JB Accepted Manuscript Posted Online 12 March 2018
J. Bacteriol. doi:10.1128/JB.00724-17

Promoter boundaries for the *luxCDABE* and *betIBA-proXWV* operons in *Vibrio harveyi* defined by the method RAIL: Rapid Arbitrary PCR Insertion Libraries

Christine M. Hustmyer[1], Chelsea A. Simpson[1], Stephen G. Olney[1], Matthew L. Bochman[2], and Julia C. van Kessel[1*]

1. Department of Biology, Indiana University, Bloomington, IN 47405
2. Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN 47405

* Corresponding author:
Email: jcvk@indiana.edu
Telephone: 812-856-2235
Fax: 812-856-5710

Running title: RAIL method for defining promoter boundaries

Keywords: reporter fusion, isothermal DNA assembly, arbitrary PCR, *Vibrio harveyi*, quorum sensing, bioluminescence

Abstract

Experimental studies of transcriptional regulation in bacteria require the ability to precisely measure changes in gene expression, often accomplished through the use of reporter genes. However, the boundaries of promoter sequences required for transcription are often unknown, thus complicating construction of reporters and genetic analysis of transcriptional regulation. Here, we analyze reporter libraries to define the promoter boundaries of the *luxCDABE* bioluminescence operon and the *betIBA-proXWV* osmotic stress operon in *Vibrio harveyi*. We describe a new method called RAIL (Rapid Arbitrary PCR Insertion Libraries) that combines the power of arbitrary PCR and isothermal DNA assembly to rapidly clone promoter fragments of various lengths upstream of reporter genes to generate large libraries. To demonstrate the versatility and efficiency of RAIL, we analyzed the promoters driving expression of the *luxCDABE* and *betIBA-proXWV* operons and created libraries of DNA fragments from these loci fused to fluorescent reporters. Using flow cytometry sorting and deep sequencing, we identified the DNA regions necessary and sufficient for maximum gene expression for each promoter. These analyses uncovered previously unknown regulatory sequences and validated known transcription factor binding sites. We applied this high-throughput method to *gfp*, *mCherry*, and *lacZ* reporters and multiple promoters in *V. harveyi*. We anticipate that the RAIL method will be easily applicable to other model systems for genetic, molecular, and cell biological applications.

Importance

Gene reporter constructs have long been essential tools for studying gene regulation in bacteria, particularly following the recent advent of fluorescent gene reporters. We developed a new method that enables efficient construction of promoter fusions to reporter genes to study gene regulation. We demonstrate the versatility of this technique in the model bacterium *Vibrio harveyi* by constructing promoter libraries for three bacterial promoters using three reporter

51 genes. These libraries can be used to determine the DNA sequences required for gene
52 expression, revealing regulatory elements in promoters. This method is applicable to various
53 model systems and reporter genes for assaying gene expression.

54 **Introduction**

55 Central to the study of bacterial physiology and development is the ability to monitor and
56 quantify gene expression. Monitoring gene expression is greatly aided through the use of gene
57 reporter fusions. Transcriptional and translational fusion constructs facilitate single-cell and
58 population-wide gene expression investigations to study the influence of regulatory factors,
59 perform genetic screens, and visualize protein localization patterns. Typically, such reporters
60 are cloned downstream of regulatory promoters or genes of interest and introduced into a model
61 bacterial system, either on replicating plasmids or integrated into the genome. Numerous
62 reporter genes have traditionally been used to assay gene expression, such as *lux* (bacterial
63 luciferase), *lacZ* (β-galactosidase), *phoA* (alkaline phosphatase), *bla* (β-lactamase), and *cat*
64 (chloramphenicol acetyltransferase) (1, 2). However, the advent of more modern techniques has
65 allowed for the use of fluorescent proteins such as green fluorescent protein (GFP) for these
66 studies without the need for substrates or specialized media (1-4).

67 To adequately and efficiently study the expression pattern of a particular gene, the
68 defined regulatory region controlling promoter activity must be known. The region upstream of
69 the promoter driving *luxCDABE* transcription in *Vibrio harveyi* is an example of a locus with a
70 large and undefined regulatory region, which has limited studies of gene regulation. This
71 particular promoter is of interest because it drives expression of the bioluminescence genes with
72 >100-fold increase in transcription and >1000-fold increase in bioluminescence production
73 under activating conditions (*i.e.*, quorum sensing) (5-10). It was previously suggested that the
74 *lux* promoter requires ~400 bp upstream of the translation start site and ~60 bp downstream of
75 the start codon for full activation of the *cat* reporter gene (9). The requirement for a large
76 promoter region is due in part to the presence of seven binding sites for the transcription factor
77 LuxR upstream of the primary transcription start site, each of which is necessary for maximal
78 activation of the promoter (5, 6, 9). The ~400-bp region of $P_{luxCDABE}$ is relatively large compared
79 to some bacterial regulatory promoters (*e.g.*, the *lac* promoter), but comparable in size to other

80 promoters with evidence of cooperative binding between transcription factors and DNA looping
81 (*e.g.*, the *araBAD* promoter) (5, 11-13). Indeed, full activation of the *luxCDABE* promoter
82 requires the transcription factor LuxR and nucleoid-associated protein Integration Host Factor
83 (IHF), and DNA looping by IHF is proposed to drive interactions between LuxR and RNA
84 polymerase for transcription activation (6).

85 Another *V. harveyi* operon that has an unknown promoter region is the *betIBA-proXWV*
86 operon in *V. harveyi*. The *betIBA-proXWV* osmoregulation genes encode proteins required for
87 the synthesis and transport of the osmoprotectant glycine betaine (14). These genes are auto-
88 regulated by the BetI repressor and activated 3- to 10-fold by LuxR (14). There are two sites in
89 the *betIBA-proXWV* promoter that have been shown to be bound by LuxR *in vitro* and *in vivo*,
90 though the role of these sites in transcriptional regulation has not yet been tested (5, 14). For
91 both the *luxCDABE* and *betIBA-proXWV* operons, the boundaries of the promoters are not
92 defined, and thus, mechanistic studies of transcriptional regulation of these operons is limited.

93 Here, we describe a new method for rapidly generating reporter plasmids that we used
94 to define promoter regions. The RAIL method (Rapid Arbitrary PCR Insertion Libraries) exploits
95 the power of arbitrary PCR and isothermal DNA assembly (IDA) to insert semi-randomized
96 fragments of promoter DNA into reporter plasmids (15-17). Using RAIL, we generated libraries
97 containing fragments of various lengths of the region upstream of the *luxCDABE* operon
98 transcriptionally fused to *gfp*. We used flow cytometry sorting to screen the library of promoter
99 fragments for reporter expression and next-generation sequencing to map the 3' boundary of
100 the *luxCDABE* promoter required for full activation. We also applied this method to two
101 additional promoter regions in *V. harveyi* (*betIBA-proXWV* and *VIBHAR_06912*), and we
102 demonstrated the versatility of the system by using two additional reporters, mCherry and β-
103 galactosidase. This approach enabled us to identify the required regions for gene expression for
104 multiple promoters and simultaneously produce usable gene reporter constructs. Our method
105 should be widely applicable to any system for which gene reporters have been established and

represents a simple and efficient technique to construct reporter fusions for molecular, genetic, and cell biology studies.

Results

Measuring transcription activation from the V. harveyi luxCDABE *promoter using fluorescent reporter fusions*

To study the mechanism of LuxR regulation of the *luxCDABE* promoter, we constructed four reporter plasmids containing various fragments of the *luxCDABE* locus transcriptionally fused to *gfp* using traditional cloning methods (Fig. 1A). Each plasmid contains the same 5' end (~400 bp upstream of the *luxC* ORF), and the 3' ends vary as follows: 1) 2 bp after the transcription start site at -26 (pJV369), 2) at the LuxC translation start site (pJV367), 3) 36 bp into the *luxC* ORF (pSO04), and 4) 407 bp into the *luxC* ORF (pJV365) (Fig. 1A). All plasmid constructs contained the seven LuxR binding sites previously found to be essential for transcriptional activation. Only plasmid pJV365 contained LuxR site H, which has previously been shown to be non-essential for activation. We chose the lengths of these fragments to investigate the requirement for the 5'-UTR, site H, and various lengths of the *luxC* ORF. We first tested LuxR activation of these reporter plasmids in *Escherichia coli* because expression of *luxR* in *E. coli* is sufficient to drive high levels of transcription of the *luxCDABE* operon (5, 10, 18), and the use of *E. coli* is more efficient for transformation. Transcription activation of the *luxCDABE* promoter was assayed in *E. coli* strains containing a second plasmid either constitutively expressing *luxR* (pKM699) or an empty vector (pLAFR2). The plasmid containing the 3' boundary 36 bp into the *luxC* ORF was highly expressed (Fig. 1B), which is consistent with a previous study using a nearly identical promoter fragment (Fig. S1A) (5). The strain containing pJV369 with the DNA fragment up to and including the transcription start site also displayed high levels of GFP. Activation was appreciably decreased (~7-fold) for the pJV367

132 strain containing the 5' untranslated region (5'-UTR) but ending at the LuxC translation start site
133 compared to the pSO04 strain (Fig. 1B). Also, the strain containing pJV365 with 407 bp of the
134 *luxC* ORF was not activated above 2-fold (Fig. 1B). A similar trend was obtained when these
135 constructs were conjugated into *V. harveyi* strains and the GFP expression was compared
136 between wild-type and Δ*luxR* strains (Fig. S1B). From these data, we conclude that a promoter
137 fragment ending 2 bp past the transcription start site is sufficient for activation. Further, we
138 revealed that varying lengths of 3' constructs fused to *gfp* produce unexpected changes in gene
139 expression across the *luxCDABE* promoter.
140
141 *The RAIL method*
142 Our observation that varying 3' ends of the *luxCDABE* promoter greatly affected gene
143 expression led us to expand our analysis of the expression profile of promoter fusions across
144 the entire locus. Therefore, we needed to construct numerous promoter fragments
145 transcriptionally fused to a fluorescent reporter. Instead of constructing each of these plasmids
146 individually, we designed a cloning technique combining the power of arbitrary PCR and IDA
147 (a.k.a., Gibson assembly) (15-17). This method enabled us to simultaneously amplify fragments
148 of varying lengths and clone them into a vector backbone to create a library in four simple steps
149 (Fig. 2). First, arbitrary primers were used in a preliminary round of PCR in conjunction with a
150 primer that specifically anneals to the promoter (Fig. 2, primer 1F). Four arbitrary primers were
151 synthesized with eight sequential random nucleotides anchored at the 3' end with two specific
152 nucleotides: AA, TT, AT, or TA (Table S1). We chose to use A-T pairs to anchor the primer due
153 to the low G+C content of *V. harveyi.* Each of these four primers also contains a linker at the 5'
154 end (Fig. 2, primer 1R). The first round of PCR produced a range of products that varied in
155 length from 100 to >3000 bp and that appeared as faint smears of products as expected for
156 random priming (Fig. 2). For some loci, no smear could be visualized by gel electrophoresis

after the first round of PCR, but this did not impact the success of the second round of amplification.

In the second step, the products from round 1 were further amplified using a nested primer (primer 2F), and a linker was added with homology to the plasmid backbone (Fig. 2). Primer 2R anneals to the linker on primer 1R. The second round of PCR using these primers was performed with the products from round 1 as templates. This second step served to increase the amount of DNA product and to add a linker to the 5' end. Each reaction in round 2 produced a smear of products that contained homology to the plasmid backbone at their 5' and 3' ends (Fig. 2). The smear of products can also be gel extracted to the desired size. In the third step, the plasmid backbone was PCR-amplified or digested by specific restriction enzymes to form a linear product (Fig. 2). In the fourth and final step, IDA was performed to clone the promoter fragments into the plasmid backbone, and the mixture was transformed into *E. coli* to obtain isolated clones (Fig. 2).

Defining the 3' boundary of the luxCDABE *operon using RAIL*

We used the RAIL method to generate a large library of plasmids with promoter fragments fused to *gfp*. This library had fixed 5'-ends and varying 3'-ends generated by combining PCR products from four arbitrary primers, as shown in Figure 2, and inserts ranging from ~50 to >1,000 bp. We screened for *gfp* activation using fluorescence-activated cell sorting (FACS). The libraries were sorted by FACS into four groups: no GFP expression, low GFP expression, medium GFP expression, and high GFP expression (Fig. 3A). The 'no GFP' pool contained cells expressing similar or lower fluorescence than the negative control strain. The 'high GFP' pool contained cells expressing similar or higher fluorescence than the positive control strain. The 'low GFP' and 'medium GFP' pools were arbitrarily chosen to collect cells in the intermediate region between 'no GFP' and 'high GFP' without any overlap between the four bins (Fig. 3A). Illumina sequencing of the plasmid DNA from these pools enabled us to visualize

183 the 3' terminal end of the region cloned into the plasmid by graphing the location of the
184 sequencing coverage (42 bp) and the 3' terminal nucleotides (Fig. 3B, Fig. S2A). From these
185 graphs, we pinpointed the boundary in the *luxCDABE* promoter required for maximum
186 expression and showed the expression profile for promoter fragments across the entire locus
187 (Fig. 3B, Fig. S2A). The plasmids containing promoter fragments that terminated at nucleotide
188 +129 (relative to +1, the start of the *luxC* ORF) were highly enriched in the 'high expression'
189 pool, and plasmids in the 'no expression' pool were specifically de-enriched in this same
190 location (Fig. 3B, Fig. S2A). The 'high expression' pool had a clear 3' boundary at +129, which
191 is 16 bp upstream of LuxR site H. Thus, a DNA fragment that terminates at +129 includes LuxR
192 sites A, B, C, D, E, F, and G (6). The observation that LuxR site H was not included in this
193 region of 'high expression' is consistent with previous findings that site H is non-essential for
194 transcription activation at high cell density in *V. harveyi* (6). We conclude from these data that
195 fragments with 3' ends longer than +129 were decreased in reporter gene expression. There is
196 also a clear edge where sequencing coverage drops off for the 'high expression' pool at -55
197 (Fig. 3B). However, the exact minimum boundary cannot be determined because we did not use
198 every combination of anchor nucleotides in the arbitrary primers. Also, within the 'high
199 expression' pool, we noted a peak of sequencing coverage that started at +36, suggesting that
200 this is the minimum length promoter sufficient for high GFP expression in this library (Fig. S2A).
201 Thus, promoter fragments with ends ranging from +36 through +129 yield maximum expression
202 levels without being detrimental. The 'medium expression' pool contained sequences that
203 terminated at +199, which is located 32 bp beyond LuxR site H (Fig. 3B). Plasmids with
204 promoter fragments that extended throughout the *luxC* ORF past +199 had low levels of
205 expression, whereas plasmids without GFP expression were limited to promoter regions
206 upstream of -55 (Fig. 3B). We conclude that long promoter fragments decrease GFP expression
207 and are not suitable reporter plasmids. We also conclude that plasmids containing promoter
208 fragments shorter than -55 are not sufficient to activate transcription. Collectively, by analyzing

209 the sequencing data of RAIL libraries, we located the DNA region that is sufficient for maximal
210 transcription activation (-393 to +36), validated previous findings that LuxR sites A through G
211 are required for activation of *luxCDABE* (6), and defined the expression profile for the
212 *luxCDABE* locus in the context of a transcriptional fusion to *gfp*.
213
214 *Defining the 3' boundary of the* betIBA-proXWV *operon using RAIL*
215 We next used the RAIL strategy to construct reporter clones for the *betIBA-proXWV*
216 operon using a different fluorescent reporter, *mCherry*. We screened the promoter clones
217 individually before using the high-throughput flow cytometry method to analyze the library to test
218 whether we could identify useful promoter clones via a small-scale screen. Approximately 40
219 plasmids were screened by restriction digest for inserts of varying sizes, and the inserts were
220 sequenced to determine the size of the inserted region. We observed that plasmids containing
221 regions shorter than the predicted transcription start sites did not show any activation compared
222 to the empty vector control strain (Fig. 4B, pCH28 as an example). However, plasmids with
223 larger regions that extended into the *betI* ORF were activated by LuxR, such as pCH50 and
224 pCH72 (Fig. 4B). Plasmids containing the entire *betI* gene did not display activation (Fig. 4B,
225 pCH75). These data show that the RAIL method can be used for small-scale screens for
226 promoter clones by individually assaying plasmids.
227 We next synthesized a large library of *betIBA-proXWV* promoter fusions to *mCherry*
228 using RAIL. It is important to note that only one arbitrary primer was used to generate this
229 library, which limited the range of PCR products across the locus. This library of clones was
230 sorted by FACS for those that maximally expressed *mCherry* (Fig. 4C). The dynamic range of
231 the *betIBA-proXWV* promoter is substantially smaller than that of *luxCDABE*, resulting in
232 approximately 3-fold difference in expression in the averages of the positive and negative
233 controls. Thus, we chose to sort cells with fluorescence levels above the negative control strain
234 (Fig. 4C). In doing so, we lost cells that exhibited intermediate fluorescence levels but could

235 therefore be assured that all cells we collected were expressing high levels of fluorescence. The
236 Illumina sequencing coverage and 3' terminal nucleotides of the DNA in the two pools was
237 graphed (Fig. 4D, Fig. S2B). Sequencing analyses revealed the minimum 3' boundary for the
238 *betIBA-proXWV* promoter to be at -13 (Fig. 4D, Fig. S2B; relative to +1, the start of the *betI*
239 ORF), suggesting that the -46 transcription start site is the primary site for this locus. The 'high
240 expression' pool contained plasmids with DNA fragments up through the first portion of the *betI*
241 ORF at +25, which then tapered off (Fig. 4D, Fig. S2B). Plasmids with fragments that extended
242 more than half-way through the *betI* gene displayed low or no expression. Collectively, these
243 data showed that similarly to the *luxCDABE* locus, transcription reporters were functional if they
244 contained DNA fragments past the 3' boundary near the transcription start site. However, longer
245 fragments extending into the ORF decreased reporter gene expression.

246

247 *Versatility of the RAIL method for cloning with other promoters and reporter genes*

248 We also successfully used the RAIL technique to generate a promoter library using the
249 *lacZ* reporter for another *V. harveyi* gene, *VIBHAR_06912*, which encodes a transcription factor.
250 *VIBHAR_06912* expression is repressed by LuxR (19), and this is likely indirect repression
251 because there are no detectable LuxR binding sites in this region (5). Using RAIL, multiple
252 clones with varying promoter lengths were generated as transcriptional fusions to *lacZ*, and
253 strains were assayed for β-galactosidase activity in *V. harveyi* (Fig. 5A). All of the plasmids with
254 long promoter lengths were repressed by LuxR in the wild-type strain compared to the Δ*luxR*
255 strain (Fig. 5B). Conversely, a plasmid with a short fragment (pJV342) showed the same level of
256 β-galactosidase activity in the wild-type strain as in the Δ*luxR* strain (Fig. 5B). We note that each
257 construct promoted transcription to different levels, even in the absence of LuxR repression.
258 This suggests that other regulatory elements in addition to LuxR affect transcription at this

259 locus. Thus, we conclude that we again generated functional promoter fusion plasmids for this
260 promoter for future studies of gene expression and regulation of *VIBHAR_06912*.
261
262 *Reporter gene affects measurement of gene expression*
263 We noted that for each of the three promoters we studied, plasmid constructs that
264 contained promoter regions that extended into the ORF of the first gene had variable levels of
265 expression. For example, the pJV365 plasmid that included 407 bp of the *luxC* gene only
266 expressed GFP ~2-fold more in the presence of LuxR than in its absence (Fig. 1B). This is in
267 contrast to plasmid pMGM115 from the Miyamoto *et al.* study that contains the full *luxC* ORF
268 and displays maximal activation of the *cat* gene (~50-fold more than truncated promoters) (9).
269 To examine these contradictory results further, we constructed plasmids containing the entire
270 *luxC* gene and its promoter region driving expression of *gfp*, *lacZ*, or *mCherry* (Fig. 6A, 6B).
271 These constructs contained the intragenic region between *luxC* and *luxD* (15 bp), and the
272 reporter gene was cloned in place of the *luxD* ORF (Fig. 6B). We observed that the *lacZ* and
273 *mCherry* plasmids were activated 16- to 20-fold, whereas the *gfp* construct was only activated
274 1.6-fold by LuxR in *E. coli* (Fig. 6C). The *gfp* (pSO05) and *mCherry* (pSO11) plasmids had
275 similar levels of activation when the plasmids were introduced into wild-type *V. harveyi*, though
276 neither were expressed maximally (Fig. S1B, S1C).
277 We hypothesized that the observed decrease in activation with longer fragments might
278 be due to instability of the transcript when the *luxC* ORF is present upstream of the *gfp* reporter.
279 Thus, we constructed *mCherry* reporter plasmids containing the same four *luxCDABE* promoter
280 fragments that were fused to *gfp* in Figure 1A and assayed these in *E. coli* (Fig. 6A). We verified
281 that the shortest region tested (2 bp past the primary transcription start site) was sufficient for
282 activation, and there was no significant difference in expression with a construct containing a
283 slightly longer promoter fragment (Fig. 6D), and the three shortest fragments were activated
284 >50-fold (Fig. 6D). However, as seen with GFP, the plasmid with the longest promoter fragment

(*e.g.*, pJV366 with 407 bp of the *luxC* ORF) yielded a significantly lower level of mCherry expression (Fig. 6D, activated 17-fold), which was similar to the construct containing the entire *luxC* ORF (Fig. 6C, pSO11, activated 17-fold). A similar trend was observed in *V. harveyi* for these mCherry plasmids (Fig. S1C). Thus, we conclude that constructs containing long fragments indeed decrease expression of downstream reporters, and for some of these large decreases occur (*i.e.*, *gfp*). This result is not observed with expression of the *luxCDABE* operon *in vivo*; the expression levels of each of the five genes in the operon are similar and do not differ by more than 2-fold from one another (as determined by microarray analysis) (19).

To examine these results, we measured transcript levels of *gfp* for several P_{luxC} reporter plasmids in *E. coli*. The relative transcript levels of *gfp* from qRT-PCR measurements were high for the three plasmids containing short regions of the *luxCDABE* promoter, but as seen with GFP expression measurements, levels of *gfp* transcripts significantly dropped ~43-fold in a strain containing the pJV365 plasmid containing 407 bp into the *luxC* ORF compared to a strain containing pJV369 (Fig. 6E). The levels of *gfp* transcripts were significantly decreased in pJV365 compared to all the other plasmids tested with shorter promoter fragments. Thus, we conclude that the decrease in GFP expression in the pJV365-containing strain is due to a decrease in transcript levels, which may be caused either by transcript instability or a decrease in transcription initiation or elongation in plasmids with long promoter fragments. We did not observe a significant decrease in *gfp* transcript levels with pJV367 as observed with GFP expression (Fig. 1B), suggesting that the decrease in GFP expression may be due to constraints at the post-transcriptional or translational level. These results indicate that testing multiple promoter fusions is beneficial for identifying a promoter-reporter fusion that functions *in vivo* to mimic expression from the native locus.

Discussion

We have developed the RAIL method for rapid construction of promoter fusion plasmids and demonstrated that this approach can be applied to multiple promoters and reporter genes. The RAIL strategy can be used to quickly generate a few reporters or to create large libraries of promoter fusions for high-throughput analysis of the regions that drive transcription activation. The method requires simple cloning steps, and once the system is designed for a particular plasmid backbone, only two locus-specific primers are needed. For our plasmid backbone, we designed arbitrary primer sets for creating fusions to *gfp*, *mCherry*, and *lacZ* that can be used with any gene locus (Table S1), and these primers can be easily modified for use in any plasmid with a reporter gene.

Our library sets revealed several important findings with regard to the expression profiles for the *luxCDABE* and *betIBA-proXWV* promoters. First, we validated previous work describing the requirement for LuxR binding sites in these promoters (5, 6, 9, 14). Second, we identified the promoter region that is required for high levels of transcription activation for these two promoters. We did not resolve the 3' boundary to a specific nucleotide locus in these experiments because we did not use every combination of anchor nucleotides in the arbitrary primers and restricted our analysis to combinations of A and T pairs. To acquire complete coverage, a full set of random primers with every combination of nucleotides as anchors should be used. However, with this resolution we clearly found a marked difference in plasmids containing various fragments of the promoters such that we could identify the region sufficient for maximum gene expression. Even when only one arbitrary primer was used, as in the case of the *betI* library, we were still able to determine regions of maximal regulation, but with less coverage. Therefore, we have shown that the RAIL method is applicable for small-scale studies, in which perhaps only one primer is used for quicker analysis, or large-scale studies, where a combination of arbitrary primers will result in higher coverage of the promoter region and produce higher resolution. Further, arbitrary primer design in these studies were limited to terminal A-T combinations due to low G+C content in *V. harveyi* (~45%). In GC-rich organisms,

336 we propose that arbitrary primers should instead be designed with terminal GC combinations for
337 more precise anchoring. Smaller fragments may be sufficient to drive the same level of gene
338 expression, which can be tested with the full series of anchor nucleotides in the arbitrary
339 primers. Further, the minimum 5' end of the promoters in this study are not known, so 5' ends
340 were chosen several hundred basepairs upstream of the ORF (~400 – 1000bp). Future studies
341 could use the same approach to map the 5' boundary of these two promoters, which is a
342 separate but intriguing question.

343 Third, our data conclusively demonstrate that there is no requirement for the region
344 downstream of the transcription start site for full activation of the *luxCDABE* promoter. This
345 finding is important because a previous study by Miyamoto *et al.* also tested promoter regions
346 for *luxCDABE* via a *cat* promoter (9). Among the various constructs tested in that study, the
347 pMGM127 plasmid contains a region truncated slightly upstream of the -26 transcription start
348 site (the specific 3' end is undefined in the article) and the pMGM116 plasmid includes a 3' end
349 at +61 relative to the *luxC* start codon (Fig. S1A). The shorter promoter in pMGM127 shows no
350 transcription activation, whereas the longer promoter in pMGM116 had full activation of the *cat*
351 reporter (9). These data and other observations have led to an anecdotal hypothesis in the field
352 that there is an element downstream of the transcription start site that is required for full
353 activation of the *luxCDABE* promoter. Our data refute this hypothesis because the pJV369
354 plasmid does not include the 5'-UTR and is maximally activated in both *E. coli* and *V. harveyi*.

355 Finally, our analysis of various promoter-reporter fusion plasmids demonstrated that not
356 all reporter fusions are created equal and suggests that testing various reporter constructs for
357 each gene of interest is beneficial to finding the optimal reporter for downstream assays. We
358 noted that plasmid constructs with long fragments of the *luxCDABE* and *betIBA-proXWV*
359 promoters that included sections of the first ORF in the operon were substantially decreased in
360 expression, and we showed that this is effective at the transcript level for *luxC-gfp* fusions (Fig.
361 6E). However, we also noted that the strains containing the pJV367 plasmid that had a

362 decrease in GFP fluorescence did not exhibit a decrease in *gfp* transcript levels (Fig. 6E). This
363 result implies that the 7-fold decrease in GFP fluorescence is due to post-transcriptional or
364 translational effects, such as mRNA secondary structure that may block translation initiation.
365 We have focused attention on our results for the multiple plasmids with long *luxCDABE*
366 promoter fragments that show significantly decreased levels of reporter expression (pJV365,
367 pJV366, pSO05, pSO10, and pSO11). Our qRT-PCR analysis showed that transcript levels
368 were significantly decreased for pJV365 compared to its counterpart plasmids with shorter
369 promoter fragments. These data suggest that decreased reporter expression for all the other
370 long promoter plasmids may also be due to decreased transcript levels. There are at least two
371 possible reasons why the pJV365 plasmid has decreased transcript levels. One possibility is
372 that transcripts generated with fragments of the *luxCDABE* operon fused to the *gfp* gene may
373 fold into unstable secondary structure and be subject to degradation. However, we suspect that
374 this explanation is unlikely to be the cause of low expression for every plasmid with a fragment
375 longer than +129, as we would predict that at least some would be stable. A second possibility
376 is that LuxR binding to site H is acting as a roadblock to transcription elongation, which results
377 in the abrupt drop in GFP expression for plasmids containing promoter regions that terminate
378 after site H. Previously, we showed that scrambling site H does not decrease LuxR activation of
379 β-galactosidase expression in a *luxC-lacZ* reporter plasmid under conditions in which LuxR is
380 maximally expressed at high cell densities in *V. harveyi* (6). However, the results of our
381 expression profiling experiment in *E. coli* with the *luxCDABE* promoter library suggest that
382 plasmids that contain LuxR site H have decreased levels of transcription activation and are
383 strictly in the 'low GFP expression' pool (Fig. 3B). LuxR has an extremely high affinity for site H
384 with a K_d of 0.6 nM, one of the tightest LuxR binding affinities in the genome (5). Thus, it is
385 curious why LuxR binds at this locus with no apparent activation defect when tested at high cell
386 density in *V. harveyi*.

387 Protein roadblocks have been described in bacteria and eukaryotes that hinder
388 transcription, and elongation factors aid in transcription elongation through these roadblocks by
389 various mechanisms (*e.g.*, Mfd in *E. coli*) (20). In addition, when multiple RNAP molecules are
390 initiated from the same promoter, these trailing RNAP complexes can "push" a stalled RNAP
391 through a roadblock (21). Thus, it is possible that higher levels of transcription initiation of the
392 *luxCDABE* promoter in *V. harveyi* at high cell densities drive transcription elongation through
393 site H, whereas lower levels of LuxR at low cell densities in *V. harveyi* or in our synthetic *E. coli*
394 system are not sufficient to push through the LuxR site H roadblock. LuxR concentrations are
395 low in the cell at low cell densities, and thus, the relatively few LuxR molecules likely bind to the
396 highest affinity sites, such as site H in the *luxC* ORF. As cells grow to high cell densities, LuxR
397 levels accumulate (19, 22, 23) and enable LuxR binding to other sites, which drives high levels
398 of transcription initiation and may relieve binding of LuxR to site H to allow RNA polymerase
399 elongation. Alternatively, the roadblock might be relieved by restructuring of the DNA
400 architecture at the locus. Because we have already shown that IHF binds to multiple places at
401 the *luxCDABE* region and its binding is positively cooperative with LuxR, this DNA bending may
402 play a role in removing transcription roadblocks. We also observed a sharp difference between
403 the 'medium expression' pool and 'low expression' pool just downstream of the LuxR site H (Fig.
404 3B), suggesting that there may be yet another roadblock in this region. Future studies should
405 elucidate the role of LuxR binding sites within ORFs in *V. harveyi*, which are observed
406 throughout the genome (5).
407 In conclusion, the RAIL method offers a rapid and efficient method to obtain libraries of
408 reporter fusions that can be used for various studies of gene expression and regulation. Often in
409 bacterial genetics, researchers attempt to create promoter fusions by cloning a reporter gene in
410 place of the translation start site, and this would have yielded suboptimal reporters for the
411 *luxCDABE* promoter. Anecdotally, and as we experienced with the *betIBA-proXWV* and
412 *luxCDABE* promoters, one often needs to construct multiple reporter fusions to identify a

413 promoter region that drives gene expression mimicking native locus gene expression. Thus, our
414 method is more efficient by generating numerous clones in a single cloning experiment. We
415 envision use of the RAIL method for numerous other purposes, such as creating functional GFP
416 protein fusions for studying protein localization, identifying *cis*-regulatory sequences in
417 promoters (*e.g.*, protein binding sequences), inserting affinity tags for purification strategies, and
418 identifying highly expressed soluble constructs for protein purification. Finally, this method
419 should be applicable to any model organism for which genetic cloning techniques have been
420 established.
421
422 **Materials and Methods**
423
424 *Bacterial strains and media*
425 *E. coli* strains S17-1λpir, DH10B, and derivatives (Table S2) were used for cloning and
426 *in vivo* assays. *E. coli* strains were grown shaking at 275 RPM at 37°C in lysogeny broth (LB),
427 augmented with 10 µg/mL chloramphenicol and 10 µg/mL tetracycline when required. The *V.*
428 *harveyi* BB120 is strain ATCC BAA-1116, which was recently reassigned to *Vibrio campbellii*
429 (24). It is referred to as *V. harveyi* throughout this manuscript for consistency with previous
430 literature. BB120 and derivatives (Table S2) were grown at 30°C shaking at 275 RPM in LB
431 Marine (LM) medium supplemented with 10 µg/mL chloramphenicol when required. LM is
432 prepared similarly to LB (10 g tryptone, 5 g yeast extract) but with 20 g NaCl instead of 10 g
433 used in LB.
434
435 *Molecular methods*
436 Oligonucleotides (Table S1) were purchased from Integrated DNA Technologies. All
437 PCR reactions were performed using Phusion HF polymerase (New England BioLabs) or iProof
438 polymerase (BioRad). Restriction enzymes, enzymes for isothermal DNA assembly (15), and

439 dNTPs were obtained from New England BioLabs. DNA samples were visualized on 1%
440 agarose gels. Standard cloning methods and primers for the single plasmid constructs listed in
441 Table S3 are available upon request. Standard sequencing of single plasmid constructs was
442 conducted by ACGT, Inc. and Eurofins Genomics. To measure the expression levels of
443 fluorophore reporter plasmids, *E. coli* and *V. harveyi* strains were grown overnight at 30°C
444 shaking at 275 RPM. Strains were diluted 100-fold in growth media and selective antibiotics in
445 96-well plates (black with clear bottom), covered with microporous sealing tape (USA Scientific),
446 and incubated shaking at 30°C at 275 RPM for 16-18 h. Fluorescence and OD_{600} from strains
447 expressing *mCherry* and *gfp* were measured using either a BioTek Synergy H1 or Cytation plate
448 reader. Miller assays were conducted as previously described (6). RNA extraction and qRT-
449 PCR were performed and analyzed as described (14) with primers listed in Table S1 on a
450 StepOne Plus Real-Time PCR machine (Applied Biosystems). Transcript levels were
451 normalized to the level of expression of the internal standard *recA*, and the standard curve
452 method was used for data analysis. The error bars on graphs represent the standard deviations
453 of measurements for at least three biological samples. Statistical analysis was performed using
454 GraphPad Prism version 7.0c. Additional information about statistical analyses pertinent to each
455 result set are included in the figure legends.
456
457 *RAIL: construction of promoter libraries by arbitrary PCR*
458 The arbitrary PCR method was adapted from Schmidt *et al.* (25) with several
459 modifications. The first round of PCR was conducted using two primers: a forward primer
460 specific to the promoter of interest and a reverse primer for random DNA amplification (Fig. 2,
461 primers 1F and 1R, respectively). The 1R primer includes a priming sequence, followed by eight
462 random nucleotides ('N'), and terminating in two defined nucleotide anchors, either AT, TA, TT,
463 or AA (Table S1). For PCR round 1, ~10-100 ng/µl of genomic DNA from *V. harveyi* BB120 or a
464 plasmid containing the region of interest was added as the template. The reaction included 200

465 µM dNTPs, 250 µM primers, 5% DMSO, 0.5 µl Phusion polymerase, and 1X Phusion buffer.
466 Cycling parameters were as follows and as previously published (25): an initial denaturation at
467 95°C for 5 min, then 5 cycles of 95°C for 30 s, 25°C for 30 s, and 72°C for 2.5 min, followed by
468 30 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 2.5 min, and a final extension step of
469 72°C for 10 min. PCRs were purified using the GeneJet PCR Purification Kit (Thermo Scientific)
470 and eluted in 30-50 µl of elution buffer. The second round of PCR used primers 2F and 2R (Fig.
471 2). The forward primer (2F) included 30 nt homology to the plasmid backbone for IDA and a
472 sequence specific to the promoter of interest that is nested downstream of the 1F primer. The
473 reverse primer (2R) included the priming sequence that is identical to that of primer 1R. To
474 perform PCR round 2, 5 µl of the purified DNA from round 1 was used as the template. These
475 reactions also included 200 µM dNTPs, 250 µM primers, 5% DMSO, 0.5 µl Phusion polymerase
476 and 1X Phusion buffer. Cycling parameters were as follows: an initial denaturation at 95°C for 5
477 min, then 35 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 2.5 min, and a final extension
478 step of 72°C for 10 min. PCR products were separated by agarose gel electrophoresis,
479 visualized by UV transillumination, and the products were gel extracted to desired target size
480 using a GeneJet Gel Extraction Kit (Thermo Scientific).
481 Cloning of arbitrary PCR inserts into the plasmid backbone was performed using IDA as
482 described (15). Library inserts were incubated in IDA reactions with 100 ng of plasmid
483 backbone, and these reactions were transformed into electrocompetent *E. coli* Electromax
484 DH10B cells (ThermoFisher) and plated on media with selective antibiotics. DNA from individual
485 colonies was first screened by restriction digest and sequenced to confirm that inserts of the
486 desired size were incorporated. For generation of libraries for sorting, >50,000 colonies were
487 collected from plates, mixed in LB selective media, and the culture stored at -80°C. DNA
488 extracted from this library was transformed into electrocompetent *E. coli* S17-1λpir cells
489 containing a plasmid expressing *luxR* (pKM699). After this second transformation step, >50,000

490 colonies were collected from plates, mixed in LB selective media, and the culture stored at -
491 80°C.
492
493 *Flow cytometry sorting*
494 A BD FACSAria II was used to sort *E. coli* cells based on expression of the fluorescent
495 reporter. In preparation for the sort, the library culture from the frozen stock was grown in 50 mL
496 of selective medium at 30°C shaking to OD_{600} = 0.5. DNA was extracted from this culture as the
497 'input' DNA sample for sequencing. Control cultures were used to set the sorting gates. The
498 positive controls for maximal expression were strains containing pKM699 (expressing *luxR*) and
499 a plasmid construct that demonstrated high levels of expression for either P_{luxC} (pJV369, P_{luxC}-
500 *gfp* positive control plasmid) or P_{betI} (pCH50, P_{betI}-*mCherry* positive control plasmid). The
501 negative controls were strains containing pKM699 and pJS1194 (*gfp* negative control plasmid)
502 or pCH76 (*mCherry* negative control plasmid), which lack promoters in front of *gfp* or *mCherry*,
503 respectively. The library culture was sorted into bins with >10,000 cells per bin. For the P_{luxC}
504 sort, there were four bins: no GFP expression, low GFP expression, medium GFP expression,
505 or high GFP expression. For the P_{betI} sort, there were two bins: high mCherry expression or no
506 mCherry expression. The sorted cell cultures were incubated in 5 mL of selective medium
507 shaking at 30°C at 275 RPM and grown to stationary phase. Cultures were stored at -80°C, and
508 the DNA was extracted for sequencing using the GeneJet Miniprep kit (Thermo Scientific).
509
510 *Illumina library preparation and sequencing*
511 DNA extracted from all sorted samples and input controls was purified over a Performa
512 DTR gel filtration cartridge (Edge Biosystems). Next, the DNA was sheared using a Covaris
513 S220 in 6 x 16 mm microtubes to average sizes of 400 bp and analyzed on an Agilent 2200
514 TapeStation using D1000 ScreenTape. The sheared DNA samples (1 μg) were each treated

515 with terminal deoxynucleotidyl transferase (TdT) in a tailing reaction using TdT (Promega) and a
516 mixture of dCTP and ddCTP (475 μM and 25 μM final concentrations, respectively) to generate
517 a poly-C tail. The reactions were incubated at 37°C for 1 h, heat-inactivated at 75°C for 20 min,
518 and cleaned over a Performa DTR gel filtration cartridge. Next, two rounds of PCR were
519 performed to attach sequences for Illumina sequencing (Table S1). In round 1, a C-tail specific
520 primer olj376 and a gene-specific primer (CH060 for the *gfp* library, and CH063 for the *mCherry*
521 library) were used to amplify the promoter fragments that were cloned into the plasmid during
522 the RAIL method. This PCR reaction used Taq polymerase (NEB), with the following cycling
523 conditions: an initial denaturation at 95°C for 2 min, then 24 cycles of 95°C for 30 s, 58°C for 30
524 s, and 72°C for 2 min, and a final extension step of 72°C for 2 min. The round 2 PCR used a
525 nested gene specific primer (CH061 for the *gfp* library, and CH064 for the *mCherry* library) to
526 provide added specificity and also to append the linker sequence needed for Illumina
527 sequencing. The second primer in the reaction contained different barcodes for each sample to
528 enable the libraries to be pooled and sequenced simultaneously (Table S1; BC37-44). The
529 round 2 PCRs were performed using Taq polymerase, and cycling as follows: an initial
530 denaturation at 95°C for 2 min, then 12 cycles of 95°C for 30 s, 52°C for 30 s, and 72°C for 2
531 min, and a final extension step of 72°C for 2 min. These final PCR products were examined by
532 DNA gel electrophoresis, at which point a smear of products was visible on the gel.
533 Sequencing was performed on a NextSeq 500 using a NextSeq 75 reagent kit using
534 42bp x 42bp paired-end run parameters and gene specific primers (CH062 for the *gfp* library,
535 and CH065 for the *mCherry* library). Reads were checked with FastQC (v0.11.5; Available
536 online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc) to ensure that the quality of
537 the data was high and that there were no noteworthy artifacts associated with the reads. The R1
538 read from paired-end Illumina NextSeq reads were quality and adapter trimmed with
539 Trimmomatic (v0.33) (26) using the following parameters: ILLUMINACLIP:<adapters>:3:20:6

LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:25. Reads were high quality with 80-85% of the reads surviving trimming. Reads were mapped against the *Vibrio campbellii* ATCC BAA-1116 genome (including molecules NC_009777, NC_009783, and NC_009784) using bowtie 2.3.2 (27) and visualized using JBrowse (version 1.10.12) to analyze the alignment to the *betI* and *luxC* gene regions. Analysis of the promoter-seq data hinged on contrasting high, medium, low, and no expressing cells to the null distribution. The null distribution was determined by sequencing cells collected without any sorting applied. The reads associated with the null, high, medium, low, and no expression datasets for the *luxC* library (or null, high, and no expression datasets for the *betI* library) were normalized by transforming the data into the fraction of bases covered, which was defined as the depth of coverage at a particular base divided by the total depth of coverage over the particular promoter region (*betI*: bases 1361141-1362440 on NC_009784; *luxC*: bases 1424774-1426907 on NC_009784). Analysis was performed by plotting the $\log_2$ ratio of the observed fraction of bases covered (for high, medium, low or no expression cell collections) over the expected (null) distribution.

Acknowledgments

We thank Ankur Dalia for assistance with the design of the Illumina sequencing protocol and helpful discussions. We also thank James Ford and Douglas Rusch from the Indiana University Center for Genomics and Bioinformatics for Illumina sequencing experiments, bioinformatics analysis, and assistance with graphing the results. We also wish to acknowledge Christiane Hassel from the Indiana University Flow Cytometry Core Facility for assistance with FACSAria II analysis and sorting. We thank Ryan Fritts for assistance with statistical analyses. We appreciate the use of Jason Tennessen's BioTek Cytation plate reader, and we thank Ryan Chaparian and Blake Petersen for assistance with experiments. This work was supported by the National Institutes of Health (NIH) Grant 1R35GM124698-01 to JVK and start-up funds from Indiana University to JVK and MLB.

References

1. Rosochacki SJ, Matejczyk M. 2002. Green fluorescent protein as a molecular marker in microbiology. Acta Microbiol Pol 51:205-16.
2. Jiang T, Xing B, Rao J. 2008. Recent developments of biological reporter technology for detecting gene expression. Biotechnol Genet Eng Rev 25:41-75.
3. Andersen JB, Sternberg C, Poulsen LK, Bjorn SP, Givskov M, Molin S. 1998. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. Appl Environ Microbiol 64:2240-6.
4. Arnone MI, Dmochowski IJ, Gache C. 2004. Using reporter genes to study cis-regulatory elements. Methods Cell Biol 74:621-52.
5. van Kessel JC, Ulrich LE, Zhulin IB, Bassler BL. 2013. Analysis of activator and repressor functions reveals the requirements for transcriptional control by LuxR, the master regulator of quorum sensing in Vibrio harveyi. MBio 4.
6. Chaparian RR, Olney SG, Hustmyer CM, Rowe-Magnus DA, van Kessel JC. 2016. Integration host factor and LuxR synergistically bind DNA to coactivate quorum-sensing genes in Vibrio harveyi. Mol Microbiol 101:823-40.
7. Chatterjee J, Miyamoto CM, Zouzoulas A, Lang BF, Skouris N, Meighen EA. 2002. MetR and CRP bind to the Vibrio harveyi lux promoters and regulate luminescence. Mol Microbiol 46:101-11.
8. Meighen EA. 1991. Molecular biology of bacterial bioluminescence. Microbiol Rev 55:123-42.
9. Miyamoto CM, Smith EE, Swartzman E, Cao JG, Graham AF, Meighen EA. 1994. Proximal and distal sites bind LuxR independently and activate expression of the Vibrio harveyi lux operon. Mol Microbiol 14:255-62.
10. Swartzman E, Silverman M, Meighen EA. 1992. The luxR gene product of Vibrio harveyi is a transcriptional activator of the lux promoter. J Bacteriol 174:7490-3.
11. Reznikoff WS. 1992. The lactose operon-controlling elements: a complex paradigm. Mol Microbiol 6:2419-22.
12. Dunn TM, Hahn S, Ogden S, Schleif RF. 1984. An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. Proc Natl Acad Sci U S A 81:5017-20.
13. Galagan J, Lyubetskaya A, Gomes A. 2013. ChIP-Seq and the complexity of bacterial transcriptional regulation. Curr Top Microbiol Immunol 363:43-68.
14. van Kessel JC, Rutherford ST, Cong JP, Quinodoz S, Healy J, Bassler BL. 2015. Quorum sensing regulates the osmotic stress response in Vibrio harveyi. J Bacteriol 197:73-80.
15. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343-5.
16. Welsh J, McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213-8.

17. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531-5.
18. Showalter RE, Martin MO, Silverman MR. 1990. Cloning and nucleotide sequence of luxR, a regulatory gene controlling bioluminescence in Vibrio harveyi. J Bacteriol 172:2946-54.
19. van Kessel JC, Rutherford ST, Shao Y, Utria AF, Bassler BL. 2013. Individual and Combined Roles of the Master Regulators AphA and LuxR in Control of the Vibrio harveyi Quorum-Sensing Regulon. J Bacteriol 195:436-43.
20. Park JS, Marr MT, Roberts JW. 2002. E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. Cell 109:757-67.
21. Epshtein V, Toulme F, Rahmouni AR, Borukhov S, Nudler E. 2003. Transcription through the roadblocks: the role of RNA polymerase cooperation. EMBO J 22:4719-27.
22. Waters CM, Bassler BL. 2006. The Vibrio harveyi quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers. Genes Dev 20:2754-67.
23. Teng SW, Wang Y, Tu KC, Long T, Mehta P, Wingreen NS, Bassler BL, Ong NP. 2010. Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. Biophys J 98:2024-31.
24. Lin B, Wang Z, Malanoski AP, O'Grady EA, Wimpee CF, Vuddhakul V, Alves Jr N, Thompson FL, Gomez-Gil B, Vora GJ. 2010. Comparative genomic analyses identify the Vibrio harveyi genome sequenced strains BAA-1116 and HY01 as Vibrio campbellii. Environ Microbiol Rep 2:81-89.
25. Schmidt KH, Pennaneach V, Putnam CD, Kolodner RD. 2006. Analysis of gross-chromosomal rearrangements in Saccharomyces cerevisiae. Methods Enzymol 409:462-76.
26. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.
27. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-9.

Figure Legends

Figure 1. Promoter fusion plasmids for the *luxCDABE* genes. (A) Diagram of the regions of the *luxCDABE* promoter present in various plasmids listed. LuxR binding sites (LuxR BS) are shown as gray boxes, and letters correspond to the LuxR binding site name A through H. Transcription start sites are indicated by black arrows. The LuxC translation start site is shown

as +1. Lengths of constructs are shown relative to the LuxC translation start site. (B) Relative GFP expression per OD_{600} (GFP/OD_{600}) is shown for *E. coli* strains containing plasmids with varying *luxCDABE* promoter fragments fused to *gfp* as indicated in (A). The strains also contain either a plasmid constitutively expressing LuxR (pKM699) or an empty vector (pLAFR2). Relative expression was calculated by dividing the values for the pKM699-containing strain by the pLAFR2-containing strain. Different letters (a, b, c) indicate significant differences between strains ($p<0.05$; one-way analysis of variance [ANOVA] followed by Tukey's multiple-comparison test on log-transformed data; $n=3$).

Figure 2. Schematic of the RAIL method for constructing libraries of promoter fusions. In PCR round 1, primers 1F and 1R were used to amplify a range of products specific to the *luxC* promoter. The 1R primers have eight random ('N') nucleotides incorporated, are anchored by two nucleotides (AA, AT, TA, or TT), and contain a linker (shown in blue). In PCR round 2, the 2F and 2R primers were used to further amplify and add a linker to the products. Primer 2F anneals just downstream of 1F and contains a linker (shown in green). Primer 2R anneals to the linker region of primer 1R. The linear plasmid backbone was prepared either by restriction digest or PCR. The library of products was inserted into the linear plasmid backbone via IDA using the homologous sequences present in the two linker regions (green and blue). The final plasmids contain fragments of the *luxC* promoter fused to the reporter. Gel images shown are examples of products from PCR rounds 1 and 2 for the *luxCDABE* locus using arbitrary R1 primers JCV1135-1138 and F1 primer SO71 in round 1 (Table S1). The products of the round 1 reactions were used in round 2 with primers SO72 and JCV1139 (Table S1).

Figure 3. Flow cytometry sorting and next-generation sequencing of the $P_{luxCDABE}$-*gfp* library. (A) FACS GFP expression data from three *E. coli* cultures: 1) a negative control strain containing an empty vector control (pJS1194) and a plasmid expressing LuxR (pKM699), 2) a

673 positive control strain containing a $P_{luxCDABE}$-*gfp* reporter plasmid (pJV369) and pKM699, and 3)
674 the $P_{luxCDABE}$ library of plasmids in *E. coli* containing pKM699. Gates (boxes) indicate the cells
675 sorted (% of total population) into four bins: no GFP expression, low GFP expression, medium
676 GFP expression, and high GFP expression. Data were presented using FlowJo software. (B)
677 Genomic regions associated with differential expression of the *luxCDABE* locus. The nucleotide
678 coverage of the reads (42 bp) is shown for different populations of cells with distinct levels of
679 GFP reporter expression as indicated in the legend. Data are graphed as the nucleotide position
680 (x-axis) versus the $\log_2$-ratio of observed coverage density divided by the expected coverage
681 density (as determined by the read counts observed in the total library; y-axis). Areas with
682 positive values in $\log_2$ observed/expected coverage densities indicate an enrichment of
683 sequence reads in that region, and areas with negative values indicate lower than expected
684 frequency reads. The thick black bar indicates the location of the *luxC* ORF. The locations of
685 LuxR binding sites are indicated by white boxes. The locus to which the 2F primer anneals is
686 indicated by an arrow.
687
688 **Figure 4. Expression data from the $P_{betIBA\text{-}proXWV}$-*mCherry* library.** (A) Diagram of the regions
689 of the *betIBA-proXWV* promoter present in various plasmids. LuxR binding sites (LuxR BS) and
690 the BetI binding site (BetI BS) are shown as gray and white boxes, respectively. Putative
691 transcription start sites are indicated by black arrows. The BetI translation start site is shown as
692 +1. Lengths of constructs are shown relative to the BetI translation start site. (B) Relative
693 mCherry expression per OD_{600} (mCherry/OD_{600}) is shown for *E. coli* strains as calculated by
694 dividing the values for the pKM699 strain by the pLAFR2 strain. The strains contained plasmids
695 with varying *betIBA-proXWV* promoter fragments fused to *mCherry* as indicated in (A). Different
696 letters (a, b, c, d) indicate significant differences between strains (p<0.05; one-way ANOVA
697 followed by Dunnett's multiple-comparison test; n=3). (C) FACS data showing mCherry
698 expression for three *E. coli* cultures: 1) a negative control strain containing pCH76 and pKM699,

2) a positive control strain containing pCH50 and pKM699, and 3) the $P_{betIBA\text{-}proXWV}$ library in *E. coli* containing pKM699. Gates indicate the cells sorted (% of total population) into two bins: no mCherry expression and high mCherry expression. (D) Genomic regions associated with differential expression of the *betIBA-proXWV* locus. The nucleotide coverage of the reads (42 bp) is shown for different populations of cells with distinct levels of or mCherry reporter expression as indicated in the legend. Data are graphed as the nucleotide position (x-axis) versus the $\log_2$-ratio of observed coverage density divided by the expected coverage density (as determined by the read counts observed in the total library; y-axis). Areas with positive values in $\log_2$ observed/expected coverage densities indicate an enrichment of sequence reads in that region, and areas with negative values indicate lower than expected frequency reads. The thick black bar indicates the location of the *luxC* ORF. The locations of LuxR and BetI binding sites are indicated by white boxes. The locus to which the 2F primer anneals is indicated by an arrow.

Figure 5. Promoter-reporter fusion plasmids for the *VIBHAR_06912* promoter. (A) Diagram of the regions of the *VIBHAR_06912* promoter present in various plasmids. The putative transcription start site is indicated by a black arrow. The VIBHAR_06912 translation start site is shown as +1. Lengths of constructs are shown relative to the VIBHAR_06912 translation start site. (B) Modified Miller units are shown for wild-type (BB120) and Δ*luxR* (KM669) strains containing various plasmids as indicated in (A). Asterisks (*) indicate significant differences between wild-type and Δ*luxR* strains for the various plasmids ($p<0.05$; two-way ANOVA followed by Sidak's multiple-comparison test on log-transformed data; $n=3$; ns, not significant).

Figure 6. Measuring transcription of plasmids with long promoter regions fused to reporters. (A) Diagram of the regions of the *luxCDABE* promoter present in various constructs. LuxR binding sites (LuxR BS) are shown as gray boxes. Transcription start sites are indicated

724 by black arrows. The LuxC translation start site is shown as +1. Lengths of constructs are
725 shown relative to the LuxC translation start site. Plasmids containing *mCherry*, *gfp*, or *lacZ*
726 fusions are indicated. (B) Diagram of plasmids containing *luxCDABE* promoters and the *luxC*
727 ORF fused to *gfp* (pSO05), *lacZ* (pSO10), or *mCherry* (pSO11). Each construct contains the 15-
728 bp sequence between *luxC* and *luxD* as shown. (C) Relative expression of reporters (GFP,
729 mCherry, and LacZ) is shown for *E. coli* strains containing either a plasmid constitutively
730 expressing LuxR (pKM699) or an empty vector (pLAFR2). Relative expression was calculated
731 by dividing the values for the pKM699-containing strain by the pLAFR2-containing strain. LacZ
732 expression was determined by modified Miller assays, and GFP and mCherry expression were
733 assayed using a plate reader. Different letters (a, b) indicate significant differences between
734 strains ($p<0.05$; one-way ANOVA followed by Tukey's multiple-comparison test on log-
735 transformed data; $n=3$). (D) Relative expression of mCherry (mCherry/OD_{600}) is shown for *E.*
736 *coli* strains containing either pKM699 or pLAFR2, calculated as described in (C). Different letters
737 (a, b, c) indicate significant differences between strains ($p<0.05$; one-way ANOVA followed by
738 Tukey's multiple-comparison test on log-transformed data; $n=3$). (E) Relative transcript levels of
739 *gfp* determined by qRT-PCR for *E. coli* strains containing either pKM699 or pLAFR2, calculated
740 as described in (C). Different letters (a, b) indicate significant differences between strains
741 ($p<0.05$; one-way ANOVA followed by Tukey's multiple-comparison test on log-transformed
742 data; $n=6$).
743