# Evaluation of *Caulobacter crescentus* strain fitness at the genomic scale during growth in natural freshwater

Kristy L. Hentchel[1], Leila M. Reyes Ruiz[1], Aretha Fiebig[1], Patrick D. Curtis[2], Maureen L. Coleman[3*], Sean Crosson[1,4*]

[1]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA.

[2]Department of Biology, University of Mississippi, University, MS 38677, USA.

[3]Department of the Geophysical Sciences, University of Chicago, Chicago, IL 60637, USA.

[4]Department of Microbiology, University of Chicago, Chicago, IL 60637, USA.

[*]To whom correspondence should be addressed: E-mail: mlcoleman@uchicago.edu, scrosson@uchicago.edu; Phone: (+1) 773-834-1926

Running title: Defining strain fitness in a natural environment

Key words: *Caulobacter crescentus*, transposon sequencing (TnSeq), barcode sequencing (BarSeq), fitness, freshwater, Lake Michigan

## ABSTRACT

Bacterial genomes are the products of selection in fluctuating, spatially heterogeneous, and often nutrient-poor environments. Thus bacterial gene function is best understood in this natural context, but replicating such conditions in the lab is challenging. Here we used a transposon sequencing approach to globally measure the fitness consequences of gene disruption in the free-living freshwater bacterium *Caulobacter crescentus*, cultivated in complex peptone broth, a defined mineral medium (M2-xylose), or in littoral waters of Lake Michigan. Transposon insertions in genes involved in amino acid and nucleotide biosynthesis and in transport resulted in large fitness defects in both lake water and M2-xylose relative to a complex peptone broth, though defects were more extreme in M2-xylose in most cases. In lake water, strain fitness was increased by insertions in genes required for flagellum biosynthesis and reduced by insertions in genes involved in holdfast biosynthesis. We further uncovered numerous hypothetical and uncharacterized genes for which disruption resulted in reduced fitness in either lake water, M2-xylose, or both. At the genomic scale, the fitness profile of the strain pool cultivated in lake water was more similar to that in complex peptone broth than in M2-xylose. Individual gene and strain fitness scores were significantly more variable in lake water than in M2-xylose, presumably owing to variability in water conditions collected on different days. This study advances our understanding of the genetic determinants of *Caulobacter* growth in a complex natural environment, and provides a new avenue to study gene function in a free-living freshwater bacterium.


## IMPORTANCE

Bacteria live in complex environments that are variable over time and space. Bacterial genomes encode the molecular machinery that allows the cell to interact with its environment and with other organisms. Much of this machinery remains uncharacterized, in part because it is difficult to simulate natural conditions in the laboratory. Using the model bacterium *Caulobacter crescentus*, we developed an

49 approach to study bacterial gene function in conditions mimicking the organism's natural freshwater
50 habitat. We identified dozens of genes that are beneficial or detrimental in this natural lake water
51 environment compared to typical laboratory conditions. Our approach will help illuminate the
52 functions of many genes that are missed using standard cultivation approaches, and is broadly
53 applicable to diverse organisms and environments.
54

55 **INTRODUCTION**

56 Microbial cells interact with their environment on the microscale, which has enormous
57 physicochemical complexity. This complexity arises in part from spatial heterogeneity, which is well
58 recognized in biofilms, soils, and host-associated habitats (1-3). Free-living aquatic bacteria also
59 encounter strong chemical gradients that can appear as ephemeral patches, arising from algal exudates,
60 sinking particles, or lysis events (4, 5) and they may have to cope with prolonged periods of nutrient
61 scarcity. In addition, bacteria in natural environments face interspecies interactions, protistan
62 predators, and viruses, as well as physical gradients such as temperature and light. These biotic and
63 abiotic factors, acting together in complex combinations, have driven myriad bacterial adaptations that
64 enable survival and reproduction in heterogeneous natural environments.

65 Despite the complexity faced by bacterial cells in situ, studies on microbial physiology and
66 gene function have relied on simplified experimental conditions. Hence, it is not surprising that a large
67 fraction of bacterial genomes remain uncharacterized. However, recently developed methods that
68 combine transposon mutagenesis with next-generation DNA sequencing technologies now enable the
69 phenotypic assessment of thousands to millions of distinct mutant strains. Such transposon sequencing
70 (Tn-Seq) approaches (6) have been used to interrogate bacterial gene function in defined in vitro
71 conditions or in animal models of infection, and discover new genes involved in specific biological
72 processes, including secretion, sporulation, and conjugation (7). More recently, transposon
73 mutagenesis approaches in which each transposon carries a unique 20-bp barcode sequence have been
74 developed (8); each insertion site is associated with a short barcode, and the abundance of all the
75 mutant strains in the pool can be assessed by a simple amplification and sequencing of the barcode.

76 Here, we used a barcoded Tn-Seq approach to identify genes affecting fitness in a freshwater
77 oligotrophic bacterium, *Caulobacter crescentus* strain CB15 (hereafter referred to as *Caulobacter*),
78 cultivated in filtered and unfiltered freshwater of Lake Michigan, Illinois, USA. As a well-

79 characterized and genetically tractable freshwater oligotroph that was originally isolated from a
80 freshwater pond in California, USA, in 1960 (9), *Caulobacter* is uniquely suited for this investigation.
81 However, *Caulobacter* is typically grown in a complex medium consisting of dilute peptone and yeast
82 extract (PYE) or in a defined medium consisting of mineral salts and a single carbon source (M2X)
83 (10), neither of which adequately represents natural freshwater. PYE is replete with peptides, amino
84 acids, and a range of carbon sources, and the defined medium contains abundant macronutrients and
85 enables dense culture growth, but requires that *Caulobacter* synthesizes all cellular building blocks
86 from salts and a simple sugar. Natural freshwaters, by contrast, contain an undefined, complex mixture
87 of organic and inorganic nutrient sources. Labile compounds such as amino acids and simple
88 carbohydrates are turned over rapidly and thus never accumulate to high concentrations. We predicted
89 that genes that are dispensable in PYE or M2X medium would provide important functions when cells
90 are subjected to the physicochemical and biological complexity of natural freshwater. A previous Tn-
91 Seq study has examined the genes required for *Caulobacter* growth in PYE (11), providing a reference
92 point for comparison to natural freshwater (see Supplemental Material).

93 Here we report that, based on global mutant fitness scores, Lake Michigan water is more
94 similar to PYE than to M2X defined medium. Growth in M2X defined medium poses larger fitness
95 constraints on the mutant pool than growth in either PYE or Lake Michigan water. We uncovered
96 fitness costs and benefits for genes involved in amino acid and nucleotide biosynthesis, motility, and
97 adhesion across the distinct cultivation conditions. Surprisingly, there were few significant differences
98 in the fitness of mutant strains cultivated in filtered and unfiltered lake water. Though all lake water
99 experiments were conducted within a 2-week time period, we observed day-to-day variability in strain
100 fitness scores, which likely reflects temporal heterogeneity of the lake environment. This variability in
101 strain fitness scores in lake water collected on separate days contrasts with low variability of fitness
102 scores for cell in replicate defined media conditions.

103

104 **RESULTS**

105 *Growth of* Caulobacter *is substantive in waters from two Laurentian Great Lakes*

106 To identify *Caulobacter* mutants with altered fitness in a bona fide freshwater environment, it was first
107 necessary to test whether we could measure substantive growth of *Caulobacter* in natural freshwater.
108 Lake Michigan, one of the five Laurentian Great Lakes, represents the kind of dynamic oligotrophic
109 freshwater ecosystem that is inhabited by *Caulobacter* spp. (12, 13). We investigated the growth
110 dynamics of wild-type *Caulobacter* (**Fig. 1**) in water collected from the littoral zone of Lake Michigan.
111 With no additional supplementation, filtered (0.1 μm) lake water supported *Caulobacter* growth to a
112 maximal density of approximately $5 \times 10^5$ CFU/mL (**Fig. 1A**). From an initial inoculum of $2.5 \times 10^4$
113 CFU/mL, *Caulobacter* doubled 4–5 times at a rate of approximately 0.14 $hr^{-1}$ or a doubling time of 5
114 hours. Similar growth rates were observed in unfiltered lake water. Supplementation with 0.1% xylose
115 increased the maximal density by about 10-fold, while addition of 23 μm $K_2HPO_4$ had no effect (**Fig.**
116 **1A**), implying that carbon, but not phosphorus, limits *Caulobacter* growth in Lake Michigan water. For
117 comparison, we also assayed *Caulobacter* growth in water collected from Lake Superior and found a
118 similar growth yield (**Fig. 1B**). Supplementation with either 23 μm $K_2HPO_4$ or 0.1% xylose did not
119 significantly enhance *Caulobacter* growth, but together 0.1% xylose and 23 μM $K_2HPO_4$ enhanced
120 growth by more than 10-fold, suggesting that both carbon and phosphorus limit growth in Lake
121 Superior. By comparison, *Caulobacter* grew to a density of $3 \times 10^9$ CFU/mL in PYE broth or in
122 defined M2X medium (**Fig. 1C**). It is noteworthy that these high population densities attained in
123 laboratory media are unsustainable: in both M2X and PYE, cell density decreased by 2–3 orders of
124 magnitude after 2 days of cultivation, and long-term culture viability was more compromised in M2X
125 than in PYE. By contrast, the lower population densities achieved in lake water were stable over the
126 course of 1 week (**Fig. 1**). This result is consistent with a report by Poindexter describing *Caulobacter*
127 species that tolerated prolonged nutrient scarcity with little loss of viability, a physiological feature of

128 the genus that has been leveraged for their enrichment from water and soil samples (14). Based on our
129 results, we chose to carry out a genetic analysis of mutant *Caulobacter* strains carrying transposon
130 insertions in unsupplemented water from Lake Michigan.
131
132 *A genomic-scale approach identifies* Caulobacter *mutants with altered growth and survival in lake*
133 *water*
134 We sought to test whether different genes and pathways are required for *Caulobacter* growth in a
135 natural freshwater medium compared to defined M2X or complex PYE medium. To this end, we
136 constructed two different transposon mutant libraries using Tn5, which inserts randomly throughout
137 the genome, and Tn-Himar, which inserts specifically in TA dinucleotides. The *Caulobacter* genome is
138 67% G+C, and TA dinucleotides occur on average every 82 bp. Each transposon in the pool of Himar
139 transposons contains a unique 20-bp barcode sequence (Wetmore et al 2015). This barcode approach
140 (BarSeq) has the benefit of requiring the sites of insertion to be mapped only once. Thereafter, the
141 abundance of each Tn-Himar mutant strain in a complex pool can be quantified by amplifying and
142 sequencing the barcode, which uniquely identifies the insertion site (8). Both transposon libraries were
143 constructed by growing cells in PYE. Thus, insertions in genes essential for growth in this condition
144 are not represented in either library. Transposon pool statistics for Tn-Himar and Tn5 are shown in
145 **Table 1**.
146 We cultivated the *Caulobacter* Tn5 pool in PYE and filtered lake water (0.1 μm). Though Tn5
147 is capable of insertion at almost every position in the genome, our Tn5 library had lower site saturation
148 than our Tn-Himar library, which limited the statistical power to identify significant fitness effects (15,
149 16). We calculated mutant fitness and gene essentiality for all genes (**Table S3**) and identified 55
150 genes for which Tn5 disruption significantly diminished or enhanced growth in lake water relative to
151 PYE (adjusted p-value cutoff < 0.01) (**Table S4**). Given the limited statistical power of the Tn5
152 dataset, we primarily focused our analyses on the Tn-Himar dataset, with the Tn5 data providing useful

153 validation.

154 The *Caulobacter* Tn-Himar mutant pool was cultivated in the following conditions: 1) complex
155 PYE medium, 2) defined M2X medium, 3) filtered lake water, and 4) unfiltered lake water. Replicate
156 lake water treatments were performed using water collected from Lake Michigan on 4 different days in
157 December 2016 at Promontory Point, Hyde Park, Chicago, Illinois, USA (**Table 2**). The Tn-Himar
158 library contained an estimated $2 \times 10^6$ strains, of which $7 \times 10^4$ passed the criteria for barcode mapping
159 (8). To capture the diversity of our mutant pool, we initiated the selections with the same total number
160 of cells ($2.5 \times 10^7$) in each treatment and aimed for 4–5 doublings in late exponential growth. Given
161 that the maximum density of *Caulobacter* cultivated in Lake Michigan water is approximately $10^6$
162 cells/mL (**Fig. 1**), volumes of 2.5 L were required for lake water treatments. Cells were harvested by
163 filtration after 64 h of growth. For PYE and M2X treatments, cells were grown in 1.5 mL volumes for
164 10 and 20 h, respectively, and harvested by centrifugation after approximately five doublings. Initial
165 and final cell densities of replicate experiments are shown in **Fig. S1**.

166 After harvest, barcodes were amplified by PCR and sequenced. Sequences were analyzed using
167 the approach outlined by Wetmore and colleagues to quantify the consequences on fitness of disrupting
168 each gene under each cultivation condition relative to cultivation in complex PYE medium. Strain
169 fitness can be viewed as the $\log_2$ of the ratio of barcode abundance in lake water (or M2X) to barcode
170 abundance in the 'control' PYE condition (8). Given the average 20–30-fold increase in cell number
171 during growth of the mutant pool, a Tn-Himar insertion strain that did not grow at all would be
172 expected to have a fitness score in the range of -4 to -5. More extreme (lower) fitness scores may be
173 expected for strains that did not survive cultivation. The observed fitness scores in this experiment are
174 consistent with these expectations (**Fig. 2**). The complete list of Tn-Himar BarSeq gene fitness values
175 is presented in **Table S2.**

176 To evaluate the validity of the BarSeq analysis approach, we examined fitness scores for strains
177 with insertions in xylose utilization genes after cultivation in M2X. Genes in the *xylXABCD* operon are

178 required for xylose utilization (17, 18). As expected, insertions in these genes generated fitness scores
179 of -3.6 to -6.6 when the pool was cultivated in M2X (**Fig. 3A, Table S2**). Disruption of *xylR*, which
180 functions as a transcriptional repressor of the xylose operon (17), resulted in a positive fitness score in
181 M2X relative to PYE, consistent with the idea that derepression of the xylose utilization genes is
182 advantageous when xylose is the sole carbon source. Disruption of the *xylXABCD* genes had little to no
183 consequence on fitness in lake water, which contains a range of carbon sources beyond xylose;
184 disruption of *xylR* resulted in a modest fitness decrease in lake water relative to PYE (**Fig. 3A)**.
185

186 *Variability in strain fitness across replicates is higher in lake water than in defined medium*

187 M2X medium has a relatively stable, defined composition. Lake water, by contrast, is heterogeneous
188 over time and, particularly in the case of unfiltered water, in space (because of variations in e.g.
189 dissolved solutes, particles or grazers). We collected water on four separate occasions within a 2-week
190 period for the four biological replicates and observed greater variability in strain fitness scores across
191 the four lake water replicates than in M2X (**Fig. 2A & 3C, and Table S5**). This variability may reflect
192 changing environmental conditions, as differences in temperature, precipitation, and wind-driven
193 mixing could influence nutrient availability and microbial community composition, which could in
194 turn alter the functional importance of particular genes. For example, more similar fitness profiles were
195 obtained using water collected Nov 30 and Dec 9 (replicates 1 and 3) compared to Dec 6 and Dec 12
196 (replicates 2 and 4) (**Fig. 3B**). Significant snowfall events preceded both Dec 6 and Dec 12 water
197 collections, but not the Nov 30 and Dec 9 collections, suggesting a possible weather-related driver of
198 lake water strain fitness variability. The fitness of individual mutants could, therefore, help
199 discriminate genes whose functions are consistently important from genes that are exploited under
200 transient conditions.
201

202 *Tn-Himar strain fitness defects are more extreme in defined medium than in lake water*

203 Transposon disruption of genes required for amino acid biosynthesis, nucleotide biosynthesis,
204 lipopolysaccharide biosynthesis, and nucleotide sugar biosynthesis resulted in extreme (fitness score <
205 -4) growth defects in M2X (**Fig. 3, 4, S3–S5, and Tables S2–S5**). Given that growth in M2X medium
206 requires de novo biosynthesis of diverse monomers and intermediates, many of which are supplied
207 exogenously in the reference PYE condition, these results were expected. In many cases, strains with
208 severe fitness defects in M2X also had reduced growth in lake water, though the fitness costs of
209 disruption were less severe. We controlled the number of doublings (approximately 4–5) across all
210 conditions, so the more pronounced negative fitness scores in defined medium compared to lake water
211 cannot be explained by differences in the number of doublings. Instead, these differences imply that
212 lake water is more similar to PYE than M2X is to PYE, in terms of the metabolic demands it imposes
213 on cells; this result is supported by principal component analysis across the different growth conditions
214 (**Fig. 3A**). Indeed, we expect that the waters of Lake Michigan carry a diverse array of metabolites and
215 carbon and nitrogen substrates for growth. In this complex milieu, genes may be dispensable if their
216 products are supplied exogenously, or if alternative pathways exist. Growth in defined media,
217 however, provides fewer opportunities for substrate uptake and functional substitution. Surprisingly,
218 we did not observe large differences in strain growth between unfiltered and 0.1μm filtered lake water
219 conditions (**Fig. 3C**). This suggests that other microorganisms present in the lake water did not have
220 major effects on strain growth given our cultivation approach and experimental timescale.
221
222 *Pathway analysis of BarSeq fitness data in lake water and defined medium*
223 To further explore pathways conferring differential fitness between lake water and artificial media, we
224 chose to focus on genes whose disruption induced the largest fitness effects relative to the PYE
225 reference condition, namely fitness scores ± 3 standard deviations (± 3σ) from the mean (< -1.2 and >
226 +1.2). Based on this criterion, we identified 83 and 82 genes in the filtered and unfiltered lake water
227 conditions, respectively, and 213 genes in the defined M2X medium (**Table S7**). Of these 238 total

228 genes, 40 genes were shared across all conditions (**Table S8**). Broad functional patterns were assessed
229 using clusters of orthologous group (COG) annotations (19) (**Fig. 4**).

230 The largest fitness effects were observed for genes in amino acid and nucleotide biosynthesis,
231 and in transport; we observed similar defects in Tn5 strains harboring disruptions in select amino acid
232 and nucleotide anabolic genes (**Tables S3, S4, S6**). Tn-Himar disruption of a set of genes encoding
233 catabolic enzymes involved in the glycine cleavage pathway and branched amino acid degradation
234 resulted in enhanced fitness scores in both the defined M2X medium and in lake water, relative to the
235 PYE control (**Fig. 3C and Tables S6, S7**). This result was supported by similar fitness trends in the
236 Tn5 dataset (**Tables S3, S4, S6**). Deletion of these catabolic genes is likely detrimental in the reference
237 PYE condition, compared to M2X or lake water; alternatively, it is possible that the loss of particular
238 catabolic functions is advantageous, even in natural freshwater. The former model is consistent with
239 transcriptional data showing that select amino acid degradation pathways — including glycine
240 cleavage, histidine, branched chain, and phenylalanine degradation — are upregulated in PYE
241 compared with M2X (20).

242 We also observed strong positive effects on fitness after disruption of genes involved in
243 motility in lake water relative to PYE (**Fig. 3 & 4**). We more carefully examined the fitness scores of
244 genes involved in synthesis and assembly of the flagellum (**Fig. 5**). The flagellum is assembled in a
245 regulated hierarchy of stages, which is well described in *Caulobacter* (21-23). Class II genes encode
246 the inner components of the flagellum, including the export apparatus, and regulatory proteins that
247 activate expression of class III and IV genes upon completed assembly of the class II structure. Class
248 III genes encode the basal body and hook structures. Completion of the class III structures activates
249 translation of the class IV genes, which encode the subunits of the flagellar filament. Thus, defects in
250 each class prevent expression of subsequent classes. Within each class of flagellar genes, we observed
251 consistent fitness effects, demonstrating the power of this method to capture even modest effects of
252 gene disruption. Disruption of class II flagellin genes conferred an advantage that was significantly

253 greater in lake water than in M2X compared to PYE, (**Fig. 5B & S2**). The effect of disruption of class
254 III genes followed similar trends with smaller magnitude effects on fitness. *Caulobacter* encodes six
255 redundant class IV flagellin genes (24), thus it was not surprising that disruption of any of the three
256 flagellin genes represented in our BarSeq collection had no effect on fitness. Disruption of the motor
257 stator gene *motA* or *motB*, which results in a fully assembled but paralyzed flagellum (25, 26), did not
258 affect fitness under our cultivation conditions. This result suggests that the fitness advantage of
259 flagellar gene disruption is not derived from energy saved in powering the flagellum, but rather in
260 energy saved in synthesizing and assembling the flagellum, or as a result of regulatory processes
261 affected by incomplete assembly of the flagellum. In the lake water cultivations, we noted appreciable
262 day-to-day variation in the fitness of each class of flagellar gene mutants (**Fig. S2**), which was
263 particularly pronounced for class III genes. Patterns in this day-to-day variability were consistent
264 across members of each class, suggesting that this variability is driven by environmental factors.

265 Fitness was further affected by the ability to synthesize the polar extracellular adhesin known as
266 the holdfast (**Fig. 5 & S2**). We systematically analyzed genes involved in synthesis, secretion, and
267 attachment of the holdfast in each cultivation condition. Most holdfast genes yield partial to complete
268 defects in holdfast development when disrupted (27); we categorized these genes as 'unique functions'
269 genes. However, two sets of holdfast biosynthesis genes have redundant functions: two Wzy-family
270 polymerase genes that function in holdfast development and three paralogs of the HfsE
271 glycosyltransferase that have genetically redundant activities in holdfast synthesis (27). Transposon
272 disruption of genes in the unique function group resulted in a modest but consistent fitness advantage
273 in M2X and a fitness disadvantage in both filtered and unfiltered lake water, relative to PYE (**Fig. 5A**
274 **& S2A**). When considered as a group, the fitness consequence for loss of holdfast was significantly
275 different between growth conditions ($p < 0.0001$) (**Fig. 5A**). Disruption of holdfast genes with
276 redundant functions, which do not result in loss of holdfast development (27), had no consequence on
277 fitness.

278

279 *Cultivation in freshwater uncovers phenotypes associated with genes of unknown function*

280 We hypothesized that many genes of unknown function play important roles in natural environmental
281 contexts but not in typical laboratory media. Of the group of genes outside the ± 3σ fitness (advantage
282 or disadvantage) score cutoff, hypothetical genes or genes of unknown function accounted for 16% in
283 filtered lake water, 15% in unfiltered lake water, and 7% in defined medium (**Table 3**). Across these
284 three conditions, a total of five hypothetical genes were shared. *CCNA_03860* was the only
285 hypothetical gene for which disruption provided a fitness benefit across all three conditions relative to
286 PYE. Further investigation of CCNA_03860 identified a conserved domain belonging to the YkuD
287 superfamily, which has been shown to have L,D-transpeptidase catalytic activity, providing an
288 alternate pathway for peptidoglycan cross-linking (28, 29). Disruption of *CCNA_01724*, *CCNA_03864*,
289 *CCNA_03909*, and *CCNA_00375* resulted in reduced fitness across all three conditions relative to
290 PYE. The majority of hypothetical genes and genes of unknown function that met the ± 3σ cutoff
291 criterion exhibited negative fitness values (92%), indicating that disruption is detrimental in these
292 conditions compared to PYE. This result supports our hypothesis that genes of unknown function play
293 important roles in supporting growth in natural environments.

294

295 **DISCUSSION**

296 *Tn-seq fitness scores can provide a readout of environmental state*

297 Bacterial genomes carry relatively little noncoding DNA. Genes that confer no fitness benefit tend to
298 decay over time (30) implying that genes that are maintained are beneficial at least under some
299 circumstances. Yet traditional microbial cultivation approaches often fail to yield discernable mutant
300 phenotypes for many classes of genes. To understand the functional capacity of bacterial genomes, it is
301 necessary to interrogate gene function in more environmentally relevant contexts, embracing spatial

302 and temporal heterogeneity. The genomic-scale fitness analysis of *Caulobacter* transposon mutants
303 reported in this study provides new understanding of genes that affect growth in a *bona fide* freshwater
304 environment. Relative to a complex peptone medium (PYE), disruption of genes involved in amino
305 acid biosynthesis, nucleotide biosynthesis, lipopolysaccharide biosynthesis, and nucleotide sugar
306 biosynthesis, and other classes of metabolic genes results in large fitness defects in natural freshwater
307 from Lake Michigan. However, fitness effects were variable across lake water replicates, using water
308 collected across several days; this variability likely reflects physicochemical and biological variability
309 in the lake and suggests an important role for transient response genes in fluctuating environments.
310

311 *Consequences on strain fitness of disrupting genes that function in motility and attachment in*
312 *freshwater*

313 The energetic cost of flagellar biosynthesis and motility is well established (31, 32). Our data
314 indicate that transposon disruption of genes required for the synthesis of the single polar flagellum of
315 *Caulobacter* enhanced fitness in lake water relative to PYE medium (**Fig. 5**). This is consistent with a
316 *Salmonella* Tn-Seq study that revealed a fitness advantage in strains with disrupted flagellar genes
317 (33). Notably, we found that fitness effects were not uniform across all flagellar genes: disruption of
318 class II genes, which has the greatest impact on flagellar gene expression, also led to greater effects on
319 fitness, compared to class III and IV genes. The fitness enhancement in lake water is not related to the
320 energy consumed by motor rotation, as strains with insertions in the *motA* and *motB* stator genes,
321 which assemble a full but non-rotary flagellum (26), showed no fitness difference (**Fig. 5).** We
322 conclude that the relative fitness advantage of flagellar gene disruption is related to the cost of
323 biosynthesis of flagellar proteins. It seems certain that over longer cultivation timescales, and in more
324 spatially complex environments, the *Caulobacter* flagellum provides a fitness advantage, as flagellar
325 genes are maintained in natural freshwater environments.

326 Our data reveal that disruption of genes required for holdfast biosynthesis is disadvantageous
327 when strains are cultivated in lake water relative to PYE. This fitness cost was evident in both filtered
328 (particle-free) and unfiltered lake water relative to PYE (**Fig. 5**), suggesting that the effect is not due to
329 adhesion to particles in the medium. Instead, it is possible that the holdfast confers a growth advantage
330 by enabling adherence to the flask surface, where polymeric nutrients concentrate to form conditioning
331 films (34, 35).
332
333 *Genetic evidence that a complex medium is a more environmentally relevant growth condition for*
334 Caulobacter *than a typical defined mineral medium*
335 Fitness defects of *Caulobacter* mutants were often more severe in a defined mineral xylose medium
336 (M2X) than in lake water, relative to PYE. This is consistent with our observation that the overall
337 fitness profiles of *Caulobacter* Tn strains cultivated in lake water more closely resemble PYE than
338 M2X, suggesting that cultivation of *Caulobacter* in dilute complex medium is a better proxy for
339 natural freshwater. *Caulobacter* belongs to a group of dimorphic prosthecate (i.e. stalked)
340 alphaproteobacteria that are often specialized for oligotrophic, dilute environments (9, 14). Indeed, the
341 inhibition of growth and stalk development due to excess nutrients was the first physiological property
342 of *Caulobacter* spp. to be described (36). Complex and defined media of varying compositions have
343 been outlined for cultivation of *Caulobacter* and related genera, but it is notable that dilute peptone
344 (less than 0.2% w/v) generally supports growth of all dimorphic prosthecate bacteria (14). This
345 observation supports the notion that the natural nutrient environment of this class of bacteria is best
346 captured by cultivation in a dilute complex medium that contains amino acids and other trace complex
347 biomolecular components. While it is likely that a more optimal defined medium exists that better
348 captures the conditions of natural freshwater, our data provide evidence that an M2-based medium
349 exerts highly specific metabolic constraints on the bacterial cell and may not be an ecologically or
350 physiologically relevant growth condition.

351

352 *An approach to study hypothetical genes and domains of unknown function in freshwater bacteria*

353 The explosion of bacterial genome sequence information has far outpaced our ability to
354 characterize gene function using traditional approaches, leading to the accumulation of thousands of
355 'unknown' protein families. Many of these families are conserved throughout the bacterial domain,
356 which is evidence that they confer a selective benefit in particular conditions. This leads to the
357 following question: under what circumstances do these conserved families provide a fitness advantage?
358 At the outset of this study, we hypothesized that many of these unknown protein families would prove
359 to be important in the natural ecological context of a bacterium. Among the genes whose disruption
360 leads to the greatest fitness effects (± 3σ) in filtered lake water relative to PYE, approximately 15% are
361 hypothetical or conserved genes of unknown function (**Tables 3, S7–S8**). The mutagenesis, cultivation,
362 and sequencing approach we describe here indicates that these genes of unknown function play an
363 important role in *Caulobacter* physiology in a natural freshwater environment. Going forward, one can
364 take advantage of lake-specific growth phenotypes to begin to define the functions of these genes in an
365 ecologically relevant context.

366

367 **MATERIALS AND METHODS**

368 *Bacterial strains and primers*. Strains used in this study and their descriptions are listed in **Table S1**.
369 Primer name, sequence, and function are also listed in **Table S1**. All primers were synthesized by
370 Integrated DNA Technologies (Coralville, IA, USA).

371

372 *Composition of laboratory media*. *Caulobacter* was grown in PYE complex medium (0.2% peptone
373 (Fisher Scientific Lot No. 005604), 0.1% yeast extract (Fisher Scientific Lot No. 166295), 0.5 mM
374 $MgSO_4$, 0.5 mM $CaCl_2$) or M2X minimal defined medium (6.1 mM $Na_2HPO_4$, 3.9 mM $KH_2PO_4$, 9.3

375 mM $NH_4Cl$, 0.5 mM $MgSO_4$, 10 μM $FeSO_4$ (EDTA chelate; Sigma Chemical Co.), 0.25 mM $CaCl_2$)
376 supplemented with 0.15% xylose (10). *Escherichia coli* strains for transposon conjugation were grown
377 in LB broth (1% peptone, 0.5% yeast extract, 0.5% NaCl). Solid growth medium was made by adding
378 1.5% agar to the above liquid media.
379
380 *Lake water collection and filtration*. Water from Lake Michigan was collected at Promontory Point,
381 Chicago, Illinois, USA (Latitude: 41.794, Longitude: -87.579). Parameters such as water temperature,
382 phosphate and nitrate/nitrite level (Aquacheck Water Quality Test Strips; Hach), and pH (pH indicator
383 strips, Fisher Scientific) were tested on-site and are listed in **Table 2**. Water collection occurred in
384 2016, on Nov 30, Dec 6, Dec 9, and Dec 12. Lake water was filtered using Nalgene™ Rapid-Flow™
385 Sterile Disposable 0.1 μm Filter Units with PES Membrane (Thermo Scientific).
386
387 *Measurement of* Caulobacter *growth in waters of Lake Michigan and Lake Superior*. Colonies of
388 *Caulobacter* were inoculated into 2 mL of PYE in glass culture tubes (13 × 100 mm) and grown
389 overnight at 30°C with shaking at 200 rpm, for a total of five biological replicates. At saturation, 1 mL
390 of culture was centrifuged at 8,000 ×*g* and washed twice in 1 mL of filtered lake water. The washed
391 pellet was resuspended in filtered lake water to a final $OD_{660}$ of 0.1, and 0.5 μL (approximately 0.5–1
392 × $10^5$ cells) was inoculated into 5 mL of filtered lake water in a glass culture tube (20 mm × 150 mm)
393 in technical duplicate. Cultures were grown at 30°C with shaking at 200 rpm. At various time intervals
394 over the course of 48 h, 20 μL of culture was removed to perform 10-fold serial dilutions for colony
395 enumeration. A total of 5 μL of each dilution was spotted onto PYE agar plates, which were incubated
396 at 30°C for 2 days. Growth was monitored by enumeration of colony forming units (CFUs).
397
398 *Construction of a barcoded Tn-Himar (BarSeq)* Caulobacter *mutant library*. The recipient strain
399 (*Caulobacter*) was grown overnight in 2 mL of PYE at 30°C with shaking at 200 rpm. This starter

400 culture was used to inoculate 20 mL of PYE and grown at 30°C overnight with shaking at 200 rpm
401 until saturated. The donor *E. coli* strain (APA752, gift from Adam Deutschbauer, University of
402 California-Berkeley, USA), carrying the pKMW3 (kanamycin resistant) Himar transposon vector
403 library (8), was inoculated into 20 mL of LB containing kanamycin (30 μg $mL^{-1}$) and diaminopimelate
404 (300 μM) and grown overnight at 37°C with shaking at 200 rpm. To mate the barcoded transposon
405 pool into *Caulobacter*, both the recipient strain and the donor strain were centrifuged at 8,000 ×*g* for 2
406 min and resuspended in a total volume of 500 μL of PYE medium. The cultures were mixed at a 10:1
407 ratio of recipient to donor and mixed by gentle pipetting. The mixed culture was centrifuged again at
408 8,000 ×*g*, and the supernatant decanted. The cells were resuspended in 30 μL of PYE, spotted onto a
409 PYE agar plate containing diaminopimelate (300 μM), and incubated overnight at 30°C. After growth,
410 the mating spot was scraped from the plate and resuspended in 6.5 mL of PYE. This suspension was
411 spread evenly (500 μL per plate) over 14 large (150 × 15 mm) PYE agar plates containing 25 μg $mL^{-1}$
412 kanamycin and incubated for approximately 3 days at 30°C. Cells were harvested from all the plates
413 and inoculated into 400 mL of PYE containing 5 μg $mL^{-1}$ kanamycin. This cell mixture was grown at
414 30°C with shaking at 200 rpm for three doublings. Cells were centrifuged at 8,000 ×*g*, resuspended in
415 70 mL of PYE containing 15% glycerol, and stored as 1 mL aliquots at -80°C.
416
417 *Mapping of the sites of Tn-Himar insertion in the* Caulobacter *BarSeq library*. Genomic DNA was
418 extracted using guanidium thiocyanate as previously described (37). DNA quality and quantity was
419 measured using a $NanoDrop^{OneC}$ (Thermo Scientific). The quality of extracted genomic DNA from the
420 *Caulobacter* library was also checked by using a Bioanalyzer (Agilent) prior to the sequencing run.
421 The DNA was sheared (~300 bp fragments), cleaned with a standard bead protocol, end-repaired and
422 A-tailed, and a custom double-stranded Y adapter was ligated. The custom adapter was prepared by
423 annealing Mod2_TS_Univ (ACG CTC TTC CGA TC*T) and Mod2_TruSeq (phos-GAT CGG AAG
424 AGC ACA CGT CTG AAC TCC AGT CA) as described (8). The sheared fragments containing

425 transposon fragments were enriched by PCR using the primers Nspacer_BarSeq_pHIMAR (5′ ATG
426 ATA CGG CGA CCA CCG AGA TCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT
427 NNN NNN CGC CCT GCA AGG GAT GTC CAC GAG 3′) and P7_MOD_TS_index1 (5′ CAA GCA
428 GAA GAC GGC ATA CGA GAT CGT GAT GTG ACT GGA GTT CAG ACG TGT GCT CTT 3′)
429 using GoTaq® Green Master Mix according to the manufacturer's protocol in a 100-µL volume with
430 the following cycling conditions: 94°C for 2 min, 25 cycles at 94°C for 30 s, 65°C for 20 s, and 72°C
431 for 30 s, followed by a final extension at 72°C for 10 min. After a second bead cleanup, the
432 *Caulobacter* BarSeq library was sequenced using a standard Illumina sequencing primer on an
433 Illumina HiSeq2500 at the University of Chicago Genomics Facility with a 150-bp single-end read.
434 The locations of Himar transposon insertions were aligned and mapped using BLAT (38), and unique
435 barcode sequences were associated with their corresponding genome insertion location using a custom
436 Perl script (MapTnSeq.pl). Sets of barcodes that consistently map to one location in the genome were
437 identified using a custom Perl script (DesignRandomPool.pl). These scripts have been described by
438 Wetmore and colleagues (8) and are available at https://bitbucket.org/berkeleylab/feba.
439
440 *Cultivation of the BarSeq transposon library in Lake Michigan water, and harvest of genomic DNA*
441 *after cultivation*. A total of 2 mL of the *Caulobacter* BarSeq library was inoculated into 18 mL of
442 complex medium (PYE). The culture was split into two tubes (20 × 150 mm) with a volume of 10 mL
443 each and left to outgrow in a cell culture roller drum (Fisher Scientific) at 30°C for 4 h at speed 4. The
444 tubes were moved to an incubator set at 30°C with shaking at 200 rpm, and the cultures were grown for
445 an additional 2 h. Cultures were combined and centrifuged for 20 min at 3,000 ×*g* at 4°C. The cell
446 pellet was resuspended and washed in 10 mL of filtered lake water, and centrifuged again at 3,000 ×*g*
447 for 20 min at 4°C. The resulting pellet was resuspended in 5 mL of filtered lake water, and the $OD_{660}$
448 measured. Flasks containing filtered or unfiltered lake water (2.5 L total volume, 4 replicates per
449 condition per experiment) were inoculated with the washed library with the aim of an initial starting

450 concentration of approximately 2.5 × $10^7$ total cells per flask. Flasks were incubated at 30°C with
451 shaking at 150 rpm. At 0 and 64 h, an aliquot of culture was removed to perform serial dilutions and
452 spotted on PYE agar plates. Growth was monitored by enumeration of CFUs on plates after 2 days of
453 incubation at 30°C. After ~64 h of growth, cells were collected by filtration using an Express Plus®
454 Membrane 0.22 μM filter (Millipore). Filters were stored at -80°C until needed. To mimic saturating
455 conditions with the same number of doublings in defined M2X and complex PYE laboratory medium
456 as in lake water, we inoculated the cultures at a concentration that after five doublings (the estimated
457 number of doublings in lake water), the cultures reached saturation. Cells were pelleted at 10,000 ×g
458 for 1 min and stored at -20°C. Genomic DNA from all samples was extracted using guanidium
459 thiocyanate as previously described (37), with the exception that the lake water samples were lysed
460 directly from the filters they were collected on. DNA quality and quantity was measured using a
461 NanoDrop$^{OneC}$ (Thermo Scientific).
462
463 *PCR amplification of Tn-Himar barcodes and sequencing*. PCR amplification for each sample was
464 performed as previously described (8) using a standard reaction protocol for Q5 DNA polymerase
465 (New England BioLabs) with the primers BarSeq_P1 (reverse primer, 5′ AAT GAT ACG GCG ACC
466 ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNN GTC GAC
467 CTG CAG CGT ACG 3′) and 1 of 16 forward primers (BarSeq_P2_IT001 to BarSeq_P2_IT016, listed
468 in **Table S1**) containing unique 6-bp TruSeq indexes that were sequenced using a separate index
469 primer. Cycling conditions were as follows: 98°C for 4 min followed by 25 cycles of 30s at 98°C, 30s
470 at 55°C, and 30s at 72°C, followed by a final extension at 72°C for 5 min. PCR products per medium
471 condition were mixed together for PCR purification using GeneJET PCR Purification Kit (Thermo
472 Scientific). Purified samples were run on a 2.5% agarose gel to confirm correct product size (~200 bp).
473 A total of 10 μL per purified PCR product was pooled, assessed for quality, and quantified using a
474 Bioanalyzer. The amplified barcodes from the reference (PYE) and treatment (M2X, unfiltered lake

475 water, and filtered lake water) were sequenced on an Illumina HiSeq4000 at the University of Chicago
476 Genomics Facility, multiplexing all 16 samples in one lane. Tn-Himar 150-bp single-end read raw
477 sequence data for transposon insertion site mapping and the 50-bp single-end PCR amplicon barcode
478 sequence data (for BarSeq analysis) have been deposited in the NCBI Sequence Read Archive under
479 BioProject accession PRJNA429486; BioSample accession SAMN08348121; SRA accession
480 SRP128742.
481
482 *Analysis of Tn-Himar strain fitness*. We followed the fitness calculation protocol of Wetmore and
483 colleagues (8), using scripts available at https://bitbucket.org/berkeleylab/feba. Briefly, the total count
484 of each barcode in each sample was calculated using a Perl script (MultiCodes.pl) and, from this table
485 of barcodes, strain fitness was calculated using an R script (FEBA.R). The fitness of each strain was
486 calculated as a normalized $\log_2$ ratio of barcode counts in the treatment sample to counts in the PYE
487 reference sample. The fitness of genes was calculated as the weighted average of strain fitness values,
488 the weight being inversely proportional to a variance metric. Successful gene fitness calculations
489 required at least 3 reads per strain and 30 reads per gene. Insertions in the first 10% or last 10% of a
490 gene were not considered in gene fitness calculations.
491 Genes for which Tn-Himar disruption resulted in significant growth defects or growth enhancement in
492 the assayed conditions relative to PYE complex medium were determined by calculation of a mean
493 fitness value and also a mean t-value per gene.
494 To assess the distribution of fitness scores, we calculated the standard deviation of each
495 condition using the frequency distribution of the mean fitness value of each gene (filtered lake water =
496 0.41, unfiltered lake water = 0.40, defined medium = 1.1). When the outlier region of the defined
497 medium dataset (< -2.5) was removed, the calculated standard deviation was 0.36; therefore, a standard
498 deviation of 0.4 was chosen and applied to all conditions. Genes with a mean fitness value ± 3σ from

499 the mean (less than -1.2 and greater than +1.2) were selected for further examination. The complete
500 data set of fitness values for each condition is listed in **Table S2**.

501 We examined t-values, the weighted fitness value of a gene divided by a variance metric (as
502 previously described (8)), to provide a metric to assess the significance of fitness values (see **Table S9**
503 for a full list of t-values for each gene). Mean t-values ± the standard deviation for each experimental
504 condition are as follows: M2X (-0.03 ± 1.11), Lake Michigan water filtered (0.01 ± 1.04), Lake
505 Michigan water unfiltered (0.01 ± 1.05). The complete data set of fitness values for each condition is
506 listed in **Table S2**. Mean fitness value and mean t-values plus or minus standard deviation (for the
507 complete *Caulobacter* genome) are presented in **Table S5**.

508

509 *Analysis of* Caulobacter *Tn5-seq fitness.*A *Caulobacter* Tn5 insertion library containing $3 \times 10^5$
510 independent insertions was constructed as previously described (39). The lake water fitness experiment
511 for the Tn5 library (from Lake Michigan water collected in April 2016) was performed similarly to the
512 Tn-mariner (BarSeq) library experiments with the following exceptions: A total of 200 µL of the
513 *Caulobacter* Tn5 library was inoculated into 20 mL of PYE for the initial outgrowth for a total of 5 h.
514 Library outgrowth was inoculated into 2 L for the complex medium (PYE) and unfiltered lake water
515 samples, and 2 replicates of 2 L each for the filtered lake water samples. Lake water cultures were
516 filtered after a total of approximately 60 h of growth, while the complex medium condition was filtered
517 after 12 h to approximate the same number of doublings. However, our complex medium condition
518 doubled over 6 times, versus 4 doublings for the lake samples, resulting in 2 orders of magnitude of
519 growth in the complex medium condition and 1 order of magnitude of growth for each lake water
520 condition.

521 A standard library preparation was performed at the University of Chicago Genomics Facility
522 to ligate standard Illumina adapters to our sheared and sized selected samples. A nested PCR approach
523 was used on these samples to specifically amplify transposon-containing DNA fragments for

524 sequencing. A low cycle PCR amplification for each sample was first performed using a standard
525 reaction protocol for KOD Xtreme™ Hot Start Polymerase with 5% DMSO and 0.3 μM primer using
526 the primers F1 (5′ ACC TAC AAC AAA GCT CTC ATC AAC C 3′) and P7 (5′ CAA GCA GAA
527 GAC GGC ATA CGA 3′) (39). Cycling conditions were as follows: 95°C for 90 s, 5 cycles of 95°C
528 for 15 s 68°C for 30 sec, and 72°C for 30 s; 95°C 15 s, 13 cycles of 55°C 30 s 72°C 30 s, followed by a
529 final extension at 72°C for 5 min. Samples were treated with ExoSAP-IT™ PCR product cleanup
530 reagent (Thermo Fisher Scientific) according to manufacturer's protocol. A second PCR step was
531 performed with the transposon specific primer containing the adapter sequence using KOD Xtreme™
532 Hot Start Polymerase with 5% DMSO and 0.3 μM primer in a 62.5-μL reaction volume using the
533 primers Tn5-left (5′ AAT GAT ACG GCG ACC ACC GAG ATC TAC ACA TCC TCT AGA GTC
534 GAC CTG CAG GCA TGC AAG C 3′) and P7 (5′ CAA GCA GAA GAC GGC ATA CGA 3′) (39).
535 Cycling conditions were as follows: 95°C for 3 min, 12 cycles of 95°C for 30 s 55°C for 30 sec, and
536 72°C for 30 s, followed by a final extension at 72°C for 5 min. Products were run on an 1% agarose gel
537 to confirm products of appropriate size (~200 bp). After standard bead cleanup, the samples were
538 sequenced on an Illumina HiSeq2500 at the University of Chicago Genomics Facility using a custom
539 sequencing primer TnSeq-out (5′ CCT GCA GGC ATG CAA GCT TCA GGG TTG AGA TGT GTA
540 3′) (39).

541 Fitness analysis was performed as previously described (16) using the TRANSIT software
542 (available at https://github.com/mad-lab/transit). We used the comparative analysis functions of
543 TRANSIT to compare our complex medium and lake water conditions. For this, TRANSIT utilizes a
544 permutation test, which compares and quantifies differences in sequencing read counts in a gene
545 between two different conditions to determine if they are statistically significant (16). The
546 comprehensive data set from this analysis is available in **Table S3**. Genes with an adjusted p value
547 <0.01 are listed in **Table S4**. **Table S6** shows the shared gene list between the Tn5 and Tn-mariner
548 approaches used in this study. The raw Tn5-seq data are deposited in the NCBI sequence read archive

549 under BioProject accession PRJNA429486; BioSample accession SAMN08348191; SRA accession
550 SRP128742.
551

552 **ACKNOWLEDGEMENTS**

553 The authors have declared no competing interests. This work was supported by UChicago BIG grant to
554 S.C. and M.C. K.L.H was supported by an NIH Ruth Kirschstein Postdoctoral Fellowship (F32
555 GM122242) and a Chicago Biomedical Consortium Postdoctoral Grant (FP064244-01-PR). L.R.R was
556 supported by the NIH Molecular and Cellular Biology Training Grant (T32 GM007183). P.D.C. is
557 supported an NSF CAREER award (1552647); he began the Tn5 library construction in the laboratory
558 of Dr. Yves V. Brun at Indiana University. We thank the members of the Crosson laboratory for
559 helpful discussions, Adam Deutschbauer (University of California-Berkeley) for the *E. coli* APA752
560 strain, and David Hershey for construction of the *Caulobacter crescentus* CB15 Himar transposon
561 library. We also thank Pieter Faber and Abhilasha Cheruku from the University of Chicago Genomics
562 Facility for helpful discussions.
563

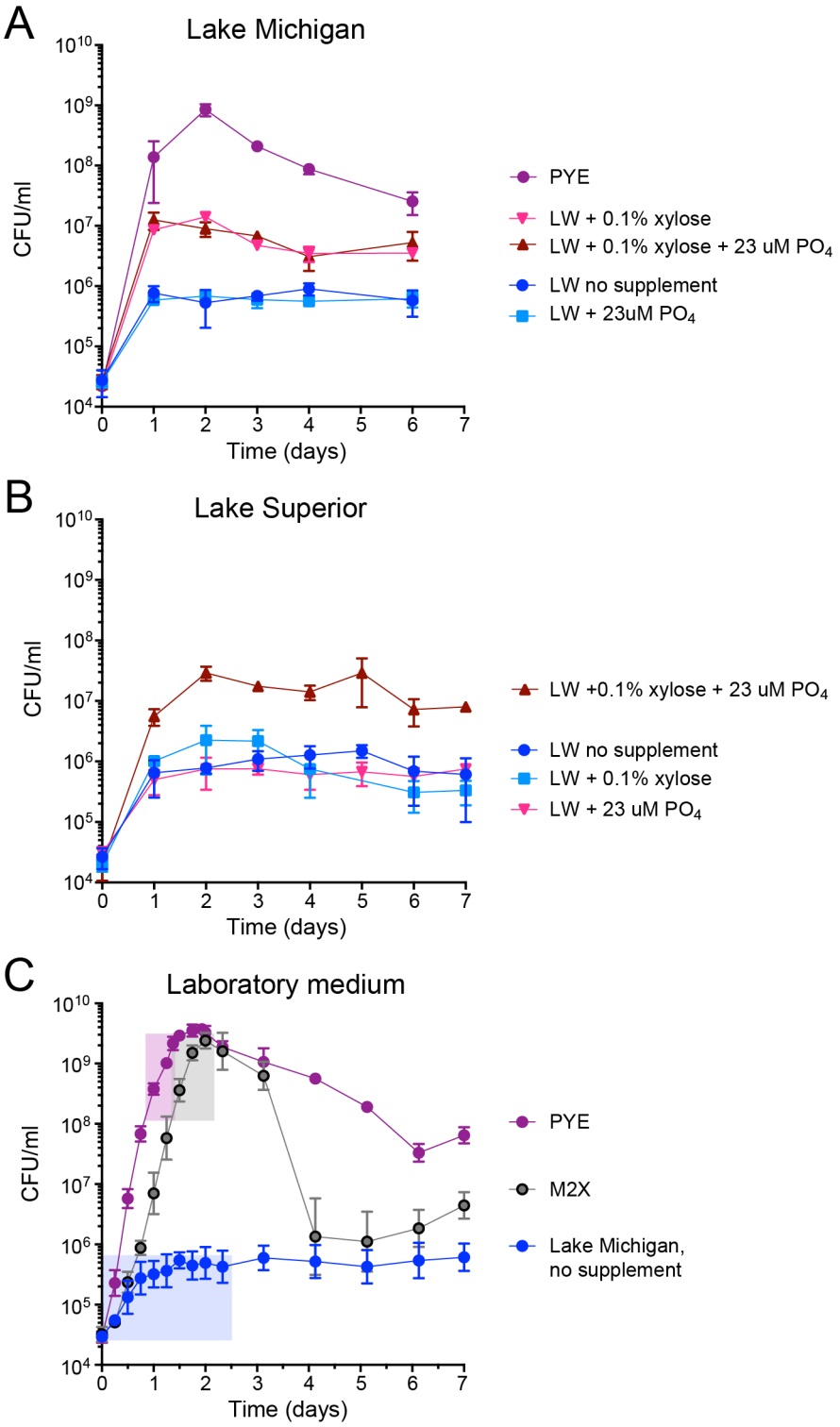
564 **REFERENCES**


565 1. **Cao M, Goodrich-Blair H.** 2017. Ready or Not: Microbial Adaptive Responses in Dynamic
566 Symbiosis Environments. J Bacteriol **199:**pii: e00883-00816.
567 2. **van Gestel J, Vlamakis H, Kolter R.** 2015. Division of Labor in Biofilms: the Ecology of Cell
568 Differentiation. Microbiol Spectr **3:**MB-0002-2014.
569 3. **Vos M, Wolf AB, Jennings SJ, Kowalchuk GA.** 2013. Micro-scale determinants of bacterial
570 diversity in soil. FEMS Microbiol Rev **37:**936-954.
571 4. **Azam F, Malfatti F.** 2007. Microbial structuring of marine ecosystems. Nat Rev Microbiol
572 **5:**782-791.
573 5. **Stocker R.** 2012. Marine microbes see a sea of gradients. Science **338:**628-633.
574 6. **van Opijnen T, Camilli A.** 2013. Transposon insertion sequencing: a new tool for systems-
575 level analysis of microorganisms. Nat Rev Microbiol **11:**435-442.
576 7. **Kwon YM, Ricke SC, Mandal RK.** 2016. Transposon sequencing: methods and expanding
577 applications. Appl Microbiol Biotechnol **100:**31-43.
578 8. **Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J,**
579 **Butland G, Arkin AP, Deutschbauer A.** 2015. Rapid quantification of mutant fitness in
580 diverse bacteria by sequencing randomly bar-coded transposons. MBio **6:**e00306-00315.
581 9. **Poindexter JS.** 1964. Biological Properties and Classification of the Caulobacter Group.
582 Bacteriol Rev **28:**231-295.
583 10. **Ely B.** 1991. Genetics of Caulobacter crescentus. Methods Enzymol **204:**372-384.
584 11. **Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ,**
585 **McAdams HH, Shapiro L.** 2011. The essential genome of a bacterium. Mol Syst Biol **7:**528.
586 12. **Lee PO, McLellan SL, Graham LE, Young EB.** 2015. Invasive dreissenid mussels and
587 benthic algae in Lake Michigan: characterizing effects on sediment bacterial communities.
588 FEMS Microbiol Ecol **91:**1-12.
589 13. **McMillan L, Stout R.** 1977. Occurrence of *Sphaerotilus*, *Caulobacter*, and *Gallionella* in Raw
590 and Treated Water. J Am Water Works Assoc **69:**171-173.
591 14. **Poindexter JS.** 2006. Dimorphic Prosthecate Bacteria: The Genera Caulobacter, Asticcacaulis,
592 Hyphomicrobium, Pedomicrobium, Hyphomonas and Thiodendron, p 72-90. *In* Dworkin M,
593 Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (ed), The Prokaryotes: Volume 5:
594 Proteobacteria: Alpha and Beta Subclasses doi:10.1007/0-387-30745-1_4. Springer New York,
595 New York, NY.
596 15. **Chao MC, Abel S, Davis BM, Waldor MK.** 2016. The design and analysis of transposon
597 insertion sequencing experiments. Nat Rev Microbiol **14:**119-128.
598 16. **DeJesus MA, Ambadipudi C, Baker R, Sassetti C, Ioerger TR.** 2015. TRANSIT--A
599 Software Tool for Himar1 TnSeq Analysis. PLoS Comput Biol **11:**e1004401.
600 17. **Stephens C, Christen B, Watanabe K, Fuchs T, Jenal U.** 2007. Regulation of D-xylose
601 metabolism in Caulobacter crescentus by a LacI-type repressor. J Bacteriol **189:**8828-8834.
602 18. **Stephens C, Christen B, Fuchs T, Sundaram V, Watanabe K, Jenal U.** 2007. Genetic
603 analysis of a novel pathway for D-xylose metabolism in Caulobacter crescentus. J Bacteriol
604 **189:**2181-2185.
605 19. **Wolf YI, Makarova KS, Yutin N, Koonin EV.** 2012. Updated clusters of orthologous genes
606 for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer.
607 Biol Direct **7:**46.
608 20. **Hottes AK, Meewan M, Yang D, Arana N, Romero P, McAdams HH, Stephens C.** 2004.
609 Transcriptional profiling of Caulobacter crescentus during growth on complex and minimal
610 media. J Bacteriol **186:**1448-1461.

611 21. **Ardissone S, Viollier PH.** 2015. Interplay between flagellation and cell cycle control in
612 Caulobacter. Curr Opin Microbiol **28:**83-92.
613 22. **Benson AK, Wu J, Newton A.** 1994. The role of FlbD in regulation of flagellar gene
614 transcription in Caulobacter crescentus. Res Microbiol **145:**420-430.
615 23. **Xu H, Dingwall A, Shapiro L.** 1989. Negative transcriptional regulation in the Caulobacter
616 flagellar hierarchy. Proc Natl Acad Sci U S A **86:**6656-6660.
617 24. **Faulds-Pain A, Birchall C, Aldridge C, Smith WD, Grimaldi G, Nakamura S, Miyata T,**
618 **Gray J, Li G, Tang JX, Namba K, Minamino T, Aldridge PD.** 2011. Flagellin redundancy
619 in Caulobacter crescentus and its implications for flagellar filament assembly. J Bacteriol
620 **193:**2695-2707.
621 25. **Ely B, Croft RH, Gerardot CJ.** 1984. Genetic mapping of genes required for motility in
622 Caulobacter crescentus. Genetics **108:**523-532.
623 26. **Johnson RC, Ely B.** 1979. Analysis of nonmotile mutants of the dimorphic bacterium
624 Caulobacter crescentus. J Bacteriol **137:**627-634.
625 27. **Toh E, Kurtz HD, Jr., Brun YV.** 2008. Characterization of the Caulobacter crescentus
626 holdfast polysaccharide biosynthesis pathway reveals significant redundancy in the initiating
627 glycosyltransferase and polymerase steps. J Bacteriol **190:**7219-7231.
628 28. **Kumar P, Kaushik A, Lloyd EP, Li SG, Mattoo R, Ammerman NC, Bell DT, Perryman**
629 **AL, Zandi TA, Ekins S, Ginell SL, Townsend CA, Freundlich JS, Lamichhane G.** 2017.
630 Non-classical transpeptidases yield insight into new antibacterials. Nat Chem Biol **13:**54-61.
631 29. **Bielnicki J, Devedjiev Y, Derewenda U, Dauter Z, Joachimiak A, Derewenda ZS.** 2006. B.
632 subtilis ykuD protein at 2.0 A resolution: insights into the structure and function of a novel,
633 ubiquitous family of bacterial enzymes. Proteins **62:**144-151.
634 30. **Mira A, Ochman H, Moran NA.** 2001. Deletional bias and the evolution of bacterial
635 genomes. Trends Genet **17:**589-596.
636 31. **Martinez-Garcia E, Nikel PI, Chavarria M, de Lorenzo V.** 2014. The metabolic cost of
637 flagellar motion in Pseudomonas putida KT2440. Environ Microbiol **16:**291-303.
638 32. **Smith DR, Chapman MR.** 2010. Economical evolution: microbes reduce the synthetic cost of
639 extracellular proteins. MBio **1:**10.1128/mBio.00131-00110.
640 33. **Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell**
641 **DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK.** 2009. Simultaneous assay of
642 every Salmonella Typhi gene using one million transposon mutants. Genome Res **19:**2308-
643 2316.
644 34. **Loeb GI, Neihof RA.** 1975. Marine Conditioning Films, p 319-335, Applied Chemistry at
645 Protein Interfaces, vol 145. American Chemical Society.
646 35. **Schneider RP, Leis A.** 2003. Conditioning Films in Aquatic Environments, Encyclopedia of
647 Environmental Microbiology doi:10.1002/0471263397.env036. John Wiley & Sons, Inc.
648 36. **Loeffler F.** 1890. Weitere Untersuchungen uber die Beizung und Farbung der Geisseln bei den
649 Bakterien. Centralbl Bakteriol Parasitenkd **7:**625-639.
650 37. **Pitcher DG, Saunders NA, Owen RJ.** 1989. Rapid extraction of bacterial genomic DNA with
651 guanidium thiocyanate. Lett Appl Microbiol **8:**151-156.
652 38. **Kent WJ.** 2002. BLAT--the BLAST-like alignment tool. Genome Res **12:**656-664.
653 39. **Curtis PD, Brun YV.** 2014. Identification of essential alphaproteobacterial genes reveals
654 operational variability in conserved developmental and cell cycle systems. Mol Microbiol
655 **93:**713-735.
656

657

658




659

660 **Figure 1. Growth of *Caulobacter* in laboratory medium, and supplemented or unsupplemented**
661 **water from two Great Lakes.** Overnight cultures washed with lake water (LW) were inoculated into
662 5 mL of filtered water from Lake Michigan (A) or Lake Superior (B). Growth assays in water
663 supplemented with carbon (0.1% w/v xylose) and/or phosphorus (23 μM $K_2HPO_4$) as indicated;
664 growth was monitored every 24 hours by enumeration of colony forming units (CFUs) by dilution
665 plating. Lake water growth is compared to growth in a laboratory peptone yeast extract (PYE)
666 medium. Data represent mean ± standard deviation of 5 replicates per condition. (C) Fine scale growth

667 of *Caulobacter* in PYE, M2-xylose defined medium (M2X), and filtered Lake Michigan water. Cells
668 were grown as in A and B and monitored by enumerating CFUs. Data represent mean ± standard
669 deviation of 5 replicates per condition. Boxes represent the approximate region of the growth curve
670 (cell density and incubation time) in which the barcoded Tn-Himar mutant library (BarSeq) pools were
671 cultivated.
672

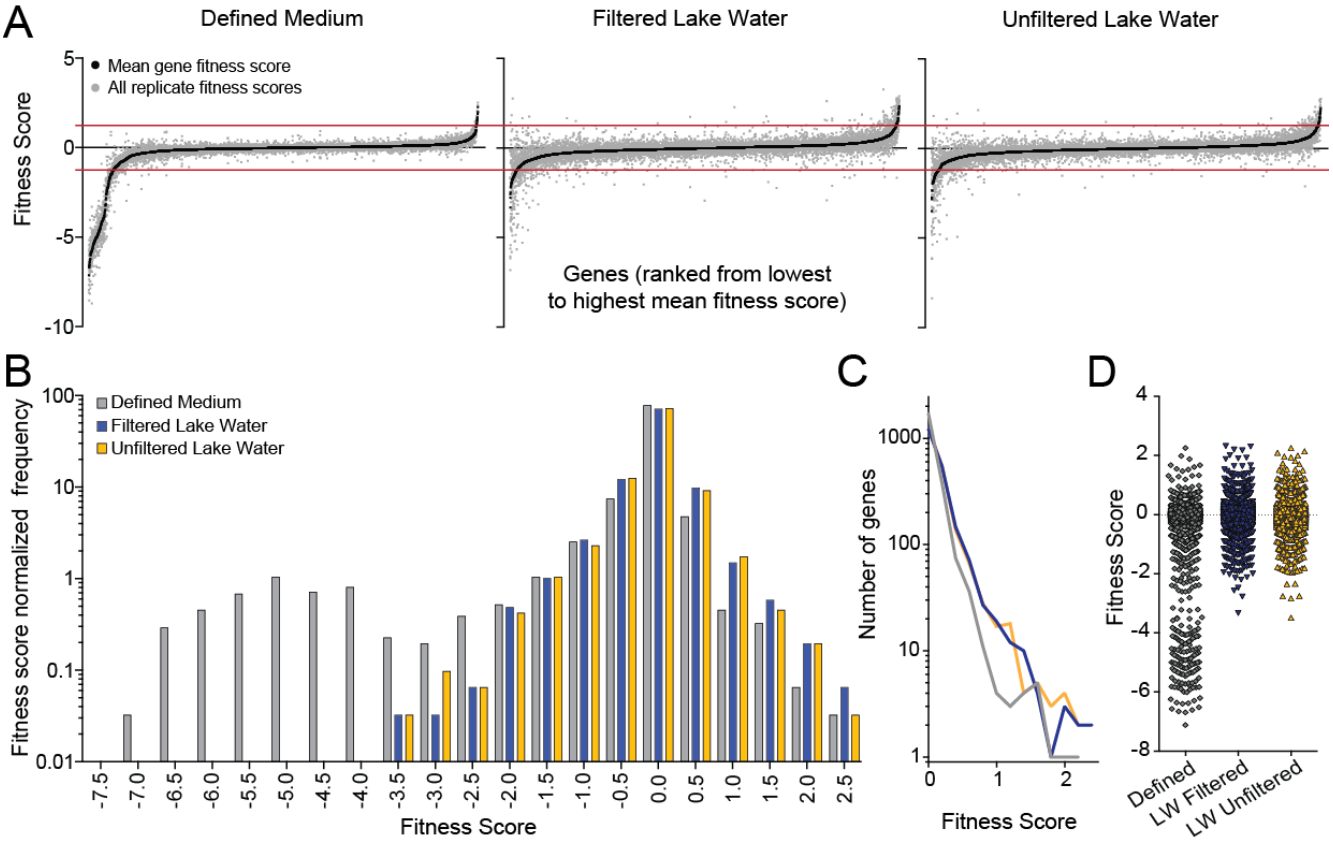



**Figure 2. *Caulobacter* gene fitness score summary after cultivation in defined medium, filtered, or unfiltered Lake Michigan water.** (A) Rank ordered mean fitness scores of each scorable *Caulobacter* gene across each of the four replicate experiments for each growth condition is plotted; black = mean fitness score; gray = independent replicate fitness scores. Red lines represent ± 3σ from the mean score of the entire dataset (which is approximately zero). (B) Distributions of mean gene fitness scores for each condition: defined M2X medium (gray), filtered Lake Michigan water (blue), and unfiltered Lake Michigan water (yellow). (C) Average gene fitness values between 0 and +2.5 plotted for each condition; defined M2X medium (gray), filtered Lake Michigan water (blue), and unfiltered Lake Michigan water (yellow). (D) Average gene fitness scores plotted for each of the three cultivation conditions: defined M2X medium (gray), filtered Lake Michigan water (blue), and unfiltered Lake Michigan water (yellow).

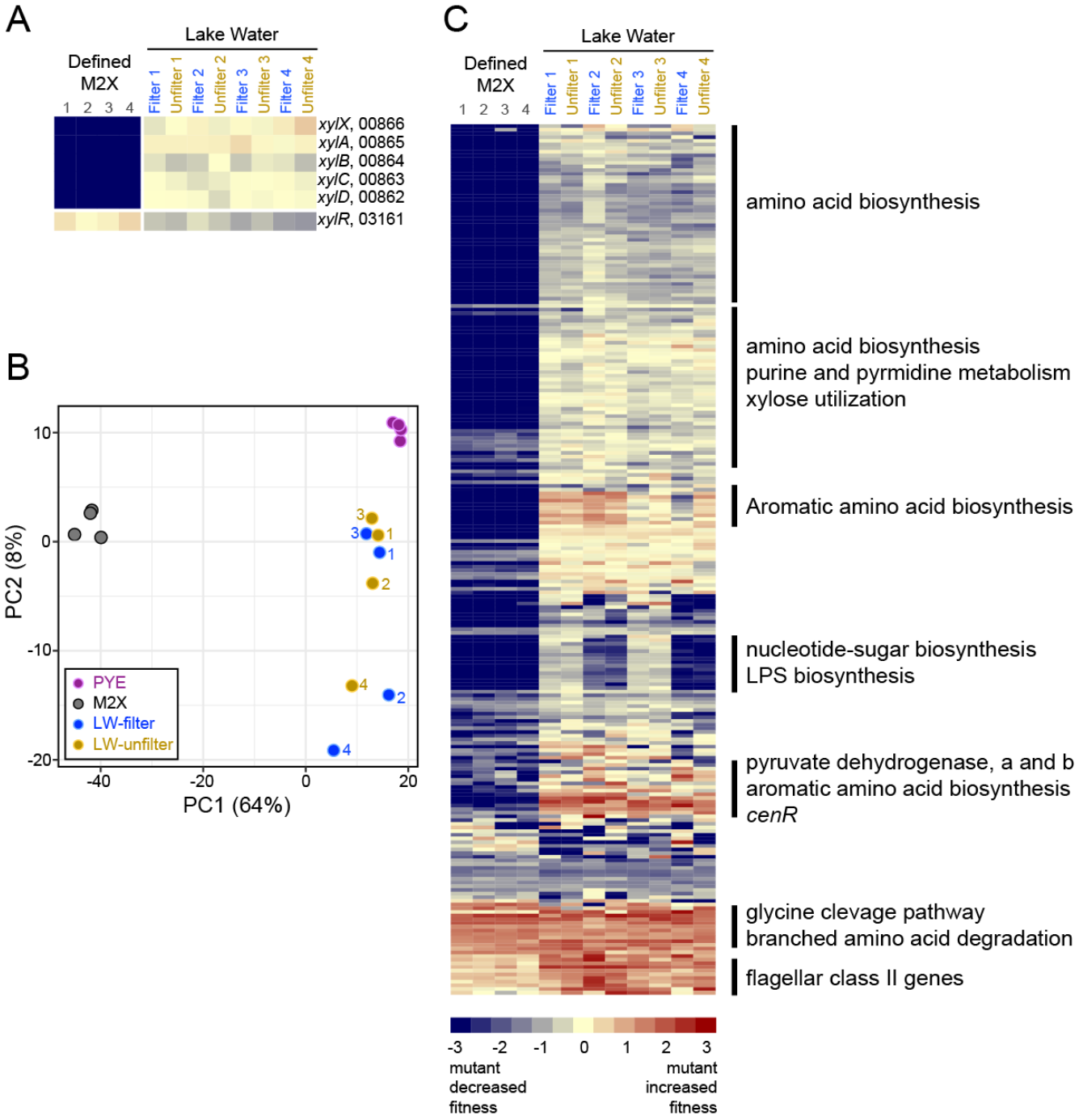



**Figure 3. Functional summary of mutant strains with diminished or enhanced fitness in minimal defined medium and Lake Michigan water.** (A) Heatmap of gene fitness scores for six xylose utilization genes (*xylX*, *xylA*, *xylB*, *xylC*, *xylD*, *xylR*) from each replicate experiment of cultivation of the Tn-Himar mutant library in M2X defined medium, filtered lake water, and unfiltered lake water. Scale bar is shown in C. (B) Principle component (PC) analysis (PCA) of genomic-scale fitness values for the barcoded *Caulobacter* Tn-Himar mutant library cultivated in complex peptone yeast extract (PYE) medium (reference set), defined M2X medium, filtered Lake Michigan water, and unfiltered Lake Michigan water. The plot shows PCA values for the four replicate datasets for each cultivation condition. Percent of variance contributed by the first two PCs is noted on the axes. (C) Heatmap of

701 gene fitness scores for genes with mean fitness scores higher than 1.2 or lower than -1.2 (± 3σ fitness
702 score cutoff) in at least one cultivation condition. Genes are clustered, and fitness scores for each
703 replicate experiment are presented as in B. Functions of genes within particular regions of this 238
704 gene cluster are noted; this entire cluster is split into three fully annotated clusters in Figures S3–S5.
705

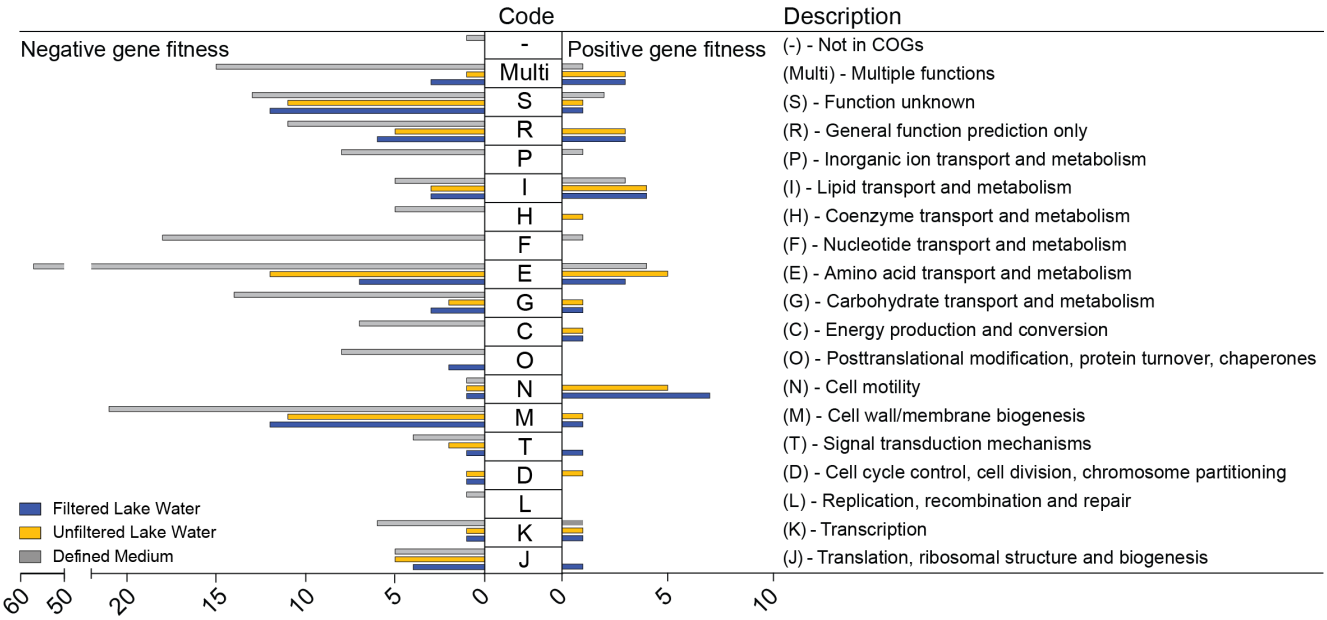
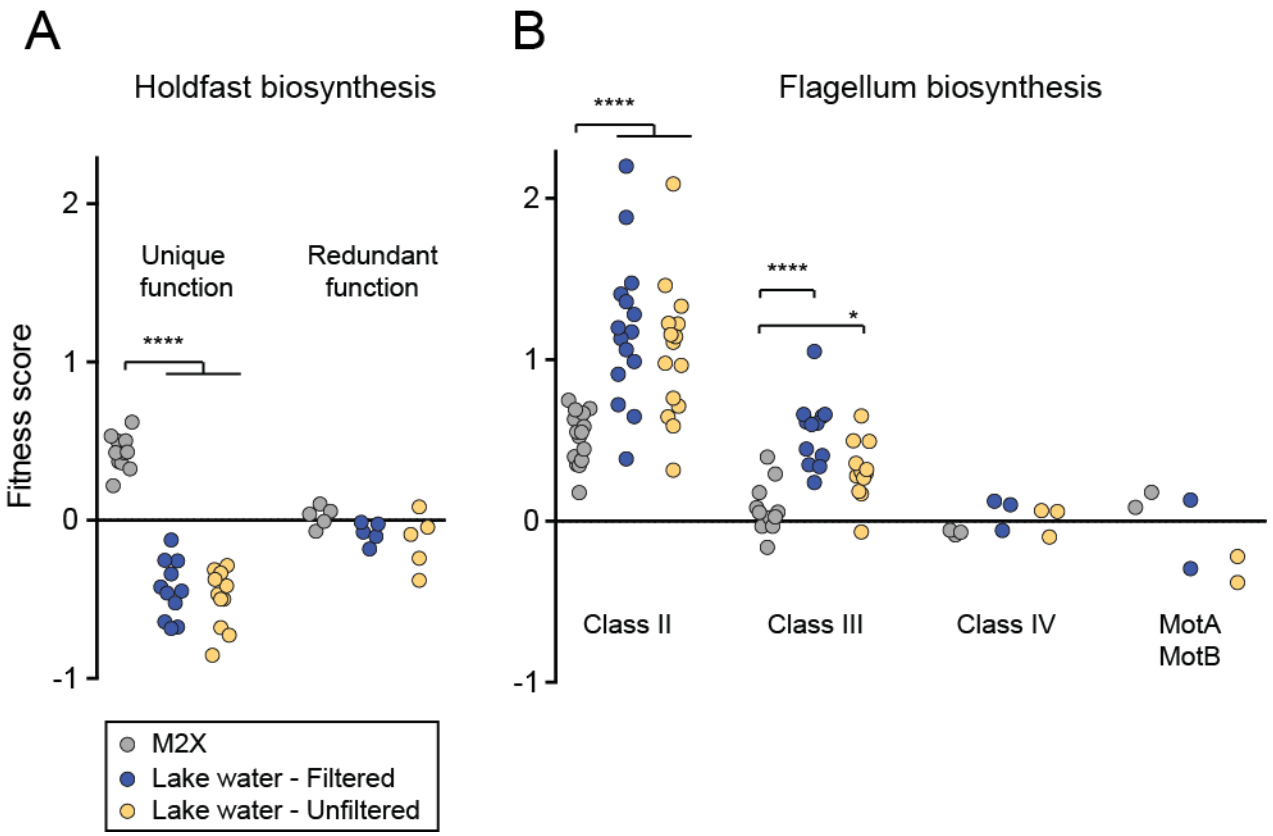



**Figure 4. COG-based functional annotation.** The analysis includes genes with fitness scores of absolute value greater than 1.2 (± 3σ fitness score cutoff). Each gene was assigned a cluster of orthologous group (COG) functional category (https://www.ncbi.nlm.nih.gov/COG/). The number of genes in each COG category is plotted; genes with negative fitness values (left) and genes with positive fitness values (right).

713




714

715 **Figure 5. Genes with functions in flagellum and holdfast biosynthesis influence fitness in Lake**
716 **Michigan water.** Mean fitness scores of genes involved in (A) holdfast biosynthesis (14 genes) and
717 (B) flagellum biosynthesis (36 genes) are plotted for each cultivation condition (four independent
718 growth replicates): defined M2X medium (gray); filtered lake water (blue); and unfiltered lake water
719 (yellow). Holdfast genes with unique function (i.e. single gene deletions have a holdfast defect) and
720 redundant function (i.e. single mutants have no holdfast defect) are shown separately. Class II, Class
721 III, and Class IV genes in the flagellar hierarchy and motA/motB stator complex genes are also shown
722 separately. Clusters of holdfast and flagellum fitness score data, with individually annotated genes, are
723 shown in Figure S2. One-way ANOVA was applied to assess differences in fitness scores between
724 cultivation groups within the same gene class; Tukey's post test (**** $p < 0.0001$; * $p < 0.05$).

725

726
727
728 # TABLES

729 **Table 1:** Transposon library statistics

| Library | Unique insertion sites | Total TA sites | Percent sites hit | Average transposons per ORF | Mean reads |
|---|---|---|---|---|---|
| Tn5 | 115,788 | N/A | 2.9 | 30 | 90–150 (per Tn) |
| Tn-Himar (BarSeq) | 46,395 | 49,437 | 93.8 | 24 | 7612 (per gene) |

730

731 **Table 2:** Analysis of Lake Michigan water used for barcoded Tn-Himar fitness experiments

| Water collection* | Nov. 30 | Dec. 6 | Dec. 9 | Dec. 12 |
|---|---|---|---|---|
| Date of experiment | Dec. 2 | Dec. 6 | Dec. 9 | Dec. 12 |
| Water temperature (°C) | 7.5 | 7 | 3 | 2 |
| Air temperature (°C) | 7.8 | 4.4 | -3.3 | -7.2 |
| pH | 5.8 | 5.8 | 5.8 | 5.8 |
| Phosphate | 5 ppm | 5 ppm | 5 ppm | 5 ppm |
| Nitrate/Nitrite** | undetectable | undetectable | undetectable | undetectable |

*Water collection occurred in 2016 at Promontory Point, Hyde Park, Chicago, Illinois, USA.
**Limit of detection: Nitrite 0.15 mg/L, Nitrate 1 mg/L

732

733

734 **Table 3:** Average fitness scores for hypothetical genes and genes of unknown function with fitness
735 scores ± 3σ from the mean (bold and shaded) in at least one condition.
736

| Locus number | | Defined | Filtered LW | Unfiltered LW |
|---|---|---|---|---|
| CCNA_01219 | No conserved domains | **1.68** | -0.70 | 0.78 |
| CCNA_03860 | COG3786, L,D-peptidoglycan transpeptidase, YkuD superfamily | **1.33** | **1.53** | **1.41** |
| CCNA_03909 | No conserved domains | **-1.20** | **-1.76** | **-1.84** |
| CCNA_00375 | No conserved domains | **-1.21** | **-2.57** | **-2.83** |
| CCNA_03864 | DUF3576, pfam12100 | **-1.71** | **-1.32** | **-1.57** |
| CCNA_01724 | COG4649, TPR_21 pfam09976 | **-2.58** | **-1.25** | **-1.28** |
| CCNA_02875 | No conserved domains | **-1.54** | **-2.19** | -0.39 |
| CCNA_00927 | No conserved domains | **-2.56** | **-2.11** | -0.67 |
| CCNA_00913 | No conserved domains | -0.28 | **-2.48** | **-1.21** |
| CCNA_00895 | No conserved domains | -0.19 | **-2.14** | **-1.44** |
| CCNA_03273 | COG4944, DUF1109, pfam06532 NrsF, anti-sigF | **-1.63** | -0.39 | -0.27 |
| CCNA_00519 | No conserved domains | **-1.66** | -0.37 | -0.61 |
| CCNA_03692 | COG1975, Xanthine and CO dehydrogenase maturation factor, XdhC/CoxF family | **-4.06** | -0.98 | -0.65 |
| CCNA_01178 | DUF934, pfam06073, COG3749 | **-4.36** | -0.60 | -0.64 |
| CCNA_02669 | Uncharacterized membrane protein, DUF3422, pfam11902, COG4949 | **-4.41** | 0.38 | 0.20 |
| CCNA_01176 | DUF2849, pfam11011 | **-4.98** | -0.81 | -0.84 |
| CCNA_01676 | TamB, COG2911, pfam04357 | **-6.40** | -0.52 | -0.74 |
| CCNA_01286 | No conserved domains | 0.25 | -0.25 | **-1.27** |
| CCNA_02053 | No conserved domains | -0.56 | -0.74 | **-1.34** |
| CCNA_02160 | No conserved domains | 0.26 | -1.02 | **-1.38** |
| CCNA_03015 | No conserved domains | -0.15 | 0.57 | **-1.46** |
| CCNA_03945 | No conserved domains | -0.43 | 0.24 | **-1.90** |
| CCNA_02796 | No conserved domains | -0.45 | **-1.20** | -0.52 |
| CCNA_03984 | No conserved domains | 0.01 | **-1.22** | -0.25 |
| CCNA_03883 | No conserved domains | 0.05 | **-1.25** | 0.99 |
| CCNA_03420 | No conserved domains | 0.40 | **-1.46** | 0.42 |

737
738 LW, Lake Michigan water.

739 **Supplemental Material**
740
741
742 ***Tn5 and Tn-Himar: comparing gene essentiality and the effects of gene disruption on fitness across***
743 ***studies***

744 A previous analysis of a highly saturated *Caulobacter* Tn5 transposon library revealed a set of
745 genes that are required for growth in complex PYE medium (1); approximately 14% of genes in the
746 genome were deemed essential. The total genome insertion coverage was lower in the Himar library
747 described here than in the Tn5 dataset of Christen et al (2011), as Tn-Himar inserts specifically into
748 TA dinucleotide sites (with 67% GC content, TA sites are relatively limited in the *Caulobacter*
749 genome). Genes for which we failed to detect Tn-Himar insertions (**Table S12**) were largely consistent
750 with essential genes reported by Christen et al (2011), with exceptions likely due to differential
751 coverage of Tn5 versus Tn-Himar mutagenesis and differences in metrics used to define essentiality. A
752 comparison of the essential genes defined by Christen et al (2011) and by our Tn5-seq and Tn-Himar
753 fitness studies is presented in **Table S5**.

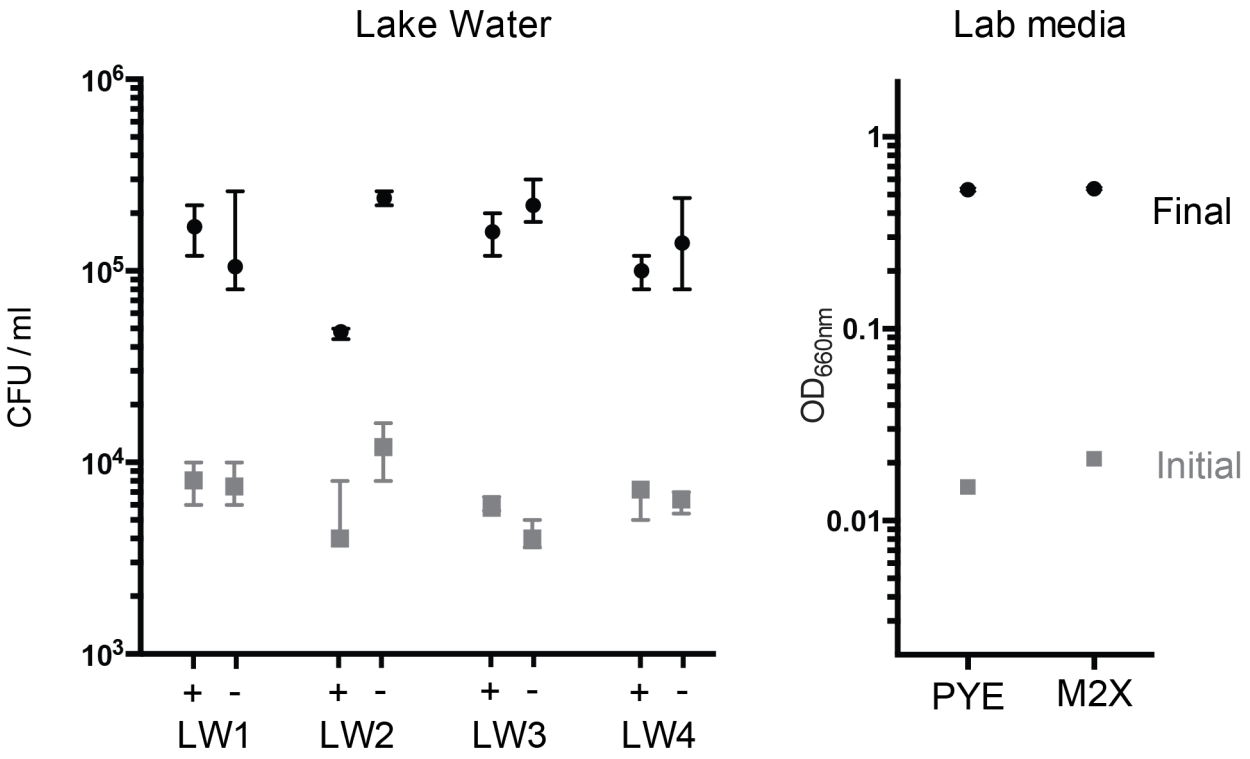
754 We have uncovered evidence for gene disruptions that both enhanced or reduced strain fitness
755 in lake water and M2X relative to PYE. Such results are consistent for a number of genes across both
756 the Tn5 and Tn-Himar datasets. Disruption of genes encoding three metabolic enzymes, a class C β-
757 lactamase family protein (*CCNA_00255*), transaldolase (*CCNA_03729*), and methylcrotonyl-CoA
758 carboxylase (*CCNA_02250*), enhanced *Caulobacter* fitness in Lake Michigan water relative to PYE
759 using both Tn5 and Tn-Himar approaches (**Table S6**). The functional role of the β-lactamase family
760 protein is unclear; the transaldolase and methylcrotonyl-CoA carboxylase may affect the
761 anabolic/catabolic balance by modulating the pentose phosphate pathway and branched chain amino
762 acid catabolism, respectively. Indeed, transcription of methylcrotonyl-CoA carboxylase is significantly
763 upregulated in PYE relative to M2X in the published dataset of Hottes et al (2004), providing evidence
764 that this activity is more important in complex PYE medium than in defined medium.

765 The Tn5 and Tn-Himar approaches identified a shared set of 26 genes for which disruption
766 resulted in decreased fitness in lake water relative to PYE. The largest effects were observed for
767 multiple genes involved in leucine, proline, homoserine, asparate, and glutamate biosynthesis. Again,
768 these data are consistent with transcriptomic data in which amino acid biosynthesis genes are
769 upregulated in M2X relative to PYE (2). Of note, disruption of the anthranilate synthase
770 (*CCNA_01972*) gene, required for tryptophan biosynthesis, resulted in a significant reduction in fitness
771 of *Caulobacter* cultivated in the presence of other microorganisms in unfiltered lake water, whereas
772 strains grown in filtered water did not have a growth defect. A full comparison of fitness data from the
773 Tn-Himar BarSeq and Tn5 approaches is presented in **Tables S6–S8**.

774 **References**
775

776 1. **Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ,**
777 **McAdams HH, Shapiro L.** 2011. The essential genome of a bacterium. Mol Syst Biol **7:**528.
778 2. **Hottes AK, Meewan M, Yang D, Arana N, Romero P, McAdams HH, Stephens C.** 2004.
779 Transcriptional profiling of Caulobacter crescentus during growth on complex and minimal
780 media. J Bacteriol **186:**1448-1461.
781
782

783
784
785
786 **SUPPLEMENTAL FIGURES AND LEGENDS**
787
788




789
790 **Figure S1:** Numbers of barcoded Tn-Himar mutants increase 20–30-fold (on average) during
791 cultivation in Lake Michigan water and laboratory media. Initial (gray) and final (black) cell densities
792 for each Tn-Himar pool growth experiment. In lake water growth experiments, approximately $10^7$
793 mutant cells were inoculated into each of the three flasks of 2.5 L of filtered (+) or unfiltered (-) water
794 from Lake Michigan collected on 4 different days. Initial and final CFUs were enumerated for each
795 flask. In complex (PYE) and defined (M2X) laboratory media, approximately $10^7$ mutant cells were
796 inoculated into 1.5 mL of medium. Initial and final $OD_{660}$ of four replicates were measured. Mean ±
797 range is presented.
798

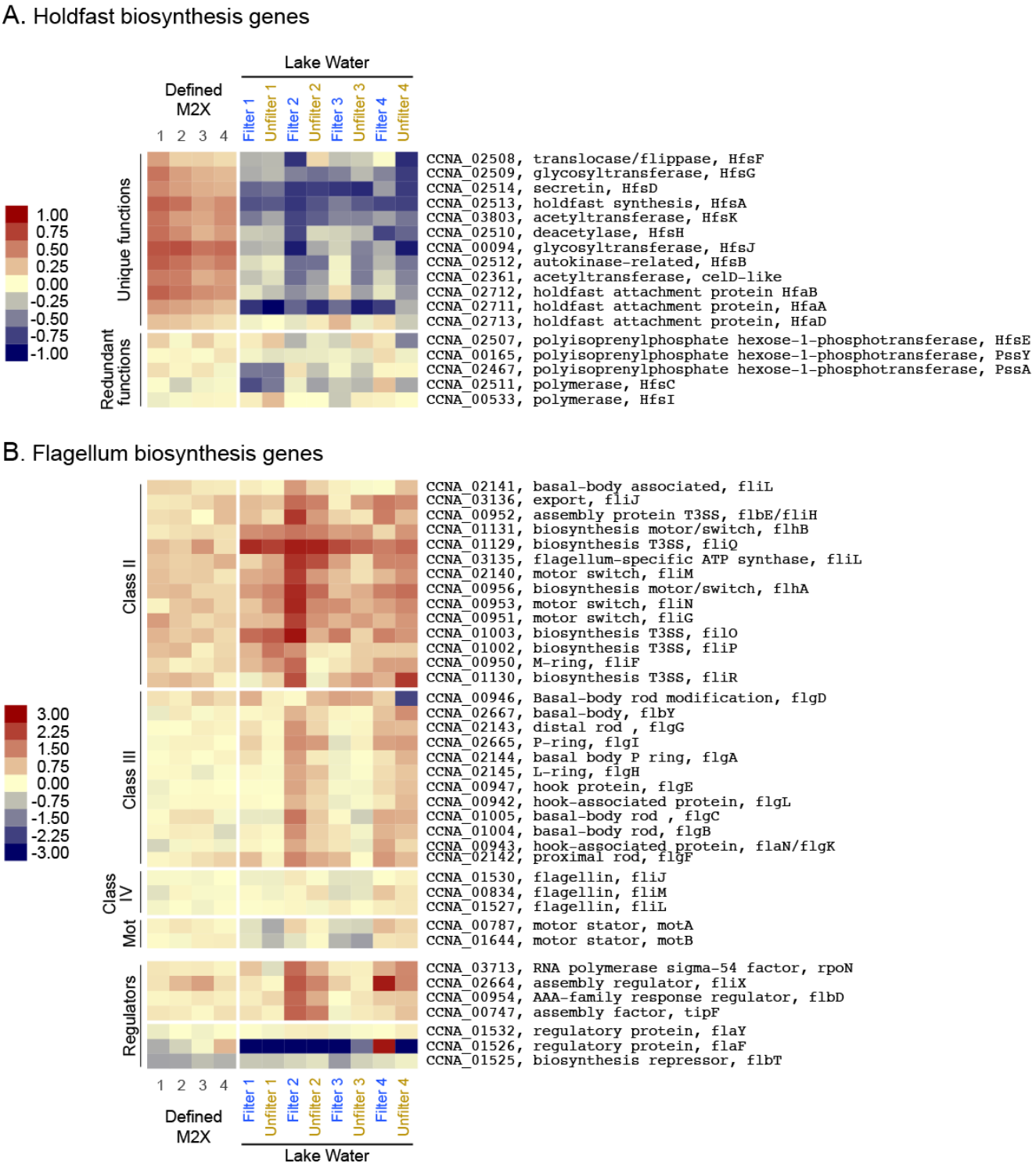



**Figure S2**. Heatmap of fitness scores for genes with established functions in (A) holdfast biosynthesis and (B) flagellum biosynthesis and motility calculated after cultivation in minimal defined M2X medium and in Lake Michigan water filtered or unfiltered. Holdfast genes with unique functions (i.e. single gene deletions have a holdfast development defect) and redundant functions (i.e. single mutants have no holdfast defect), as previously defined (27), are shown separately. Class two (II), class three

808 (III) and class four (IV) flagellar genes are marked on the cluster in panel B. Stator mot genes and
809 flagellar regulators are also marked. Scale bars are shown for (A) and (B).
810

**Figure S3.** Heatmap of fitness scores for genes shown in Fig. 3C, with gene names visible.

**Figure S4.** Heatmap of fitness scores (continued) for genes shown in Fig 3C, with gene names visible.

**Figure S5.** Heatmap (continued) of fitness scores for genes shown in Figure 3C, with gene names visible.

818
819
820 SUPPLEMENTAL TABLES
821 Table S1. Strains and primers used in this study
822 Table S2. Tn-Himar gene fitness values
823 Table S3. Tn5 library gene fitness values
824 Table S4. Tn5 library gene fitness, adjusted p-value <0.01
825 Table S5. Data sets for Christen et al 2011, our Tn5 library, and Tn-Himar library arranged by gene
826 Table S6. Shared genes between Tn5 and Tn-Himar fitness experiments
827 Table S7. Tn-Himar library genes that are ±3 standard deviations from the mean
828 Table S8. Tn-Himar library list of genes shared between minimal medium and lake conditions that are
829 ±3 standard deviations from the mean
830 Table S9. Values of t-test from BarSeq analysis for all genes that met the threshold for fitness
831 calculation outlined
832 Table S10. Genes for which Tn-Himar disruption results in a specific advantage (positive fitness or t
833 value) or disadvantage (negative fitness or t value) in M2-xylose defined medium
834 Table S11. Genes for which Tn-Himar disruption results in a specific advantage (positive fitness or t
835 value) or disadvantage (negative fitness or t value) in filtered or unfiltered Lake Michigan water, but
836 not in M2-xylose defined medium (relative to complex PYE medium)
837 Table S12. Genes in *Caulobacter* that were not hit by Tn-Himar
838
839