocal. Defining a genome-wide
208 selective sweep as a reduction in diversity across the genome in one GBH population compared to
209 the average of the control populations, only 3/11 MAGs showed any evidence of a genome-wide
210 selective sweep. We further hypothesized that GBH would impose stronger selection on MAGs
211 predicted to be GBH-sensitive at the beginning of the experiment. We classified MAGs as GBH-
212 sensitive or resistant based on their EPSPS allele. Of the 11 MAGs, five were classified as Class I
213 (GBH-sensitive), four were classified as Class II (GBH-resistant), and two could not be classified
214 because the EPSPS gene was not present in the annotation (**Figure 1, Table 1**). There were no
215 apparent differences in genetic diversity after the GBH pulse between predicted GBH-sensitive
216 MAGs (**Figure 4A & 4B**), resistant MAGs (**Figure 4C & 4D**), or unclassified MAGs (**Figure
217 S3**), none of which showed evidence for complete, repeatable genome-wide selective sweeps
218 driven by GBH.

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



219 **Figure 4. Population diversity of GBH-sensitive and resistant MAGs**. (A & C) Each row in
220 each heat map represents a nucleotide position which is polymorphic or fixed on a non-reference
221 allele in at least one pond. The bars are coloured based on the reference allele frequency. Reference
222 allele frequency at each genome position is calculated as the proportion of reads mapping to that
223 site in the reference MAG that match the reference allele. Rows are ordered by the mean reference
224 allele frequency across all ponds. (B & D) Total number of polymorphic sites in the MAG
225 population divided by the MAG genome size (bp) x 10^6. (A & B) Predicted GBH-sensitive MAGs.
226 (C & D) Predicted GBH-resistant MAGs.

227

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

228 **Identifying genetic targets of selection**

229 Regardless of whether genome-wide selective sweeps occur, our data provide the opportunity to
230 identify the likely genetic targets of GBH selection. It is typically challenging to identify the targets
231 of selection after a genome-wide sweep, because the targets of selection are genetically linked to
232 other hitchhiking mutations (42, 43). Our replicated study design alleviates this challenge because
233 selected mutations could sweep independently in different replicate ponds (although this is not
234 guaranteed to occur if the exact same genome fixes in replicate ponds). However, classic genome-
235 wide sweeps appear to be rare in our experiment (**Figures 4 and S3**), and soft or gene-specific
236 sweeps may be more common – both of which facilitate finding the targets of selection. We used
237 four different approaches to identify genes targeted by GBH selection: (1) those with large changes
238 in SNV frequency between GBH and control ponds, (2) those with consistently reduced SNV
239 density between GBH and control ponds, (3) those with consistently reduced numbers of SNVs
240 between GBH and control ponds, and (4) those with large changes in gene copy number between
241 GBH and control ponds (Methods).

242 In the first approach, we calculated the difference between the mean reference allele frequency at
243 each nucleotide position in control ponds and GBH ponds. We found that all populations had genes
244 containing at least one large SNV frequency shift (range of 10 - 1,393 genes per MAG with a shift
245 in reference allele frequency of 0.5 or more; **Table 2, Table S3**). Pooled across all MAGs, we
246 found that these genes with shifts in SNV frequency were significantly enriched in several
247 functional categories involved in metabolism (COG categories G, E, I, P, and Q; Fisher's exact
248 test, $p < 0.05$, FDR < 0.1; **Figure 5**). We also found a significant depletion of SNV frequency
249 changes in genes from categories J (translation, ribosomal structure and biogenesis), K
250 (transcription), N (cell motility), and X (mobilome). These results point to certain gene functions
251 consistently targeted by selection across all populations.

252

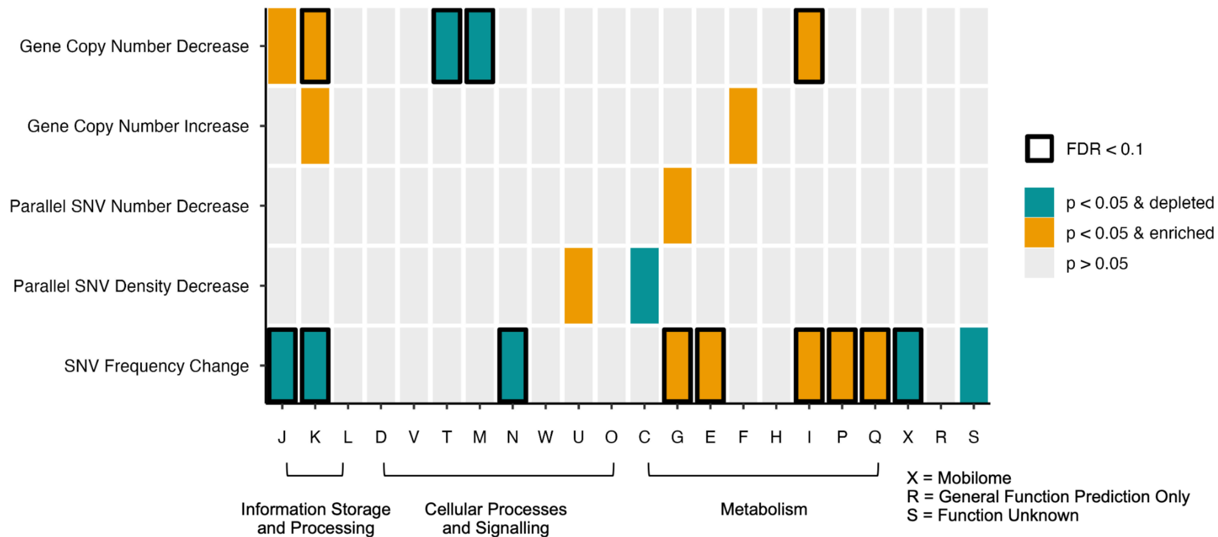
bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

253 **Table 2.** Summary of the number of genes identified as targets of selection. Predicted sensitivity
254 and resistance to glyphosate are indicated by (S) and (R) respectively.

Species (MAG)	EPSPS Class	SNV Frequency Shift	Parallel SNV Density Decrease	Parallel SNV Number Decrease	Gene Copy Number Increase	Gene Copy Number Decrease
Burkholderiaceae 1	I (S)	484	2	0	100	137
Burkholderiaceae 2	I (S)	76	107	7	52	112
Verrucomicrobiae	I (S)	1398	41	111	162	413
Flavobacteriales 1	I (S)	424	17	2	21	22
Flavobacteriales 2	I (S)	10	0	0	83	57
Roseomonas_A	II (R)	669	2	3	183	183
Prosthecobacter	II (R)	1025	10	4	80	132
Sphingorhabdus_B	II (R)	1248	1	0	524	1034
Luteolibacter	II (R)	73	85	18	172	140
Erythrobacter	Unclassified	666	0	0	105	108
Bosea sp001713455	Unclassified	1089	8	0	151	147

255
256
257 Next, we identified genes in GBH-exposed populations with less diversity compared to the rest of
258 the genome, as expected in a gene targeted by a selective sweep. Nine of the 11 populations had
259 at least one gene (range of 1 - 107) with a repeatable (parallel across replicate ponds) decrease in
260 SNV density in each GBH-control comparison (**Table 2, Table S4**). Further, six of these
261 populations had at least one gene (range of 2 - 111 genes) with a significant SNV number decrease
262 in each GBH-control comparison using an existing method (44) which identifies genes with
263 differential SNV counts between treatments (**Table 2, Table S5**). These parallel decreases in both
264 SNV density and number provide strong evidence for selection since they occur repeatedly across
265 replicates. These genes were enriched in carbohydrate transport and metabolism and intracellular
266 trafficking, secretion, and vesicular transport (COG categories G and U, Fisher's exact test, $p <$
267 0.05). Even if not significant after multiple hypothesis correction (FDR > 0.1), COG category G
268 was also enriched in large SNV frequency changes, supporting its importance as a target of GBH
269 selection (**Figure 5**).

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



270

271 **Figure 5. Functional categories enriched or depleted for genes identified as targets of**
272 **selection.** For each method used to identify possible targets of selection, a Fisher's exact test was
273 performed for each COG category between genes identified as targets of selection and all other
274 genes present in the combined MAGs. Bars are coloured by significance and odds ratio (< 1 =
275 depleted or > 1 = enriched). Bars outlined in black pass our false discovery rate (FDR) threshold
276 of < 0.1. COG categories: J = Translation, ribosomal structure and biogenesis, K = Transcription,
277 L = Replication, recombination, and repair, D = Cell cycle control, cell division, chromosome
278 partitioning, Y = Nuclear structure, V = Defense mechanisms, T = Signal transduction
279 mechanisms, M = Cell wall/membrane/envelope biogenesis, N = Cell motility, Z = Cytoskeleton,
280 W = Extracellular structures, U = Intracellular trafficking, secretion, and vesicular transport, O =
281 Posttranslational modification, protein turnover, chaperones, C = Energy production and
282 conversion, G = Carbohydrate transport and metabolism, E = Amino acid transport and
283 metabolism, F = Nucleotide transport and metabolism, H = Coenzyme transport and metabolism,
284 I = Lipid transport and metabolism, P = Inorganic ion transport and metabolism, Q = Secondary
285 metabolites biosynthesis, transport and catabolism, X = Mobilome: prophages, transposons, R =
286 General function prediction only, S = Function unknown

287

288 Lastly, we identified genes that differed in copy number between control and GBH ponds. For
289 each gene, we compared its relative depth of coverage in controls to GBH ponds (Methods). We
290 found that all populations had genes that increased or decreased by > 0.5 gene copies between
291 control and GBH ponds (**Table 2, Table S6**). Genes that decreased in copy number were
292 significantly enriched in transcription (K) and lipid transport and metabolism (I) and depleted in
293 signal transduction (T) and cell wall/membrane biogenesis (M) (**Figure 5**).

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

294 Overall, genes involved in metabolism were frequently identified as targets of selection by
295 different methods, while genes involved in cellular processes and signalling were rarely identified
296 (**Figure 5**). Meanwhile, translation and transcription (J and K) tended to decrease in gene copy
297 number under GBH treatment. Selection on amino acid transport and metabolism, along with
298 reduced copy number of transcription and translation genes, is mainly driven by MAGs classified
299 as GBH-sensitive (**Figure S4**). This could be due to GBH-sensitive species, but not resistant
300 species, being under selection for changes in amino acid metabolism – which contains the
301 shikimate pathway targeted by glyphosate. Thus, MAGs with inferred GBH resistance or
302 sensitivity at the beginning of the experiment may have experienced different genetic targets of
303 selection.

304 ## Discussion

305 Previous work has shown GBHs like Roundup alter the composition of aquatic bacterial
306 communities (2, 36). However, how individual bacterial populations within a community evolve
307 in response to GBH is unknown. More generally, our understanding of microbial evolution in
308 nature has been hindered by the lack of studies combining experimentally controlled selective
309 pressures with measures of within-species diversity (evolution) in the context of a diverse
310 community (45). To fill this gap, we quantified within-species diversity in bacterioplankton
311 exposed to GBH in controlled setup freshwater mesocosm ponds. We found that genome-wide
312 selective sweeps were rare and often incomplete over the time scale of the experiment. Rather, all
313 of the 11 populations studied contained measurable and often substantial genetic diversity (on the
314 order of hundreds to thousands of SNVs per Mbp, consistent with standing strain-level diversity)
315 in both GBH-treated and control ponds. Potential selective sweeps in one pond were often not
316 observed in a replicate pond. This lack of repeatability raises doubt about whether a sweep was
317 driven by selection or drift.

318 There are several explanations for the lack of genome-wide selective sweeps documented in this
319 study, which are not mutually exclusive. First, a four-week time scale might have been too short
320 to observe sweeps proceeding to completion. One other potential genome-wide sweep documented
321 in a lake took years to complete (31) so it is plausible that sweeps in natural environments require
322 much longer to complete than the few dozen generations shown in simplified evolutionary models
323 (46). Second, it is possible that the community-level response to selection is dominated by species

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

324 sorting, with pre-existing GBH-resistant taxa flourishing and sensitive taxa declining, giving little
325 opportunity for evolutionary responses within these species. A previous experiment in the same
326 mesocosm system showed that GBH substantially restructures the bacterioplankton community
327 and reduces diversity (2), which we confirmed in this study. It is therefore plausible that ecological
328 responses dominated evolutionary responses in this context and time scale. Third, recombination-
329 driven gene-specific selective sweeps or soft selective sweeps may be more common than genome-
330 wide selective sweeps in our experiment. We detected signals of selection in numerous genes,
331 including parallel purges of diversity in GBH-treated ponds but not controls. Gene-specific or soft
332 selective sweeps are plausible mechanisms to explain these patterns. Using short-read
333 metagenomic data, it is difficult to infer recombination; therefore direct evidence for
334 recombination-driven gene-specific selective sweeps is lacking. Given the relatively high levels of
335 within-species standing genetic diversity, which is typical in a natural community, adaptive
336 mutations could spread on distinct genetic backgrounds ("strains") in a soft selective sweep,
337 resulting in the maintenance of diversity without high levels of recombination. Future work using
338 long-read sequencing or whole genomes from isolated bacteria will be needed to resolve the
339 relative influence of soft and gene-specific selective sweeps.

340 Despite the lack of evident genome-wide selective sweeps, we identified genes with consistent
341 changes in allele frequency or copy number in GBH ponds, consistent with natural selection acting
342 on specific cellular pathways in multiple species. Glyphosate binds EPSPS and prevents aromatic
343 amino acid synthesis; we therefore expected natural selection on EPSPS and surrounding amino
344 acid metabolic pathways. We classified each species as putatively resistant or sensitive to
345 glyphosate at the beginning of the experiment based on known mutations in the EPSPS gene. We
346 expected GBH-sensitive populations to be under stronger selection than GBH-resistant
347 populations, but we did not observe any notable differences in their genome-wide diversity
348 between treatments. However, we did find evidence that the functional categories of genes under
349 selection differed between GBH-sensitive and resistant populations. Although *aroA*, the gene
350 encoding EPSPS, was not identified as a target of selection in any species, genes involved in amino
351 acid metabolism (COG category E) were significantly enriched in SNV frequency changes in
352 response to GBH across all species – an effect that is driven by GBH-sensitive species. This
353 suggests that amino acid metabolism pathways surrounding EPSPS might be specifically under
354 selection in species with a GBH-sensitive EPSPS allele. Other GBH-sensitive species might be

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

355 selected for slower growth, as has been observed in *B. subtilis* facing nutrient starvation (47). As
356 evidence for this, we found that GBH selected for a decreased copy number of transcription and
357 translation genes in GBH-sensitive MAGs, but not in GBH-resistant MAGs. Together with
358 previous findings that GBH selects for multidrug efflux pumps in bacterioplankton communities
359 (36), our results highlight that targets of selection can be diverse in natural communities, not
360 always centered on canonical resistance mutations.

361 Why *Prosthecobacter* and *Sphingorhabdus_B* — both GBH-resistant MAGs based on their EPSPS
362 allele — would be targeted by GBH selection is unclear. The *Prosthecobacter* population
363 experienced fixation of many alternate alleles in one GBH replicate but not the other. Such an
364 inconsistent response suggests the dominance of drift over GBH-driven selection. In
365 *Sphingorhabdus_B* where the starting population at time point 1 is known, the number of SNVs
366 (including many at low frequency) increased along with the number of fixations in one GBH
367 replicate. One hypothesis is that GBH interferes with the cell in unknown ways, beyond the
368 canonical EPSPS target. Alternatively, selection on GBH-resistant species could be indirect. For
369 example, GBH causes shifts in community composition, which could subsequently impose new
370 selective pressures on these species through competition, cross feeding, or by creating novel niches
371 (8, 19). Importantly, the potential soft sweep in *Sphingorhabdus_B* was only observed in one pond,
372 with the MAG decreasing below the limit of detection in the other pond after the GBH pulse. This
373 again suggests that stochastic dynamics (drift) could dominate selection in a species close to the
374 limit of local extinction.

375 Further research will be needed to quantify selective sweep dynamics over longer time scales, and
376 for a larger number of taxa. Ten of the MAGs analyzed were present at very low or undetectable
377 abundances at time point 1, preventing us from tracking changes in diversity over time. This is
378 unlikely to affect our conclusions regarding genome-wide selective sweeps (or the lack thereof)
379 because each pond was filled with water from the same lake at the same time, making it improbable
380 that the initial diversity within each MAG varied substantially across ponds. Nevertheless,
381 additional time-series – over longer time scales – would provide more robust support for our
382 conclusions. With bacterial doubling in nature roughly every 1-25 hours (48), the pond populations
383 likely evolved for approximately 28-700 generations after the GBH pulse. Depending on the
384 strength of selection, this could be sufficient time to detect changes in allele frequencies (including

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

385 fixation) for fast-growing, but perhaps not slow-growing species. While the 11 populations we
386 tracked encompass both GBH-sensitive and resistant EPSPS classes, they are far from representing
387 the full bacterial diversity present in Lac Hertel. Unfortunately, many MAGs that were recovered
388 in multiple control ponds and time points were absent from GBH ponds after the GBH pulse, which
389 prevented us from including them in the analysis. This observation is likely due to the selective
390 pressure imposed by GBH, where MAGs sensitive to GBH drastically decrease in abundance after
391 the GBH pulse and are not captured at our sequencing depth. While we do not see strong evidence
392 for complete genome-wide selective sweeps across the 11 populations we tracked, it remains
393 unknown if sweeps occurred in any of the other 304 populations that were too rare to detect using
394 our shotgun metagenomic approach.

395 Our study represents an advance in our understanding of evolution in natural environments. By
396 applying a known selective pressure to semi-natural communities, we were able to go beyond
397 documenting sweep-like patterns with unknown causes, and move toward attributing these patterns
398 to selection. In our limited sample of species, and over short time scales, we can conclude that
399 genome-wide selective sweeps are rare in response to GBH, a stressor with clear community-level
400 effects. Even if genome-wide sweeps are rare, we show that GBH imposes selection on several
401 categories of genes, including those involved in amino acid metabolism, transcription and
402 translation – particularly in GBH-sensitive species. This highlights the potentially unexpected
403 evolutionary consequences of agrochemical runoff into freshwater ecosystems. Future work will
404 be needed to determine if genome-wide selective sweeps are more common in response to different
405 selective pressures, in rarer taxa, and over longer time scales.

406 # Methods

407 ## Experimental design, sampling, and sequencing

408 In this experiment, nine "ponds" (mesocosms) were filled with approximately 1000L of water
409 originating from Lac Hertel, a pristine lake with no known prior herbicide or pesticide
410 contamination, at Gault Nature Reserve in Mont-Saint-Hilaire, Quebec. We sampled the nine
411 ponds five times over 8-weeks between July 16th, 2021 and September 10th, 2021. Four out of
412 nine ponds received two pulses of the glyphosate-based herbicide, Roundup, in the form of
413 Roundup Super Concentrate Grass and Weed Control (reg. no. 22759; Bayer). One control pond

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

414 received a single pulse of phosphorus (K_2PO_4) to control for Roundup as a phosphorus source (49).
415 The four remaining control ponds did not receive any treatment. On day 0, the four GBH ponds
416 received a 15 mg/L GBH pulse and the phosphorus control pond received a 320 µg/L pulse of
417 phosphorus (K_2PO_4). On day 28, the four GBH ponds received a 40 mg/L GBH pulse. GBH
418 concentrations were calculated based on the concentration of glyphosate acid. We collected 1L of
419 water at five timepoints: day 0 (pre-GBH pulse), 7 & 28 (after 15 mg/L GBH pulse), and 35 & 56
420 (after 40 mg/L GBH pulse). For each sample, we filtered 250 mL of water through a 0.22µm pore
421 size polyethersulfone membrane (Sigma-Aldrich) to collect the bacterial community for
422 metagenomic sequencing. The filters were stored at -80 °C prior to DNA extraction. DNA was
423 extracted using the DNeasy PowerWater Kit (QIAGEN), libraries were prepared with the
424 NEBNext Ultra II DNA Library Prep kit (New England Biolabs), and sequenced on an Illumina
425 NovaSeq6000 S4 v.1.5 with 150bp paired-end reads.

426 **Metagenomic assembly, binning, and classification of metagenome-assembled genomes**
427 **(MAGs)**

428 Metagenomic reads were trimmed with trimmomatic v0.39 (50) to remove illumina adapters and
429 discard low quality reads. Next, we co-assembled metagenomic reads by pond, pooling the five
430 timepoints for each pond, with MEGAHIT v1.2.9 (51, 52). Following the anvi'o metagenomics
431 workflow, we used anvi'o v7 (53) to generate a contig database for each pond's co-assembly,
432 identify genes with Prodigal v2.6.3 (54), and annotate each database with single copy gene
433 taxonomy with HMMER (55). Next, metagenomic reads from each timepoint were mapped to the
434 co-assembly of their corresponding pond with bowtie2 v2.4.4 (56) using default parameters.
435 Samtools v1.15 (57) was used to convert the SAM output to BAM format and sort and index the
436 BAM file. We used anvi'o to profile contigs longer than 2500 bp in each pond's contig database
437 by calculating coverage and single nucleotide variants across samples using the BAM files. We
438 created a merged contig profile for each pond and used CONCOCT v1.0.0 (58) to bin contigs
439 based on nucleotide composition, kmer frequencies, and coverage across samples. This resulted in
440 1536 bins across the nine ponds. Bins with > 70% completion were manually refined using the
441 anvi'o interactive interface to remove contigs that were incorrectly binned (53). The 606 bins that
442 had > 70% completion and < 10% redundancy were considered MAGs and were dereplicated at a
443 98% average nucleotide identity (ANI) with anvi'o's dereplication command using fastANI v1.32

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

444 (59) resulting in 315 non-redundant MAGs. Taxonomic classification of MAGs was done using
445 the Genome Taxonomy Database (GTDB) Release 214 (60) with GTDB-Tk v2.1.0 (61, 62).

446 **Gene prediction, annotation, and SNV calling**

447 We used prodigal v2.6.3 (54) in metagenomic mode to predict protein coding genes in our database
448 of 315 non-redundant MAGs. Functional annotation of genes was done using eggNOG mapper
449 v2.1.12 (63, 64) using DIAMOND (65) to align sequences. MAGs with an annotated EPSPS gene
450 (*aroA*) were classified as class I (GBH-sensitive) or class II (GBH-resistant) using the online
451 classifier "*EPSPSClass* server" (39) which looks for specific amino acid markers within the gene.
452 Next, metagenomic reads from each timepoint and pond were competitively mapped to our MAG
453 database using bowtie2 v2.4.4 (56) with default parameters. Mapping outputs were converted to
454 BAM format and indexed with samtools v1.16.1 (57). To increase the coverage for SNV calling,
455 BAM files from the two time points after the 15 mg/L GBH pulse (day 7 and day 28) were merged
456 using samtools v1.16.1 (57) and are henceforth referred to as time point 2. Similarly, the two time
457 points after the 40 mg/L pulse (day 35 and day 56) were merged and are referred to as time point
458 3. For each MAG used for further analysis, to control for coverage bias in SNV detection, we
459 subsampled mapped reads with samtools v1.18 (57) to match the coverage of the lowest coverage
460 pond. We used inStrain v1.8.0 (66) to identify SNVs in each MAG population at the three time
461 points. We set a minimum MAPQ of 2 to discard multi-mapped reads, a minimum ANI of reads
462 mapping to the reference database of 95%, and a minimum position coverage of 5x to call a SNV.
463 We discarded SNVs within 100 bp of contig edges and SNVs with a position depth greater than 3
464 times the MAG coverage or less than ⅓ of the MAG coverage. These depth filters remove both
465 low-coverage regions and high-coverage regions containing likely mismapped reads, using
466 established cutoffs (8, 67). We further removed SNVs with more than two alternative alleles. The
467 inStrain command "IS.get('covT')" was used to extract the depth at each position to differentiate
468 between positions where the reference frequency was 1 (and thus depth at that position was not
469 reported in the output) and positions with a depth too low to call a SNV (< 5x).

470 **Species richness**

471 To estimate the species richness in each pond after the 15 mg/L GBH pulse, we first subsampled
472 metagenomic reads from each pond with samtools v1.18 (57) to match the pond with the lowest

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

473 number of reads at time point 2. Next, we used CoverM (68) to estimate the number of MAGs in
474 our database of 315 MAGs that were present with at least 1X average depth and a 50% breadth of
475 coverage. We performed a Wilcoxon Rank Sum test in R version 4.3.3 (69) to determine if GBH
476 treated ponds had lower species richness (number of MAGs present) than control ponds. For each
477 MAG with at least 4x coverage, we plotted the number of polymorphic sites in that MAG versus
478 the species richness for each pond the MAG was present in. To determine if higher species richness
479 in a pond is correlated with higher polymorphic sites within a MAG, we plotted a linear trendline
480 for each MAG.

481 **Identifying genes with differential shifts in SNV frequency between treatments**

482 We identified genes with large shifts in SNV frequency between GBH and control ponds. For each
483 SNV, we calculated the mean reference frequency in control ponds and the mean reference
484 frequency in GBH ponds. We calculated the difference between reference allele frequencies in
485 control and GBH ponds, and considered as 'large' any absolute difference greater or equal to an
486 arbitrary threshold of 0.5. These SNVs were further filtered to remove SNVs where there was
487 overlap in the reference allele frequency between control and GBH ponds: if the average GBH
488 pond reference allele frequency is low compared to controls, all GBH ponds must have a reference
489 allele frequency lower than all control ponds. This assures that the direction of allele frequency
490 change is consistent across treatments. Next, we determined which genes contain at least one SNV
491 with a differential shift of at least 0.5 in SNV frequency between treatments. Note that some of
492 these genes may contain SNVs with opposing directional shifts.

493 **Identifying genes with a decrease in diversity between treatments**

494 Genes experiencing sweeps are expected to be purged of genetic diversity as an adaptive allele
495 rises in frequency. To identify such genes, we focused on those with a reduced number of SNVs
496 during GBH treatment compared to controls. We identified genes with fewer SNVs regardless of
497 their frequency. For this analysis, we only included genes with at least 3x coverage in every pond
498 and excluded genes with a coverage greater than 3 times the MAG coverage as this could be from
499 read donating from other species which share the same gene (67). We calculated the SNVs density
500 (per Mbp) for each gene and then calculated the difference in each pond-pond comparison. For
501 each comparison, we selected the top 5% of genes which have a decrease in SNV density. From

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

502 this list we selected genes which have a SNV density decrease in each GBH-control comparison
503 (which we term a parallel SNV density decrease). We removed genes that were in the top 5% of
504 any GBH-GBH or control-control comparison since these should not be relevant to GBH
505 adaptation.

506 As a complement to this analysis, we used an existing method which identifies genes with
507 differential SNV counts between treatments (44). This method assumes a Poisson distribution of
508 mutations among genes (44). Briefly, for each MAG, we compared the total number of SNVs in a
509 gene in one pond to the number of SNVs in that gene in another pond. For each pair of ponds, we
510 identified genes with a significant SNV number difference between ponds. Next, we determined
511 which genes had a significant decrease in SNV number in GBH ponds compared to controls in
512 every GBH-control comparison (which we term a parallel SNV number decrease).

513 **Gene copy number variation**

514 For each gene, we estimated the number of gene copies to identify gene copy number differences
515 between treatments. Gene copy number was calculated as the gene's depth of coverage divided by
516 the average MAG depth of coverage. Genes with a copy number greater than 3 (likely enriched in
517 mismapped reads) in any pond were excluded from the analysis. For each remaining gene we
518 compared the average copy number in control ponds to the average copy number in GBH ponds.
519 We considered genes with a copy number difference of at least 0.5 between control and GBH
520 ponds to be notable. We further removed genes if there was any overlap in copy number between
521 any control and GBH pond (e.g. if GBH copy number average is lower than the control average,
522 all GBH ponds had to have a copy number lower than all control ponds).

523 **COG function enrichment analysis**

524 We performed a gene set enrichment analysis using the COG database (70) on the five sets of
525 genes identified as potential targets of selection: (1) SNV frequency changes, (2) parallel SNV
526 density decreases, (3) parallel SNV number decreases, (4) gene copy number increases, and (5)
527 gene copy number decreases. Genes in each MAG were annotated with COG IDs using eggNOG
528 mapper v2.1.12 (63, 64) and COG categories from the 2020 database update (71). We used R
529 version 4.3.3 (69) to perform a Fisher's exact test with each gene list to determine if any COG
530 category was significantly enriched for potential targets of selection compared to a background set

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

531 of all genes from the 11 MAGs. A false discovery rate (FDR) correction was applied to account
532 for multiple comparisons. In addition to the overall enrichment test, we performed separate tests
533 for inferred GBH-sensitive and resistant MAGs.

534 **Data Availability**

535 Metagenomic sequences from each sample are available on NCBI under BioProject
536 PRJNA1161687.

537 **Code Availability**

538 Scripts for all analyses are available at https://github.com/emderrick/LEAP_sweeps.

539 **Acknowledgements**

540 We thank Delaney Barth for her assistance in the field and in the lab, Michelle Gros for her
541 assistance in the field, the McGill Genome Centre sequencing platform, and members of the
542 Shapiro lab for their feedback on the project. This study was supported by an NSERC Discovery
543 Grant to B.J.S. LEAP was constructed under the leadership of Andrew Gonzalez, with funding
544 from the Canadian Foundation for Innovation for which we are grateful.

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Falkowski PG, Fenchel T, Delong EF. The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science. 2008 May 23;320(5879):1034–9.
2. Barbosa da Costa N, Fugère V, Hébert MP, Xu CCY, Barrett RDH, Beisner BE, et al. Resistance, resilience, and functional redundancy of freshwater bacterioplankton communities facing a gradient of agricultural stressors in a mesocosm experiment. Mol Ecol. 2021;30(19):4771–88.
3. Romero F, Acuña V, Sabater S. Multiple Stressors Determine Community Structure and Estimated Function of River Biofilm Bacteria. Appl Environ Microbiol. 2020 Jun 2;86(12):e00291-20.
4. Schwartz DJ, Langdon AE, Dantas G. Understanding the impact of antibiotic perturbation on the human microbiome. Genome Med. 2020 Sep 28;12(1):82.
5. Motta EVS, Raymann K, Moran NA. Glyphosate perturbs the gut microbiota of honey bees. Proc Natl Acad Sci U S A. 2018 Oct 9;115(41):10305–10.
6. Nogales B, Lanfranconi MP, Piña-Villalonga JM, Bosch R. Anthropogenic perturbations in marine microbial communities. FEMS Microbiol Rev. 2011 Mar 1;35(2):275–98.
7. Hesse E, Luján AM, O'Brien S, Newbury A, McAvoy T, Soria Pascual J, et al. Parallel ecological and evolutionary responses to selection in a natural bacterial community. Proc Natl Acad Sci. 2024 Sep 3;121(36):e2403577121.
8. Madi N, Chen D, Wolff R, Shapiro BJ, Garud NR. Community diversity is associated with intra-species genetic diversity and gene loss in the human gut microbiome. Mitri S, Weigel D, Bajić D, editors. eLife. 2023 Feb 9;12:e78530.
9. Shapiro BJ. Reuniting ecology and evolution. Environ Microbiol Rep. 2019;11(1):13–4.
10. Zaborskytė G, Hjort K, Lytsy B, Sandegren L. Convergent within-host evolution alters key virulence factors in a Klebsiella pneumoniae clone during a large hospital outbreak. bioRxiv; 2024.

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

11. Young BC, Wu CH, Gordon NC, Cole K, Price JR, Liu E, et al. Severe infections emerge from commensal bacteria by adaptive evolution. Holden MT, editor. eLife. 2017 Dec 19;6:e30637.
12. Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, et al. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nat Genet. 2014 Jan 1;46(1):82–7.
13. Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis MR, et al. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat Genet. 2011 Dec 1;43(12):1275–80.
14. Goyal A, Bittleston LS, Leventhal GE, Lu L, Cordero OX. Interactions between strains govern the eco-evolutionary dynamics of microbial communities. Weigel D, editor. eLife. 2022 Feb 4;11:e74987.
15. Wolff R, Shoemaker W, Garud N. Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome. mBio. 2023 Feb 21;14(2):e02502-22.
16. García-García N, Tamames J, Linz AM, Pedrós-Alió C, Puente-Sánchez F. Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. ISME J. 2019 Dec;13(12):2969–83.
17. Padfield D, Vujakovic A, Paterson S, Griffiths R, Buckling A, Hesse E. Evolution of diversity explains the impact of pre-adaptation of a focal species on the structure of a natural microbial community. ISME J. 2020 Nov 1;14(11):2877–89.
18. Schluter D, Pennell MW. Speciation gradients and the distribution of biodiversity. Nature. 2017 Jun;546(7656):48–55.
19. Madi N, Vos M, Murall CL, Legendre P, Shapiro BJ. Does diversity beget diversity in microbiomes? Weigel D, Kemen E, Wolfe BE, editors. eLife. 2020 Nov 20;9:e58999.
20. Cohan FM. Bacterial Species and Speciation. Syst Biol. 2001 Aug 1;50(4):513–24.
21. Cohan FM, Perry EB. A Systematics for Discovering the Fundamental Units of Bacterial Diversity. Curr Biol. 2007 May;17(10):R373–86.

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

22. Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev. 2011 Sep 1;35(5):957–76.

23. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, et al. Population Genomics of Early Events in the Ecological Differentiation of Bacteria. Science. 2012 Apr 6;336(6077):48–51.

24. Shapiro BJ, Polz MF. Microbial Speciation. Cold Spring Harb Perspect Biol. 2015 Oct;7(10):a018143.

25. Garud NR. Understanding soft sweeps: a signature of rapid adaptation. Nat Rev Genet. 2023 Jul;24(7):420–420.

26. Harris KB, Flynn KM, Cooper VS. Polygenic Adaptation and Clonal Interference Enable Sustained Diversity in Experimental Pseudomonas aeruginosa Populations. Mol Biol Evol. 2021 Dec 1;38(12):5359–75.

27. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nat Rev Genet. 2013 Dec 1;14(12):827–39.

28. Liu Z, Good BH. Dynamics of bacterial recombination in the human gut microbiome. PLOS Biol. 2024 Feb 8;22(2):e3002472.

29. Sakoparnig T, Field C, van Nimwegen E. Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species. eLife. 2021 Jan 8;10:e65366.

30. Wolff R, Garud NR. Pervasive selective sweeps across human gut microbiomes. bioRxiv; 2023.

31. Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 2016 Jul;10(7):1589–601.

32. Rohwer RR, Kirkpatrick M, Garcia SL, Kellom M, McMahon KD, Baker BJ. Bacterial ecology and evolution converge on seasonal and decadal scales. bioRxiv; 2024.

33. Brovini EM, Cardoso SJ, Quadra GR, Vilas-Boas JA, Paranaíba JR, Pereira R de O, et al. Glyphosate concentrations in global freshwaters: are aquatic organisms at risk? Environ Sci

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Pollut Res. 2021 Nov 1;28(43):60635–48.

34. CCME. Canadian Water Quality Guidelines for the Protection of Aquatic Life - Glyphosate. Winnipeg, Manitoba; 2012.

35. Liao H, Li X, Yang Q, Bai Y, Cui P, Wen C, et al. Herbicide Selection Promotes Antibiotic Resistance in Soil Microbiomes. Mol Biol Evol. 2021 Jun 1;38(6):2337–50.

36. Barbosa da Costa Naíla, Hébert Marie-Pier, Fugère Vincent, Terrat Yves, Fussmann Gregor F., Gonzalez Andrew, et al. A Glyphosate-Based Herbicide Cross-Selects for Antibiotic Resistance Genes in Bacterioplankton Communities. mSystems. 2022 Mar 10;7(2):e01482-21.

37. Duke SO, Powles SB. Glyphosate: a once-in-a-century herbicide. Pest Manag Sci. 2008 Apr 1;64(4):319–25.

38. Schönbrunn E, Eschenburg S, Shuttleworth WA, Schloss JV, Amrhein N, Evans JNS, et al. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proc Natl Acad Sci U S A. 2001 Feb 13;98(4):1376–80.

39. Leino L, Tall T, Helander M, Saloniemi I, Saikkonen K, Ruuskanen S, et al. Classification of the glyphosate target enzyme (5-enolpyruvylshikimate-3-phosphate synthase) for assessing sensitivity of organisms to the herbicide. J Hazard Mater. 2021 Apr 15;408:124556.

40. Staub JM, Brand L, Tran M, Kong Y, Rogers SG. Bacterial glyphosate resistance conferred by overexpression of an E. coli membrane efflux transporter. J Ind Microbiol Biotechnol. 2012 Apr 1;39(4):641–7.

41. Hove-Jensen B, Zechel DL, Jochimsen B. Utilization of Glyphosate as Phosphate Source: Biochemistry and Genetics of Bacterial Carbon-Phosphorus Lyase. Microbiol Mol Biol Rev. 2014 Mar 5;78(1):176–97.

42. Shapiro BJ, Polz MF. Ordering microbial diversity into ecologically and genetically cohesive units. Trends Microbiol. 2014 May 1;22(5):235–47.

43. Shapiro BJ, David LA, Friedman J, Alm EJ. Looking for Darwin’s footprints in the microbial world. Trends Microbiol. 2009 May 1;17(5):196–204.

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

44. Shoemaker WR, Lennon JT. Predicting Parallelism and Quantifying Divergence in Microbial Evolution Experiments. mSphere. 2022 Feb 9;7(1):e00672-21.

45. Lenski RE. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J. 2017 Oct;11(10):2181–94.

46. Barghi N, Schlötterer C. Distinct Patterns of Selective Sweep and Polygenic Adaptation in Evolve and Resequence Studies. Genome Biol Evol. 2020 Jun 1;12(6):890–904.

47. Gray DA, Dugar G, Gamba P, Strahl H, Jonker MJ, Hamoen LW. Extreme slow growth as alternative strategy to survive deep starvation in bacteria. Nat Commun. 2019 Feb 21;10(1):890.

48. Gibson B, Wilson DJ, Feil E, Eyre-Walker A. The distribution of bacterial doubling times in the wild. Proc R Soc B Biol Sci. 2018 Jun 13;285(1880):20180789.

49. Hébert MP, Fugère V, Gonzalez A. The overlooked impact of rising glyphosate use on phosphorus loading in agricultural watersheds. Front Ecol Environ. 2019;17(1):48–56.

50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

51. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015 May 15;31(10):1674–6.

52. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016 Jun 1;102:3–11.

53. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015 Oct 8;3:e1319.

54. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010 Mar 8;11(1):119.

55. Eddy SR. Accelerated Profile HMM Searches. PLOS Comput Biol. 2011 Oct

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

20;7(10):e1002195.

56. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr 1;9(4):357–9.

57. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

58. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014 Nov 1;11(11):1144–6.

59. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018 Nov 30;9(1):5114.

60. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022 Jan 7;50(D1):D785–94.

61. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020 Mar 15;36(6):1925–7.

62. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022 Dec 1;38(23):5315–6.

63. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019 Jan 8;47(D1):D309–14.

64. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021 Dec 1;38(12):5825–9.

65. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using

bioRxiv preprint doi: https://doi.org/10.1101/2024.09.17.613573; this version posted September 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

DIAMOND. Nat Methods. 2021 Apr 1;18(4):366–8.

66. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat Biotechnol. 2021 Jun 1;39(6):727–36.

67. Garud NR, Good BH, Hallatschek O, Pollard KS. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLOS Biol. 2019 Jan 23;17(1):e3000102.

68. Aroney STN, Newell RJP, Nissen J, Camargo AP, Tyson GW, Woodcroft BJ. CoverM: Read coverage calculator for metagenomics. Zenodo; 2024.

69. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2024.

70. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000 Jan 1;28(1):33–6.

71. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 2021 Jan 8;49(D1):D274–81.