CRAGE-RB-PI-seq enables transcriptional profiling of rhizobacteria during plant-root colonization

Tomoya Honda[1,2,§], Sora Yu[2], Dung Mai[1], Leo Baumgart[1], Gyorgy Babnigg[3], & Yasuo Yoshikuni[1,2,4-7,§]

[1]The US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3]Biosciences Division, Argonne National Laboratory, Lemont, IL 60439, USA
[4]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[5]US Department of Energy Center for Bioenergy and Bioproducts Innovation (CABBI), Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[6]Global Institution for Collaborative Research and Education, Hokkaido University, Hokkaido 060-8589, Japan
[7]Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Tokyo 183-8538, Japan
[§]Correspondence: thonda@lbl.gov, yyoshikuni@lbl.gov

Abstract

Plant roots release a wide array of metabolites into the rhizosphere, shaping microbial communities and their functions. While metagenomics has expanded our understanding of these communities, little is known about the physiology of their members in host environments. Transcriptome analysis via RNA sequencing is a common approach to learning more, but its use has been challenging because plant RNA masks bacterial transcripts. To overcome this, we developed randomly-barcoded promoter-library insertion sequencing (RB-PI-seq) and combined it with chassis-independent recombinase-assisted genome engineering (CRAGE), using *Pseudomonas simiae* WCS417 as a model rhizobacterium. This method enables targeted amplification of barcoded transcripts, bypassing plant RNA interference and allowing measurement of thousands of promoter activities during root colonization. We identified promoters involved in biofilm formation, reactive oxygen species response, and root immune suppression, which all contribute to root colonization. Additionally, we discovered that transcriptional activation of xanthine dehydrogenase and a lysozyme inhibitor are crucial for evading plant immune system defenses. This framework is scalable to other bacterial species and provides new opportunities for understanding rhizobacterial gene regulation in native environments.

Introduction

Plant roots secrete nutrient-rich compounds into the surrounding soil, creating unique habitats that support diverse microbial communities[1,2]. These microbial communities, in turn, play critical roles in plant physiology by promoting growth, protecting against disease, or, in some cases, causing infection[3–5]. Over the past decade, extensive metagenomic studies have identified the members of microbial communities across various plant species, genotypes, and growth environments[6–10]. The next major challenge is to develop a framework that predicts microbial gene functions, regulation, and impacts on plant physiology. Achieving this requires in-depth knowledge of the physiology of individual microbial species within their host environments. In particular, studying key organisms such as plant-growth-promoting rhizobacteria and pathogens is critical for advancing sustainable agricultural systems.

Among plant-associated bacteria, *Pseudomonas* species have been extensively studied because of their roles in promoting plant growth and protecting against disease. In particular, *Pseudomonas simiae* WCS417 has served as a model bacterium, with traits beneficial to its hosts and a remarkable ability to colonize a wide range of plant species[11]. Previous studies have uncovered molecular mechanisms underlying these symbiotic relationships, such as siderophore production[12] and modulation of plant immune systems[13,14]. Additionally, transposon mutagenesis sequencing of *P. simiae* WCS417 identified genetic determinants that influence early-stage root colonization, including through chemotaxis and substrate utilization[15]. These findings suggest that bacteria engage in sophisticated chemical interactions with host plants during colonization. However, global profiling of these physiological adaptations remains limited.

Transcriptomics is a powerful tool for characterizing the temporal dynamics of global gene expression. In host plants, transcriptomic studies have revealed coordinated cascades of gene expression in response to microbial colonization or microbe-associated molecular patterns[16–18]. Although the host responses have been studied at both population and single-cell levels, the adaptations of rhizobacteria remain less explored. This gap is primarily due to technical challenges, as bacterial RNA is significantly outnumbered by host-plant RNA. This issue is common in plant-associated bacteria, and a similar issue is found in bacteria associated with other hosts[19–22]. To address this, various strategies have been explored, including deep sequencing[23,24], depletion of host and bacterial rRNA[23–25], enrichment of bacterial mRNA using customized probes[22,26], and physical separation of bacterial cells from host tissues through techniques such as laser microdissection[27], flow cytometry[28], and density-gradient centrifugation[26]. Alternatively, in vitro cell culturing with root exudates has been a common approach. Although the choice of approach depends on the research question, profiling transcriptomes in planta with high efficiency and scalability remains the most desirable option.

Synthetic biology offers an alternative way to profile global transcriptional activities. Using DNA barcodes to report promoter activity enables quantification through a straightforward workflow involving PCR amplification and barcode sequencing. This method has been successfully used in model prokaryotic and eukaryotic organisms, revealing unique features of promoter regulatory elements[29–32]. In the present study, we introduced a DNA-barcoded promoter library into *P. simiae* WCS417, using chassis-independent recombinase-assisted genome engineering (CRAGE)[33,34],

a versatile strain-engineering technology developed in our group. We then established an experimental workflow to characterize in planta promoter activities by selective amplification of barcoded transcripts, overcoming interference from plant RNA. We term this workflow as “CRAGE randomly-barcoded promoter-library insertion sequencing (CRAGE-RB-PI-seq),” or PI-seq in short. Using this method, we identified a set of uniquely regulated promoters involved in plant-root colonization and validated their activities and functions through a series of independent experiments. The framework we developed enables rapid characterization of transcriptional activities and can be applied to a wide range of bacterial species, providing new insights into gene function and regulation in native environments.

Results

Development of the CRAGE-RB-PI-seq workflow

To generate a barcoded promoter library, we first computationally extracted the 140 bp upstream sequences from the start codons of all annotated protein-coding sequences in the *P. simiae* WCS417 genome (**Fig.1A**). We selected a length of 140 bp because of the technical limitations of oligo pool synthesis available at the time from Twist Biosciences. To optimize versatility, we divided promoter libraries into three groups according to the distances to upstream genes: group I (≤30 bp), group II (31–139 bp), and group III (≥140 bp). Each library was synthesized, barcoded, and cloned into a CRAGE accessory vector (**Fig.1B**), then transformed into a conjugal *E. coli* donor strain.

The promoter libraries were chromosomally integrated into *P. simiae* WCS417 (strain SB599) using CRAGE (**Fig.1C**). DNA sequencing (DNA-seq) was used to link each promoter with unique barcodes (**Fig.1D**). This process successfully cloned 91% of promoters (5,040 out of 5,541), each tagged with unique barcodes (**Supplementary Data 1)**. Pooled cell libraries were cultured under various conditions, both in vitro and in planta (**Fig.1E**). Promoter activities were quantified using targeted RNA-seq and DNA-seq, normalizing RNA-derived barcode counts to DNA-derived counts (**Fig.1E**).

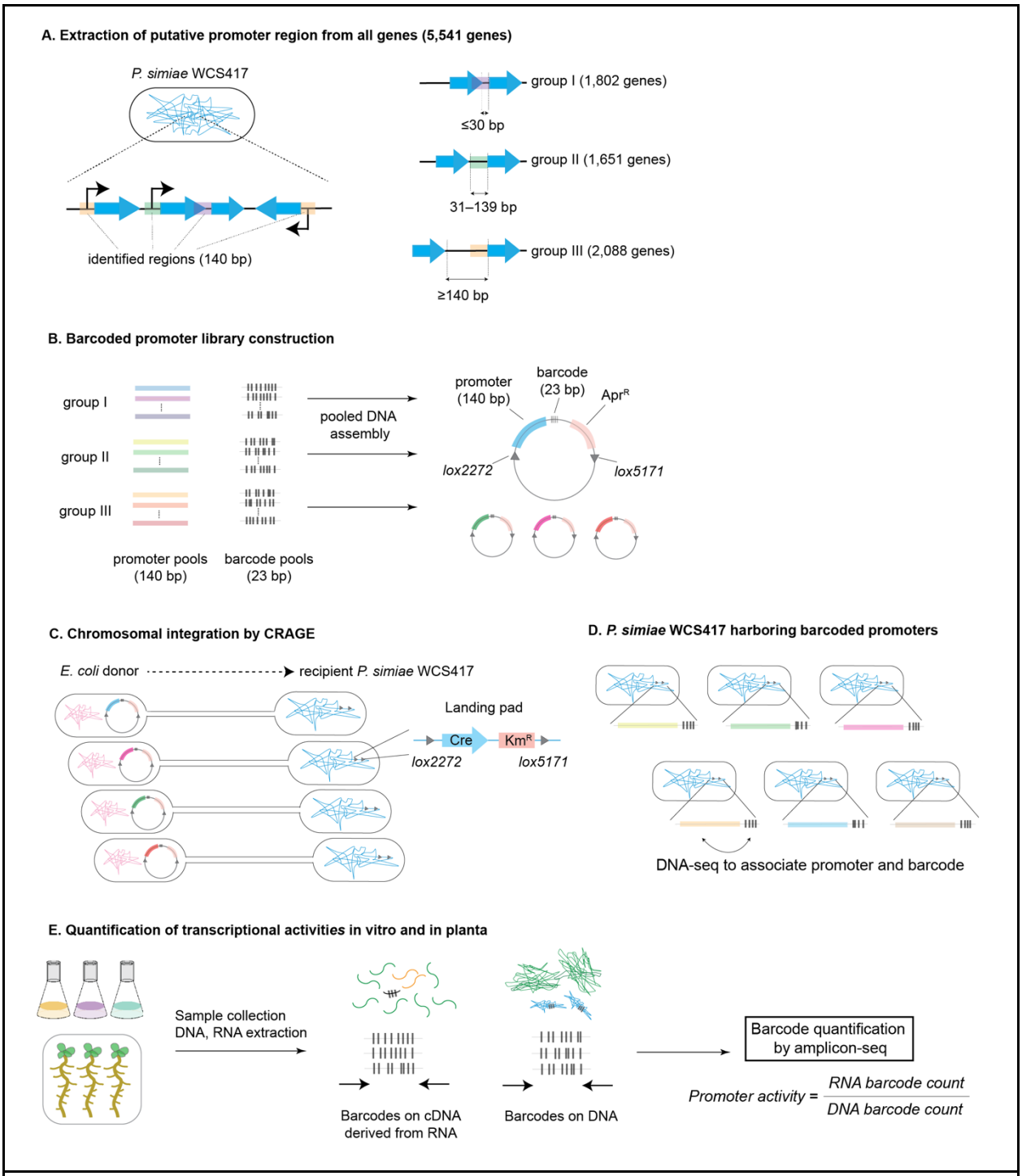

Figure 1: Development of the CRAGE-PI-seq workflow

(**A**) A total of 5,541 sequences, each 140 bp sequences immediately upstream of start codons, were computationally extracted. These promoter sequences were categorized into three groups according to the distances to upstream genes: group I (≤30 bp), group II (31–139 bp), and group III (≥140 bp). (**B**) Each promoter library was synthesized, barcoded with 23 bp sequences, and cloned into the CRAGE accessory vector (pW26_mod), which included an apramycin-resistant

marker (Apr^R) flanked by two mutually exclusive lox sites (*lox2272* and *lox5171*), necessary for recombination. (**C**) The plasmids were transformed into an *E. coli* donor strain and conjugated into *P. simiae* WCS417 (strain SB599), which harbored a CRAGE landing pad on the chromosome. Cre-mediated recombination integrated the barcoded promoters into the landing pad. (**D**) The promoter-barcode regions were amplified from the genomic DNA and sequenced via targeted DNA-seq to associate each promoter with its corresponding unique barcodes. (**E**) The transcriptional activity of promoters was characterized by growing cells under various conditions, extracting DNA and RNA, and amplifying barcode regions from the DNA and cDNA. The cDNA was synthesized from RNA by selective reverse transcription. Promoter activities were then quantified by normalizing RNA barcode counts to DNA barcode counts for each gene.

Transcriptional profiling of *P. simiae* WCS417 under various in vitro conditions

To assess the effectiveness of our approach, we first grew the pooled library (groups I–III) under five different conditions, preparing sequencing samples for both PI-seq and conventional RNA-seq (**Fig.2A**). PI-seq consistently demonstrated high library coverage and reproducibility across conditions. For example, DNA barcode analysis showed that under the reference condition (20 mM glucose), 98.8% of promoters (4,982 out of 5,040 in the library) were covered at least 50 times per million reads, and the histogram showed a narrow distribution (**Fig.2B**). The data reproducibility between biological replicates was high (Pearson's r > 0.99) (**Fig.2C**), indicating sufficient library coverage under these culture conditions.

Promoter activity — determined by normalizing RNA barcode counts to DNA barcode counts — showed a large variation within and between groups I–III, with the activity levels differing by several orders of magnitude (**Fig.2D**). In particular, group II and III promoters (green and magenta bars in the figure) exhibited overall higher activity than group I promoters (blue bars). The results of biological replicates were highly reproducible for group II and III libraries (Pearson's r > 0.83), while the results between biological replicates for group I were less reproducible (Pearson's r ~ 0.52) (**Fig.2E**), a trend noted across all growth conditions (**Fig.S1**).

We hypothesized that the lower activity of group I promoters (derived from ≤30 bp intergenic regions) might be because many promoters were part of operons, and therefore lacked active transcription start sites. To investigate this, we used Rockhopper[35] to predict operon structures in silico, incorporating *P. simiae* WCS417 genome sequence data and RNA-seq data. This analysis identified 1,149 operons comprising 3,447 genes, as well as 2,094 single isolated genes (**Fig.2F, Supplementary Data 2**). As expected, most group I promoters were part of operons, while all group III promoters corresponded to either operon start sites or isolated genes. Of the 3,243 genes (1,149 operonic + 2,094 single), 2,880 promoters were successfully cloned into our libraries. Subsequent analysis focused on these 2,880 promoters to further validate their activities.

Analysis of differentially induced promoters from PI-seq data

We next compared PI-seq data from cells grown under 20 mM glucose conditions with the data from 20 mM citrate conditions. Differential analysis identified 32 promoters that were uniquely

142 upregulated under glucose conditions, and 10 promoters that were uniquely upregulated under
143 citrate conditions (**Fig.2G**). RNA expression driven by these upregulated promoters were similarly
144 upregulated, as RNA-seq verified these patterns (**Fig.2H**). Comparison of fold changes in
145 promoter activity and RNA expression demonstrated a strong quantitative agreement between
146 PI-seq and RNA-seq, as illustrated in **Fig.2I**. Notably, there was a strong correlation between the
147 two methods for differentially induced promoters (**Fig.2I,** blue and red points, fitted with a solid
148 line: Pearson's r = 0.91).
149
150 Additional analysis of other culture conditions confirmed similar trends for both PI-seq and RNA-
151 seq (**Fig.S2**). Unsupervised clustering analysis of the PI-seq data revealed distinct clustering
152 depending on culture conditions, consistent with the pattern in RNA-seq data (**Fig.S3**).
153 Importantly, PI-seq captured physiologically relevant changes in promoter activity (**Fig.2J**). For
154 example, under glucose conditions, the two most upregulated promoters (annotated by arrows in
155 **Fig.2G**) drive genes (PS417_RS06235, PS417_RS06240) involved in glucose binding and
156 transport, while under citrate conditions, the most upregulated promoter drives an operon that
157 encodes porin and tripartite-tricarboxylate transporter-substrate-binding protein
158 (PS417_RS20775, PS417_RS20780), both functional for citrate uptake[36,37]. These findings align
159 with expected differences corresponding to the different culture conditions, and the large fold
160 changes in RNA expression were confirmed by RNA-seq (annotated in **Fig.2H**). Overall, these
161 analyses suggest that PI-seq is effective at identifying a core set of promoters that are
162 reprogrammed in response to environmental changes.
163

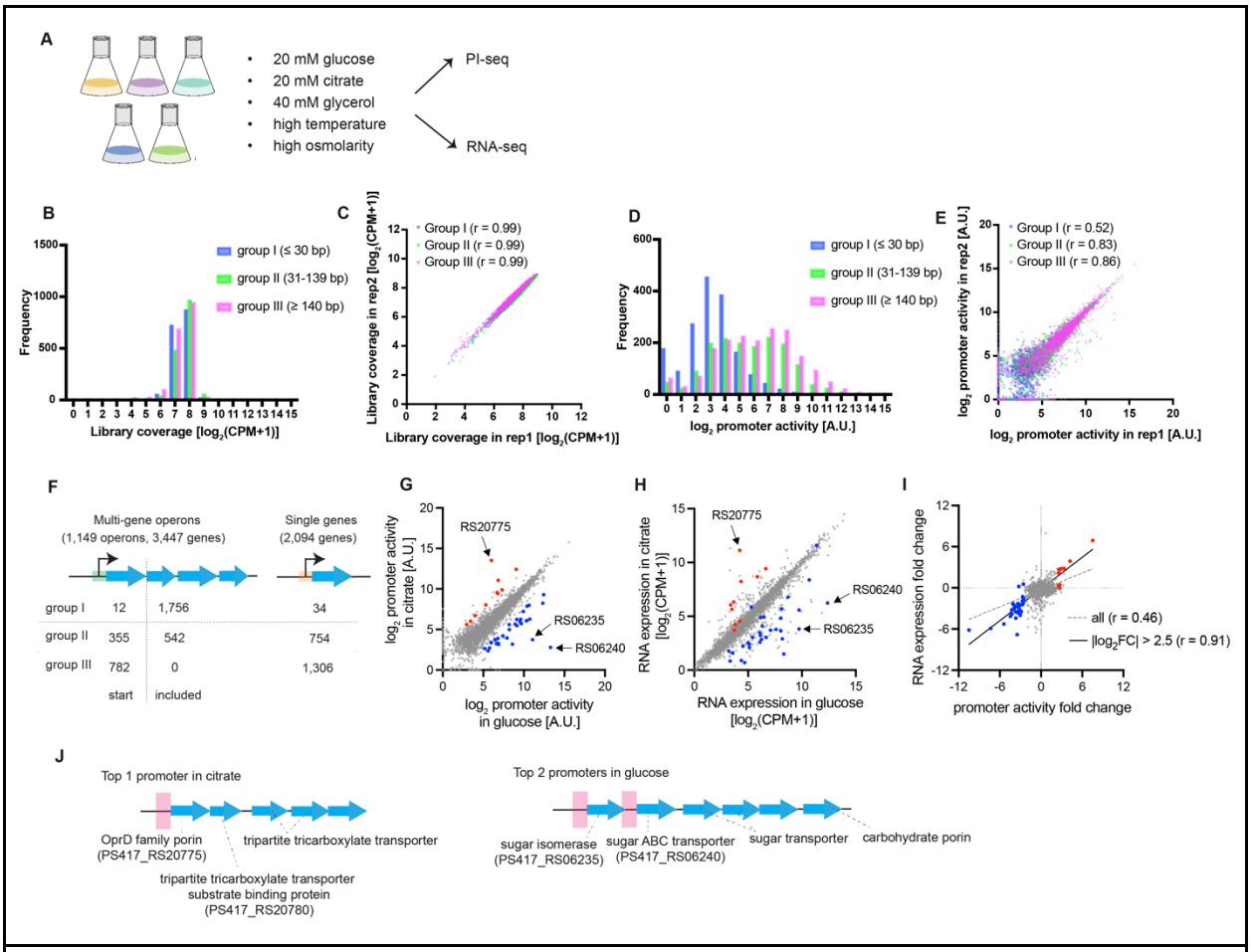

Figure 2: Characterization of promoter activities in liquid cultures.
(**A**) Experimental overview. A pooled library of groups I–III was grown under five different conditions. PI-seq and RNA-seq libraries were created for each condition (*n* = 2 biological replicates). (**B-E**) Analysis of PI-seq data for the cells grown in 20 mM glucose, used as an example. DNA barcode counts showed a narrow distribution and high coverage of the designed library (B). Comparison of the library coverage for two biological replicates showed high reproducibility (C). In contrast, RNA barcode counts normalized to DNA barcode counts showed a large variation in promoter activity. Groups II and III promoters (green and magenta bars, intergenic distance ≥ 31 bp) were more active than group I promoters (blue bars, intergenic distance ≤ 30 bp) (D). The biological replicates for groups II and III showed high reproducibility of promoter activity, while the reproducibility of group I promoter activity was weaker (E). Pearson's r correlation values for each group are shown in parentheses. (**F**) In silico prediction of operon structure using Rockhopper identified 1,149 operons and 2,094 single genes. From these 3,243 genes, 2,880 promoters were successfully cloned into the library and used to compare PI-seq and RNA-seq results. (**G**) Comparison of promoter activity under two conditions. PI-seq data showing promoter activity under 20 mM glucose (x-axis; blue points) and under 20 mM citrate (y-axis; red points) highlights upregulated promoters. Differentially activated promoters were selected based on fold changes (FCs) ($\log_2$FC > 2.5) after filtering

out low-activity promoters (log_2 promoter activity < 2). (**H**) Comparison of gene expression under 20 mM glucose and 20 mM citrate conditions using RNA-seq data. Genes driven by the promoters identified by PI-seq in (G) are highlighted in the same colors. (**I**) Correlation of expression changes between PI-seq and RNA-seq. Comparison of fold changes in expression between PI-seq (x-axis) and RNA-seq (y-axis) shows strong quantitative agreement for the promoters with large fold changes (log_2FC > 2.5), indicated by red and blue points. The fold changes were calculated by analyzing the expression under the 20 mM citrate condition divided by that under the 20 mM glucose condition for each method. Pearson's r correlation values are provided (all genes: dashed gray line; genes with log_2FC > 2.5: solid black line). (**J**) Illustrations of the most upregulated promoters and their associated downstream genes. Consistent with the experimental conditions, the promoter most upregulated under 20 mM citrate conditions drives genes encoding porin and a tripartite tricarboxylate transporter, while the two promoters most upregulated under 20 mM glucose conditions drive sugar transporter genes within the same operon. The 140 bp promoter regions are highlighted in pink.

Transcriptional profiling of *P. simiae* WCS417 during *Arabidopsis* root colonization

Using the engineered strains, we investigated *P. simiae* WCS417 promoter activity during colonization of *Arabidopsis* roots. We inoculated 10-day-old seedlings with pooled libraries from groups II and III (**Fig.3A**), omitting group I promoters because of their low activity. After 3 days of colonization, root samples were collected every 24 hours for 2 days, and promoter activity was quantified by barcode counts from extracted DNA and RNA. Despite having ~100 times fewer cells in planta (~$2x10^7$) than the number of cells in liquid cultures ($1.6x10^9$) and handling samples contaminated with ~20 mg of plant roots, we achieved high library coverage and reproducible results (**Fig.3B–E**). This underscores the utility of our approach, as conventional RNA-seq often requires deep sequencing to overcome plant RNA contamination and obtain sufficient bacterial reads (**Table S2**).

To understand the colonization dynamics, we analyzed promoter activity over time by calculating log_2 fold changes relative to day 1. We identified 74 promoters that were significantly upregulated and 50 promoters that were significantly downregulated during the colonization process (**Fig.3F**). Further examination incorporating Rockhopper analysis revealed that several promoters drive co-transcription of multiple genes, resulting in 111 upregulated and 101 downregulated genes (**Fig.3G, Supplementary Data 4**). Functional insights were gained through KEGG pathway analysis (**Fig.3G**), which revealed enrichment of a biofilm formation pathway from upregulated genes, while biomass production pathways, such as fatty acid, ribosome, and amino acid biosynthesis, were enriched from downregulated genes. Examples of these promoter activities are shown in **Fig.3H**. These findings align with previous studies[23,38,39], highlighting the physiological shift from metabolically active to adhesive and protective states during root colonization. Notably, similar responses in *P. simiae* WCS417 are partially triggered by coumarins exuded from *Arabidopsis* roots[40].

Additionally, several promoters associated with genes essential for root colonization were upregulated. These included catalase (PS417_RS27325) and DNA repair enzymes (PS417_RS21405, PS417_RS14920) that respond to reactive oxygen species (ROS)[41,42], as well as pyrroloquinoline quinone (PQQ)-dependent genes (PS417_RS14915, PS417_RS10720) that suppress root immune responses via gluconic acid production[14] (**Fig.3I**). Other upregulated promoters, potentially important for colonization, included those driving a sigma factor *rpoE* (PS417_RS20735), xanthine dehydrogenase (PS417_RS0404), and a hypothetical protein (PS417_RS02110) predicted to be an MliC (membrane-bound lysozyme inhibitor of c-type lysozyme) family protein (**Fig.3J**); we experimentally validated the latter two, as shown in **Fig.5**.

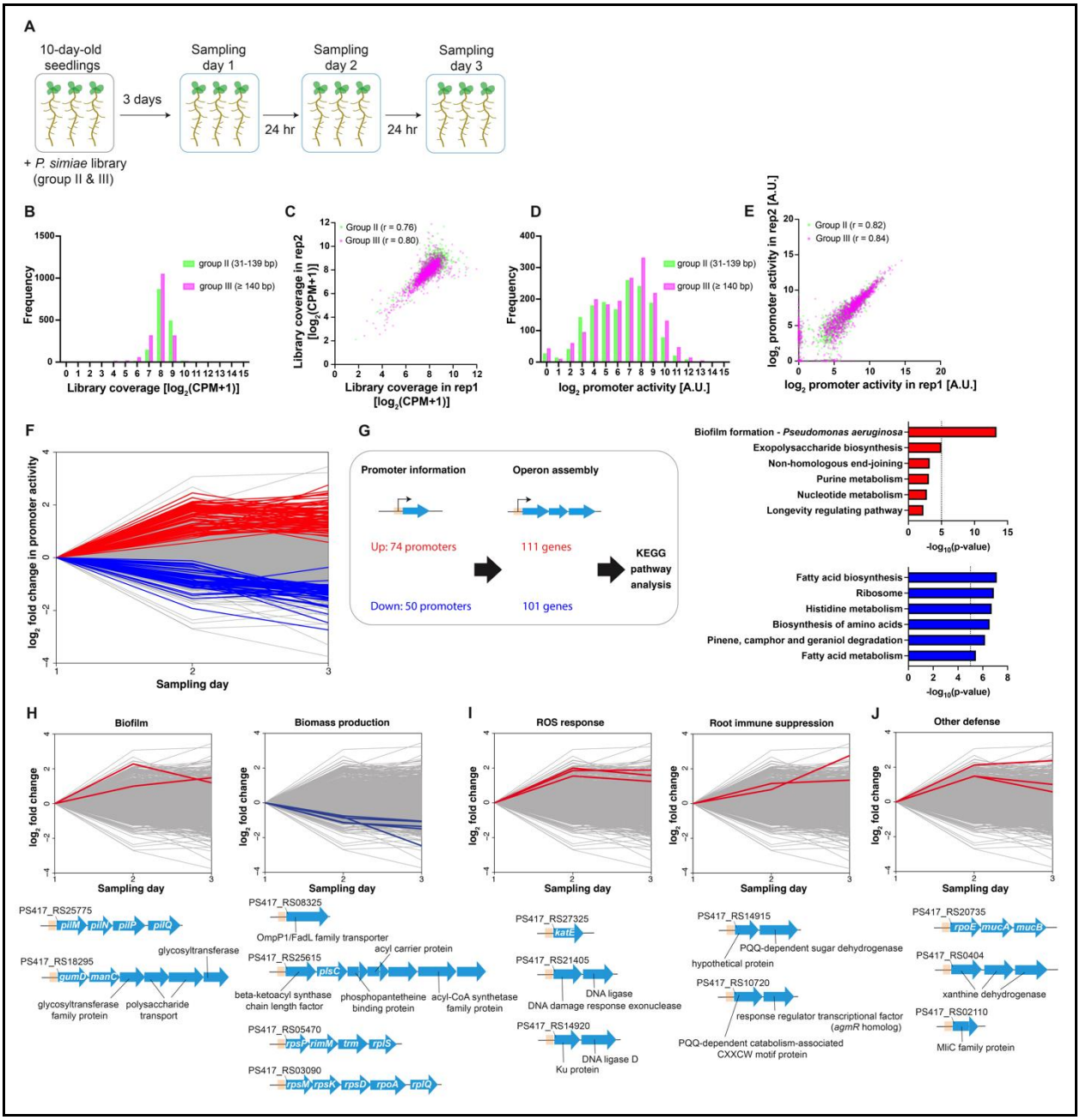

Figure 3: Characterization of promoter activities during *Arabidopsis* root colonization. (**A**) *P. simiae* WCS417 populations containing promoter libraries from groups II and III were inoculated onto phytagel plates with 10-day-old *Arabidopsis thaliana* seedlings. Group I was excluded because of low promoter activity (Fig.2). After 3 days of colonization, root samples were collected every 24 hours over the next 2 days. Barcode sequencing libraries were prepared by targeted PCR amplification from extracted RNA and DNA to assess promoter activity (n = 4 biological replicates per time point). (**B–E**) Day 1 PI-seq data from root-colonizing cells demonstrated a high library coverage (B) and reproducibility (C), a large variation (D) and high reproducibility of promoter activity across biological replicates (E). (**F**) Log_2 fold changes (FCs) in promoter activity were calculated relative to day 1. Red and blue lines mark significantly upregulated and downregulated promoters, respectively ($|log_2FC| > 1$ and Benjamini-Hochberg adjusted p-value < 0.05). (**G**) Predicted operon structures (from Fig.2F) were linked to promoters, and the corresponding genes were subjected to KEGG pathway analysis. Upregulated genes were enriched in a biofilm formation pathway (top), while downregulated genes were associated with biomass synthesis and metabolic functions (bottom). Dashed lines indicate statistical significance thresholds (adjusted p-value < 10^{-5}). (**H–J**) Examples of promoter activities related to biofilm formation and biomass production (H), ROS response and plant immune suppression (I), and other defense-related genes (J). Red lines represent upregulated promoters and blue lines represent downregulated promoters, with operon structures illustrated below and promoter regions highlighted in orange.

201

202 **Verification of PI-seq data by green fluorescent protein reporter imaging analysis**

203 To independently validate our PI-seq results, we selected five of the upregulated and three of the
204 downregulated promoters shown in **Fig.3H–J** (see **Table 1** for strain list) and constructed reporter
205 strains with green fluorescent protein (GFP) expression driven by 500 bp sequences upstream of
206 each gene, rather than by the 140 bp sequences used in the promoter library. As a control, we
207 used a strain with constitutive P*lac*-driven GFP expression[34].
208
209 Initially, we attempted to visualize GFP signals using microscopy, but variability in signal strength
210 across different parts of root tissues and between plants made data collection challenging. To
211 overcome this, we designed new constructs using the T7 expression system to amplify the GFP
212 signals (**Fig.4A**; see **Fig.S4** for construction details). This approach allowed us to quantify GFP
213 expression from whole *Arabidopsis* roots using a gel-doc imager (**Fig.4B**).
214
215 Fluorescent strains were incubated with *Arabidopsis* seedlings following the same protocol as
216 shown in **Fig.3A,** and mean fluorescent intensities were measured. Fluorescence levels in the
217 constitutive promoter strain (strain ST41) remained constant (**Fig.4C,F**), while fluorescence in the
218 five upregulated promoters (in strains ST5, ST23, ST30, ST31, ST35) increased significantly
219 (**Fig.4D,F**), validating our PI-seq data. Unexpectedly, the fluorescence levels for the
220 downregulated promoters (in strains ST19 and ST34) remained unchanged, and the level for one
221 construct (ST26) increased (**Fig.4E,F**). This may be due to residual T7 RNA polymerase and

GFP. While downregulation was more difficult to capture, these constructs confirmed the activities of the upregulated promoters, as indicated by their differences from the constitutive promoter.

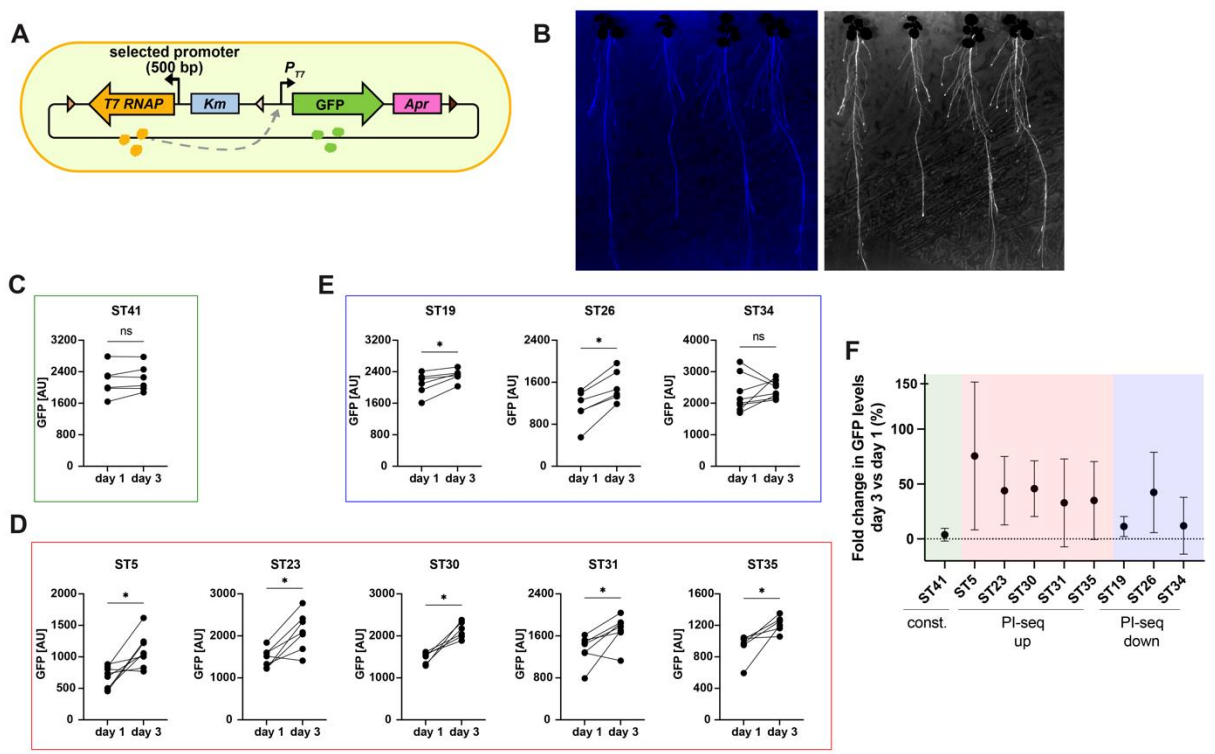

Figure 4: Verification of promoter activities via GFP reporter assay.
(**A**) GFP reporter strains were constructed to assess promoter activity, with GFP expression mediated by T7 RNA polymerase. This design allowed for quantification of GFP signals across whole *Arabidopsis* roots using a gel-doc imager. To ensure accuracy, we used 500 bp sequences upstream of each gene for reporter strain construction, rather than 140 bp sequences from the promoter library (see **Fig.S4** for details). (**B**) Example images showing four *Arabidopsis* seedlings colonized by a constitutive GFP-expressing strain (ST41). Left: RGB image, right: inverted grayscale image. (**C–E**) Fluorescent intensities were measured on days 1 and 3 of colonization, corresponding to the sampling points in **Fig. 3A**. The P*lac* promoter served as a control for constitutive expression (**C**: strain ST41). Five promoters were selected to confirm upregulated activity from PI-seq (**D**: ST5, ST23, ST30, ST31, ST35), and three were selected to confirm downregulation (**E**: ST19, ST26, ST34). Statistical analysis was performed using the Wilcoxon Signed-Rank Test (*: $P < 0.05$, ns: no significance). Each point represents a replicate. (**F**) Fold changes in fluorescence intensities on day 3 compared with intensities on day 1. Error bars represent standard deviations. While the constitutive promoter showed little change (C), the upregulated promoters exhibited significant increases (D). The downregulated promoters, however, exhibited no change or increased activity, potentially due to residual T7 RNA polymerase or GFP (E).

Promoter activity in root colonization based on PI-seq	Locus tag	Gene product based on RefSeq annotation	GFP strain tested (shown in Fig.4)	*Mutant availability	**Growth in minimum medium	Competition assay tested (shown in Fig.5)
Upregulated	PS417_RS04040	Xanthine dehydrogenase (Xdh): 2Fe-2S iron-sulfur cluster-binding protein	ST5	Yes	Yes	Tested
Upregulated	PS417_RS18295	Undecaprenyl-phosphate glucose phosphotransferase	ST23	Yes	Yes	Tested
Upregulated	PS417_RS10720	PQQ-dependent catabolism-associated CXXCW motif protein	ST30	Yes	Yes	Tested
Upregulated	PS417_RS02110	MliC family protein	ST31	Yes	Yes	Tested
Upregulated	PS417_RS14920	Ku protein	ST35	No	Not determined	Untested
Downregulated	PS417_RS03090	30S ribosomal protein S13	ST19	No	Not determined	Untested
Downregulated	PS417_RS26170	Imidazoleglycerol-phosphate dehydratase HisB	ST26	Yes	No	Untested
Downregulated	PS417_RS08325	OmpP1/FadL family transporter	ST34	Yes	Not determined	Untested

Table 1: Strains used to verify promoter activity and characterize mutant phenotypes. *Mutant availability was confirmed through the internal JGI portal, which stores data on insertion mutants created by Cole *et al*[15]. **Growth was tested in M9 minimal medium supplemented with 20 mM glucose.

226

227 ## Characterization of mutant phenotypes

228 To examine the functions of genes driven by the upregulated promoters, we obtained transposon
229 insertion mutants for four genes created by Cole *et al*[15] (see **Table 1**) and characterized their
230 metabolic and root colonization capabilities. When they were grown alone, the growth rates of all
231 mutants were similar to those of the wild-type strain in M9 minimal media supplemented with
232 glucose alone (**Fig.5A**) or with glucose plus casamino acids (**Fig.5B**). However, in a competitive
233 colonization assay, the wild-type strain significantly outcompeted the mutant strains (**Fig.5C**),
234 particularly those with mutations in undecaprenyl-phosphate glucose phosphotransferase
235 (PS417_RS18295) and PQQ-dependent catabolism-associated protein (PS417_RS10720).
236 These findings align with previous studies highlighting the role of biofilm formation[43,44] and plant

immune suppression in root colonization[14], facilitated by polysaccharide synthesis and PQQ-dependent metabolism, respectively.

The wild-type strain also outcompeted two other mutants, carrying insertion mutations in genes PS417_RS02110 and PS417_RS04040 whose roles in the root colonization ability of *P. simiae* WCS417 had not been previously reported. According to RefSeq annotation PS417_RS02110 encodes an MliC family protein. The MliC protein from *Pseudomonas aeruginosa* and *Escherichia coli* is known to function as a c-type lysozyme inhibitor, aiding in host colonization of humans and other animals[45,46]. Although *P. simiae* WCS417's MliC shares limited protein sequence identity with these homologs, AlphaFold 3[47] predicted significant structural similarities (**Fig.S5**), suggesting a similar lysozyme-inhibiting function. This is especially relevant given recent studies showing that lysozyme-like hydrolases are produced by *Arabidopsis*[48] and epiphytic fungi[49], potentially influencing microbial communities. While *Pseudomonas* species are generally lysozyme-resistant, the role of MliC may be critical in the presence of agents that destabilize the outer membrane of the bacterium, thereby increasing its susceptibility to external lysozymes[50,51]. To investigate whether MliC increases lysozyme tolerance in *P simiae* WCS417, we examined the effect of lysozyme on the wild-type strain and a mutant carrying an insertion in *mliC*. While the wild-type tolerated high lysozyme doses, the mutant showed reduced viability (**Fig.5D**), indicating that MliC contributes to *P. simiae* WCS417's protection against lysozymes from plants and from other microbes.

The second gene, PS417_RS04040, encodes xanthine dehydrogenase (XDH), whose transcriptional activation has been linked to ROS tolerance and host infection in *Ralstonia solanacearum*[52] and in *Borrelia burgdorferi*[53]. Since ROS generation is a common plant immune response, we hypothesized that XDH in *P. simiae* WCS417 plays a role in neutralizing plant-generated ROS. To test this, we examined the effect of ROS stress on the wild-type strain and a mutant carrying an insertion in *xdh* during exponential growth in M9 minimal medium with glucose. Upon exposure to 5 mM hydrogen peroxide, both wild-type and mutant strains experienced growth arrest, while the mutant showed a longer arrest by ~60 minutes compared to the wild-type strain (**Fig.5E**), supporting our hypothesis. These findings suggest XDH may contribute to root colonization by working in concert with other ROS-responsive enzymes such as catalases and DNA ligases.

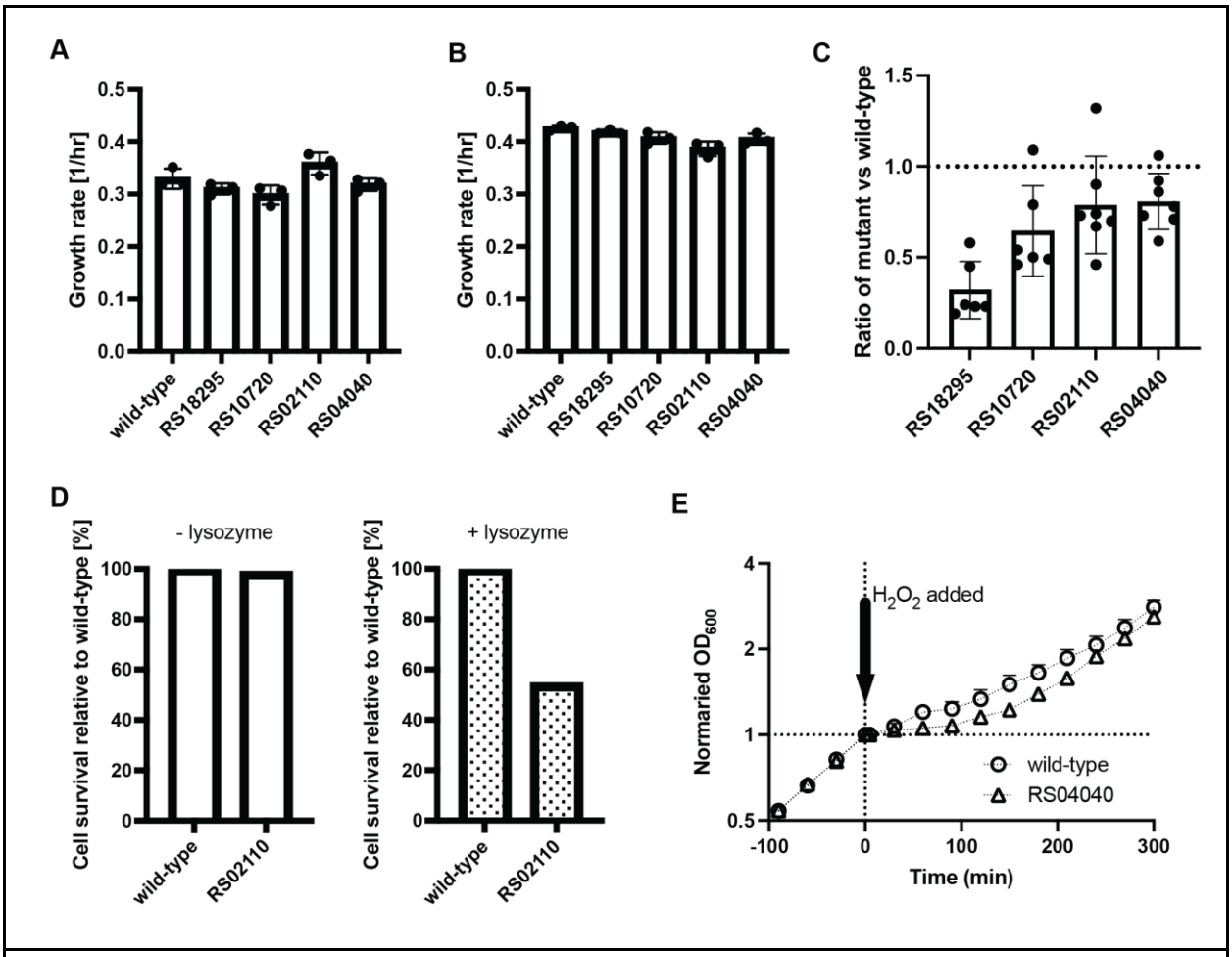

Figure 5: Mutant phenotypes in root colonization and stress tolerance.
(**A–B**) Growth rates of the wild-type strain and four insertion mutants were compared in M9 media supplemented with 20 mM glucose (A) and 20 mM glucose plus 2% casamino acids (B). The growth rates of mutant strains were indistinguishable from those of the wild-type strain under both conditions. The mutant strains are referred to by their locus tags (with "PS417_" omitted; see **Table 1** for full tag). (**C**) Root colonization competition assay with *Arabidopsis* seedlings. Wild-type and individual mutant strains were mixed 1:1 and applied to plates containing 10-day-old *Arabidopsis* seedlings. After 5 days, root samples were collected, and wild-type and mutant strain colonies were counted. The ratios of mutant to wild-type colonies are shown in the graph. All tested mutants were outcompeted by the wild-type strain, as indicated by their ratios falling below the line at y = 1.0, which represents an equal ratio. (**D**) Lysozyme assay of wild-type strain and the insertion mutant in MliC family protein. Cell viability was measured after incubating cells in PBS for 24 hr, either in the absence (left) or presence (right) of 25 mg/mL lysozyme. The mutant strain was 47% less viable than the wild-type strain was. (**E**) Oxidative stress tolerance assay to examine the function of xanthine dehydrogenase (PS417_RS04040). Wild-type and mutant strains were grown in M9 media supplemented with 20 mM glucose and challenged with 5 mM H_2O_2 at time 0. The mutant required approximately

60 minutes longer than the wild-type to restore exponential growth after oxidative stress exposure (*n* = 3 biological replicates).

Discussion

Plant-growth-promoting rhizobacteria offer a promising avenue for enhancing plant biomass and crop yields without high chemical fertilizer use. To harness their full potential, it is critical to understand their in planta physiology, including gene functions and regulatory mechanisms. However, studying these rhizobacteria in host-associated environments is challenging because of their low biomass. Traditional RNA-seq approaches for plant-colonizing bacteria often require large sample sizes and extensive rRNA depletion, followed by deep-sequencing. Even then, only about 0.5% of reads map to bacterial gene coding sequences, with a large proportion mapping to plants[23]. Further optimizations such as improved combinatorial rRNA removals[11] and physical separation of cells prior to lysis[26] are necessary to enrich bacterial mRNA when RNA-seq approaches are used.

In this study, we developed CRAGE-RB-PI-seq in the model rhizobacterium *P. simiae* WCS417 to study its transcriptional activities during *Arabidopsis* root colonization. This approach uses PCR amplification of RNA barcodes from a promoter library, enabling nearly 100% enrichment of bacterial transcripts. This is not only highly efficient but also cost-effective, particularly when the species of interest is present in low abundance. While our experiment involved only ~10^7 cells collected from ~20 mg of root per sample, the input cell amount could be reduced even further by increasing the number of PCR cycles. Moreover, this approach is adaptable to other bacteria compatible with CRAGE. Our current strain portfolio includes approximately 50 species across diverse phyla, many of which were isolated from soil environments. Barcoding also offers a distinct advantage in that it allows differentiation of homologous genes across different organisms; the difficulty in differentiating these is a frequent challenge in metatranscriptome analysis.

We selected the 140 bp sequences that were immediately upstream of individual genes as promoters because of the technical limitations of the oligo pools available at the time from Twist Biosciences. Despite this constraint, our results showed high correlation between PI-seq and RNA-seq, particularly in identifying differentially expressed genes (**Fig.2IJ** and **Fig.S2**). We expect that extending the promoter regions would result in even greater accuracy; this is now possible with new oligo pools that can harbor 240 bp sequences (Twist Biosciences). Advances in oligo synthesis technologies such as enzymatic oligo synthesis[54] and DropSynth[55] will also allow construction of longer promoter sequences, improving library quality. Additionally, results could be enhanced by refining library design to capture entire promoter regions, including regulatory elements and transcriptional start sites, which could be facilitated by 5'-enriched RNA-seq methods such as differential RNA-seq[56] and CAGE[57]. The availability of high-quality genome annotations will further contribute to the effectiveness of these libraries.

Using PI-seq, we found that many promoters associated with genes essential for root colonization (such as those involved in biofilm formation, ROS response, and plant immune suppression) were

activated, confirming previous genetic studies. Additionally, we identified previously unrecognized transcriptional changes, including the activation of the *xdh* and *mliC* genes, which further enhance the mutualistic interactions between *P. simiae* WCS417 and *Arabidopsis*. Although earlier studies have highlighted the importance of chemotaxis and flagellar motility in root colonization[15,58,59], our analysis did not detect the activation of promoters associated with these genes, possibly because of our focus on late-stage sampling, which may have missed earlier activations. As root exudates significantly influence bacterial physiology, future research could explore how specific metabolites regulate bacterial gene expression.

Notably, the strongest upregulation occurred for the promoter driving an operon encoding a PQQ-dependent catabolism-associated protein (PS417_RS10720) and a response regulator transcription factor (PS417_RS10715) from day 1 to day 3 of colonization (6.8-fold increase, **Fig.3I**). The transcription factor shares 86% sequence identity with *A*gmR (also known as PedR1*)* in *P. aeruginosa* PAO1, a key regulator of PQQ biosynthesis[60]. Although our PI-seq analysis did not detect activation of the downstream PQQ biosynthesis operon (*pqqABCDE*)—likely because *pqqA* was not annotated in the early version of the *P. simiae* WCS417 genome—we confirmed its activation through a GFP reporter assay where the *pqqA* upstream region was extracted (**Fig. S6**). PQQ is a known cofactor involved in sugar and alcohol oxidation, with the metabolic byproducts that contribute to plant immune suppression by modulating local pH[14]. PQQ can also function as a growth-promoting factor when directly applied to plants[61,62], and it stimulates production of natural products such as antibiotics in a broad range of *Actinobacteria*[63]. These findings underscore the vital role of PQQ production in the rhizosphere and its potential impact on microbial community structure and plant physiology.

Although PI-seq requires further refinement, this method provides a valuable tool for identifying regulatory elements associated with host-microbe interactions, particularly in non-model bacteria. Given the diverse roles that bacteria play in host physiology—such as biofertilization, biostimulation, and biocontrol—there is ample opportunity to uncover new molecular mechanisms. This knowledge will be instrumental in advancing the use of biologics, as well as in developing new biosensors and genetic circuits that can mediate interactions between microbes and their hosts[64–67].

Acknowledgements

We thank Zhiying Zhao, Robert Evans, William Kauffman, Jan-Fang Cheng, Benjamin Cole, Yuko Yoshinaga, and Jeffery Dangle for helpful discussions and technical suggestions. T.H. acknowledges a JSPS overseas research fellowship from the Japan Society for the Promotion of Science. Lawrence Berkeley National Laboratory is managed by University of California for DOE under contract number DE-AC02-05CH11231 (T.H. S.Y., D.M., L.B., Y.Y.). Argonne National Laboratory is managed by UChicago Argonne, LLC for DOE under contract number DE-AC02-06CH11357 (GB). This work was funded by the secure biosystem design (S.Y., G.B., Y.Y.) and the bioimaging programs (T.H., G.B., Y.Y.) of the U.S. Department of Energy, Office of Biological and Environmental Research under Contract No. DE-AC02-05CH11231 and DE-AC02-06CH11357. This work (proposal: 10.46936/10.25585/60001279), conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231 (T.H., D.M., L.B., Y.Y.). This project was also supported in part by the U. S. Department of Energy, Office of Science, through the Biomolecular Characterization and Imaging Sciences Program, Office of Biological and Environmental Research, under FWP 39156 (T.H., S.Y., Y.Y., G.B.).

Author Contribution

T.H. and Y.Y. designed the study. T.H. performed the experiments, with contribution by S.Y., who constructed GFP reporter strains and characterized mutant phenotypes. T.H. and Y.Y. developed the computational pipeline. D.M. performed protein structural analysis. L.B. contributed to a plasmid construction. Y.Y and G.B. supervised the study and acquired funding. T.H. and Y.Y. wrote the manuscript. All authors reviewed and approved the final version.

Competing Interest Statement

Lawrence Berkeley National Laboratory has filed an international patent application related to high-throughput characterization of bacterial promoters from their host environments on behalf of the Regents of the University of California, on which T.H. and Y.Y. are the named inventors (PCT/US2023/031771). Y.Y. is a co-founder and has financial interest in Quorum Bio, Inc. All other authors declare no competing interest.

Data availability

All sequencing data generated and analyzed in the current study will be deposited in an online archive. Supplementary data that support this study are provided as downloadable Excel files.

References

1. Zhalnina, K. *et al.* Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* **3**, 470–480 (2018).
2. McLaughlin, S., Zhalnina, K., Kosina, S., Northen, T. R. & Sasse, J. The core metabolome and root exudation dynamics of three phylogenetically distinct plant species. *Nat. Commun.* **14**, 1649 (2023).
3. Jansson, J. K., McClure, R. & Egbert, R. G. Soil microbiome engineering for sustainability in a changing environment. *Nat. Biotechnol.* **41**, 1716–1728 (2023).
4. Haney, C. H., Samuel, B. S., Bush, J. & Ausubel, F. M. Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nat. Plants* **1**, 15051 (2015).
5. Wang, N. R. *et al.* Commensal Pseudomonas fluorescens Strains Protect Arabidopsis from Closely Related Pseudomonas Pathogens in a Colonization-Dependent Manner. *mBio* **13**, e02892-21 (2022).
6. Acharya, S. M. *et al.* Fine scale sampling reveals early differentiation of rhizosphere microbiome from bulk soil in young Brachypodium plant roots. *ISME Commun.* **3**, 1–9 (2023).
7. Huang, A. C. *et al.* A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science* **364**, eaau6389 (2019).
8. Levy, A. *et al.* Genomic features of bacterial adaptation to plants. *Nat. Genet.* **50**, 138–150 (2018).
9. Lundberg, D. S. *et al.* Defining the core Arabidopsis thaliana root microbiome. *Nature* **488**, 86–90 (2012).
10. Lebeis, S. L. *et al.* Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**, 860–864 (2015).
11. Pieterse, C. M. J. *et al.* Pseudomonas simiae WCS417: star track of a model beneficial rhizobacterium. *Plant Soil* **461**, 245–263 (2021).

402 12. Berendsen, R. L. *et al.* Unearthing the genomes of plant-beneficial Pseudomonas model
403 strains WCS358, WCS374 and WCS417. *BMC Genomics* **16**, 539 (2015).
404 13. Van Wees, S. C., Van der Ent, S. & Pieterse, C. M. Plant immune responses triggered
405 by beneficial microbes. *Curr. Opin. Plant Biol.* **11**, 443–448 (2008).
406 14. Yu, K. *et al.* Rhizosphere-Associated Pseudomonas Suppress Local Root Immune
407 Responses by Gluconic Acid-Mediated Lowering of Environmental pH. *Curr. Biol.* **29**, 3913-
408 3920.e4 (2019).
409 15. Cole, B. J. *et al.* Genome-wide identification of bacterial plant colonization genes. *PLOS*
410 *Biol.* **15**, e2002860 (2017).
411 16. Stringlis, I. A. *et al.* Root transcriptional dynamics induced by beneficial rhizobacteria
412 and microbial immune elicitors reveal signatures of adaptation to mutualists. *Plant J.* **93**,
413 166–180 (2018).
414 17. Bjornson, M., Pimprikar, P., Nürnberger, T. & Zipfel, C. The transcriptional landscape of
415 Arabidopsis thaliana pattern-triggered immunity. *Nat. Plants* **7**, 579–586 (2021).
416 18. Verbon, E. H. *et al.* Cell-type-specific transcriptomics reveals that root hairs and
417 endodermal barriers play important roles in beneficial plant-rhizobacterium interactions. *Mol.*
418 *Plant* **16**, 1160–1177 (2023).
419 19. Humphrys, M. S. *et al.* Simultaneous Transcriptional Profiling of Bacteria and Their Host
420 Cells. *PLOS ONE* **8**, e80597 (2013).
421 20. Marsh, J. W., Humphrys, M. S. & Myers, G. S. A. A Laboratory Methodology for Dual
422 RNA-Sequencing of Bacteria and their Host Cells *In Vitro*. *Front. Microbiol.* **8**, (2017).
423 21. Westermann, A. J., Barquist, L. & Vogel, J. Resolving host–pathogen interactions by
424 dual RNA-seq. *PLOS Pathog.* **13**, e1006033 (2017).
425 22. Donaldson, G. P. *et al.* Spatially distinct physiology of Bacteroides fragilis within the
426 proximal colon of gnotobiotic mice. *Nat. Microbiol.* **5**, 746–756 (2020).
427 23. Andrés-Barrao, C. *et al.* Coordinated bacterial and plant sulfur metabolism in

428 Enterobacter sp. SA187–induced plant salt stress tolerance. *Proc. Natl. Acad. Sci.* **118**,
429 (2021).
430 24. Vannier, N. *et al.* Genome-resolved metatranscriptomics reveals conserved root
431 colonization determinants in a synthetic microbiota. *Nat. Commun.* **14**, 8274 (2023).
432 25. Kumar, N. *et al.* Efficient Enrichment of Bacterial mRNA from Host-Bacteria Total RNA
433 Samples. *Sci. Rep.* **6**, 34850 (2016).
434 26. Nobori, T. *et al.* Transcriptome landscape of a bacterial pathogen under plant immunity.
435 *Proc. Natl. Acad. Sci.* **115**, E3055–E3064 (2018).
436 27. Roux, B. *et al.* An integrated analysis of plant and bacterial gene expression in symbiotic
437 root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**,
438 817–837 (2014).
439 28. Ma, Q., Bücking, H., Gonzalez Hernandez, J. L. & Subramanian, S. Single-Cell RNA
440 Sequencing of Plant-Associated Bacterial Communities. *Front. Microbiol.* **10**, (2019).
441 29. Johns, N. I. *et al.* Metagenomic mining of regulatory elements enables programmable
442 species-selective gene expression. *Nat. Methods* **15**, 323–329 (2018).
443 30. Jones, E. M. *et al.* A Scalable, Multiplexed Assay for Decoding GPCR-Ligand
444 Interactions with RNA Sequencing. *Cell Syst.* **8**, 254-260.e6 (2019).
445 31. Crook, N., Ferreiro, A., Condiotte, Z. & Dantas, G. Transcript Barcoding Illuminates the
446 Expression Level of Synthetic Constructs in *E. coli* Nissle Residing in the Mammalian Gut.
447 *ACS Synth. Biol.* **9**, 1010–1021 (2020).
448 32. Jores, T. *et al.* Synthetic promoter designs enabled by a comprehensive analysis of plant
449 core promoters. *Nat. Plants* (2021) doi:10.1038/s41477-021-00932-y.
450 33. Wang, G. *et al.* CRAGE enables rapid activation of biosynthetic gene clusters in
451 undomesticated bacteria. *Nat. Microbiol.* **4**, 2498–2510 (2019).
452 34. Wang, B. *et al.* CRAGE-Duet Facilitates Modular Assembly of Biological Systems for
453 Studying Plant–Microbe Interactions. *ACS Synth. Biol.* **9**, 2610–2615 (2020).

454 35. Tjaden, B. A computational system for identifying operons based on RNA-seq data.
455 *Methods* **176**, 62–70 (2020).
456 36. Winnen, B., Hvorup, R. N. & Saier, M. H. The tripartite tricarboxylate transporter (TTT)
457 family. *Res. Microbiol.* **154**, 457–465 (2003).
458 37. Brocker, M., Schaffer, S., Mack, C. & Bott, M. Citrate Utilization by Corynebacterium
459 glutamicum Is Controlled by the CitAB Two-Component System through Positive Regulation
460 of the Citrate Transport Genes citH and tctCBA. *J. Bacteriol.* **191**, 3869 (2009).
461 38. Pankievicz, V. C. S. *et al.* RNA-seq transcriptional profiling of Herbaspirillum
462 seropedicae colonizing wheat (Triticum aestivum) roots. *Plant Mol. Biol.* **90**, 589–603 (2016).
463 39. Nobori, T. *et al.* Dissecting the cotranscriptome landscape of plants and their microbiota.
464 *EMBO Rep.* **23**, e55380 (2022).
465 40. Yu, K. *et al.* Transcriptome Signatures in Pseudomonas simiae WCS417 Shed Light on
466 Role of Root-Secreted Coumarins in Arabidopsis-Mutualist Communication. *Microorganisms*
467 **9**, 575 (2021).
468 41. WOJTASZEK, P. Oxidative burst: an early plant response to pathogen infection.
469 *Biochem. J.* **322**, 681–692 (1997).
470 42. Ebel, J. & Mithöfer, A. Early events in the elicitation of plant defence. *Planta* **206**, 335–
471 348 (1998).
472 43. Danhorn, T. & Fuqua, C. Biofilm Formation by Plant-Associated Bacteria. *Annu. Rev.*
473 *Microbiol.* **61**, 401–422 (2007).
474 44. Rudrappa, T., Biedrzycki, M. L. & Bais, H. P. Causes and consequences of plant-
475 associated biofilms. *FEMS Microbiol. Ecol.* **64**, 153–166 (2008).
476 45. Callewaert, L. *et al.* A New Family of Lysozyme Inhibitors Contributing to Lysozyme
477 Tolerance in Gram-Negative Bacteria. *PLOS Pathog.* **4**, e1000019 (2008).
478 46. Yum, S. *et al.* Structural basis for the recognition of lysozyme by MliC, a periplasmic
479 lysozyme inhibitor in Gram-negative bacteria. *Biochem. Biophys. Res. Commun.* **378**, 244–

480 248 (2009).
481 47. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with
482 AlphaFold 3. *Nature* **630**, 493–500 (2024).
483 48. Liu, X. *et al.* Host-induced bacterial cell wall decomposition mediates pattern-triggered
484 immunity in Arabidopsis. *eLife* **3**, e01990 (2014).
485 49. Eitzen, K., Sengupta, P., Kroll, S., Kemen, E. & Doehlemann, G. A fungal member of the
486 Arabidopsis thaliana phyllosphere antagonizes Albugo laibachii via a GH25 lysozyme. *eLife*
487 **10**, e65306 (2021).
488 50. Lee, K.-M. *et al.* A Genetic Screen Reveals Novel Targets to Render Pseudomonas
489 aeruginosa Sensitive to Lysozyme and Cell Wall-Targeting Antibiotics. *Front. Cell. Infect.*
490 *Microbiol.* **7**, (2017).
491 51. Torrens, G. *et al.* Targeting the permeability barrier and peptidoglycan recycling
492 pathways to disarm Pseudomonas aeruginosa against the innate immune system. *PLOS*
493 *ONE* **12**, e0181932 (2017).
494 52. Flores-Cruz, Z. & Allen, C. Ralstonia solanacearum Encounters an Oxidative
495 Environment During Tomato Infection. *Mol. Plant-Microbe Interactions®* **22**, 773–782 (2009).
496 53. Phelan, J. P., Bourgeois, J. S., McCarthy, J. E. & Hu, L. T. A putative xanthine
497 dehydrogenase is critical for Borrelia burgdorferi survival in ticks and mice. *Microbiology* **169**,
498 001286 (2023).
499 54. Palluk, S. *et al.* De novo DNA synthesis using polymerase-nucleotide conjugates. *Nat.*
500 *Biotechnol.* **36**, 645–650 (2018).
501 55.Plesa, C., Sidore, A. M., Lubock, N. B., Zhang, D. & Kosuri, S. Multiplexed gene synthesis in
502 emulsions for exploring protein functional landscapes. *Science* **359**, 343–347 (2018).
503 56. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen
504 Helicobacter pylori. *Nature* **464**, 250–255 (2010).
505 57. Adiconis, X. *et al.* Comprehensive comparative analysis of 5′-end RNA-sequencing

506 methods. *Nat. Methods* **15**, 505–511 (2018).
507 58. De Weger, L. A. *et al.* Flagella of a plant-growth-stimulating Pseudomonas fluorescens
508 strain are required for colonization of potato roots. *J. Bacteriol.* **169**, 2769–2773 (1987).
509 59.Sivakumar, R. *et al.* Evaluation of INSeq To Identify Genes Essential for Pseudomonas
510 aeruginosa PGPR2 Corn Root Colonization. *G3 GenesGenomesGenetics* **9**, 651–661
511 (2019).
512 60. Gliese, N., Khodaverdi, V., Schobert, M. & Görisch, H. AgmR controls transcription of a
513 regulon with several operons essential for ethanol oxidation in Pseudomonas aeruginosa
514 ATCC 17933. *Microbiology* **150**, 1851–1857 (2004).
515 61. Choi, O. *et al.* Pyrroloquinoline Quinone Is a Plant Growth Promotion Factor Produced
516 by Pseudomonas fluorescens B16. *Plant Physiol.* **146**, 657–668 (2008).
517 62. Carreño-López, R., Alatorre-Cruz, J. M. & Marín-Cevada, V. Pyrroloquinoline quinone
518 (PQQ): Role in Plant-Microbe Interactions. in *Secondary Metabolites of Plant Growth*
519 *Promoting Rhizomicroorganisms: Discovery and Applications* (eds. Singh, H. B., Keswani, C.,
520 Reddy, M. S., Sansinenea, E. & García-Estrada, C.) 169–184 (Springer, Singapore, 2019).
521 doi:10.1007/978-981-13-5862-3_9.
522 63. Wang, X. *et al.* Elucidation of genes enhancing natural product biosynthesis through co-
523 evolution analysis. *Nat. Metab.* 1–14 (2024) doi:10.1038/s42255-024-01024-9.
524 64. Boo, A. *et al.* Synthetic microbe-to-plant communication channels. *Nat. Commun.* **15**,
525 1817 (2024).
526 65. Pini, F. *et al.* Bacterial Biosensors for in Vivo Spatiotemporal Mapping of Root Secretion.
527 *Plant Physiol.* **174**, 1289–1306 (2017).
528 66. Haskett, T. L. *et al.* Engineered plant control of associative nitrogen fixation. *Proc. Natl.*
529 *Acad. Sci.* **119**, e2117465119 (2022).
530 67.Toju, H. *et al.* Core microbiomes for sustainable agroecosystems. *Nat. Plants* **4**, 247–257
531 (2018).