A broad-host-range event detector: expanding and quantifying performance across bacterial species

Nymul Khan[1], Enoch Yeung[2], Yuliya Farris[1], Sarah J. Fansler[1], Hans C. Bernstein[1,3,4]*

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA; [2]Data Science and Analytics Group, Pacific Northwest National Laboratory, Richland, WA, USA; [3]The Gene and Linda Voiland School of Chemical Engineering and Bioengineering, Washington State University, Pullman, WA, USA; [4]Faculty of Biosciences, Fisheries and Economics, UiT - The Arctic University of Norway, Tromsø, Norway; [5]The Arctic Centre for Sustainable Energy, UiT - The Arctic University of Norway, Tromsø, Norway

***Correspondence**: Hans C. Bernstein, UiT Norges arktiske universitet, BFE-fak., Postboks 6050 Langnes, 9037 Tromsø, NORWAY*;Phone: +47 776 46 114; Emial:* Hans.C.Bernstein@uit.no

Running Title: Non-traditional synthetic biological hosts

Keywords: Integrase, Memory, Chassis, Pseudomonas, Escherichia coli Nissle

ABSTRACT

Modern microbial biodesign relies on the principle that well-characterized genetic parts can be reused and reconfigured for different functions. However, this paradigm has only been successful in a limited set of hosts, mostly comprised from common lab strains of *Escherichia coli*. It is clear that new applications – such as chemical sensing and event logging in complex environments – will benefit from new host chassis. This study quantitatively compared how a chemical event logger performed across multiple microbial species. An integrase-based sensor and memory device was operated by two representative soil Pseudomonads – *Pseudomonas fluorescens* SBW25 and *Pseudomonas putida* DSM 291. Quantitative comparisons were made between these two non-traditional hosts and two bench-mark *Escherichia coli* chassis including the probiotic Nissle 1917 and common cloning strain DH5α. The performance of sensor and memory components changed according to each host, such that a clear chassis effect was observed. These results were obtained via fluorescence from reporter proteins that were transcriptionally fused to the integrase and down-stream recombinant region and via data-driven kinetic models. The *Pseudomonads* proved to be acceptable chassis for the operation of this event logger and actually outperformed the common *E. coli* DH5α in many ways. This study advances an emerging frontier in synthetic biology that aims to build broad-host-range devices and understand the context by which different species can execute programmable genetic operations.

INTRODUCTION

Synthetic biology is built on the concept that complex biological behaviors can be programmed using relatively simple modules of biological parts. While the field of microbial biodesign has seen major advances, the overwhelming majority of parts have only been tested in model organisms. To date, we know little about how even our most standard genetic devices will perform in microbial hosts beyond common laboratory strains of *Escherichia coli* or *Saccharomyces cerevisiae*. This represents a major knowledge gap and limitation in the field. While useful for the development and demonstration of capabilities under stable laboratory conditions, these species do not survive well in many real-world applications. Most traditional microbial hosts have limited metabolic potential, preferring substrates such as simple sugars that are typically not available in environments relevant to the next generation of synthetic biology applications such as event detection within soils, built environments or the human gut. Therefore, programmable genetic devices must be expanded into new, non-traditional chassis that are already evolved to operate in complex, dynamic environments.

One of the most common biodesign principles is that well-characterized genetic parts – e.g., promoters, UTRs and transcription factors – can be reused and reconfigured to program different functions. Some of the benchmark examples are given by the toggle switch[1], repressilator[2] and previous demonstrations of integrase-based recording devices[3-6]; all of which were exclusively demonstrated in *E. coli*. These devices have laid the foundation for more applied microbial sensor-regulator-actuator devices that have been developed to detect/report signals from the mammalian gut[7-8] and chemical threats[9-10]; yet even these advanced examples relied solely on the genetic tractability of *E. coli*. Synthetic biologists are keen to harness new non-traditional hosts such as *Psuedomands*[11-13]; yet, successful transplantation of broad-host-range genetic devices across multiple bacterial species has remained elusive, until now.

Here we present a study that demonstrates how a relatively simple chemical event logger performs across multiple microbial hosts. We chose to comparatively quantify each component of an integrase-based sensor/memory device between two *Pseudomonas* species – *Pseudomonas fluorescens* SBW25 (*Pf*) and *Pseudomonas putida* DSM 291(*Pp*) – along with two more standard *Escherichia coli*

strains including the probiotic Nissle 1917 (*EcN*) and common cloning strain DH5α (*Ec*). The event detector was expressed from each host as the same sequence on an identical broad-host range expression vector. Here, we show that chemical event logging device can be ported across multiple species, including two Pseudomonads that open new chemical sensing/logging applications soil and plant-associated environments. The performance for each component of the device depended on the host – it was subject to a strong chassis effect. Hence, study presents a new broad-host range event logging system, which advances to a rapidly growing frontier in synthetic biology aimed at engineering devices that can function across multiple species and environments[14-15].

RESULTS

The broad-host range device and its components. A two-state chemical event logger was built and quantitatively compared across each host to determine the chassis effect on performance. The device was built with specific sensor and memory components (Fig. 1a). The sensor apparatus consisted of an IPTG inducible P_{lac} promoter driving the expression of the Bxb1 serine-integrase. The Bxb1 gene was transcriptionally fused to a green fluorescent protein (GFP) to monitor the sensor's output. The *lacI*^{q} transcription factor, controlling the induction of P_{lac}, was driven by the constitutive promoter, $P_{lacI}{}^{q}$. The memory element had two potential states that depended on the Bxb1 integrase. The initial state or "*off state*" was maintained be the constitutive P_{tac} promoter in the reverse orientation from its intended open reading frame, followed by a unique barcode DNA sequence. This construct was enclosed by the *attB* and an *attP* recombination sites recognized by Bxb1. The induced state or "*on state*" was controlled by the formation of a mature Bxb1 dimer, DNA binding and tetramer formation[16], and the respective recombination of *attB* and *attP*. This process re-oriented the constitutive P_{tac} to drive expression of a red fluorescent protein (RFP). A permanent digital memory output was stored by the orientation of the barcode. The performance of the sensor and logger elements were measured by the respective GFP and RFP signals. The entire device was built into a single contig and cloned into the broad-host-range vector pBBR1MCS2[17] (Fig. 1b).

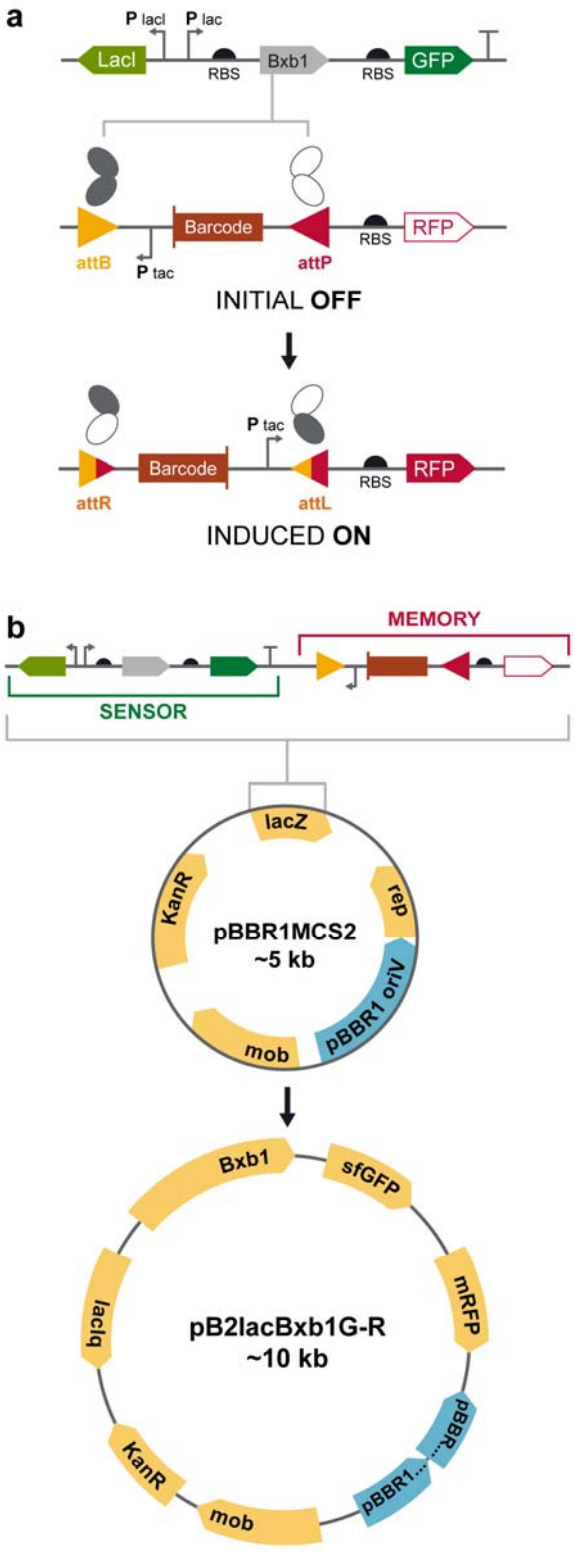


Figure 1. The broad-host range event logger and its modes of operation. (a) Expression of the Bxb1 integrase was controlled by the P_{lac} promoter and respective IPTG concentrations. Mature Bxb1 proteins dimerize and bind to DNA recombination sites (*attB* and *attP*). The dimers on the two ends of the recording element join to form a tetramer in a synaptic event that folds the DNA in the process. DNA strands are exchanged when Bxb1 monomers trade positions, flipping the internal region, which contains both a barcoded digital recorder and a constitutive P_{tac} promotor. The new sites, *attL* and *attR* – formed in the process of DNA flipping – can no longer stay attached to the Bxb1 dimers because of altered sequence, therefore release the dimers in an irreversible digital recording process. GFP and RFP reporter genes were transcriptionally fused onto the IPTG inducible sensor and recording recombination sites. **(b)** Map of the device with the sensor and memory components cloned in place of the *lacZ* gene on the *pBBR1MCS2* broad-host-range vector.

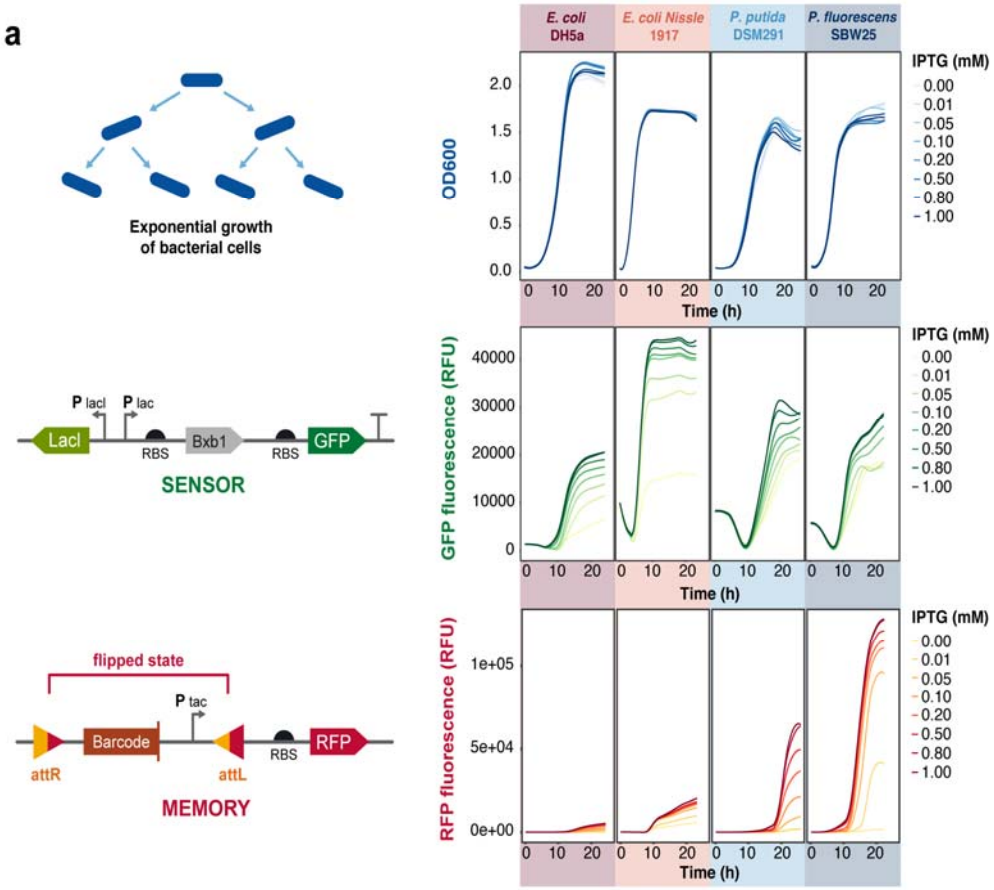
Quantifying the chassis effect. The device was operational across each of the four species. Performance of each component – sensor and logger – was assayed at eight different IPTG concentrations (0 – 1 mM) by measuring the respective mean GFP and RFP signals (Fig. 2). Total growth was also measured simultaneously via optical density (OD600 nm). The specific growth rates (μ) showed that each species – except *EcN* ($\mu = 1.087 \pm 0.017\ h^{-1}$) – had very similar growth under these conditions: *Ec* ($0.427 \pm 0.016\ h^{-1}$), *Pp* ($0.439 \pm 0.0436\ h^{-1}$) and *Pf* ($0.508 \pm 0.019\ h^{-1}$) (Supplemental Fig. S1a). The standard deviations represent the variation in growth rates across all IPTG treatments and ranged from approximately 1.5% to 10% of mean values. The induction strength of the device – controlled by IPTG concentration – showed no effect on the specific growth rate. While this indicates that there was little additional metabolic load with respect to IPTG induction, the plasmid encoded device itself imposed a significant metabolic burden on both *Pseudomonas* hosts. This was apparent from the wild-type growth rates of these species, measured at $0.759 \pm 0.017\ h^{-1}$ and $1.127 \pm 0.012\ h^{-1}$ for *Pf* and *Pp* respectively (Supplemental Fig. S1b); much higher than the respective engineered strains. In contrast, there was very little change in specific growth rate of the engineered *E. coli* hosts compared to their respective wild-types ($0.412 \pm 0.012\ h^{-1}$ and $0.979 \pm 0.022\ h^{-1}$ for *Ec* and *EcN* respectively). This was likely due to the fact that both *Pp* and *Pf* had considerably higher levels of RFP as compared to each *E. coli* host.

The sensor apparatus showed a higher dynamic response to IPTG induction when expressed from *E. coli* as compared to *Pseudomonas* hosts (Fig. 2a). Both *Pseudomonas* species had almost identical responses to IPTG and a higher degree of basal expression of GFP (measured at in the absence of IPTG) as compared to each *E. coli* host. We initially thought that this was due to the natural fluorescence of each *Pseudomonas* species near the 509 nm maximum emission of GFP. However, fluorescence measurements of the wildtype hosts proved too low to account for this anticipated basal activity. Hence, this was attributed to leaky expression P_{lac} within both *Pseudomonas* hosts. The dynamic ranges of the sensor component were much narrower in the *Pseudomonas* hosts (RFU_{max} within range of 10000) as compared to *Ec* and *EcN*, which exhibited ranges of 20000 and 45000, respectively. This was anticipated from the

design of the device as it used P_{lac}, which is optimized for tight transcriptional response and high dynamic ranges in *E. coli*. However, *Ec* – a common cloning strain and chassis for synthetic biology applications – showed the lowest performance with respect to operating the sensor component of the device as shown by the maximum GFP fluorescence achieved (RFU_{max} at 1 mM IPTG of about 20000) (Fig. 2a).

While the sensor portion of the device behaved predictably in each chassis, the memory apparatus – measured via RFP fluorescence – performed differently and showed a significant chassis effect. Interestingly, *Ec* was the lowest performing host from this study, both in terms of dynamic range and maximum fluorescence. In fact, it had a dynamic range and maximum RFU about 24 times lower than the best performing host, *Pf* (Fig. 2a). In general, the memory component performed much better in the *Pseudomonas* hosts as compared to each *E. coli* host; *Pp* and *Pf* achieved respectively 3- and 6-times the maximum RFP fluorescence than *EcN* (Fig. 2b).

Another important performance metric for cross-chassis comparison was the time scale of induction (Fig. 2b). This was quantified by the activation coefficient, which is the time for fluorescence to reach half maximum at a given IPTG concentration[18]. These had very different profiles than maximum relative fluorescence measurements. *EcN* showed the fastest induction time, but was largely flat; the half saturation time at 0.01 mM IPTG (6.17 ± 0.17 h) was similar to that at 1 mM IPTG (6.26 ± 0.32 h). For others, the activation coefficient decreased with increasing IPTG concentration and approached a minimum value at the highest IPTG concentrations. Minimum GFP half saturation times were 11 ± 0.17 h, 12.34 ± 0.096 h and 15.09 ± 0.22 h for *Pf*, *Ec* and *Pp* respectively at 1 mM IPTG. RFP fluorescence followed a similar pattern, corresponding directly to induction of P_{lac} by IPTG. This result is evidence that the expression of Bxb1 – thus P_{lac} strength – was the rate limiting step in the process from induction by IPTG to DNA flipping.



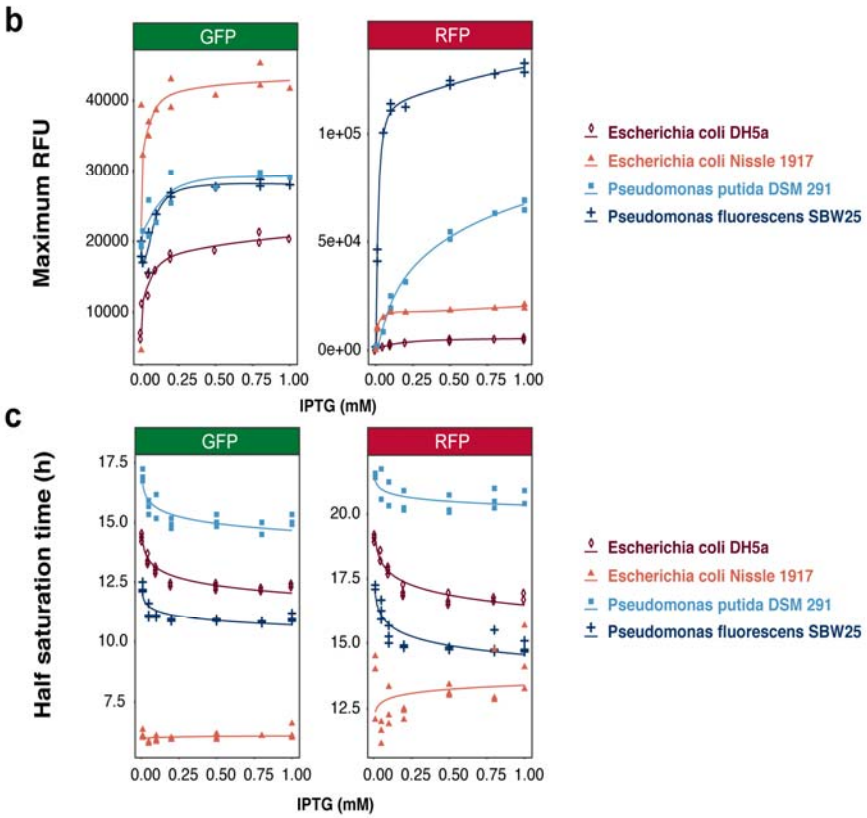


Figure 2. Quantifying the chassis effect. (a) The top four panels represent cell growth (OD600 nm), at different IPTG concentrations, of *E. coli* DH5α, *E. coli* Nissle 1917, *P. putida* DSM 291 and *P. fluorescens* SBW25 from left to right respectively. The middle row of graphs represents the fluorescence output of the GFP reporter that was transcriptionally fused to the sensor element; the bottom row represents the fluorescence output from the RFP that was fused to the memory component. **(b)** The maximum GFP and RFP fluorescence at different IPTG concentrations. **(c)** The activation coefficient shown as half-saturation times of GFP and RFP at different IPTG concentration (> 0 mM IPTG).

144

Population-based comparisons. Population-level measurements for the memory component of the device showed that while the *Pseudomonas* hosts exhibited stronger RFP output signals, the stability of the device was better in each *E. coli* chassis. This result was obtained via flow cytometry measurements taken from each host over five time points sampled over 26 h at five distinct IPTG concentrations. The results showed bimodal distribution of each host at the initial time point, prior to IPTG induction. This was likely a result from basal levels of RFP expression in a portion of the cells held in the 'off' state. However, while the 0 mM IPTG treatments showed this bimodal distribution remaining essentially constant for *Ec* and *EcN*, both *Pp* and *Pf* shifted significantly after ~12 h to favor an increasing number of cells fluorescing RFP signal (Fig. 3). This represents a false trigger of the event logger after extended periods of time and is consistent with the interpretation of leaky expression of the P_{lac} promoter observed by the *Pseudomonas*-specific GFP-Bxb1 transcriptional fusion signals (Fig. 2a). The maximum performance period of the memory device – as assayed by the time corresponding to maximum RFP signal from the flow cytometer measurements – was between 10 and 14 h for each host with the exception of *Pp*, which showed a peak at the 26 h measurement. This observation is consistent with stationary growth (post-log-phase), as observed from the optical density measurements. Hence the dampened RFP signal as well as the apparent device instability in *Pseudomonas* was only observed during stationary growth and/or slight losses in cell density by 20 h, as seen for *Pp* (Fig. 2a).

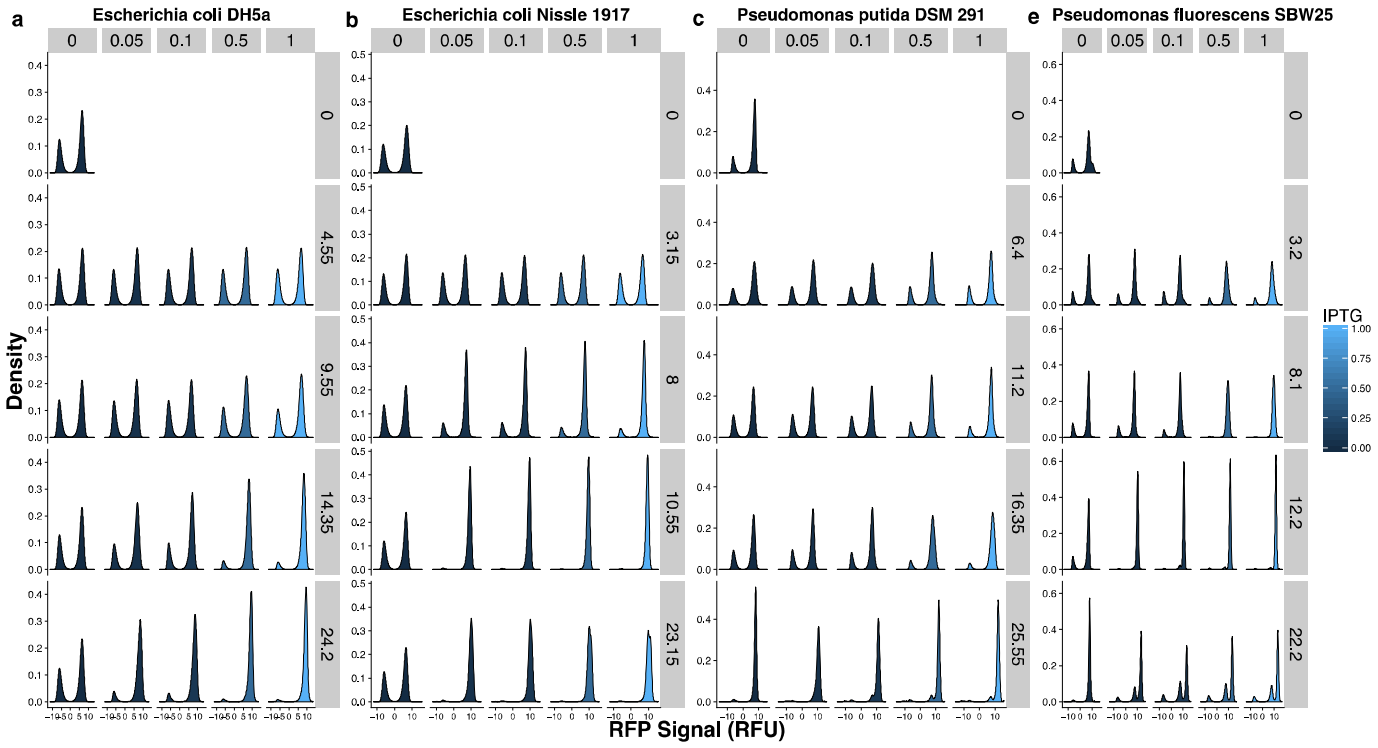


Figure 3. Population-based measurements of RFP reporting of signal recording across hosts. (a) *Escherichia coli* DH5α, **(b)** *Escherichia coli* Nissle 1917**, (c)** *Pseudomonas putida* DSM 291 and **(d)** *Pseudomonas fluorescence* SBW25**.** The rows in each of the panels correspond with sampling times in hours and columns are the IPTG concentrations given in mM. The first row in each plot shows data from the induction state from each experiment (0 h).

Simulations and performance metrics across chassis. A kinetic model was formulated to help quantify how individual components of the event logger performed across each chassis. Similar to the physical construction of the device the model was broken out into the respective sensor and memory component categories. The sensor part of model included expressions that accounted for IPTG induction of the Bxb1 integrase (given as *I*) and GFP as shown by Equations 1 and 2.

$$\frac{dI}{dt} = P - \mu * I - D * I \qquad \text{(Eq. 1)}$$

$$\frac{dGFP}{dt} = P - \mu * GFP - D * GFP \qquad \text{(Eq. 2)}$$

$$P = P_A + P_B ln([IPTG] + P_C) \qquad \text{(Eq. 3)}$$

The relationship between IPTG concentration and promoter activity is given by Equation 3, where P represents the activity of P_{lac}. It was found that this type of behavior adequately described the fluorescence output from the device. D is the protein degradation constant and μ is the specific growth rate, which accounts for dilution effects incurred by cell growth.

The memory component of the device was modeled by Equations 4 and 5, where PB is the fraction of un-flipped DNA and LR is the fraction of flipped DNA; k_{flip} is the rate constant for integrase-mediated recombination (flipping). We assumed that the plasmid copy number of each host was equivalent and that there existed an un-flipped induction state at the beginning of the experiment.

$$\frac{dPB}{dt} = -k_{flip} * PB * I^4 \quad \text{(Eq. 4)}$$

$$LR = 1 - PB \quad \text{(Eq. 5)}$$

The rate constant, k_{flip} encompasses three-time steps: 1) the time required for the integrase (I) to form a tetramer; 2) the time required for binding of the tetramer to the DNA (PB) and 3) the DNA flipping event. The overall readout from the memory component of the device was given by expression of RFP (Equation 6), which is analogous to Eq. 2 describing GFP, with the exception of the non-inducible *tac* promoter (P_{RFP}).

$$\frac{dRFP}{dt} = P_{RFP} - \mu * RFP - D * RFP \quad \text{(Eq. 6)}$$

This model adequately explained the operation of the genetic device and fitted distinct parameters for each respective host. This was especially evident by the degree to which the model could be fit to the GFP and RFP time series data; each respective output from the sensor and memory components of the device (Fig. 4ab). However, the model's ability to capture the dynamics of DNA flipping was variable between each of the hosts (Fig. 4c). We observed significant scatter derived from the qPCR assays that were

designed to measure the orientation of the barcoded DNA associated with the digital memory read-out. These data were collected to determine the fraction of flipped DNA. The noise in the measurement likely resulted from variability the plasmid recovery and purification from each host. Yet, the model still conveyed the overall the pattern for which IPTG induction instigated barcode flipping for each of the hosts and did a reasonably good job at fitting most of the data derived from each time series measurement.

Overall, the models helped show that the *Pseudomonas* hosts had favorable kinetics for operating this device despite the fact the genetic parts have been largely developed and optimized in *E. coli*. Surprisingly, the simulations showed that *Ec* – the bench mark chassis – had the most disfavorable kinetics. The probiotic strain, *EcN* showed the strongest ability to operate the sensor component of the device and was unique with respect to the suite of hosts tested in this study. This could be ascertained by combining the modeled predictions with experimental measurements. For instance, the simulated promoter strength of P_{lac} (see Eq. 3) showed that *Ec, Pp* and *Pf* are aligned with similar profiles, while the values of *P* for *EcN* were estimated to be about 4 times higher for any given concentration of IPTG. This result was consistent with independent and direct measurements of specific growth rates (Supplemental Fig. S1); *EcN* showed the fastest specific growth rate and should therefore have the highest dilution of expressed GFP protein leading to decreased fluorescence. Yet, *EcN* also showed the highest fluorescence. Thus, the strength of P_{lac} would need to be much higher to account for these opposing effects – as shown by the model.

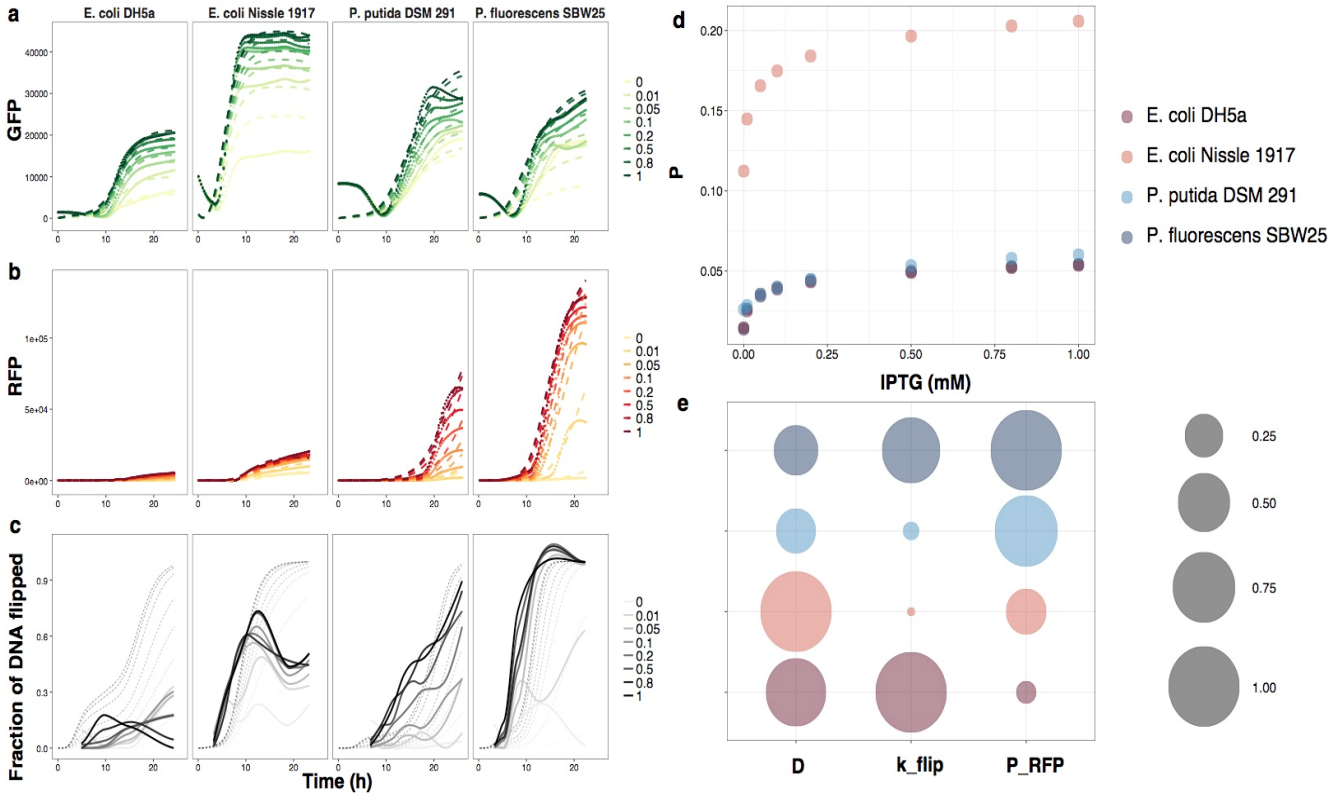


Figure 4. Comparative kinetics show chassis-dependent performance. The model outputs (dotted lines) of **(a)** the sensor component as fit by GFP fluorescence data; **(b)** the memory component as fit by RFP fluorescence; and **(c)** the memory component as fit by the fraction of DNA flipped (qPCR measurements). Each simulated time trace (panels a through c) are overlaid on the respective data typed used to parameterize the model (solid lines). **(d)** The promoter strength of P_{lac} (given in the model as P) calculated from estimated model parameters, P_A, P_B and P_C, plotted against IPTG concentration. **(e)** Comparison of simulated kinetic parameters for the protein degradation constant (D); flipping constant (k_{flip}) and strength of the constitutive P_{tac} promoter (P_{RFP}) compared across each host.

DISCUSSION

Synthetic biologists commonly (re-)discover that even the most well-characterized genetic parts often will not function in a predictable manner when taken out of the context from which they were originally characterized. For any given host, this unpredictability can arise from interference between genetic parts that have been introduced as well as cellular noise inherent to the native biological system[19-20]. Yet, the degree to which these factors are influenced by the biology of any given microbial species requires that the same genetic parts be used and compared across multiple hosts. Here, we showed that an identical genetic device can be ported across multiple microbial species and that its performance is host-dependent. We chose to deploy a relatively simple event logger and experimental design, which has enabled this

study to demonstrate emerging capability of broad-host-range genetic devices. While it is clear that more species and devices need to be tested before a broad-host-range capability matures, this early step is an important contribution towards alleviating our current dependency and limitations on very small subset of model microbes.

Despite the fact that *E. coli* DH5α is a common tool for the design and implementation of modern genetic devices, we found that in many ways it was the least ideal host tested in this study. In fact, a primary finding was that – compared to *Ec* – the two *Pseudomonas* species (*Pp* and *Pf*) showed reasonable potential as chassis for chemical event logging even though the majority of previously published reports on the parts used to build the device have only considered *E. coli*[3-4, 6, 21]. This is a promising result as these and closely related *Pseudomonads* are known to have tremendous metabolic potential for the synthesis of novel compounds[22-23], consuming complex substrates[24-25] and persisting in a wide range of habitats that include soils, plant tissues and marine ecosystems[26-29]. Expanding the synthetic parts list for *Pseudomonas* species will undoubtedly enable new biotechnological applications that should include chemical sensing and event logging in complex natural environments.

E. coli Nissle 1917 was also chosen for comparative analysis in this study because it served as an intra-species comparator to *Ec*. It was also chosen because of its growing importance in the biodesign community based on the fact that it is a commonly used probiotic[30] and highly genetically tractable. Researchers are rapidly uncovering many exciting opportunities to use *EcN* and other probiotic-hosts as programable therapeutic agents and/or diagnostic tools for human health[31-33]. In some cases, differences in intra-species performance – within *E. coli* strains – exceeded inter-species variability. This was somewhat unexpected and specifically evident from comparisons made on the sensor component of the device, which performed better in *EcN* as compared to *Ec* and both *Pseudomonas* hosts. This was specifically evident by comparing the kinetics associated with the sensor apparatus and indicates that *EcN* maintained the tightest control and largest dynamic ranges of the IPTG inducible components of the device.

Kinetic parameters estimated from the model provided a good quantitative comparison of biological properties that cannot be easily measured (Fig. 4e). For instance, the degradation constant, *D*, is

found to be fairly similar in all the hosts, which is hardly surprising considering the standardized growth conditions and similar growth rates. Estimates of the flipping rate constant (k_{flip}), however, were highly variable. In contrast to *D*, which is more indicative of cellular physiology, k_{flip} is more representative of the device-specific kinetics. This parameter depends on a number of biological factors such as codon usage, transcription, translation, protein folding as well as the efficiency of Bxb1-mediated recombination. Based on the model-enabled predictions, *EcN* stood out with a very small k_{flip} values; about 89 times smaller than the largest value attributed to *Pf*. Its high transcription rate (given by estimates of *P*) and low k_{flip} account for the observation of fast DNA flipping after initial induction followed by relatively immediate saturation (Fig. 4c). The k_{flip} values of Ec and *Pp* are moderate but the reason for the higher value of *Ec* relative to *Pp* is still somewhat uncertain since the fraction of DNA flipped is higher in *Pp* than *Ec*. The fourth parameter, P_{RFP}, is the measure of the strength of P_{tac} promoter and varies in the same way as RFP fluorescence. This promoter was actually found to work better – as assayed by the strength or RFP fluorescence – in the *Pseudomonads* than *E. coli* species.

Integrated data and kinetic modelling approaches are useful for quantifying and comparing performance across hosts. One limitation, however, was that our approach contained few species-specific physiological parameters. The exception to this is the specific growth rate (μ). Although the hosts in this study all showed similar growth rates, the specific growth rate should prove to be an important consideration when evaluating the performance of a device as hosts and growth conditions change. It was also interesting that we were able to observe and simulate dynamics in the device's performance while the cells were in stationary phase. Often, experimental observations made on engineered devices are only contextualized during log growth phases. However, future applications such as chemical event logging in dynamic environments will be better served by understanding how chassis/device pairs may function through lag, log and stationary phases of growth. This is a point that shall require more deserving attention in future studies.

The field of microbial biodesign is keen to harness new, non-traditional hosts for synthetic biology applications. Some significant advancements towards programing genetic devices – including

sensors – have already been shown in other non-traditional microbial hosts. Of specific note are previous success shown in a human gut microbe *Bacteroides thetaiotaomicron*[34] and a suite of proteobacteria isolated from a bee gut microbiome[35]. Here in this current study, we have advanced an emerging concept of broad-host-range genetic devices. While this is certainly a new frontier, some notable examples have preceded this current report including a study by Kushwana and Salis that that presented the concept of "portable power supplies" between species and demonstrated that some genetic parts can ported between *E. coli*, *P. putida* and *Bacillus subtilis*[14]. Another important avenue has been the pursuit of broad-spectrum genetic parts such as the promoters presented in a study from Yang et al. that are operational between *E. coli*, *B. subtilis* and *S. cerevisiae*[15]. The efforts to date – including our current study – have only considered a relatively small set of microbes. Future developments on cross-chassis devices may encounter new technical hurdles as the taxonomic diversity of hosts are expanded. Once harnessed, the concept of broad-host-range genetic devices should also bring new species-specific-applications. The major technical hurdle that will need to be overcome for developing chemical sensing capabilities will be the discovery or engineering of genetic components with specificity for analytes of real-world interest. The current suite of commonly used transcriptional factors and inducible promoters are clearly limited. New parts discovery and characterization efforts are sorely needed to advance the current state of microbial biodesign.

Conclusions. We quantified the chassis effect of an integrase-based chemical event logger across multiple species and two different Genera – *Pseudomonas* and *Escherichia*. The performance of sensor and memory components changed according to each host as ascertained via integrated experimental measurements and outputs from kinetic models. Specifically, *EcN* – a common probiotic bacterium – showed the tightest control and most stability of the sensor apparatus that regulated expression of the Bxb1 integrase. Both *Pseudomonas* hosts showed greater RFP output signals that corresponded with Bxb1-mediated recombination in the memory component of the device. A primary finding of this study was that – compared to *Ec* – the two *Pseudomonas* species (*Pp* and *Pf*) showed reasonable potential as

chemical event logging chassis. This study advances an emerging frontier in synthetic biology that aims to build broad-host-range devices and understand the context by which different microbial species can execute programmable genetic operations.

METHODS

Bacterial strains and cultivation. The bacterial strains used in this study includes *Escherichia coli* DH5α (New England Biolabs), *Escherichia coli* Nissle 1917 (isolated from probiotic Mutaflor capsule), *Pseudomonas fluorescens* SBW25 and *Pseudomonas putida* DSM 291 (DSMZ). All bacteria were cultured in Lauria-Bertani medium at 30 ^{o}C.

Transformation. *E. coli DH5α* was transformed using standard chemical transformation protocol. For the other species, electrocompetent cells were prepared as follows: overnight cultures were diluted 1:100 into 200 mL LB medium and grown to OD600 nm of about 0.3-0.4 (mid-log phase); cultures were harvested and spun down in four 50 mL centrifuge tubes at 5000 x g and the supernatant was discarded; cell pellets were resuspended in 15% glycerol, combined into one 50 mL centrifuge tube and collected via centrifugation at 5000 x g. This wash cycle was repeated twice and the final cell pellet was resuspended in 1 mL 15% glycerol for electroporation. The cells were then transformed by electroporation at 12500 V/cm (200 Ω and 25 μF) in 1 mm cuvettes. The entire protocol was successfully carried out at room temperature as outlined in the Tu et al. 2016 study[36]. The efficiency, in general, was found to be higher in the room-temperature methods than in conventional ice-cold methods.

Plate reader and cytometry assays. For each bacteria, three positive transformants were grown, passaged twice at 30 ^{o}C and then assayed in a 24-well plate at eight different IPTG concentrations (0, 0.01, 0.05, 0.1, 0.2, 0.5, 0.8 and 1 mM). Each well of the plate contained 1.8 mL LB + kanamycin (50 μg/mL) + IPTG. A Synergy H1 (Biotek, Winooski, VT) was used as the fluorescent plate reader for all assays. 1 μL samples were collected from each of the wells at various stages of growth and analyzed via

flow cytometry (Novocyte, ACEA biosciences, San Diego, CA). Simultaneously, 100 µL samples were also collected and frozen at -80 ^{o}C. Plasmid DNA was later extracted from the frozen samples and used for real time qPCR assays to measure the fraction of device flipped.

Real time quantitative PCR. Real time PCR was performed by ARQ Genetics LLC (Bastrop, TX) on the BioRad CFX384 Real Time System (BioRad, Hercules, CA) using assays specific for each plasmid. All of the plasmid DNA was extracted with a Zyppy – 96 Plasmid Miniprep kit (Zymo, Irvine, CA) following the next steps: all of the stains were grown at 30 °C and harvested at 3-5 h intervals; the DNA was quantified by performing PicoGreen assay on the Biotek Synergy H1 (Biotek, Winooski, VT) and reactions were diluted to matching concentrations. Each reaction well contained 5 µL of TaqMan Universal Master Mix II (Applied Biosystems, Waltham, MA), 2 µL of each sample template and 0.5 µL of each specific plasmid assay in a reaction volume of 10 µL. Cycling conditions were as follows: 95 °C for 10 min for polymerase activation, followed by 40 cycles of 95 °C for 15 seconds and 63 °C for 1 min. Data analysis was performed using CFX Manager software from BioRad, version 3.1. The experimental Cq (cycle quantification) was calibrated against the standard curve for each plasmid orientation.

Numerical simulation. The system of ODEs were solved numerically using the 'deSolve' package[37] in R[38]. The results were fitted to experimental data to estimate the five kinetic rate constants (P_A, P_B, P_C, D and k_{flip}) for each of the species in the model. The specific growth rate (μ) was calculated from measured OD600 nm at each time point and given as an input to the numerical solver.

Reproducible analysis and data deposits. The raw data for this study along with and the sequence information for the genetic construct pB2lacBxb1G-R (shown in Fig. 1b) are available from the Open Science Framework (OSF) under the name "A broad-host-range event detector: data and models" (DOI 10.17605/OSF.IO/J295C) at https://osf.io/j295c/. This OSF project also contains the R Markdown scripts

356 that can be used to reproduce all of the analyses, statistics and technical graphs presented in this
357 manuscript.
358
359 **CONFLICT OF INTEREST**
360 The authors have no conflict of interest to declare.
361
362 **ACKNOWLEDGMENTS**
363 This research was supported by funding from the National Security Directorate Seed Initiative, a
364 Laboratory Directed Research and Development (LDRD) Program of Pacific Northwest National
365 Laboratory (PNNL). PNNL is operated for the DOE by Battelle under contract no. DE-AC05-76RLO-
366 1830. Specific acknowledgments are given to Rose Perry at PNNL and John Repass at Genetics for
367 technical assistance with graphics and real time PCR. The authors would also like to thank Drs. Victoria
368 Hsiao and Richard Murray for kindly sharing materials and knowledge.

REFERENCES CITED

1. Gardner, T. S.; Cantor, C. R.; Collins, J. J., Construction of a genetic toggle switch in Escherichia coli. *Nature* **2000,** *403* (6767), 339.
2. Elowitz, M. B.; Leibler, S., A synthetic oscillatory network of transcriptional regulators. *Nature* **2000,** *403* (6767), 335.
3. Hsiao, V.; Hori, Y.; Rothemund, P. W.; Murray, R. M., A population-based temporal logic gate for timing and recording chemical events. *Mol Syst Biol* **2016,** *12* (5), 869.
4. Yang, L.; Nielsen, A. A.; Fernandez-Rodriguez, J.; McClune, C. J.; Laub, M. T.; Lu, T. K.; Voigt, C. A., Permanent genetic memory with> 1-byte capacity. *Nat Meth* **2014,** *11* (12), 1261.
5. Siuti, P.; Yazbek, J.; Lu, T. K., Synthetic circuits integrating logic and memory in living cells. *Nat Biotech* **2013,** *31* (5), 448.
6. Shur, A.; Murray, R. M., Proof of concept continuous event logging in living cells. *bioRxiv* **2018**, 225151.
7. Riglar, D. T.; Giessen, T. W.; Baym, M.; Kerns, S. J.; Niederhuber, M. J.; Bronson, R. T.; Kotula, J. W.; Gerber, G. K.; Way, J. C.; Silver, P. A., Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. *Nat Biotech* **2017,** *35* (7), 653.
8. Kotula, J. W.; Kerns, S. J.; Shaket, L. A.; Siraj, L.; Collins, J. J.; Way, J. C.; Silver, P. A., Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc Nat Acad Sci* **2014,** *111* (13), 4838-4843.
9. Wang, B.; Barahona, M.; Buck, M., A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosens Bioelectron* **2013,** *40* (1), 368-376.
10. Stocker, J.; Balluch, D.; Gsell, M.; Harms, H.; Feliciano, J.; Daunert, S.; Malik, K. A.; Van der Meer, J. R., Development of a set of simple bacterial biosensors for quantitative and rapid measurements of arsenite and arsenate in potable water. *Environ Sci Technol* **2003,** *37* (20), 4743-4750.
11. Tan, S. Z.; Reisch, C. R.; Prather, K. L., A Robust CRISPRi Gene Repression System in Pseudomonas. *J Bacteriol* **2018**, JB. 00575-17.
12. Nikel, P. I.; Martínez-García, E.; De Lorenzo, V., Biotechnological domestication of pseudomonads using synthetic biology. *Nat Rev Microbiol* **2014,** *12* (5), 368.
13. Nikel, P. I.; Chavarría, M.; Danchin, A.; de Lorenzo, V., From dirt to industrial applications: Pseudomonas putida as a synthetic biology chassis for hosting harsh biochemical reactions. *Curr Opin Chem Biol* **2016,** *34*, 20-29.
14. Kushwaha, M.; Salis, H. M., A portable expression resource for engineering cross-species genetic circuits and pathways. *Nat Commun* **2015,** *6*, 7832.
15. Yang, S.; Liu, Q.; Zhang, Y.; Du, G.; Chen, J.; Kang, Z., Construction and characterization of broad-spectrum promoters for synthetic biology. *ACS Synth Biol* **2017,** *7* (1), 287-291.
16. Ghosh, P.; Pannunzio, N. R.; Hatfull, G. F., Synapsis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites. *J Mol Biol* **2005,** *349* (2), 331-348.
17. Kovach, M. E.; Elzer, P. H.; Hill, D. S.; Robertson, G. T.; Farris, M. A.; Roop, R. M.; Peterson, K. M., Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **1995,** *166* (1), 175-176.
18. Alon, U., *An introduction to systems biology: design principles of biological circuits*. CRC press: 2006.
19. Hooshangi, S.; Thiberge, S.; Weiss, R., Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Nat Acad Sci* **2005,** *102* (10), 3581-3586.
20. Lou, C.; Stanton, B.; Chen, Y.-J.; Munsky, B.; Voigt, C. A., Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat Biotech* **2012,** *30* (11), 1137.

21. Siuti, P.; Yazbek, J.; Lu, T. K., Engineering genetic circuits that compute and remember. *Nat Protoc* **2014,** *9* (6), 1292.
22. Poblete-Castro, I.; Becker, J.; Dohnt, K.; Dos Santos, V. M.; Wittmann, C., Industrial biotechnology of Pseudomonas putida and related species. *Appl Microbiol Biotechnol* **2012,** *93* (6), 2279-2290.
23. Di Gioia, D.; Luziatelli, F.; Negroni, A.; Ficca, A. G.; Fava, F.; Ruzzi, M., Metabolic engineering of Pseudomonas fluorescens for the production of vanillin from ferulic acid. *J Biotechnol* **2011,** *156* (4), 309-316.
24. Clarke, P. H., The metabolic versatility of pseudomonads. *Anton Van Leeuw* **1982,** *48* (2), 105-130.
25. Nelson, K.; Weinel, C.; Paulsen, I.; Dodson, R.; Hilbert, H.; Martins dos Santos, V.; Fouts, D.; Gill, S.; Pop, M.; Holmes, M., Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environ Microbiol* **2002,** *4* (12), 799-808.
26. Rainey, P. B.; Bailey, M. J., Physical and genetic map of the Pseudomonas fluorescens SBW25 chromosome. *Molec Microbiol* **1996,** *19* (3), 521-533.
27. Silby, M. W.; Cerdeño-Tárraga, A. M.; Vernikos, G. S.; Giddens, S. R.; Jackson, R. W.; Preston, G. M.; Zhang, X.-X.; Moon, C. D.; Gehrig, S. M.; Godfrey, S. A., Genomic and genetic analyses of diversity and plant interactions of Pseudomonas fluorescens. *Genome Biol* **2009,** *10* (5), R51.
28. Compeau, G.; Al-Achi, B. J.; Platsouka, E.; Levy, S., Survival of rifampin-resistant mutants of Pseudomonas fluorescens and Pseudomonas putida in soil systems. *Appl Environ Microb* **1988,** *54* (10), 2432-2438.
29. Isnansetyo, A.; Kamei, Y., Bioactive substances produced by marine isolates of Pseudomonas. *J Ind Microbiol Biot* **2009,** *36* (10), 1239-1248.
30. Schultz, M.; Burton, J., Escherichia coli Nissle 1917. In *The Microbiota in Gastrointestinal Pathophysiology*, Elsevier: 2017; pp 59-69.
31. Slomovic, S.; Pardee, K.; Collins, J. J., Synthetic biology devices for in vitro and in vivo diagnostics. *Proc Nat Acad Sci* **2015,** *112* (47), 14429-14435.
32. Danino, T.; Prindle, A.; Kwong, G. A.; Skalak, M.; Li, H.; Allen, K.; Hasty, J.; Bhatia, S. N., Programmable probiotics for detection of cancer in urine. *Sci Transl Med* **2015,** *7* (289), 289ra84-289ra84.
33. Sedlmayer, F.; Fussenegger, M., Synthetic biology: A probiotic probe for inflammation. *Nat Biomed Eng* **2017,** *1* (7), 0097.
34. Mimee, M.; Tucker, A. C.; Voigt, C. A.; Lu, T. K., Programming a human commensal bacterium, Bacteroides thetaiotaomicron, to sense and respond to stimuli in the murine gut microbiota. *Cell Syst* **2015,** *1* (1), 62-71.
35. Leonard, S. P.; Perutka, J.; Powell, J. E.; Geng, P.; Richhart, D. D.; Byrom, M.; Kar, S.; Davies, B. W.; Ellington, A. D.; Moran, N. A., Genetic engineering of bee gut microbiome bacteria with a toolkit for modular assembly of broad-host-range plasmids. *ACS Synth Biol* **2018,** *7* (5), 1279-1290.
36. Tu, Q.; Yin, J.; Fu, J.; Herrmann, J.; Li, Y.; Yin, Y.; Stewart, A. F.; Müller, R.; Zhang, Y., Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency. *Sci Rep* **2016,** *6*, 24648.
37. Soetaert, K.; Petzoldt, T.; Setzer, R. W., Solving differential equations in R: package deSolve. *J Stat Software* **2010,** *33*.
38. Team, R. C., R: A language and environment for statistical computing. **2013**.

FIGURE LEGENDS

Figure 1. The broad-host range event logger and its modes of operation. (a) Expression of the Bxb1 integrase was controlled by the P_{lac} promoter and respective IPTG concentrations. Mature Bxb1 proteins dimerize and bind to DNA recombination sites (*attB* and *attP*). The dimers on the two ends of the recording element join to form a tetramer in a synaptic event that folds the DNA in the process. DNA strands are exchanged when Bxb1 monomers trade positions, flipping the internal region, which contains both a barcoded digital recorder and a constitutive P_{tac} promotor. The new sites, *attL* and *attR* – formed in the process of DNA flipping – can no longer stay attached to the *Bxb1* dimers because of altered sequence, therefore release the dimers in an irreversible digital recording process. GFP and RFP reporter genes were transcriptionally fused onto the IPTG inducible sensor and recording recombination sites. **(b)** Map of the device with the sensor and memory components cloned in place of the *lacZ* gene on the *pBBR1MCS2* broad-host-range vector.

Figure 2. Quantifying the chassis effect. (a) The top four panels represent cell growth (OD600 nm), at different IPTG concentrations, of *E. coli* DH5α, *E. coli* Nissle 1917, *P. putida* DSM 291 and *P. fluorescens* SBW25 from left to right respectively. The middle row of graphs represents the fluorescence output of the GFP reporter that was transcriptionally fused to the sensor element; the bottom row represents the fluorescence output from the RFP that was fused to the memory component. **(b)** The maximum GFP and RFP fluorescence at different IPTG concentrations. **(c)** The activation coefficient shown as half-saturation times of GFP and RFP at different IPTG concentration (> 0 mM IPTG).

Figure 3. Population-based measurements of RFP reporting of signal recording across hosts. (a) *Escherichia coli* DH5α, **(b)** *Escherichia coli* Nissle 1917**, (c)** *Pseudomonas putida* DSM 291 and **(d)** *Pseudomonas fluorescence* SBW25**.** The rows in each of the panels correspond with sampling times in hours and columns are the IPTG concentrations given in mM. The first row in each plot shows data from the induction state from each experiment (0 h).

Figure 4. Comparative kinetics show chassis-dependent performance. The model outputs (dotted lines) of **(a)** the sensor component as fit by GFP fluorescence data; **(b)** the memory component as fit by RFP fluorescence; and **(c)** the memory component as fit by the fraction of DNA flipped (qPCR measurements). Each simulated time trace (panels a through c) are overlaid on the respective data typed used to parameterize the model (solid lines). **(d)** The promoter strength of P_{lac} (given in the model as P) calculated from estimated model parameters, P_A, P_B and P_C, plotted against IPTG concentration. **(e)** Comparison of simulated kinetic parameters for the protein degradation constant (D); flipping constant (k_{flip}) and strength of the constitutive P_{tac} promoter (P_{RFP}) compared across each host.