



## Supplementary Information for Massively Parallel Screening of Synthetic Microbial Communities

Jared Kehe, Anthony Kulesa, Anthony Ortiz, Cheri Ackerman, Sri Gowtham Thakku, Daniel Sellers, Seppe Kuehn, Jeff Gore, Jonathan Friedman, Paul C. Blainey

Corresponding author: Paul C. Blainey  
Email: pblainey@broadinstitute.org

### This PDF file includes:

- Supplementary text
  - Section S1 to S17
  - Supplementary Discussion
- Figs. S1 to S26
- Tables S1 and S2
- Captions for Datasets S1 and S2 (attached separately)
- References for SI reference citations

### Other supplementary materials for this manuscript include the following:

Datasets S1 and S2

## **Supplementary Information Text**

This supplementary text contains extended technical descriptions of all methods and their validation. It also contains detailed descriptions pertaining to the setup of the screen, methods of analysis, interpretation of results, and additional discussion.

### **Section S1. kChip design and fabrication**

All kChips possessed the following features: (1) An array of microwell geometries of one or more predetermined values of  $k$ , the number of droplets each microwell was designed to receive, *e.g.* a full set of identical microwells like  $k = 2$  (“Full kChip”) (as described in the experiments of **Fig. 2**) or varied microwells containing  $k = \{1:7;19\}$  (as described in the screen of **Fig. 3**); (2) internal posts within these microwells (**Fig. S2, S3**) designed to (a) control the number of droplets entering a microwell by reducing overfilling (via droplets squeezing into a microwell) and underfilling (via droplets exiting microwells due to the oil flow associated with kChip loading), and (b) inhibit the entry of large droplets inherent to the droplet pool (*i.e.* a low-pass size filter); (3) a series of 30 90- $\mu\text{m}$  deep moat-like engravings designed to trap small droplets (*i.e.* a high-pass size filter) (**Fig. S2D**); and (4) a loading slot into which droplets are injected via micropipette (**Fig. S1, S2D, Section S5**).

All kChips were designed in AutoCAD (Autodesk). Designs began with a hexagonal array of triangles that would ultimately become internal posts (**Fig. S2A, S3A**). The optimal spacing between posts (*i.e.* the diameter of the space encircled by posts) was determined based on the choice of medium and concentration of fluorosurfactant (RAN Biotech 008 FluoroSurfactant), which were shown to affect the size of droplets produced by a droplet generator (Bio-Rad QX200) (**Section S2**), and by extension, droplet grouping and merging performance. For droplets of minimal medium (MM) made with 2% w/w fluorosurfactant, which produced droplets of  $135 \pm 5 \mu\text{m}$  diameter, this spacing parameter was determined to be 148.2  $\mu\text{m}$  (**Fig. S2A**).

A microwell boundary was then drawn to enclose a subset of posts in accordance with the desired  $k$  (**Fig. S2B**). These geometries were modular in their expansion: With each incorporation of two additional posts via enlargement of this enclosure, the grouping capacity increased by one droplet. With internal posts included, microwell droplet grouping appeared agnostic to  $k$  for  $k = \{1:7\}$  (>90% grouping as desired) (**Fig. S3B**). The generalizability of the approach was maintained for  $k = 19$  microwells, with only a small decrease in grouping performance. We posit that a geometry to capture  $k$  droplets could be inductively generated for any value of  $k$  by enclosing posts in this manner. By comparison, we saw a strong drop-off in grouping performance with  $k$  if no internal posts were included (<50% grouping correctly for  $k \geq 4$  with a strong dependence on  $k$ ) (**Fig. S3B**).

Microwells were arrayed with 50- $\mu\text{m}$  inter-microwell spacing (**Fig. S2C**). Because different  $k$  microwells have different sizes, the density of the arrays also depended on  $k$  (**Fig. 1B**). A user-defined organization of different microwells within the kChip boundaries (6.2 cm x 7.2 cm) defined the final kChip structure (**Fig. S2D**). All kChips were also equipped with a series of 30 90- $\mu\text{m}$  wide moat-like engravings (“slots”). The slots served as a high-pass size filter by trapping small droplets. Droplets around the desired size (~135  $\mu\text{m}$ , **Fig. S2A**, **Fig. S24**) remained unaffected. Given the deformability of droplets, the conservative cutoff of 90  $\mu\text{m}$  was chosen for the filtering size to ensure that no droplets of desired size were trapped. The slots were spaced 50  $\mu\text{m}$  apart from each other and 400  $\mu\text{m}$  from the onset of the microfluidic field. The slots were inset 3 mm from the edge of the kChip due to the observation that air bubbles can enter slots that extend to the edge of the kChip.

Photomasks were generated from AutoCAD designs (FineLine Imaging). kChip designs were then fabricated to 110-120  $\mu\text{m}$  feature height using photolithography on silicon wafers (Microchem SU8-2050). Microwells produced from this feature height were found to best trap droplets in a monolayer, as deeper features can allow droplets to stack causing loading of an undesired number of droplets. These wafers were then embedded into custom molds to create PDMS (Dow Corning Sylgard) kChips by soft lithography with consistent thickness (0.635 cm) and droplet-loading slot location and size. The side of the kChip that

contained microwell features was then coated with 1.5  $\mu$ m parylene C by vapor deposition (Paratronics) to inhibit water loss from droplets and stiffen the kChip to prevent interior collapse during droplet loading (**Section S5**).

## Section S2. Microbial culture input preparation

All bacterial cultures underwent an initial “starter phase”, whereby glycerol stocks of non-fluorescent environmental isolates (**Section S10**) and fluorescently labeled strains (**Section S8**) were inoculated into 525  $\mu$ L (2-mL-deep 96-well plate) and 4 mL (15-mL culture Falcon tube) of Lysogeny broth (LB) medium, respectively (30°C, 220 RPM, 16 hr). Inoculations from glycerol stocks were conducted via pin replicator (sterilized via 70% v/v ethanol bath and heat treatment between inoculations).

A subsequent “preculture phase” (30°C, 220 RPM, 24 hr) began with washing all cultures in our custom minimal medium (MM) two times. Cultures were then normalized to a starting optical density (OD<sub>600</sub>) of 0.01 in MM + 0.5% w/v glucose.

MM consisted of 1X M9 salts (Teknova), 1X trace metals (Teknova), 0.1 mM calcium chloride, and 2 mM magnesium sulfate. We additionally added 0.05% w/v bovine serum albumin (BSA) to MM to improve the retention of fluorescent dyes used in color codes and, presumably, other small molecules (1) (**Section S4**). This is necessary because droplets are not solid compartments, and the fluorosurfactant at the interface forms reverse-micelles in the oil phase. It is a well known phenomenon that hydrophobic small molecules, such as the fluorescent dyes used for color coding, can partition into these reverse-micelles such that they are depleted from the droplets (2, 3). We found that the quality of color code signals was improved in MM droplets when 0.05% w/v BSA was added (with no apparent improvements using 1% w/v BSA). We additionally measured the growth of a panel of seven fluorescent strains on panel of five carbon sources with and without BSA. Growth curves appeared highly similar when 0.05% w/v BSA or 0.5% w/v BSA was added (**Fig. S24A**). Notably, instances in which a culture was unable to grow on a given carbon source were maintained despite the addition of BSA (*i.e.* BSA caused no false positives by serving as an alternative

carbon source). We therefore included 0.05% w/v BSA when making MM droplets. We additionally found that BSA affected the size of droplets produced (Bio-Rad QX200 cartridges). We generated a custom flow cell apparatus with small flow channels to enable simultaneous assessment of droplet sizes produced from various media and fluorosurfactant concentrations (**Fig. S24B, C**). Droplets were labeled with Alexa Fluor 647 to enable detection. We found that (1) increasing fluorosurfactant from 0.5% w/w to 2% w/w increased droplet size across both LB and MM; (2) MM droplets were inherently more polydisperse than LB droplets; and (3) the incorporation of 0.05% w/v BSA decreases droplet size. These results informed kChip microwell dimensions (**Section S1**), with geometries chosen to accommodate droplets with 135- $\mu\text{m}$  diameter (**Fig. S2A**), *i.e.* the size of MM droplets made with 0.05% w/v BSA and 2% w/w fluorosurfactant.

The “experimental phase” began by washing cells three times in MM to remove residual glucose and normalizing, typically, to a starting OD<sub>600</sub> of 0.02 (or ~20 cells/droplet depending on the strain) in MM + 0.05% BSA. After adding the color codes (**Section S4**), droplets of the normalized cultures were produced, pooled, and loaded onto the kChip (**Section S5**).

We found that increases in the microwell size and the number internal posts as  $k$  increased produced no gross effects on the growth rate (**Fig. S3C**) or final yield (**Fig. S3D**) of the model organism *H. frisingense* used in our screen (**Section S12**). We detected minimal evaporation comparing droplets at 24 and 72 hrs (**Fig. S25**), and expect this has little effect on microbe growth given our observation that different microbes reach saturation in droplets and 96-well plate cultures on similar timescales (**Fig. S5, S6**).

### Section S3. Carbon source input preparation

A bank of kChip-deployable growth substrates was developed from which libraries could be drawn for use in screening (**Table S2**). Carbon compounds in this bank met the following criteria: (1) the compounds were soluble at 2% w/v in water; (2) the solutions were emulsifiable using Bio-Rad QX200 cartridges; (3) the integrity of the color code signals were maintained despite the presence of the carbon compound (or impacts could be mitigated by careful selection of the fluorescent dye concentrations).

#### **Section S4. kChip input preprocessing**

If droplets containing microbial cultures were to be generated, cell cultures were first normalized to the desired starting density in MM + 0.5% w/v BSA (**Section S2**). If droplets containing carbon sources were to be generated, 2% w/v carbon source in water was prepared and sterilized 24-48 hrs in advance (**Section S3**). Droplets could also be generated containing both a microbial culture and carbon source prepared in this manner (as in the Hf-GFP facilitation screen outlined in **Fig. 3A**).

In order to identify the contents of a given droplet, every unique input (*e.g.* a strain or environmental condition) received a “color code”, or unique ratio of three fluorescent dyes (standardized to a total final dye concentration of 1  $\mu$ M or 10  $\mu$ M)—prior to generating droplets (**Fig. S4**). Each set of three dyes collectively labeled each specific input. These three dyes were chosen among Alexa Fluor 488, Alexa Fluor 555, Alexa Fluor 594, and Alexa Fluor 647, all of which have distinct excitation and emission spectra (**Section S7**). The selection of fluorescent dyes ensured that interference with any fluorescence-based assay signals (such as GFP, YFP, or resorufin) was avoided (**Section S8, Section S9**). The broad agreement in substrate-specific growth between droplets (which included color codes) and 200- $\mu$ L cultures in 96-well plates (which had no color codes) (**Fig. 2C, S6**) indicated that inclusion of the encoding dyes did not confer any gross growth defects on the 10 strains tested.

#### **Section S5. Droplet making and kChip loading**

After microbial inputs were normalized (**Section S2**) and/or carbon sources prepared (**Section S3**), all inputs were preprocessed, *i.e.* color codes and BSA were added as necessary (**Section S4**). Droplets were then produced on a Bio-Rad QX200 Droplet Generator (which generated roughly 20,000 ~1-nL emulsifications prepared per 20  $\mu$ L input for eight inputs at a time; three minutes per 8-input cartridge). The continuous phase was a fluorocarbon oil (3M Novec 7500). For droplet making, 2% w/w fluorosurfactant (RAN Biotech 008 FluoroSurfactant) was added to stabilize droplets.

Droplets were pooled in equal proportions to prepare a total of 200  $\mu$ L of droplet suspension, or ~200,000 droplets (*e.g.* for a set of 16 inputs,  $200/16 = 12.5 \mu$ L of droplet suspension of each input would be pooled and mixed via micropipette), which was then injected into a custom built kChip loading apparatus in a single dispensing step (**Fig. S1**). The loading apparatus consisted of two acrylic pieces. The bottom piece held in place a piece of custom-cut glass (Brain Research Laboratories; 1.2 mm thickness) made hydrophobic via pretreatment with Aquapel. The top side of the kChip, which was not coated with parylene (**Section S1**), spontaneously formed a seal with the top piece of acrylic. Four neodymium magnet pairs were oriented such that the two acrylic pieces repelled each other. Working against this repulsive force, the top piece was lowered toward the bottom piece via tightening nuts until the desired standoff between the glass and kChip was attained (~500-700  $\mu$ m) to create a space for flow under the microwells (**Fig. S1C**). Via a slot in the top acrylic piece and kChip, the flow space was pre-wetted with an injection of oil (~3 mL to fill the entirety of the flow space) followed by the pooled droplets. Buoyant in the surrounding oil, the droplets were distributed around the flow space via tilting the loading apparatus. Droplet entry into microwells above was a random process whereby each microwell sampled a set of  $k$  droplets (**Section S6**). After the droplets had passed through the flow space and entered microwells, additional oil (no fluorosurfactant) was flushed through the device to wash away excess droplets and fluorosurfactant. The wingnuts were tightened further to bring the kChip into contact with the glass and limit inter-microwell crosstalk (4).

The kChip was scanned initially at 2X magnification to identify the droplets in each microwell from their color codes (**Fig. S4, Section S7**). Droplets were then merged within their microwells via exposure to an alternating current (AC) electric field (4.5 MHz, 10,000-45,000 volts) underneath the glass. The field was generated by corona treater (Electro-Technic Products), the tip of which was moved around the glass for ~10 seconds. Without application of the electric field, spontaneous merging of droplets was rare (detected as incorrectly loaded microwells in **Fig. S3**). The kChip was imaged subsequently in accordance with the relevant optical or fluorescent assay. Growth assays tested on the kChip included

monitoring constitutive fluorescent protein expression (**Section S8**), reduction of resazurin to fluorescent resorufin (**Section S9**), and phase contrast microscopy (**Supplemental Discussion – Platform challenges and limitations, Fig. S27**).

Setting up the kChip loading apparatus in preparation to receive droplets took 5-10 minutes and was completed ahead of time. The remaining setup time was ~30 minutes: Droplet making took ~3 minutes per eight inputs on the Bio-Rad QX200, droplet pooling and mixing took ~5 minutes, loading the kChip took ~5-10 minutes, and scanning the kChip took ~12-15 minutes.

## Section S6. kChip loading statistics

The appropriate number of distinct inputs  $n$  can be calculated prior to loading a kChip to ensure the desired number of replicates of a distinct composition of  $s$  droplets is attained (with a full breakdown of the realized screen size and degree of replication outlined in **Dataset S2**). The number of replicates is determined by the following factors: (1) the total number of distinct inputs,  $n$ ; (2) the number of droplets per microwell,  $k$ ; (3) the desired number of inputs per composition,  $s$  (where  $s$  is  $\leq k$ ); and (4)  $N_k$ , the total number of observed microwells filling with  $k$  droplets (**Fig. S14A**). In some cases it is desirable that  $s$  be lesser than  $k$ . For example, to measure the robustness of a pairwise effect for the composition [A + B] ( $s = 2$ ) to additional inputs in the library, all instances where the set [A + B] appears with any given additional input(s) for the communities [(A + B) + X + Y + ...] ( $k \geq 3$ ) can be measured.

We derived a formula for the probability  $p$  that all  $k$  droplets in a given microwell were unique and that a given desirable composition of size  $s$  was present among those  $k$  (**Fig. S14B**). We also considered the case where  $s = k$ , *i.e.* the desirable subset composition is the same as the number of droplets loaded per microwell, for which  $p$  reduces to  $n!/n^k$ . From here, the total expected number of replicates was determined by multiplying  $p$  by the number of observed microwells  $N_k$ .

For library sizes  $n = 8, 16, 25$ , and  $50$ , we plotted the sampling probability and mean number of replicates obtained for one of our  $k = \{1:7;19\}$  Chips for desired subset compositions  $s = 1, 2, 3$ , and  $k$  (**Fig.**

**S14B**). Notably, for the Hf-GFP facilitation screen conducted presently ( $n = 16$ ), we predicted and confirmed (**Fig. S13, Dataset S2**) that a given composition of three strains  $s = 3$  would be represented  $\sim 5$  times on average in  $k = 3$  microwells and increasingly among all communities at the higher values of  $k$ . This greater representation allowed us to inspect the robustness of three-input compositions to the presence of additional inputs in  $k \geq 4$  microwells.

From **Fig. S14C** (bottom-most plot), it is apparent that the probability of sampling a given composition at high  $k$  decreases quickly, especially for large input libraries (e.g. for a given  $k = 7$  combination and a library size of 25 inputs, there are  $\sim 2.6\text{e}6$  possible configurations, and the probability of sampling a given combination of all unique inputs approaches 0.00001). This small probability is further exacerbated by the relatively larger footprint of a  $k = 7$  microwell that results in fewer microwells per kChip compared to smaller values of  $k$  (e.g.  $\sim 13,000$   $k = 7$  microwells takes up the same amount of space as  $\sim 60,000$   $k = 1$  microwells). While one way to overcome the sampling issue is to use more kChips to generate more combinations (**Supplemental Discussion – Platform challenges and limitations**), the primary reason undersampling occurs at higher  $k$  is due to the exponential increase in combinatorial space (which has little to do with the sub-linear decrease in microwell density with  $k$ ). **Fig. S14D** (right panel) plots the expected number of replicates per Full kChip. From these plots, it's clear that to achieve  $>1$  replicate on average per combination on a single  $k = 7$  Chip, a library of 25 inputs is far too large. However, sampling among composition subsets  $s < k$  remains high for relatively small  $s$ , even for large library sizes. To achieve 20 replicates on average per  $s = 3$  combination among  $k = 7$  microwells, for example, a library of  $\sim 40$  inputs could be used. To achieve 1 replicate on average, a library of  $>100$  inputs could be used. This sort of experimental set up is useful for identifying combinations, e.g. facilitative communities consisting of 2 or 3 isolates, when the experimenter wishes to test the robustness of the facilitative effect to the presence of additional, randomly sampled microbes from the library (the analysis described in **Fig. 3D-E**).

In practice, it may be advantageous to work with kChips with an assortment of different microwell types, as we did in our screen, or kChips composed entirely of single microwell type. For each of these

cases, we calculated the maximum allowable library size to attain the desirable mean number of replicates per kChip (**Fig. S14D**).

### Section S7. Fluorescence microscopy

All fluorescence microscopy was performed using a Nikon Ti-E inverted fluorescence microscope with fluorescence excitation by a Lumencor Sola light emitting diode illuminator (100% power setting). Images were taken across up to four fluorescence channels—three for the color codes (**Section S4**) and one additional channel for fluorescence-based assays (**Section S8**, **Section S9**). Each dye (or assay signal) was detected with a different excitation wavelength generated by a collection of excitation filters (**Fig. S4**): Alexa Fluor 488 dye (or GFP or YFP expression, **Fig. 2A**) by Semrock GFP-1828A (blue excitation); Alexa Fluor 555 dye (or resazurin/resorufin, **Fig. 2B**) by Semrock SpGold-B (green excitation); Alexa Fluor 594 dye by Semrock FF03-575/25-25 [excitation filter] + FF01-615/24-25 [emission filter] (yellow excitation); and Alexa Fluor 647 dye (or *C. reinhardtii* autofluorescence, **Fig. S26**) by Semrock LF635-B (red excitation). At the image analysis stage (**Section S17**), the emission signals corresponding to each dye channel were used to identify the contents of a given droplet within each droplet grouping prior to droplet merging (**Fig. S4**). The final channel was used post-merge and at subsequent time points to quantify the assay signal.

Images were collected by a Hamamatsu ORCA-Flash 4.0 CMOS camera (exposure times range 50ms – 500ms) and 2X optical magnification (with 2X pixel binning resulting in 6.5  $\mu\text{m}/\text{pixel}$  resolution). The total scanning time for a single kChip was 12-15 minutes.

### Section S8. Fluorescently labeled microbe assay in droplets

A panel of strains constitutively expressing a fluorescent protein (GFP or YFP; either plasmid-mediated or genome integrated) was obtained (**Table S1**). We also acquired an autofluorescent eukaryotic alga (*Chlamydomonas reinhardtii* CC-503) (**Fig. S26**). We demonstrated agreement between carbon

utilization profiles for cultures in droplets and standard 96-well plate bulk cultures (SpectraMax plate reader) (**Fig. 2C, S6**). The fluorescent dyes Alexa Fluor 555, 594, and 647 were used to avoid overlap with the GFP or YFP excitation channel (*i.e.* Alexa Fluor 488 was excluded from the encoding set) (**Section S7**). For experiments involving the *C. reinhardtii* CC-503, which was autofluorescent for red excitation, the fluorescent dyes Alexa Fluor 488, 555, and 594 were used (*i.e.* Alexa Fluor 647 was excluded from the encoding set). Multiple fluorescent organisms can be monitored simultaneously. For example, we monitored *C. reinhardtii* in co-culture with GFP-labeled *E. coli* (**Fig. S26**) (for which both Alexa Fluor 488 and Alexa Fluor 647 were excluded from the encoding set).

Any kChip screen could be designed whereby the growth of a fluorescent strain or set of fluorescent strains is screened against biotic backgrounds (*i.e.* microbial communities) (**Section S2**), abiotic backgrounds (*e.g.* drug combinations, carbon source combinations) (**Section S3**), or a combination thereof.

### Section S9. Resazurin assay in droplets

In the presence of respiring microbes, the blue indicator dye resazurin is reduced to the pink, fluorescent dye resorufin (5), and therefore constitutes a colorimetric and fluorescent indicator of metabolic activity. Here we measure levels of resorufin dynamically as an assay for cell growth to characterize the growth of a given strain across an assortment of carbon sources (**Fig. 2B, S20**). In some cases, resorufin further reacts to yield non-fluorescent products, which results in a drop in resorufin fluorescence at later times.

To assess how well the resazurin assay in droplets recapitulated 96-well plate growth as assayed with the conventional OD<sub>600</sub> readout (**Fig. S8**), two sets of droplets were created. The first set of droplets (three types in total) contained microbial monocultures with no carbon source (normalized to OD<sub>600</sub> = 0.01, such that, upon merging within  $k = 2$  microwells, starting OD<sub>600</sub> = 0.005 in MM + 0.05% w/v BSA). The second set of droplets (four types in total) was generated that contained single carbon sources (1% w/v) with 80  $\mu$ M resazurin (such that, upon merging within  $k = 2$  microwells, the final carbon source

concentration was 0.5% w/v and the final resazurin concentration was 40  $\mu$ M). The fluorescent dyes Alexa Fluor 488, Alexa Fluor 594, and Alexa Fluor 647 were used to minimize interference with the excitation spectrum of resorufin (**Section S7**). These dyes were added at 10  $\mu$ M, rather than 1  $\mu$ M, to mitigate interference from the absorbance of the fluorescent dye signals by resazurin (at 80  $\mu$ M). Upon merging of droplets in  $k = 2$  microwells, strains were thereby contacted with both a carbon source and the resazurin. If the carbon source enabled growth, a fluorescent signal corresponding to the accumulation of resorufin was detected.

Cultures were kept at 21°C, and we measured their yield measured up to 50 hr. For each strain, measurements were normalized by first subtracting the background fluorescence of each carbon source (in a merged droplet containing only the carbon source and no strain), and then subtracting background resorufin fluorescence of each strain (in a merged droplet containing only the strain but no carbon source; this might be attributable to metabolic activity that is unrelated to utilization of the carbon source added to the media, *e.g.* carbon stored intracellularly). Resorufin signals were then normalized for each strain by the maximum across all carbon sources and time points.

## **Section S10. Environmental microbe isolation**

Soil samples (two ~10 cm columns of topsoil, ~1cm in diameter) were collected from Middlesex Fells Reservation in Somerville, MA (2:20pm, November 12, 2017, 2:20pm, 5.6°C). These samples were diluted in PBS within a few hours of collection (5 g soil of each vortexed in 40 mL PBS). Single strains were first isolated from streaking 70  $\mu$ L of dilutions of this mixture ( $10^{-1}$ ,  $10^{-2}$ ,  $10^{-3}$ , and  $10^{-4}$ ) on 20 different solid (agar) media (Tryptic Soy Broth (TSB) (Bacto), 1% v/v TSB, Lysogeny Broth (LB) (Bacto), 1% v/v LB, Nutrient Broth (NB) (Bacto), 1% v/v NB, M9 salts (Sigma-Aldrich) + 0.5% w/v glucose, M9 salts + 0.005% w/v glucose, M9 salts + 0.005% w/v glucose + 0.2% w/v casamino acids, M9 salts + 0.005% w/v glucose + 0.002% w/v casamino acids, M9 salts + 0.5% w/v glucose + 0.2% w/v casamino acids at pH = 4 and 5, M9 salts + 0.005% w/v glucose + 0.002% w/v casamino acids at pH = 4 and 5, Actinomycete

Isolation Agar (Teknova), Brucella Agar (Teknova), Streptomyces Medium (Teknova), Campylobacter Medium (Teknova), Bordatella Medium (Teknova), and ATCC Medium 1111 (Teknova)). Strains were selected based on the following criteria: growth in LB liquid medium of transferred colony (25°C), frozen glycerol stock revival in LB ( $OD_{600} > 0.1$ ) (30°C), and subsequent growth on M9 + 0.5% w/v glucose ( $OD_{600} > 0.1$ ) (30°C). We kept a 96-well working plate of isolates (LB, 25% v/v glycerol) stored in -80°C.

Isolates included in the Hf-GFP facilitation screen (**Fig. S9**) were selected based on robust revival from glycerol stocks and subsequent culturing steps (as outlined in **Section S2**) with prioritization of more phylogenetically distant strains (**Section S11**).

### **Section S11. 16S sequencing and phylogenetic assignment**

Sequences of the 16S rRNA gene were obtained by Sanger sequencing. Clustal X with penalizations for gap opening and gap extension of 12 and 6 was used to align the sequences (6). PhyML-SMS with default parameters was used to select GTR as the best model of nucleotide substitution, to infer the tree, and to get bootstrap support values (7). Taxonomic classifications and labels used in the phylogenetic tree (**Fig. S9**) were obtained by selecting the hit with highest S\_ab score obtained from Seqmatch (8). *Sulfolobus solfataricus*, a thermophilic archaea, was used as an outgroup species to root the tree.

### **Section S12. *H. frisingense* facilitation screen**

Microbial cultures were prepared as described in **Section S2**. The culture medium for *H. frisingense* revival from a glycerol stock also included the selection antibiotic kanamycin (30 µg/mL). Six carbon sources (galactose, glucose, fructose, raffinose, lactose, sucrose) were prepared as described in **Section S3** to a concentration of 2% w/v, which were added 1:1 with MM + 0.05% w/v BSA to produce 1X MM + 1% w/v carbon source (six unique MM each with one carbon source). A plate of 16 color codes were also prepared in advance at 50 µM total concentration (**Section S4**).

At the onset of the screen, all cultures were washed three times and resuspended in MM + 0.05% w/v BSA (no carbon source) and normalized to an initial  $OD_{600} = 0.08$ . A “droplet plate” (96-well plate) was prepared whereby each of 16 wells, corresponding to the 16 input conditions of the screen, received the following four elements: (1) 50  $\mu L$  MM + 1% w/v carbon source; (2) 25  $\mu L$  unique soil isolate at  $OD_{600} = 0.08$  (or a no-isolate control or a negative control); (3) 25  $\mu L$  Hf-GFP at  $OD_{600} = 0.08$ ; and (4) 2  $\mu L$  50  $\mu M$  color code. The two controls, a negative control and a no-isolate control, were in principle the same (just different color codes), with the negative control used to measure Hf-GFP monocultures and the no-isolate used to calculate isolate-mediated differences. The carbon source and color code were mixed in advance, though the cultures were added just before making droplets to load on a kChip.

After emulsification, each 1-nL droplet contained the following: (1) 0.5% w/v carbon source; (2) soil isolate at  $OD_{600} = 0.02$  (~20 cells) (or no-isolate control or negative control); (3) Hf-GFP at  $OD_{600} = 0.02$  (~20 cells); and (4) a set of color code (total concentration of all three fluorescent dyes at 1  $\mu M$ ). Droplets were prepared for each carbon source at a time and loaded onto a  $k = \{1:7;19\}$  Chip, *i.e.* first all sucrose-containing droplets were prepared and loaded onto the first kChip in the screen, then all lactose-containing droplets were prepared and loaded onto the second kChip, and so forth. Because there were only 16 inputs in total, there were no instances where all inputs were unique in  $k = 19$  microwells. We focused our analysis on the  $k = 1-7$  microwells, and specifically on instances where all droplets within a given microwell were unique. Throughput tradeoffs associated with this number of inputs and  $k$  values are discussed in **Supplemental Discussion – Platform challenges and limitations**.

After loading droplets, the kChip was imaged to read the color-code of each droplet (“pre-merge” scan) and droplet sets were merged within their corresponding microwells (**Section S5**). The post-merge communities contained the following: (1) 0.5% w/v carbon source; (2)  $k = 1, 2, \dots, 7$ , or 19 soil isolates collectively at  $OD_{600} = 0.02$  (or no-isolate control or negative control); (3) Hf-GFP at  $OD_{600} = 0.02$ ; and (4) 1  $\mu M$  mixed color codes (no longer serving a purpose). (All analyses in **Fig. 3** and **Fig. 4** were conducted only considering microwells filled with isolate-containing droplets and no control droplets, as summarized

in **Dataset S2**, such that the initial isolate OD<sub>600</sub> was always 0.02. That is, the final relative density of [Hf-GFP]:[total isolate] was 1:1, *i.e.* Hf-GFP made up about half of the initial biomass in the community and the total isolate content made up the remainder (**Fig. 3A**). The kChips were kept at 21°C and imaged (“post-merge” scans) at 24 hr, 48, and 72 hr (**Section S8**). A breakdown of the total number of data points is shown in **Dataset S2**.

Following the loading of each kChip in the screen, a  $k = 2$  Chip was subsequently loaded with isolate-only droplets and carbon source-only droplets with resazurin (pre-merge concentration 1% w/v carbon and 80  $\mu$ M resazurin; post-merge concentration 0.5% w/v carbon and 40  $\mu$ M resazurin) (**Section S9**). This kChip was imaged every 30 minutes for 24 hours and again at 48 and 72 hrs, enabling estimates of each isolate’s growth rate on each carbon source (**Fig. S20**).

### **Section S13. Identification of highly facilitative and robust compositions**

Two scores were used to calculate the effect size of a composition on Hf-GFP yield (“Hf-GFP median yield”) and the robustness of a composition’s effect to the presence of additional isolates (“Hf-GFP robustness”), respectively (**Fig. 3D, 3E**). To calculate “Hf-GFP median yield,” all instances of a composition appearing  $\geq 30$  times across all carbon sources were identified. This cutoff was based on the confidence with which Hf-GFP yield could be measured at different degrees of replication (where the analysis was conducted using the data collected for the experiment in **Fig. 2C**). To study the effect of different levels of replication on the reliability of our results, we downsampled data from **Fig. 2C**, and compared measurements of Hf-GFP yield in downsampled bootstrapped samples at different levels of replication against estimates using the entire dataset. We found that at 5 replicates (or, the average number of measurements per kChip when a composition appeared 30 times in the entire screen), the measurement of Hf-GFP yield corresponded strongly with that of the all replicates ( $R^2 = 0.938$ , **Fig. S12, S13**). In total, 191 compositions appeared  $\geq 30$  times and were composed of  $k = 1$  (14/14 possible combinations = 100%),  $k = 2$  (91/91 possible combinations = 100%), or  $k = 3$  isolates (86/384 possible combinations = 23.6%).

With 4.5 replicates on average per unique  $k = 3$  composition (**Dataset S2**), we expect 27 replicates on average per  $k = 3$  community across the six chips and  $111/374 = 30.6\%$  of these to be represented  $>5$  times (assuming a Poisson model centered at 27), just slightly more than what we achieved.

The median Hf-GFP yield at 72 hrs in the presence of each of these 191 compositions (across all carbon sources) was calculated (**Fig. 3E**, “Hf-GFP median yield”). These values were compared to a baseline “minimal viable yield” value of 1500 counts (one standard deviation above Hf-GFP yield in sucrose), the point at which yield was considered detectable.

To calculate “Hf-GFP robustness”, all communities where each composition appeared along with additional isolates were identified across all carbon sources. Here, we differentiate between “composition” as the exclusive subset of isolates under consideration and “communities” as all combinations including this composition and  $\geq 1$  isolate. For example, a given 3-isolate composition [A + B + C] ( $s = 3$ ) appeared as part of larger communities [A + B + C + X + ... + Z] ( $k \geq 4$ ) (**Fig. 3D**) with a predictably high number of instances (**Fig. S14**). The tenth percentile of Hf-GFP yield at 72 hr across communities containing each given composition was calculated (**Fig. 3E**, “Hf-GFP robustness”).

Compositions were uncovered that produced a wide range of facilitative effects (**Fig 3B, E**), but few performed highly by the Hf-GFP robustness score. The single isolate BuC imparted both a high facilitative effect and high robustness (**Fig. 3E**). Notably, BuC was the only isolate in our library able to grow on all six carbon sources tested (**Fig. S20, S23A**). Among compositions that included BuC, *e.g.* [BuC + Ch + Dy], we further observed small improvements to robustness (**Fig. 3E, S23B**). The composition BuC alone was the least robust of BuC-containing compositions.

The pair of isolates [BaL + Ra] also appeared among compositions where Hf-GFP yield was high and robust (**Fig. 3E, S23C**). Each carbon source provided a growth substrate for at least one of these two isolates (**Fig. S23A**). We found that the compositions BaL alone and Ra alone were both less robust than the composition [BaL + Ra] (**Fig. S23C**). We found instances where the incorporation of a third isolate, *e.g.*

in the composition [BaL + Ra + Ps], improved the magnitude and robustness of Hf-GFP facilitation (**Fig. 3E, S23C**).

We posit that robust compositions will obey the property that each carbon source can be consumed by at least one isolate in the composition, *e.g.* the composition BuC alone and the composition [BaL + Ra] (**Fig. S23A**). This may be a necessary but insufficient criterion, as compositions obeying this property may not impart high robustness. We further posit the property that specific isolates incorporated beyond such “core” groups, *e.g.* the composition [(BuC) + Ch + Dy] and the composition [(BaL + Ra) + Ps], can bolster the effect size and/or robustness further. These properties point to a design strategy for operating in very large combinatorial spaces whereby robust compositions are first identified from a set rationally selected hypothetical core groups, and then [(core group) + additional isolate(s)] can be screened for further improvements.

#### **Section S14. Validation of facilitative effects**

To validate results from the screen in larger-scale culture, cultures of Hf-GFP, BuC, BaL, and Ra were generated as described in **Section S2** (except only 4-mL culture volumes were used in the “starter phase” and “preculture phase”). The compositions Hf-GFP, Hf-GFP + BuC, Hf-GFP + BaL, Hf-GFP + Ra, and Hf-GFP + [BaL + Ra] were constructed in duplicate in 200  $\mu$ L MM containing one of each of the six carbon sources used in the screen (0.5% w/v) (as well as an even mix of all carbon sources (total 0.5% w/v) and a no-carbon control) in 96-well plates (21°C, 220 RPM). Hf-GFP and the isolates were normalized to standard starting densities [Hf-GFP OD<sub>600</sub> = 0.02]:[total isolate OD<sub>600</sub> = 0.02]. The GFP signal was monitored over five days (SpectraMax plate reader) and displayed a strong correspondence with the screening results in terms of carbon source specificity and relative size of the facilitative effect (**Fig. S16**) for all compositions and carbon sources tested.

#### **Section S15. *H. frisingense* yield vs. number of isolates analysis**

The effect of the number of unique isolates in co-culture with Hf-GFP was estimated (**Fig. 4A, S17**). For a given carbon source, the distributions of Hf-GFP yield (a.f.u.) across  $k$  values were first generated by considering all unique compositions at each value of  $k$  (*i.e.* at  $k = 3$ , composition [A + B + C] was considered, but both [A + B + B] and [A + B + C + C] were not). Medians were calculated in instances when compositions appeared  $>1$  time (and a mean was calculated if the composition was represented only 2 times). At  $k = 1$ , this amounted to 14 unique compositions (Hf-GFP in co-culture with one isolate) in the distribution; at  $k = 2$ , this amounted to  $14\text{-choose-}2 = 91$  compositions; and so forth. As  $k$  increased, the fraction of compositions represented in the data decreased relative to the total number of possible combinations, *e.g.*  $\sim 300$  compositions at the  $k = 7$  level were generated for each carbon source, even though  $14\text{-choose-}7 = 3,432$  were possible. The number of data points generated in screen as it relates to the total number of possible combinations the shown in **Dataset S2**.

To estimate Hf-GFP yield values, the distribution of yields for each  $k$  was resampled with replacement (with resampling count equal to the actual sampling count), and a median of the resampled data was calculated. This exercise was performed 100 times in each instance and the distribution the medians calculated was reported (**Fig. 4A, S17**).

Hf-GFP yield was also measured without the constraint of uniqueness of the isolates at each  $k$  (*i.e.* at  $k = 3$ , the composition [A + A + B] was included in the analysis) (**Fig. S18**). As a result, the communities with 19 inputs were included, even though there were only 14 unique isolates in the library. From this analysis, we predict that Hf-GFP yield does not change drastically with community richness beyond 7 strains (relative to the change observed between 1 and 7 strains), although this was not assessed directly.

## Section S16. Physical abiotic considerations

Control of physical environmental factors like temperature and illuminance can be incorporated into a kChip screen. To demonstrate how illuminance can be incorporated into these experiments, we screened the growth of the photoautotroph (*Chlamydomonas reinhardtii* CC-503, a naturally fluorescent

alga) and a heterotroph (*Escherichia coli* constitutively expressing GFP) in co-culture on a kChip under an array of neutral density filters (Lee Filters Gel Sheet 209 and 210) (**Fig. S26A, B**). Each filter attenuated light to a different extent. We measured the yield of the two organisms in co-cultures at 50 hr (**Fig. S26C, D**), showing the impact of initial density and light intensity on the abundance of the two organisms (**Fig. S26E**).

kChip screening can be conducted across a broad range of experimental temperatures. Droplet evaporation within the kChip, which occurs to greater extents as temperature and time increase, may limit the maximum assay duration. In the Hf-GFP facilitation screen, kChips were incubated at 21°C, and we observed only minor droplet evaporation by 72 hr (**Fig. S25**). The *C. reinhardtii/E. coli* co-culture experiment was conducted at 30°C.

### Section S17. kChip image analysis

We developed an image analysis pipeline to: (1) identify droplets as circular objects within the image; (2) decode the contents of each droplet based on the color code; (3) assign each droplet to a microwell; and (4) measure the average fluorescence of the merged droplets in each microwell.

To detect each droplet in the image, we average all fluorescence channels and apply a circular Hough transform (scikit-image) to detect circular objects with a diameter of 100–140 µm. To decode each droplet’s color code, we measure the average fluorescence intensity of each dye (a 3-dimensional vector for each droplet). These vectors are then projected onto a two-dimensional plane, eliminating systematic effects from differences in illumination intensity across the images. The DBSCAN algorithm (scikit-learn) identifies the clusters of droplets corresponding to each input condition. Optionally, a user can correct clustering errors, such as cluster collisions caused by optical activity of reagents in the input library. A quality score for each droplet can be computed based on the distance to the assigned cluster centroid. The Hungarian algorithm (scikit-learn) then maps each cluster to the predetermined centroids of each dye mixture. Pre-determined centroids can typically be set by *a priori* dye ratios.

Once identified and decoded, we compute the optimal alignment between [the centroid positions of all droplets] and [the photomask design used to construct a given kChip]. After alignment, we infer that droplets share a microwell if their centroids aligned to the same microwell in the photomask. Once droplet sets are assigned to microwells, we filter out microwells with merged droplet areas exceeding  $\pm 30\%$  of the mean merged area for a given  $k$  as an additional data quality control to account for instances where there was incomplete merging of all droplets within the microwell. For each microwell that passes this filter, we then measure the average fluorescence (*e.g.* fluorescent protein or resorfun reporter) across the merged droplet area at each timepoint.

### **Supplemental Discussion – Platform challenges and limitations**

The random combinatorial construction and optical readout of kChip assays involve several design tradeoffs.

First, the construction of all the combinations happens spontaneously in a single step, but the random nature of the approach causes dispersion in the number of replicates for each assay condition, necessitating a statistical approach to covering the desired assay at a given level of replication. The resulting throughput/replication tradeoff becomes limiting for larger combinations, *e.g.* for  $k \geq 4$  in the Hf-GFP facilitation screen, given the number of microwells available on our devices (**Fig. 3, S13**). This effect is due to the super-geometric increase in the number of possible combinations as  $k$  increases (for  $k \ll n$ ). **Section S6** discusses how to calculate the expected replication.

To aid in experimental design, **Fig. S14** plots the expected replication for many different scenarios based on the size of the input library (values of  $n$ ), the number of microwells, and the microwell size(s) (values of  $k$ ) for a given screen. For example, in the case of the Hf-GFP facilitation screen, we used a library of 16 inputs ( $n = 16$ ) and kChips designed with arrays of  $k = 1, 2, \dots, 7$ , or 19 inputs per microwell (with the number of microwells per kChip given in **Fig. 1B**). Based on these numbers, we calculate the number of expected replicates for each  $k$  (red plots in right column in **Fig. S14C**). For  $s = k = 1, 2, 3$ , and 4, the

average number of expected replicates per combination (where all inputs were unique) are  $>100$ ,  $\sim 20$ ,  $\sim 5$ , and  $<1$ , respectively. A library of  $>16$  inputs generates fewer  $s = k$  replicates at each  $k$  (orange and purple plots in **Fig. S14C**), placing an upper bound on the library size to achieve a given number of replicates per combination given a fixed number of available microwells.

While a smaller library size increases the number of replicates per combination (*e.g.* for a library of 8 inputs, blue plots in Fig. **S14C**), the screening efficiency is reduced if certain combinations are considered undesirable. For example, in the Hf-GFP facilitation screen, we were primarily interested in instances in which all inputs within a given combination were different (*e.g.* a given combination A:B:C:D:E=1:1:1:1:1 in a  $k = 5$  microwell was always considered in our analyses, but the combination A:B:C=2:2:1 in a  $k = 5$  microwell was typically excluded). In this case, a smaller library size might actually decrease the number of desirable combinations. For example, consider a small library size of 8 inputs (blue plots in **Fig. S14C**). At  $k = 7$ , there are only 8 possible unique combinations (where each input represented up to one time), but 3,432 possible non-unique combinations (where at least one input is represented greater than one time). Indeed, as  $k$  increases for a small library size, the efficiency of sampling desirable (no multi-choice) combinations decreases. In the Hf-GFP facilitation screen, we therefore chose a library size of 16 to balance this reduction in efficiency associated with small libraries while still generating a sufficiently high number of replicates for  $s = k = 1, 2$ , and 3 where all inputs were different.

In addition to the aforementioned considerations of the input library size, the number of microwells, and the microwell sizes, the overall throughput of a screen is also determined by the total number of microwells/kChips available to a screen. In the Hf-GFP facilitation screen, for example, six kChips were run in series with a single imaging system with  $\sim$ half a day of hands-on set up time ( $\sim$ 30 minutes required to set up each kChip). Depending on the desired time resolution for multiple timepoint studies (*e.g.* a 2X imaging scan takes  $\sim$ 12-15 minutes per kChip), one may wish to load greater or fewer kChips in the course of a given screen. The upper limit on the throughput of kChip screening is dictated by the number of

microwells. For screens consisting of ~100 chips (a scale we have practiced in the past (4)),  $\sim 10^6$ – $10^7$  communities could be constructed.

Second, currently the kChip platform has only been demonstrated for optical assays, such as the fluorescence imaging used here. Biochemical assays such as nucleic acid sequencing or mass spectrometry would be highly useful, and could perhaps be achieved by retrieving droplets from the array for further analysis. In addition, there is currently no provision to add reagents after the initial loading of the array, *e.g.* to add stains at the assay endpoint or to feed cells over the course of an assay. Some organisms may interfere with the optical assay, for example by secreting a fluorescent compound that interferes with the desired signal. We have not assessed phenotypes beyond growth, though “high content” morphological phenotypes such as filamentation, microcolony formation, and aggregation are likely feasible, especially with the aid of higher magnification. For example, consider the spatial organization of cultures we observed under 10X magnification images (**Fig S26D**).

Beyond fluorescence-based assays, we note the potential for phase contrast microscopy as a generalized readout that can be used to monitor microbial growth on the kChip without labeling (*e.g.* GFP expression) or addition of assay reagents (*e.g.* resazurin). We demonstrated the use of a phase contrast metric to measure *E. coli* growth on six single carbon sources (**Fig. S27**). Our phase contrast readout produced similar growth curves to those taken with other on-chip assays (GFP) and off-chip assays (GFP and OD<sub>600</sub> on 96-well plate taken via SpectraMax plate reader).

While we have only assessed the growth of aerobic microbes here, there are no inherent limitations associated with the cultivation of any particular class or microbes that are small enough to fit inside the droplets, *e.g.* anaerobes (provided that the droplets and loaded chips can be prepared and stored under anaerobic conditions).

There are additional considerations for those that wish to set up this platform. While not essential, a high-performance microscope (high numerical-aperture optics, wide field of view, and low camera dark current and read noise) is useful to the extent that throughput is not rate-limited by optics and camera

quality. Additionally, off-chip droplet production may expedited by a dedicated droplet generating instrument (*e.g.* the Bio-Rad QX200 used for these experiments) or a custom pressure manifold (4). Finally, basic proficiency with scientific image analysis (*e.g.* Matlab, Python) and data analysis for identifying and classifying droplet signals is also needed.

### **Supplemental Discussion – Ecological interpretation of results**

Data generated through kChip screening is a valuable resource to explore the underlying ecology of cellular interactions among microorganisms (bacterial, algal, and/or fungal) and their environmental dependencies.

Taking the carbon sources glucose and galactose, for which conserved glycolytic pathways are used ubiquitously across the bacterial domain, we observed pervasive competition (**Fig. S17**). In contrast, our results with the more complex oligomers (sucrose, lactose, and raffinose), on which Hf-GFP monocultures grew poorly in monoculture), are consistent with cross-feeding (**Fig. 4A**) that increases the availability of carbon accessible to Hf-GFP (9). In the case of sucrose, for example, we can speculate that enzymes produced by facilitating strains cleave sucrose into glucose and fructose, monomers that are then utilized by Hf-GFP.

We found that such facilitative effects were typically robust to community context (**Fig. S14**), suggesting that this facilitation is driven by key interactions and agnostic to the presence of additional strains. Improvement to the median yield of Hf-GFP with community richness (**Fig. 4A**) could be explained by the probability of sampling primary facilitator strains that individually facilitate Hf-GFP. In the case of sucrose, we additionally observed that the effect of a primary facilitator could be bolstered by additional strains, suggesting additive and/or higher-order effects as well (**Fig. S21, S22**). Due to this increased probability of sampling primary facilitators and these additive and/or higher-order effects, we observe that Hf-GFP yield converges across carbon sources as community size increases (**Fig. 4A, S17, S18**). Finally, the rarity of high robustness to both carbon source and community context (**Fig. 3E**) suggests that

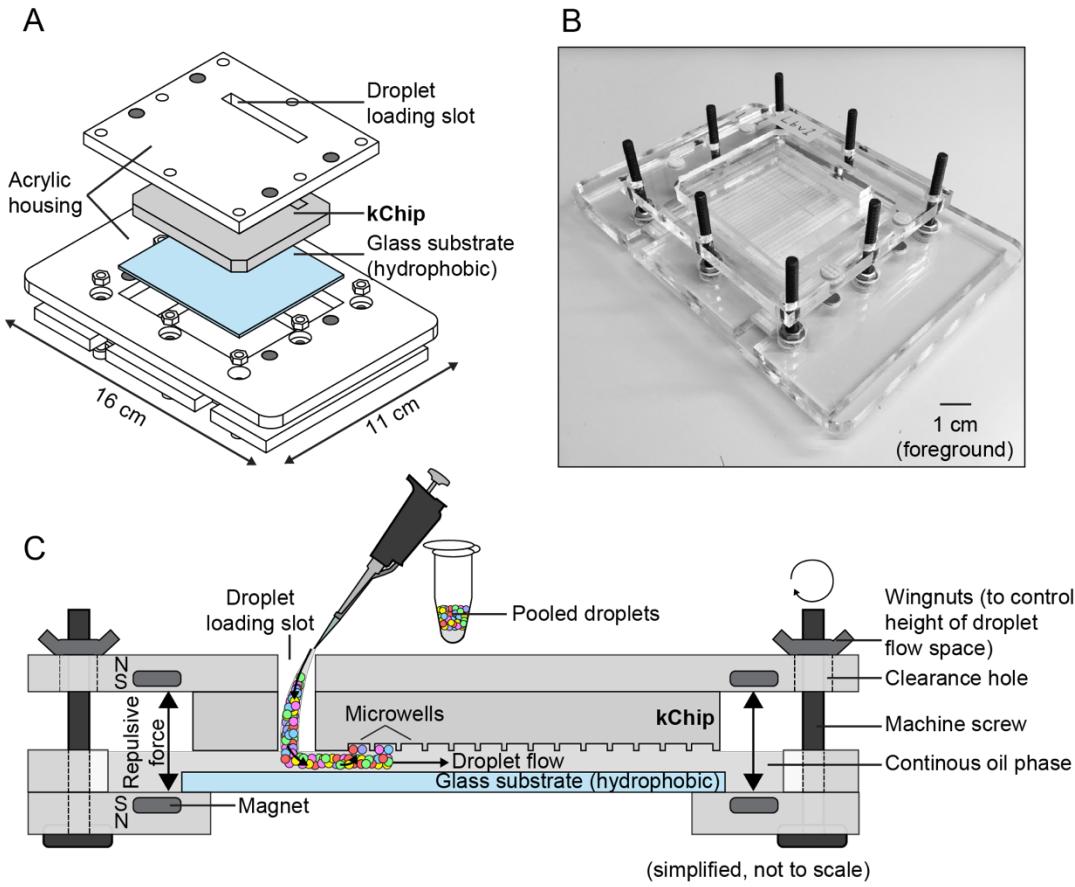
facilitative mechanisms depend highly on the environment and indicates a need for testing under many conditions to identify mechanisms and interactions robust to diverse biotic and chemical contexts.

### **Supplemental Discussion – Toward discovery of probiotics**

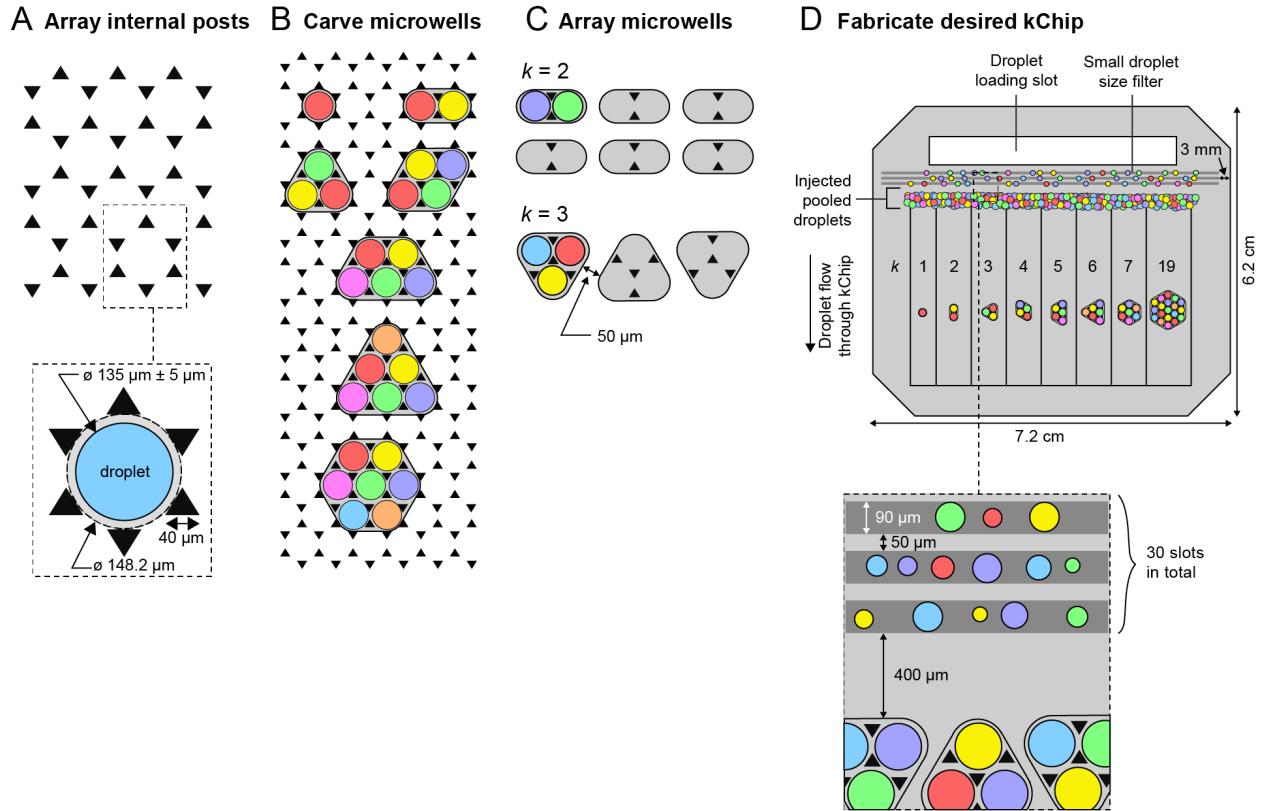
Inspired by the diversity of microbes residing in hosts (10) and the success of microbiota transplantation to counter ecological dysbiosis (11, 12), standardized interventions remain difficult to develop for a variety of reasons including a lack of mechanistic understanding and the explosion of possible strain combinations. Analogous to *in vitro* compound screening to generate therapeutic candidates, kChip screens can generate short lists of “hit” microbial cocktails that are also robust to relevant biotic and abiotic perturbations and constitute attractive candidates for validation and follow-up studies. For example, combinations of soil species can be identified that robustly facilitate plant growth-promoting rhizobacteria (PGPR), which have been shown to improve crop yields substantially (13) by providing the plant with nutrients and resisting pathogen colonization (14).

Screens to identify combinations that possess certain therapeutic functions may be particularly useful in the context of dysbiotic human microbiomes. Defined probiotics are already under development to address infections like vancomycin-resistant enterococci (15), *Clostridium difficile* (16), and *Salmonella* (17). Using the kChip, we could directly screen for strain combinations that suppress the growth of pathogens, assess the robustness of hit combinations to the presence of microbiome-like communities (as a proxy for the diversity of native microbiota among which delivered microbes may need to persist), and identify strains that inhibit suppressive capabilities of hit combinations (as an exclusionary diagnostic for use of a hit combination as a microbial therapeutic). Other potentially desirable functions of therapeutic cocktails, *e.g.* interactions with immune cells, can be hypothetically screened on the kChip platform, as it has been shown that mammalian cells can be cultivated in droplets as well (18). Finally, toward the development of prebiotics, the kChip could be used to screen environmental conditions, *e.g.* carbon sources,

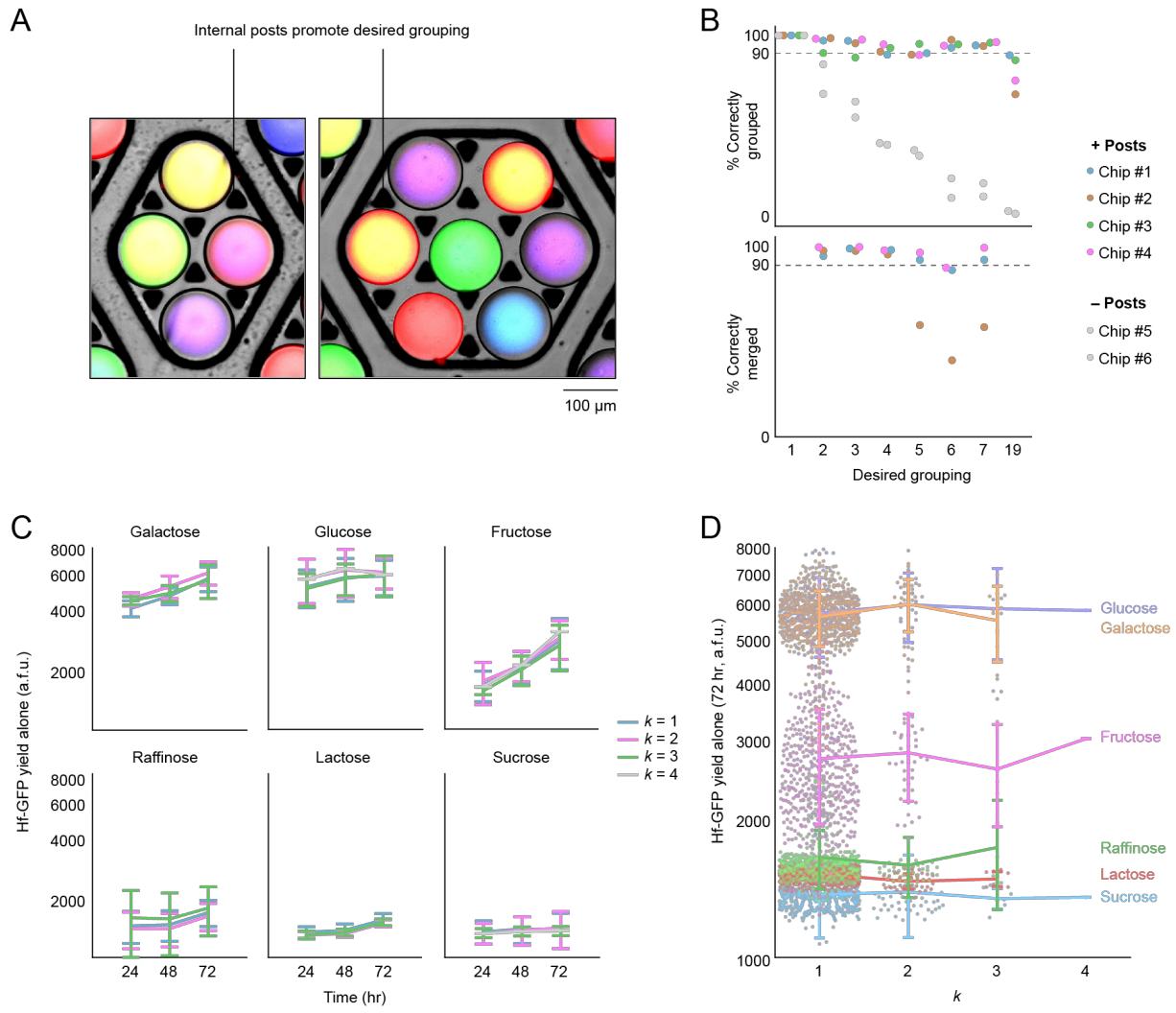
that specifically support the growth of commensal strains rather than pathogenic strains, or elicit desirable bacterial functions or immune interactions.



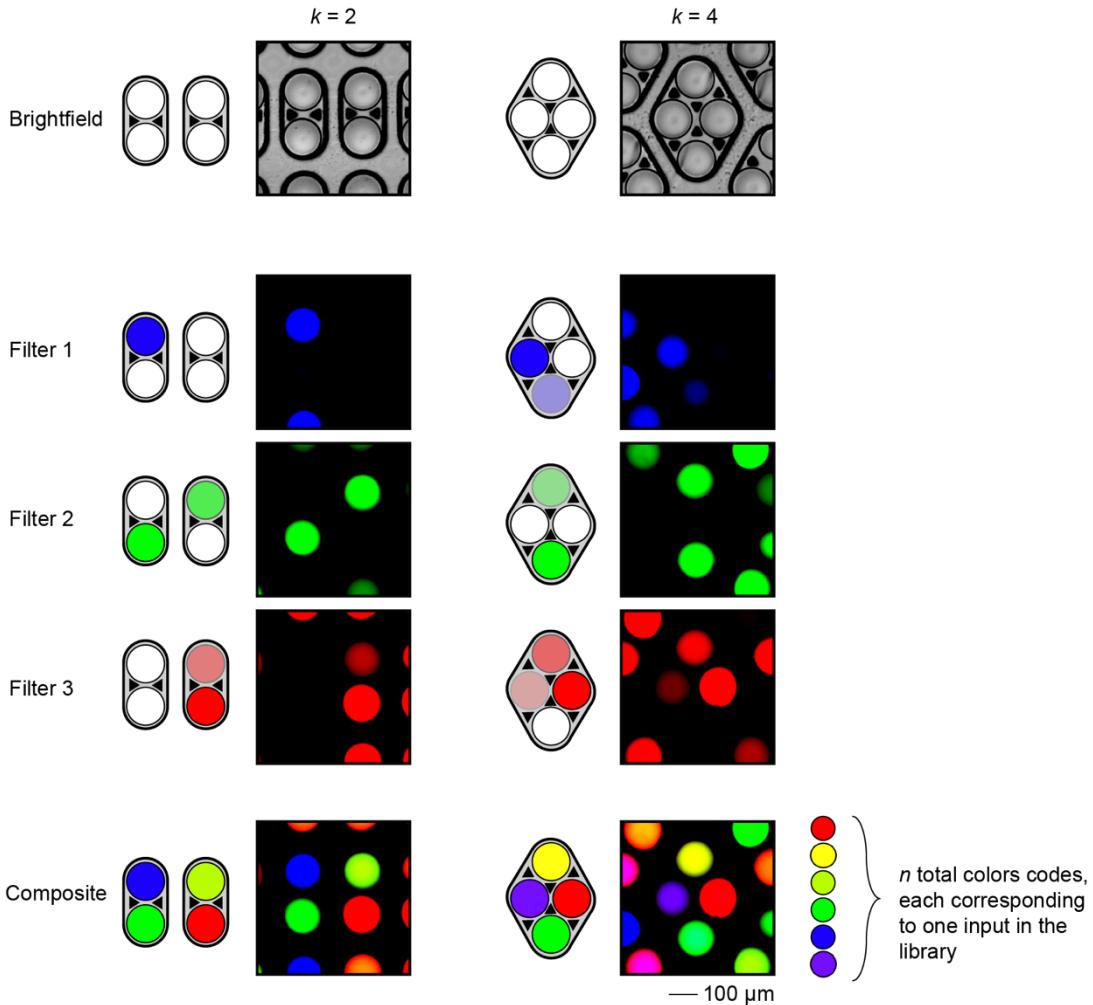
**Figure S1. A loading apparatus assists droplet loading onto kChip.** **A.** kChip microwells are loaded with droplets by suspending the kChip within a loading apparatus. This apparatus consists of an acrylic housing and hydrophobic glass substrate. The kChip naturally forms a seal with the top piece of acrylic. In its unclamped state, a flow space (~500-700  $\mu\text{m}$ ) is maintained between by a repulsive magnetic force such that droplets can flow under the kChip. Tilting the apparatus moves droplets through the flow space, and random sets of droplets spontaneously group within microwells. **B.** Photograph of the kChip loading apparatus. **C.** Side-view of loading apparatus and droplet loading procedure (not to scale).



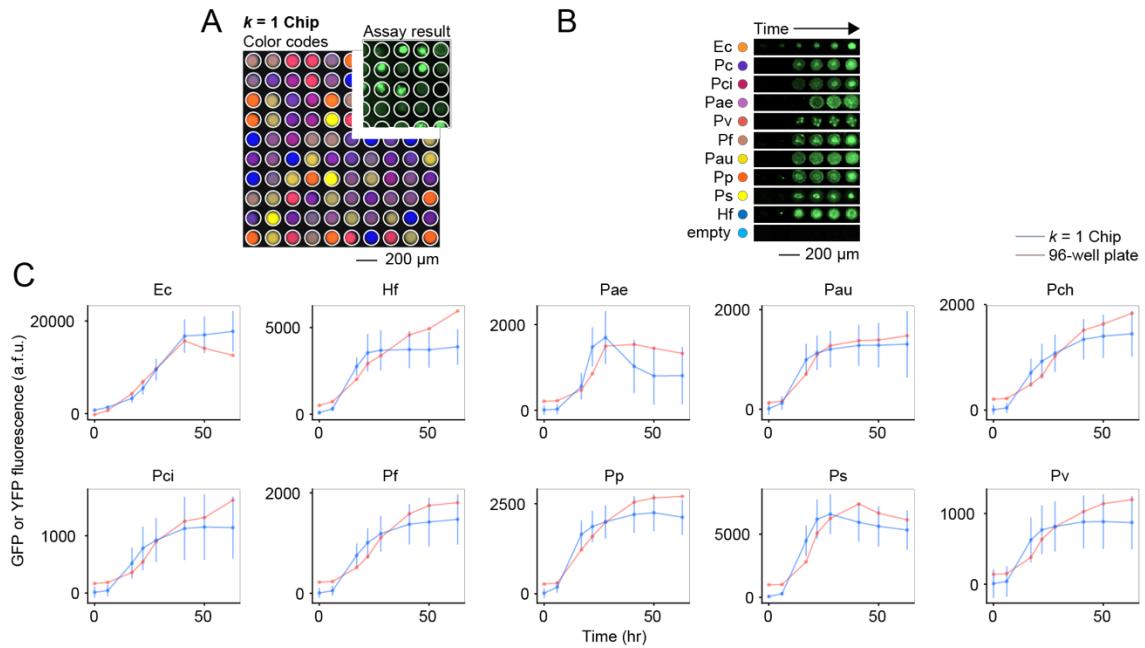
**Figure S2. kChip design strategy.** **A.** Triangular posts were arrayed such that each hexagonal arrangement enclosed a space optimized for a single droplet. The diameter of the enclosed space used in all kChip microwell types was set to 148.2  $\mu\text{m}$ , which was optimized for 135- $\mu\text{m}$  droplets, the mean droplet size of minimal medium (MM) droplets made with 0.05% BSA added to the medium and 2% stabilizing fluorosurfactant (Fig. S24). **B.** Microwell shapes were carved by encircling internal posts to achieve the desired  $k$ . These microwell designs are modular and generalizable: increasing the size of the boundary around posts produces microwells that group more droplets. **C.** Each microwell design was arrayed with  $\sim 50$   $\mu\text{m}$  inter-microwell spacing. **D.** Specific microwell arrangements on a kChip are determined by the user and can include  $\geq 1$  microwell type per kChip. A given kChip can have entirely one type (“Full kChip”), e.g. all  $k = 2$  microwells (e.g. experiments discussed Fig. 2) or an assortment of different types, e.g. subsets of  $k = \{1:7;19\}$  microwells (e.g. the Hf-GFP facilitation screen discussed in Fig. 3). All kChips also include a series of 30 90- $\mu\text{m}$  wide moat-like engravings (“slots”) designed that make up a small droplet size filter. These slots are spaced 50  $\mu\text{m}$  apart from each other and 400  $\mu\text{m}$  from the onset of the microwells. The slots are inset 3 mm from the edge of the kChip.



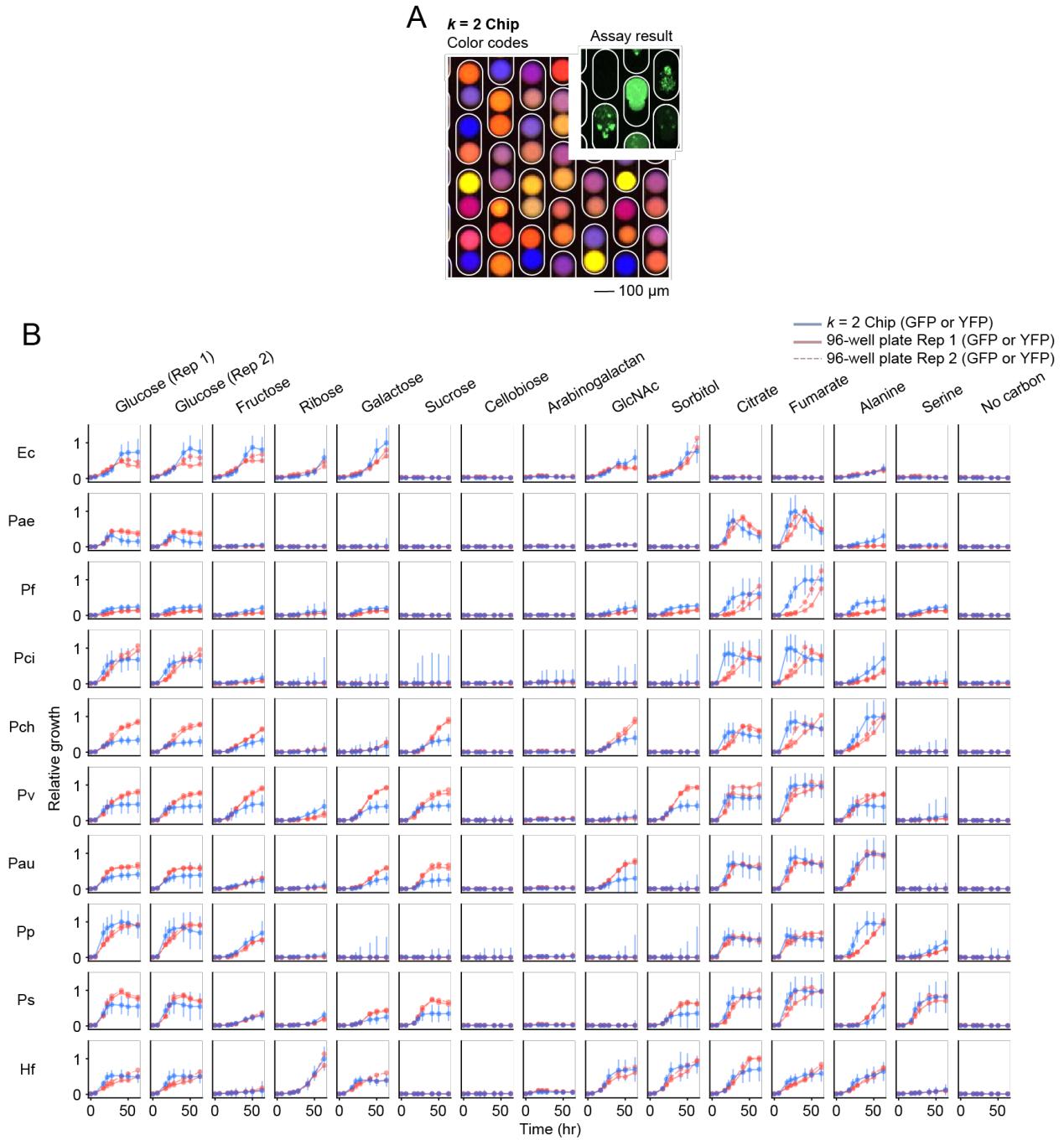
**Figure S3. Microwell geometry and internal posts promote precise droplet grouping without affecting microbial growth.** **A.** Fluorescence micrograph (10X magnification) highlighting how internal posts produce correct droplet grouping. Presumably, internal posts enable low-pass size filtering and spatial confinement of droplets within microwells. **B.** Four kChips were loaded that included internal posts (Chips #1-4) and two that did not (Chips #5-6). (Top) Over 90% of microwells fill correctly for all values of  $k = \{1:7\}$  if posts were included. If microwells did not contain internal posts, grouping performance decreased drastically as  $k$  increased. Dotted gray line = 90% grouping correctly. (Bottom) Droplets were merged in three kChips containing internal posts (Chips #1, 2, and 4). Droplets in microwells that have filled correctly will typically merge correctly. Dotted gray line = 90% merging correctly. (Merging data for  $k = 19$  was not measured due to limitations in the image data analysis.) **C, D.** In order to assess the biological effects of larger microwell geometries and the concomitant increase in internal posts, the yield of Hf-GFP in monoculture (“alone”) was measured at 24, 48, and 72 hrs (taken from the no-isolate control droplets in Hf-GFP facilitation screen dataset, which described in Fig. 3). We found no adverse effects on growth rate or 72-hr yield.



**Figure S4. Color coding of droplets is used to identify the contents of each microwell.** To identify the contents of the droplets that group within a given microwell, the kChip is scanned at 2X magnification after grouping and before merging droplets. Each fluorescent dye is excited sequentially, producing a corresponding image (dyes are chosen to avoid overlap with any fluorescence-based assays used concomitantly). Each droplet contains a unique ratio of the three fluorescent dyes such that the composite emission profile provides the droplet's unique color code. (Left) Example for  $k = 2$  microwells. (Right) Example for a  $k = 4$  microwell. (Bottom-right) The full set of colors produced maps to the  $n$  different inputs and is used to identify the combination of inputs that have grouped within each microwell.

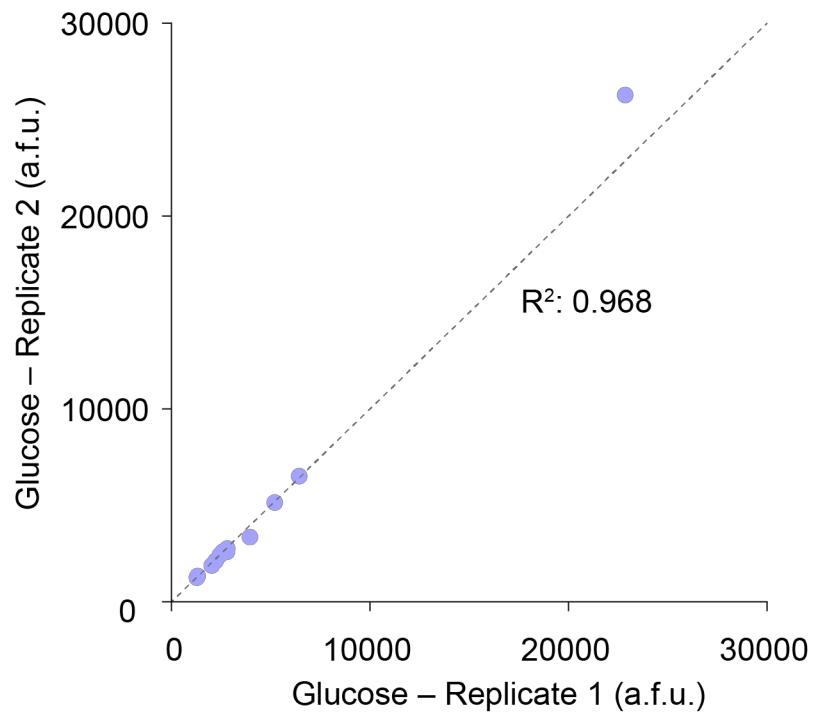


**Figure S5. Growth on glucose is similar between a  $k = 1$  Chip and a 96-well plate for 10 labeled strains.** **A.** Micrograph of droplets containing color codes and assay result (constitutive GFP or YFP expression). **B.** Example droplets showing the increase in signal for a panel of 11 inputs (10 fluorescently labeled strains + 1 “empty” medium-only control) over the course of the experiment (first 6 time points between 0 and ~40 hr shown corresponding to data points in panel C). **C.** We compared growth curves between 200- $\mu$ L cultures in 96-well plates and  $k = 1$  Chip cultures for each strain. We found that growth dynamics broadly matched between the two platforms. Data has been linearly rescaled from plates to the kChip data by computing a linear fit for each strain. Full names of labeled strains are listed in **Table S1**.

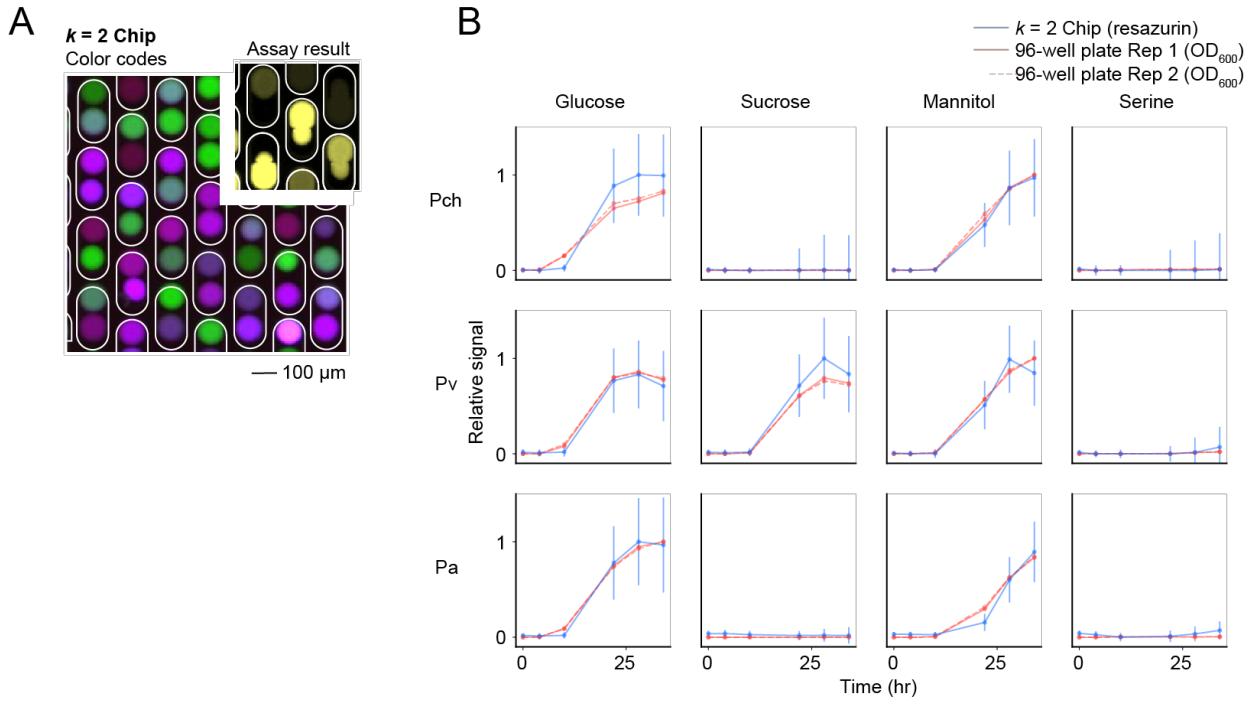


**Figure S6. Carbon utilization profiles can be obtained on *k* = 2 Chips via constitutive fluorescent protein expression and match standard culture techniques.** **A.** Example micrograph of fluorescent protein expression assay on kChip. **B.** On a *k* = 2 Chip and 200- $\mu$ L cultures in 96-well plates, we generated strain-carbon source combinations to assess between-platform agreement of growth dynamics. The  $t = 50$  hr measurements are represented in Fig. 2C. Data has been normalized in this manner: First, a second no-carbon control (not shown here) was subtracted from the growth curves. Second, all data for a given strain was divided by the maximum signal value across all carbon sources and time points. Error bars represent the standard deviation of signal across replicate microwells. Plate measurements were taken with two replicates, shown as a solid and

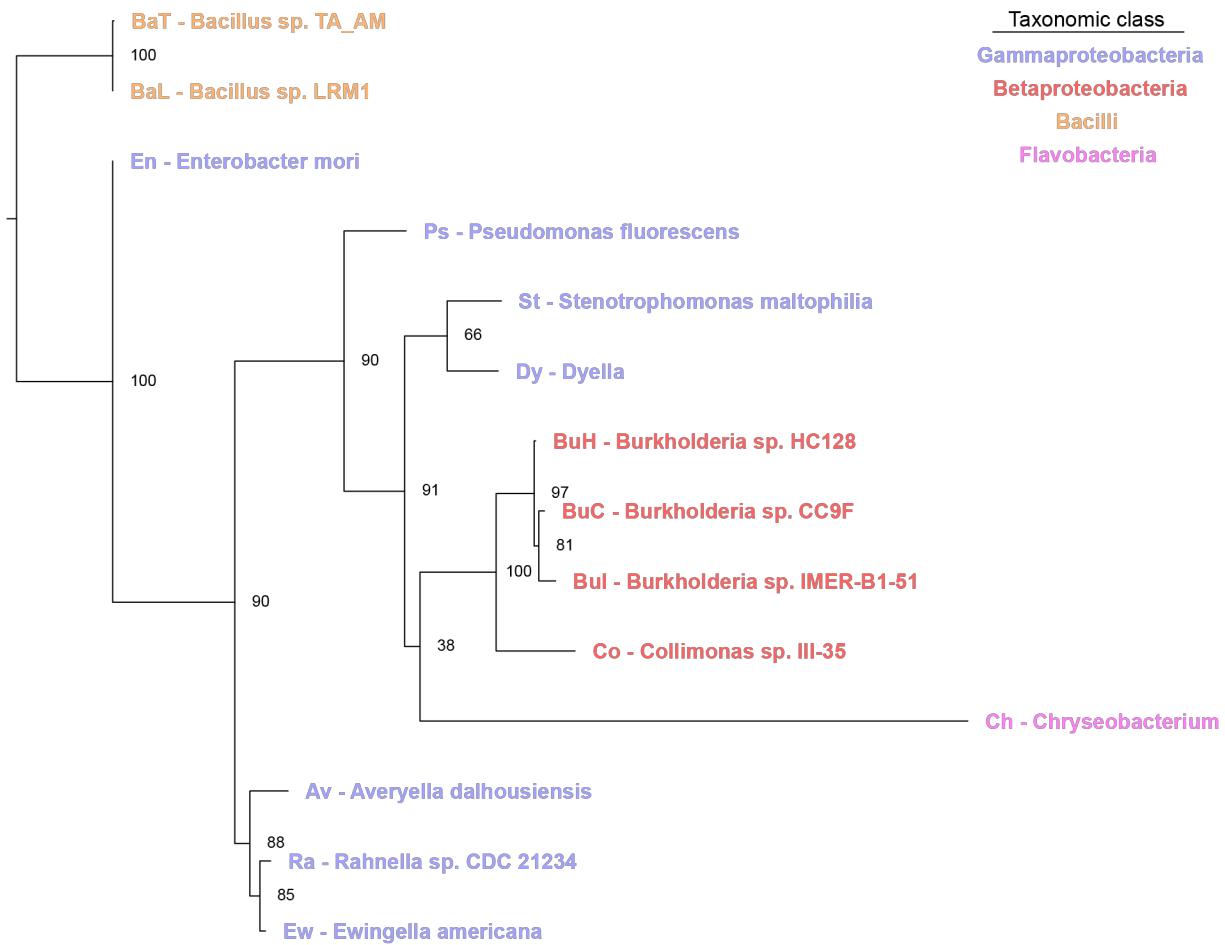
dotted line. Based on the broad agreement, we conclude that kChip screening recapitulates 96-well plate experimentation. Full names of labeled strains are listed in **Table S1**.



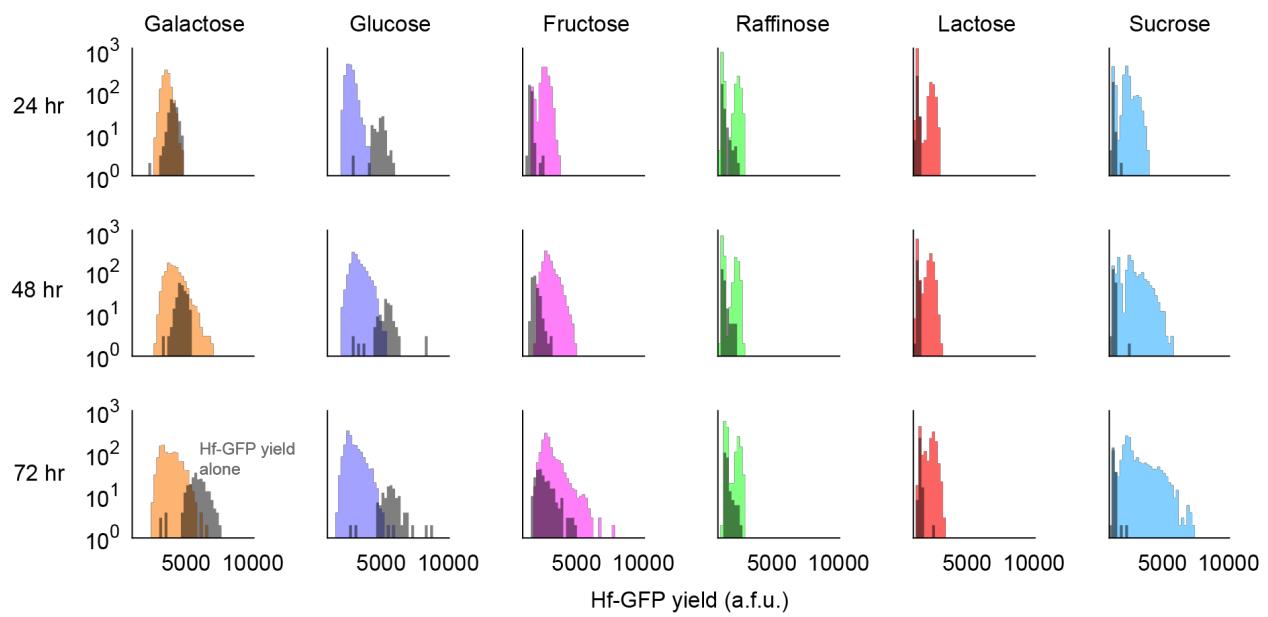
**Figure S7. Technical replicability of kChip bacterial growth assays.** To measure the technical replicability, the carbon source utilization experiment (**Fig. 2C, S6**) contained two technical replicates of glucose for each strain. A scatterplot compared the median signal obtained for each strain for each replicate of glucose at  $t = 50$  hours. Gray diagonal line =  $x=y$  line.



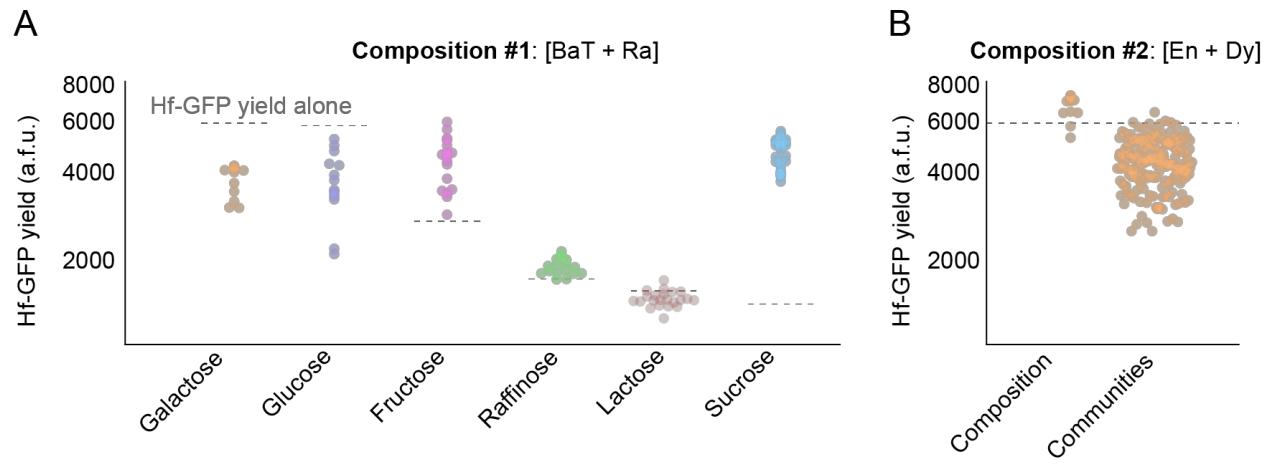
**Figure S8. Carbon utilization profiles can be attained on *k* = 2 Chips via the resazurin assay and match standard culture techniques.** **A.** Example micrograph of resazurin assay on *k*Chip. **B.** On a *k* = 2 Chip and 200- $\mu$ L cultures in 96-well plates, we generated strain-carbon source combinations to assess agreement between the resazurin assay in droplets and conventional  $OD_{600}$  measurements in plates. Endpoint measurements are represented in **Fig. 2D**. Data have been normalized in this manner: First, a no-carbon control (not shown here) was subtracted from the growth curves. Second, all data for a given strain was divided by the maximum signal value across all carbon sources and time points. Error bars represent standard deviation of droplets. Plate measurements were taken with two replicates, shown as solid and dotted lines. Based on the broad agreement, we conclude that *k*Chip screening with the resazurin assay recapitulates 96-well plate experimentation. Full names of strains used are listed in **Table S1**.



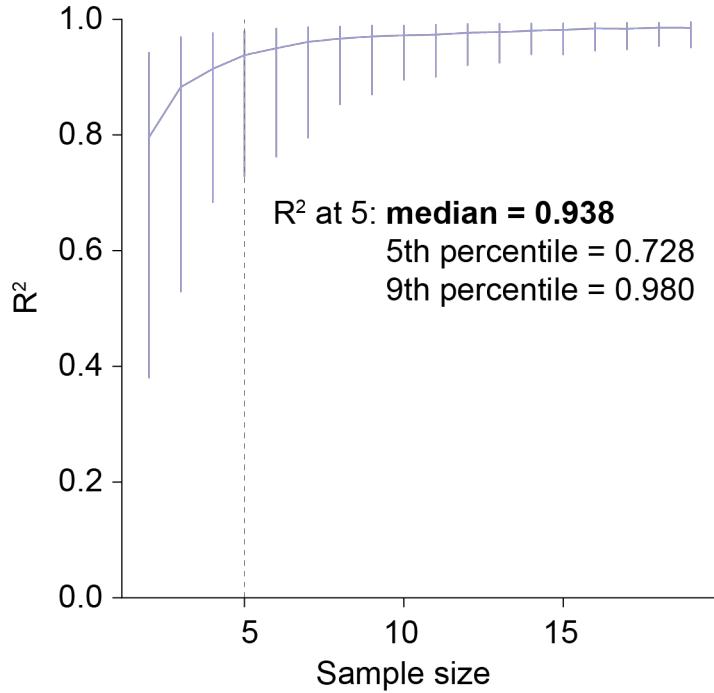
**Figure S9. Phylogenetic tree of 14 soil isolates used in *H. frisingense* facilitation screen.** Sequences of the V1 to V9 region of the 16S rRNA gene were obtained via Sanger sequencing (**Dataset S1**). Clustal X, with penalizations for gap opening and gap extension of 12 and 6, was used to align the sequences. PhyML-SMS with default parameters was used to select GTR as the best model of nucleotide substitution, to infer the tree, and to get bootstrap support values. Taxonomy classification and labels in the tree were obtained by selecting the sequence match with highest S\_ab score from Seqmatch. *Sulfolobus solfataricus*, a thermophilic archaeon, was used as an outgroup species to root the tree.



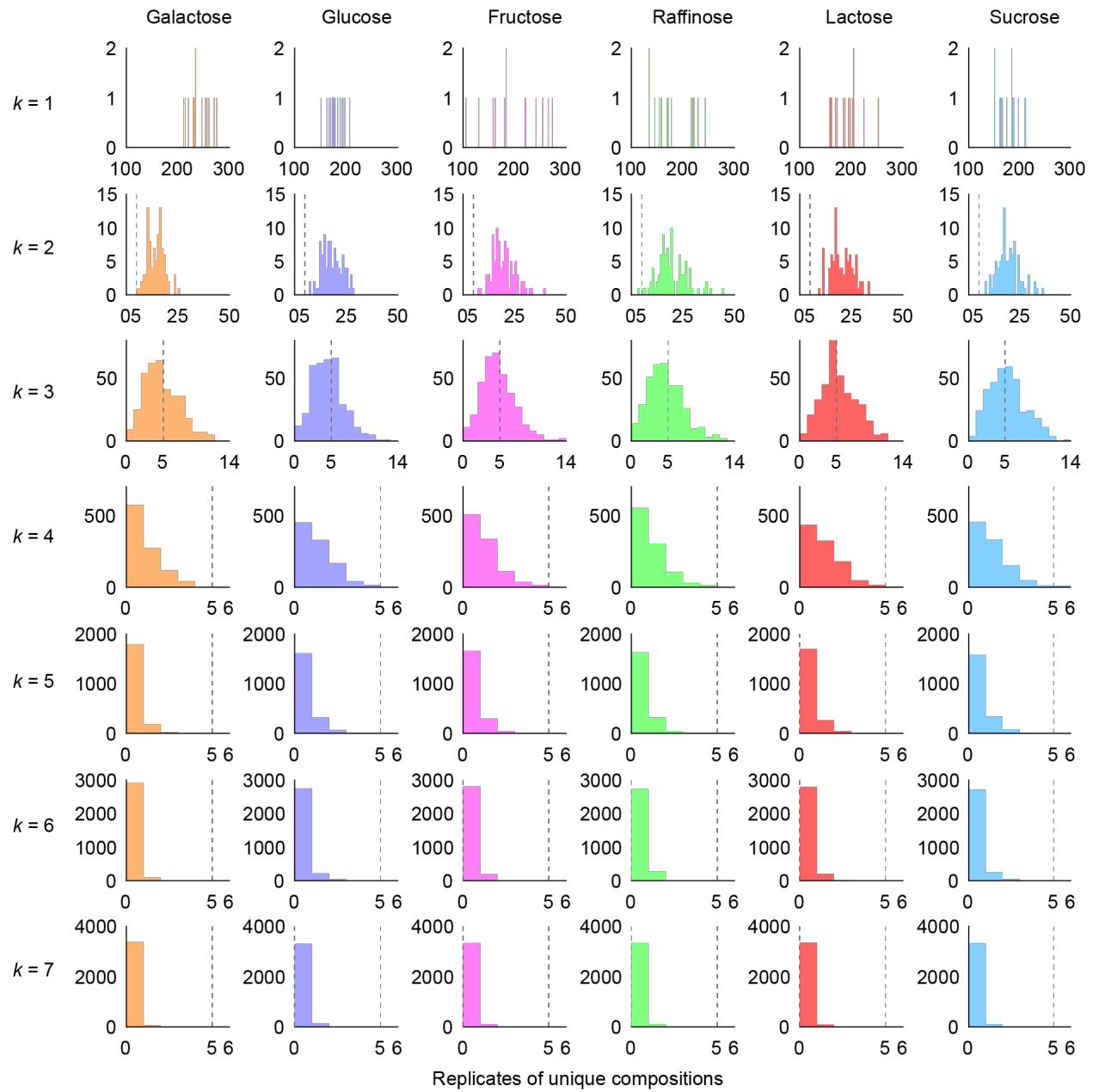
**Figure S10. Hf-GFP yield across compositions depends on carbon source and time point.** Hf-GFP yield was measured within each composition and carbon source at 24, 48, and 72 hr after droplet merging. The 72-hr data is also represented in Fig. 3C. Colored distributions = yield of Hf-GFP  $\geq 1$  isolate. Gray distributions = yield of Hf-GFP alone ( $k = 1$  microwells where droplet received no isolate). While the focus of the present analysis was yield at 72 hrs (Fig. 3, 4), time-dependent effects are evident, particularly in fructose where ubiquitous facilitation of Hf-GFP yield appears by 24 hrs.



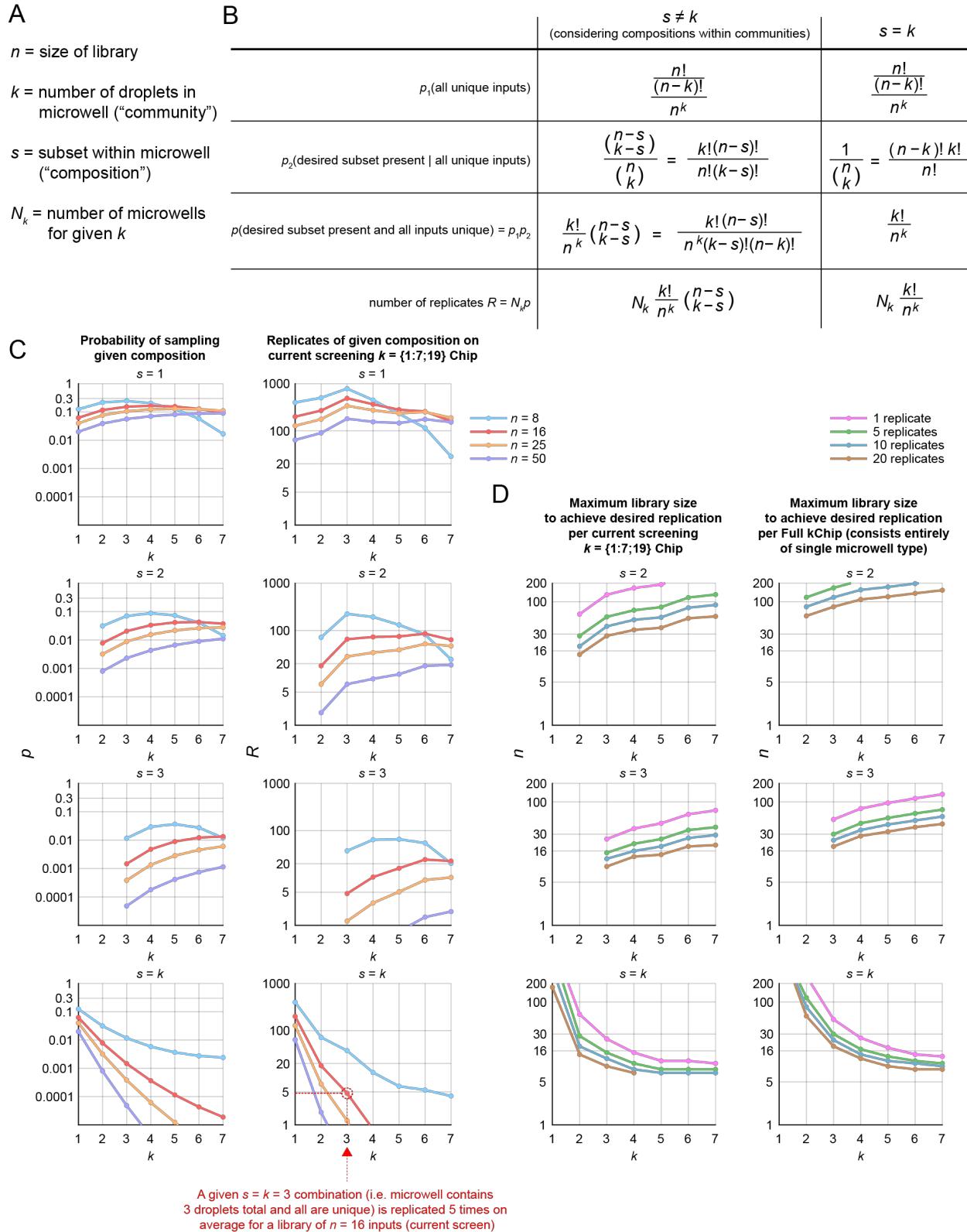
**Figure S11.** A facilitative composition may not be robust to carbon source or community context. **A.** Example  $s = k = 2$  pair of isolates with a facilitative effect on Hf-GFP that is not robust to carbon source. **B.** Example  $s = 2$  composition with a facilitative effect on Hf-GFP in galactose that is not robust to community context.



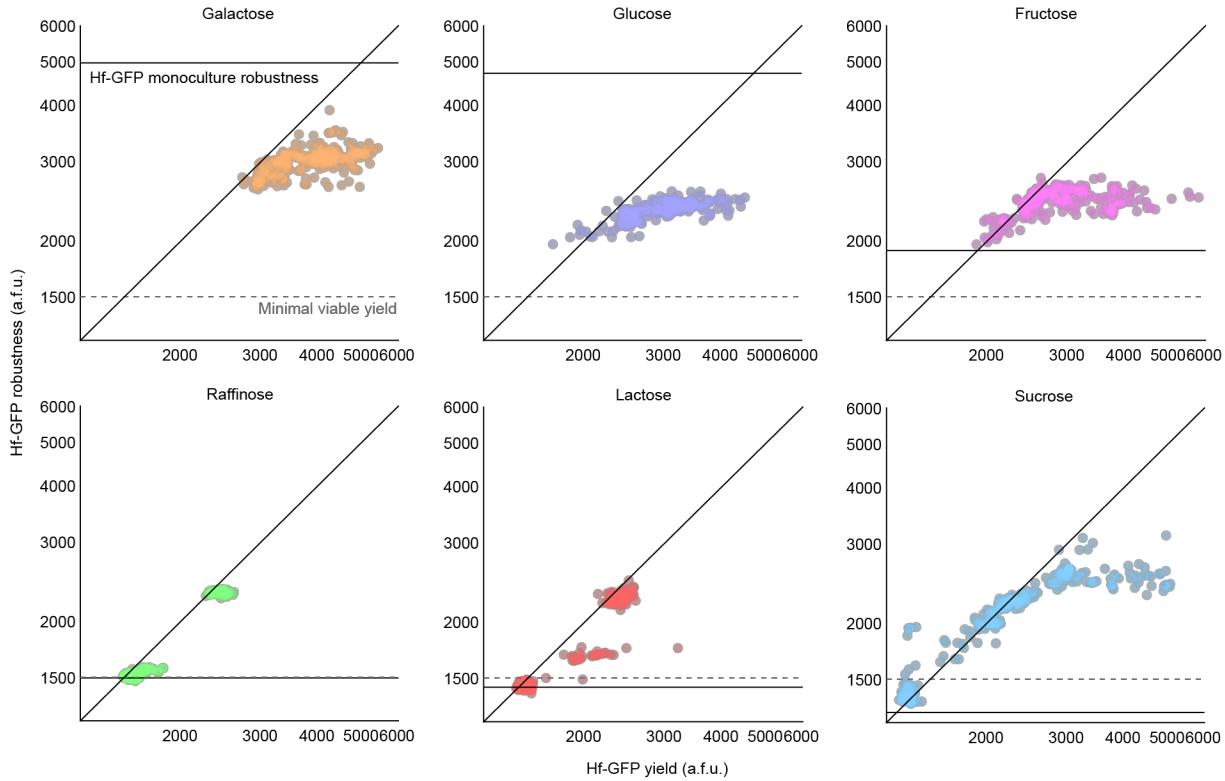
**Figure S12. Analysis of the tradeoff between microwell-replication and technical noise for Hf-GFP growth assays.** Fig. S14 shows the expected number of microwell-replicates for a given input library size. To determine the input library size for Hf-GFP facilitation screen in Fig. 3, we determined the number of microwell-replicates desired by estimating technical noise in Hf-GFP growth (Fig. S6) for given expected microwell-replication sample sizes. We resampled the data with replacement and calculated the median growth measurements for each carbon source at  $t = 63$  hours (assay endpoint). We compared these estimates with the growth values measured using all data (Fig. S6,  $t = 63$  hours), and computed an  $R^2$  value for each bootstrap sample (500 iterations). The curve shows the median  $R^2$  value, and error bars show the 5th to 95th percentile of  $R^2$  measurements from bootstrapped samples over the 500 iterations. A microwell-replication level of 5 microwells is shown as a dotted line, where the median  $R^2$  value was found to be 0.938 (5th percentile: 0.728, 95th percentile: 0.980, worst case: 0.414).



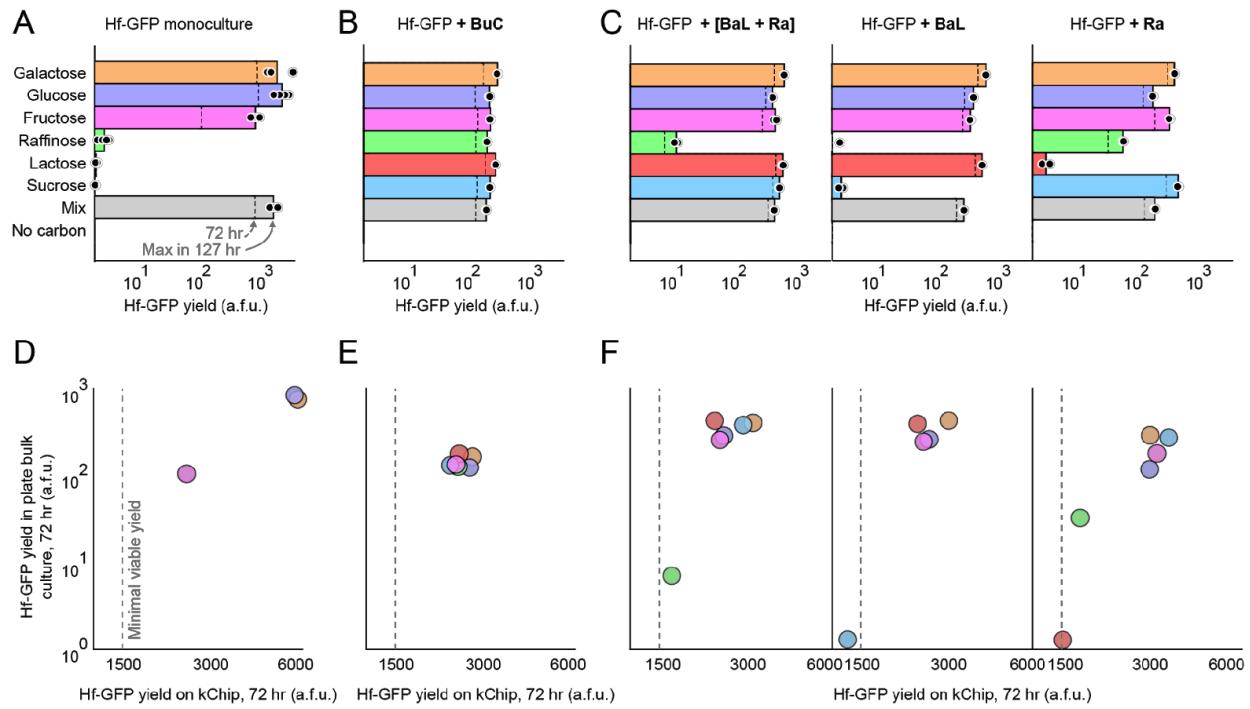
**Figure S13. Compositions that were replicated  $\geq 5$  times in this screen occurred at  $k = 1, 2$ , and  $3$ .** The number of replicates at each  $k$  was determined by the sampling (number of microwells per kChip) and library size. For the Hf-GFP facilitation screen, all  $k = 1$ , almost all  $k = 2$ , and  $\sim$ half of  $k = 3$  compositions were represented  $\geq 5$  times (**Dataset S2**). Almost no compositions for  $k \geq 4$  were represented  $\geq 5$  times. Dotted gray line = 5 replicates, the expected (**Fig. S14**) and actualized mean representation of unique compositions in  $k = 3$  microwells. The cutoff for inclusion in **Fig. 3E** was that a composition was replicated  $\geq 30$  times across the six kChips in the screen (though they did not need to be represented  $\geq 5$  times/kChip).



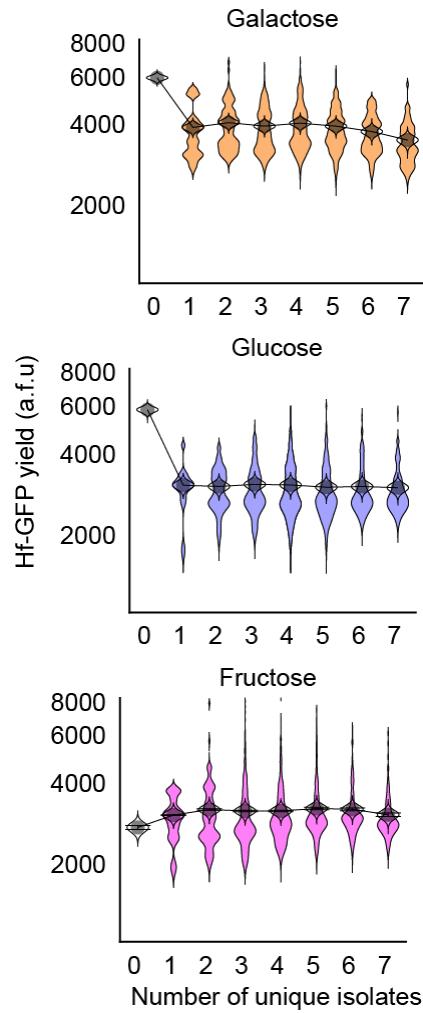
**Figure S14. Experimental setup determines the combinatorial space sampled.** **A.** Definitions of variables used. **B.** Expressions describing sampling probabilities when the desired subsets (“compositions”) under consideration are either less than total number of droplet inputs (“communities”) (in which case  $s < k$ ) or equal (in which case  $s = k$ ). **C.** (Left) Probability of sampling a given composition for different desired subsets  $s$  for different input library sizes. (Right) Expected representation of a given composition with current  $k = \{1:7;19\}$  Chips used in the Hf-GFP facilitation screen. Red line = library sized used presently ( $n = 16$ ). For  $k = 1-3$  in the Hf-GFP facilitation screen, the mean number of replicates for  $k = 1, 2$ , and  $3$  was  $\sim 200$ ,  $\sim 20$ , and  $\sim 5$ , respectively. We set 5 replicates as a threshold to measure the effect of a composition ( $s = k$ ), which excluded  $\sim$ half of  $k = 3$  compositions represented  $< 5$  times (Fig. S13, Dataset S2). Robustness of these compositions to additional isolates ( $s < k$ ) could be measured due to high replicability, which increased with  $k$ . Dotted red line = In the Hf-GFP facilitation screen, the cutoff at  $s = k = 3$  to produce a mean of 5 replicates per composition was a library of  $n = 16$  inputs. **D.** The maximum allowable library depends on the desired mean number of replicates per composition. (Left) Allowable library size with current  $k = \{1:7;19\}$  Chips. (Right) Allowable library size if using kChip consisting only of single microwell type (“Full kChip”). While the Hf-GFP facilitation screen used six kChips, we’ve previously demonstrated that screens can be feasibly completed with  $\sim 100$ s of chips (4).



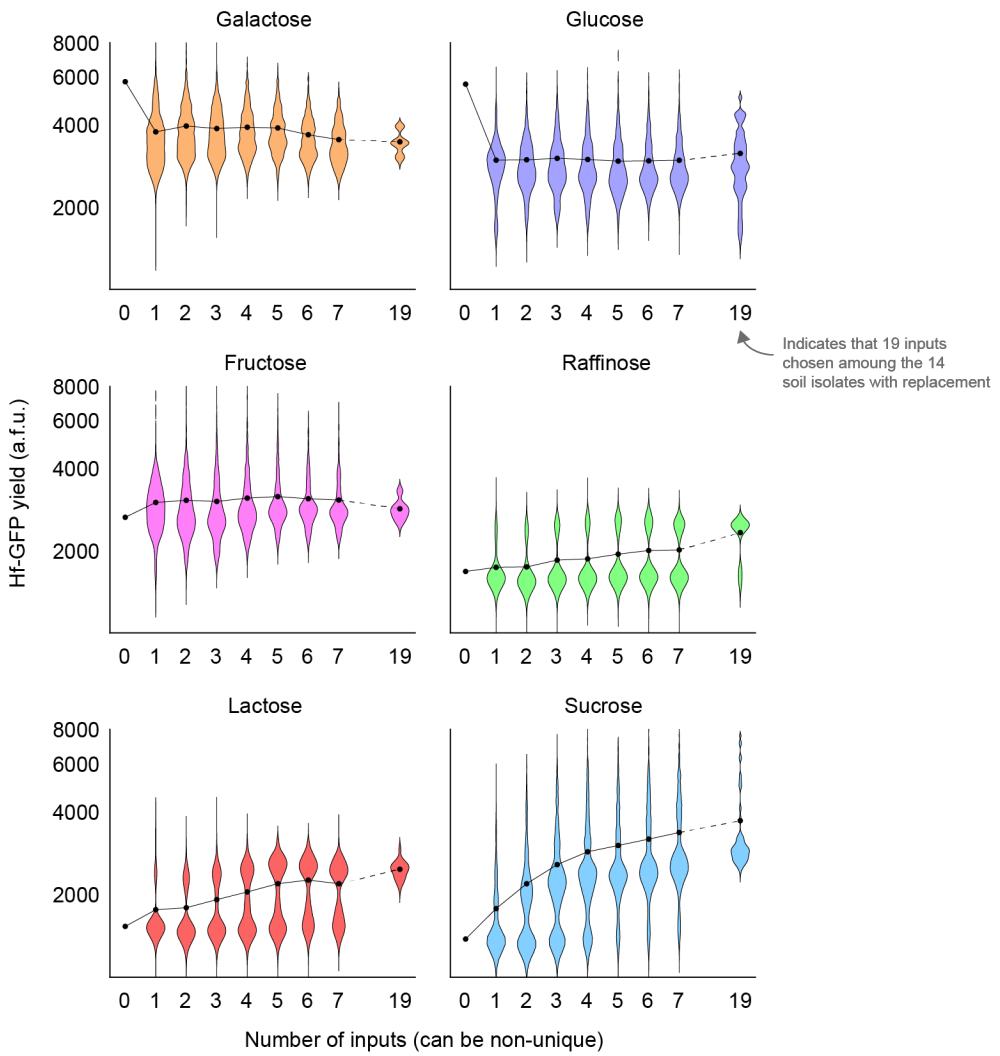
**Figure S15. The effect of facilitative composition is typically robust to additional isolates.** All data here represent compositions replicated  $\geq 5$  times (Fig. S13) separated by carbon source (with no separation by carbon source shown in Fig. 3E). “Hf-GFP yield” is the median Hf-GFP yield at 72 hr for a given composition represented  $\geq 5$  times for the given carbon source. “Hf-GFP robustness” is the 10th percentile of Hf-GFP yield for all communities containing the given composition with  $\geq 1$  additional isolates. Consistently, the facilitative effects of compositions enabling high growth of Hf-GFP are robust to the presence of additional isolates for each given carbon source. Gray dotted lines = “minimal viable growth” of Hf-GFP (1500 GFP counts, or ~one standard deviation above mean Hf-GFP yield alone in sucrose at which GFP signal indistinguishable from background). Solid horizontal lines = Hf-GFP monoculture 10th percentile (“robustness”). Black diagonal line =  $x=y$  line for reference.



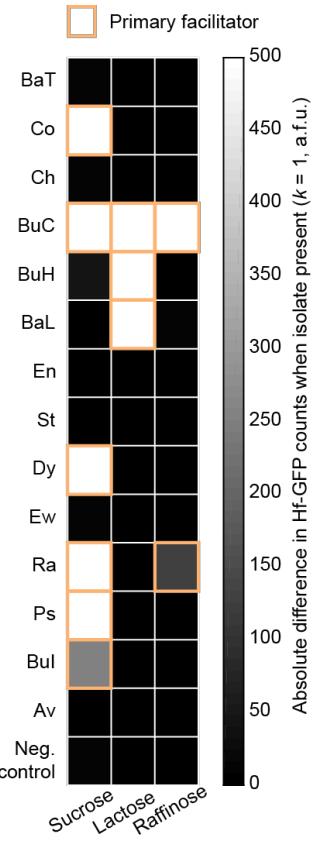
**Figure S16. The most robust compositions enabled facilitation across all carbon sources in 96-well plates bulk cultures, validating results from kChip screen.** **A-C.** Hf-GFP yields at 72 and 127 hr for 200- $\mu$ L co-cultures in 96-well plates (21°C, 220 RPM). Co-culture conditions are indicated in plot titles. Each carbon source was added to MM at 0.5% w/v. Mix = Each carbon source at 0.083% w/v such that total carbon content was 0.5% w/v. **D-F.** Comparison of raw yield values in 96-well plate bulk cultures and kChip screen. Dotted line = “Minimal viable yield” of Hf-GFP in screen. Hf-GFP yield in bulk co-culture recapitulates the result from the screen.



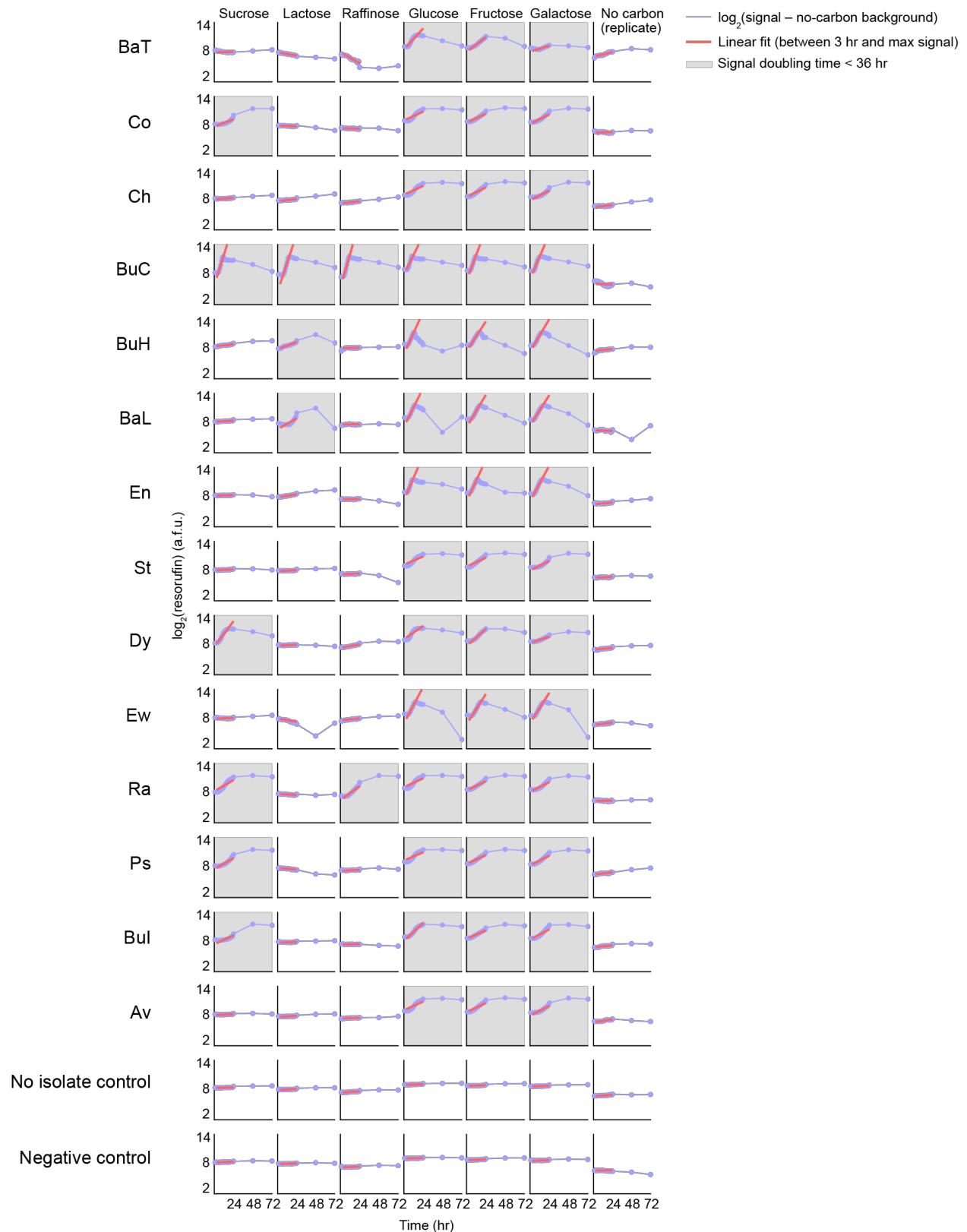
**Figure S17. Hf-GFP yield depends on community richness and carbon source.** Hf-GFP yield was measured across different numbers of unique isolates in different carbon source media (lactose, raffinose, and sucrose, the carbon sources in which it grew most poorly in monoculture, shown in Fig. 4A). In galactose and glucose, carbon sources in which Hf-GFP monocultures grew well, a suppressive effect was observed for  $\geq 1$  isolate that generally appeared agnostic to the number of unique isolates. With fructose, a carbon source for which Hf-GFP yield was greater than its yield in lactose, raffinose, and sucrose, but lesser than its yield in galactose and glucose, an initial increase yield was observed as the number of isolates increased from 1 to 3 isolates. Beyond this point, yield appeared to plateau.



**Figure S18. Hf-GFP yield was measured across 1–7 and 19 combinations of inputs.** If Hf-GFP yield is measured across all input combinations, even when the inputs within a microwell are non-unique (*e.g.* an instance of 3 inputs could include isolates [A + B + C], [A + B + B], or [A + A + A]), the qualitative trends observed when uniqueness is required (Fig. 4A, S17) are preserved for each carbon source. Hf-GFP yield does not differ substantially for 7 and 19 inputs, suggesting an agnosticism to increases in community richness in the regime of relatively high richness values. Black data points = means of distributions.

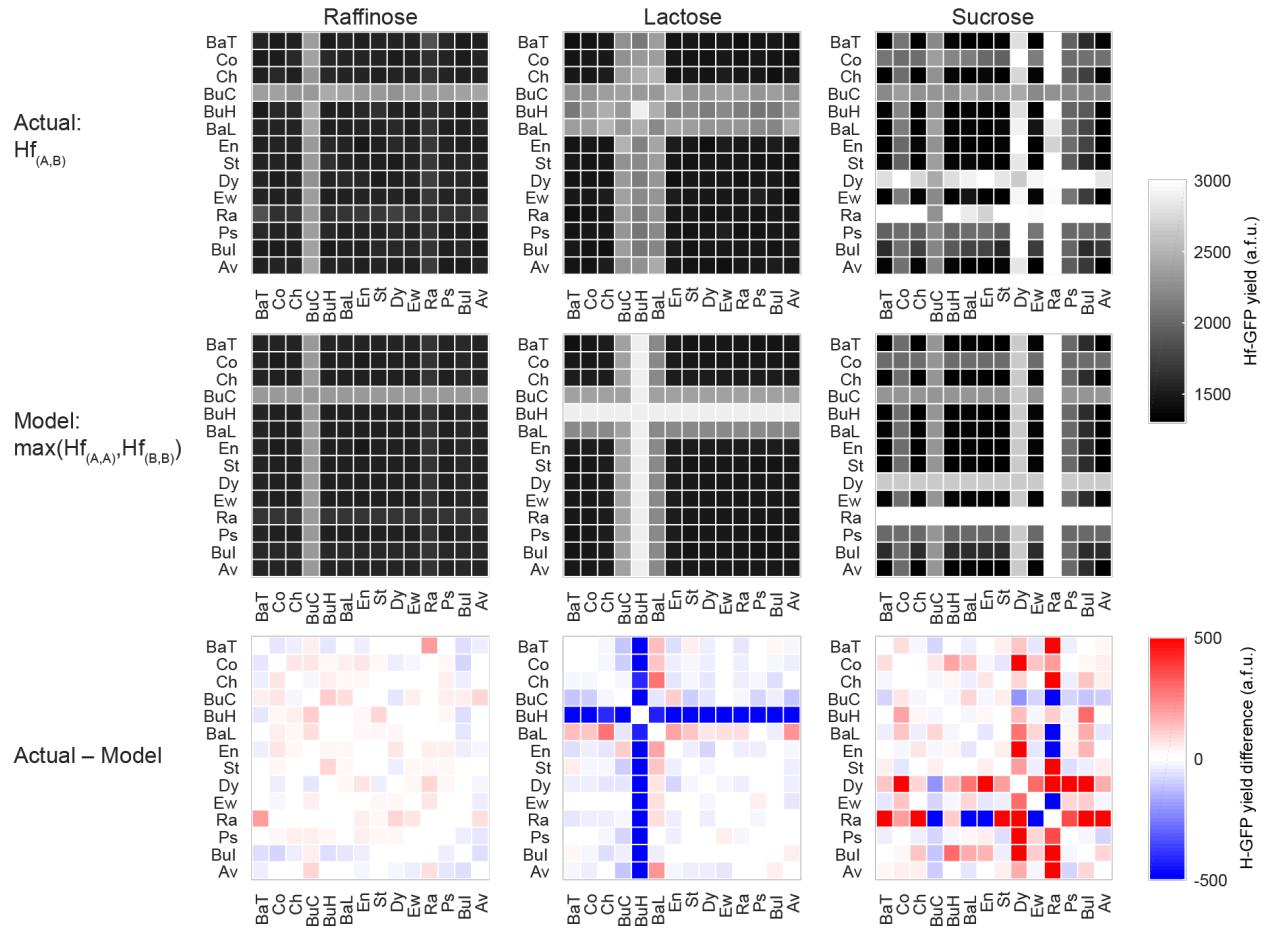


**Figure S19. Primary facilitators individually increased growth of Hf-GFP.** Primary facilitators (outlined in orange) were classified as the isolates increasing median Hf-GFP yield by an absolute difference of >100 GFP counts (a.f.u.) over Hf-GFP monoculture yield. These are the isolates left out from the gray distributions in **Fig. 4B** and highlighted for their ability to facilitate Hf-GFP in **Fig. 4C**.

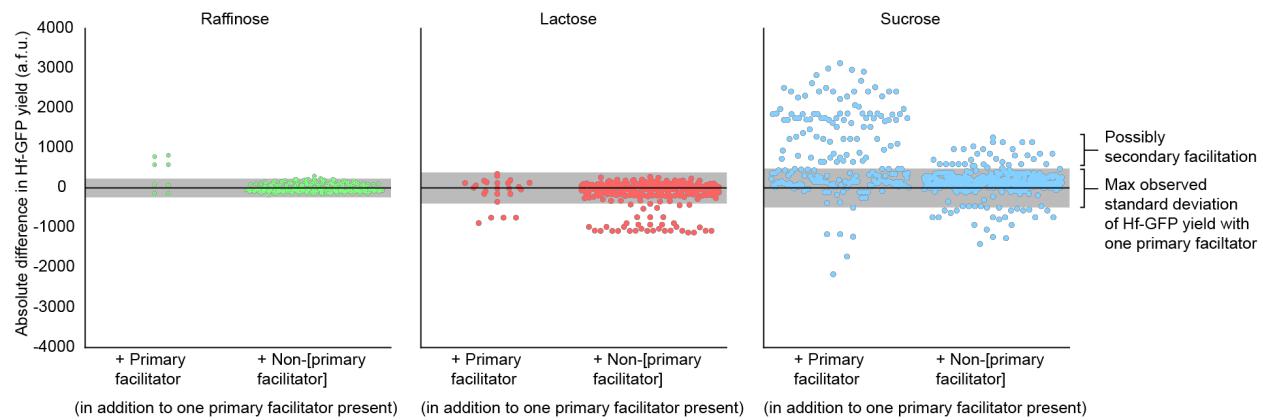


**Figure S20.** The resazurin assay was used to measure growth rate of the isolates on each carbon source. The fluorescence of resorufin was

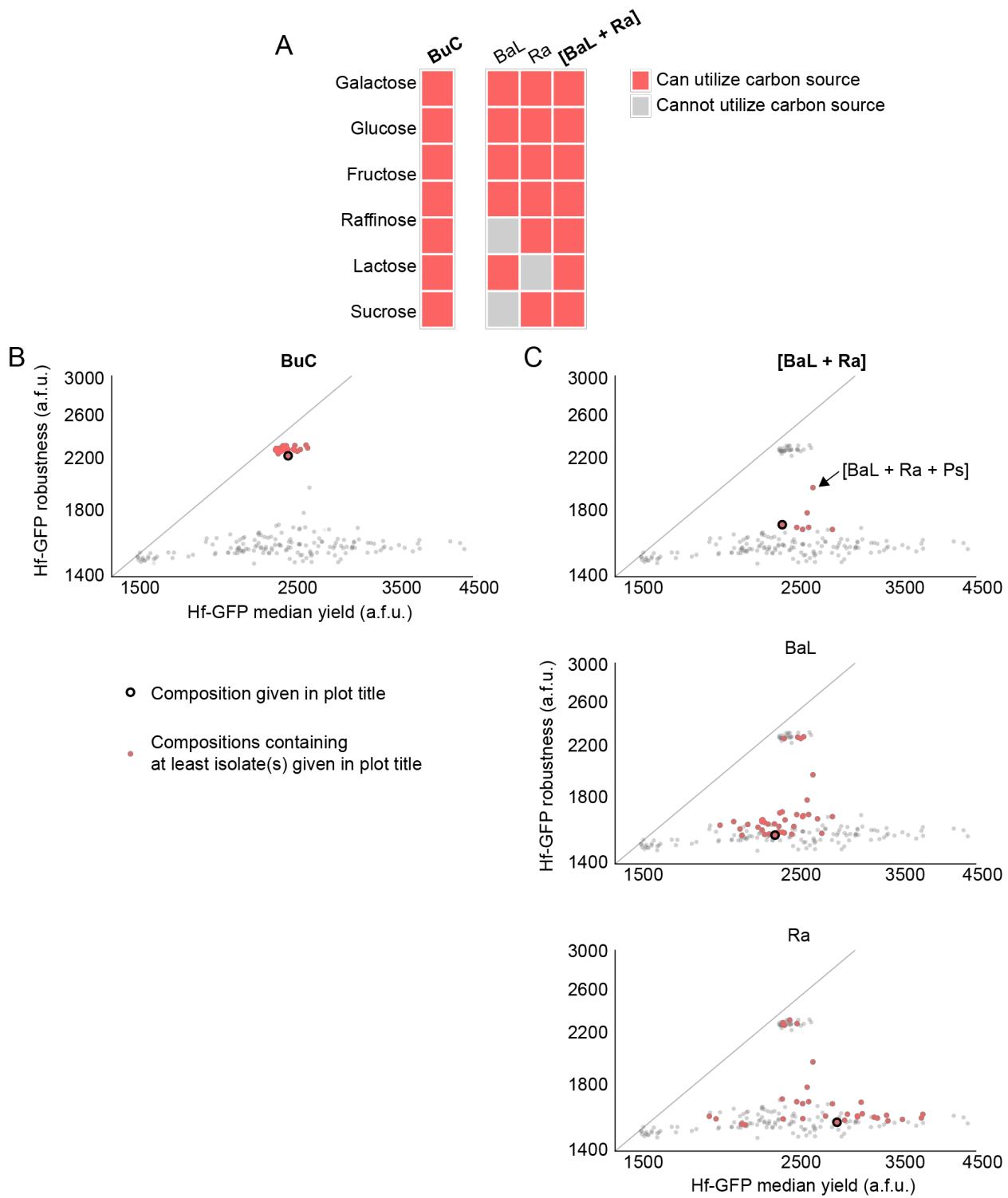
measured every 30 minutes for 24 hours and again at 48 and 72 hrs. The  $\log_2$ (background-subtracted resorufin) signal was calculated (signal used for subtraction was a no-carbon control, not shown here) and a linear fit was performed between the 3-hr time point and the maximum signal (**Fig. 4C**). Gray plots = at least one doubling has been detected by 36 hr (midpoint of the screen). A replicate no-carbon control set of droplets (shown) was also included and produced no false positives by this metric.



**Figure S21. Hf-GFP yield in the presence of two isolates can exceed its growth with each isolate individually.** (Top)  $Hf_{(A,B)}$ , the Hf-GFP yield at 72 hr in co-culture with the  $k = 2$  composition [A + B] in a given carbon source. (Middle) Simple model where Hf-GFP yield in the presence of [A + B] is determined from  $k = 2$  data as  $\max(Hf_{(A,A)}, Hf_{(B,B)})$ . (Bottom) Difference between (Top) and (Middle). Red and blue indicate degree to which  $Hf_{(A,B)}$  is greater than or less than  $\max(Hf_{(A,A)}, Hf_{(B,B)})$ , respectively. Hf-GFP often exceeds the model's expectation in sucrose.

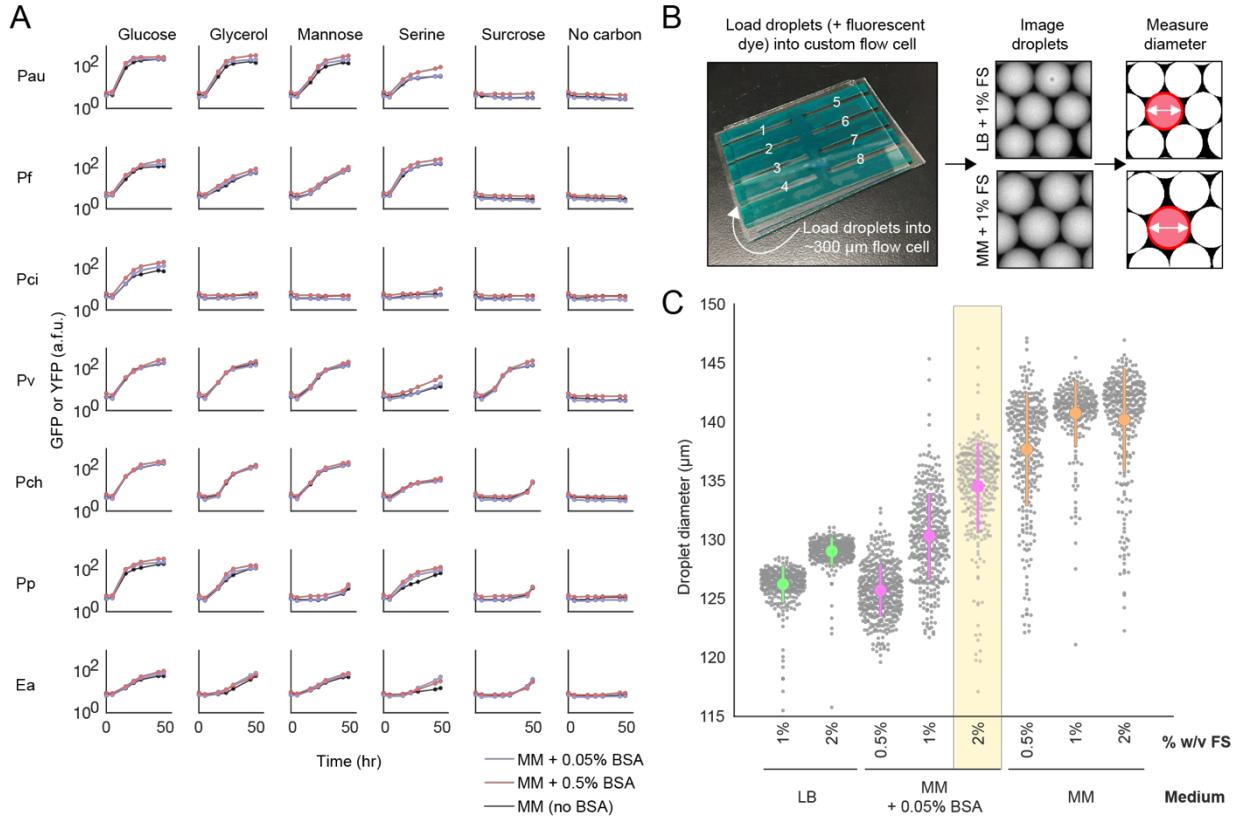


**Figure S22. With one primary facilitator already present, Hf-GFP yield can further improve with additional primary facilitators and non-[primary facilitators] in a sucrose medium.** Hf-GFP yield is enabled by the presence of a single primary facilitator in a composition (Fig. 4C). With one primary facilitator present, the addition of a second primary facilitator (left distributions) or non-[primary facilitator] (right distributions) can mediate this size of the facilitation. This effect is often positive in a sucrose medium, especially when the second isolate is a primary facilitator. We can identify isolates, tentatively, as “secondary facilitators” if they are non-[primary facilitators] but appear to exert a positive effect on Hf-GFP yield in the presence of a primary facilitator. The compositions analyzed here consist of all single isolate and pairwise isolate subsets of the 3-isolate compositions represented  $\geq 5$  times on each given carbon source. Black horizontal line = no effect. Gray shading = the maximum standard deviation among Hf-GFP yield in the presence of a single primary facilitator in the given carbon source, as an approximation for noise.

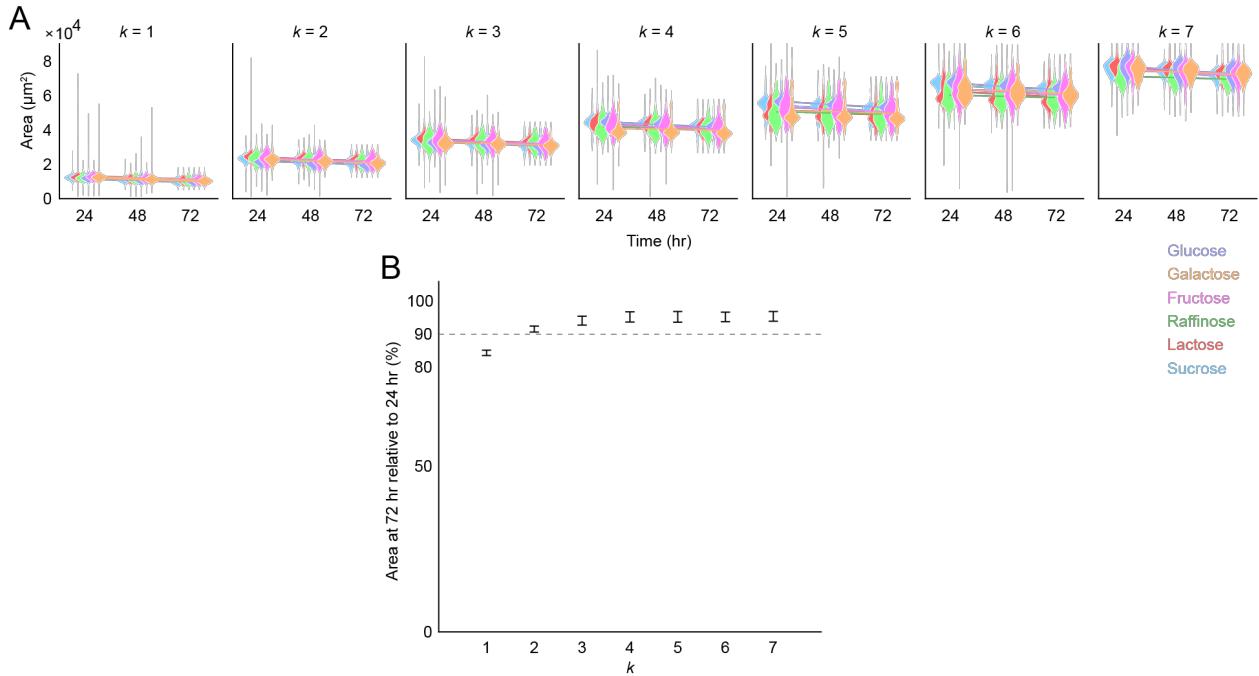


**Figure S23. The two most robust compositions consist of “core” groups of primary facilitators among which all carbon sources can be utilized.** **A.** BuC alone and the pair [BaL + Ra] can utilize all carbon sources tested as indicated by the resazurin assay (**Fig. S20**). **B.** Compositions that contain at least BuC consistently imparted relatively strong Hf-GFP median yield and the strongest robustness (explanation of these scores

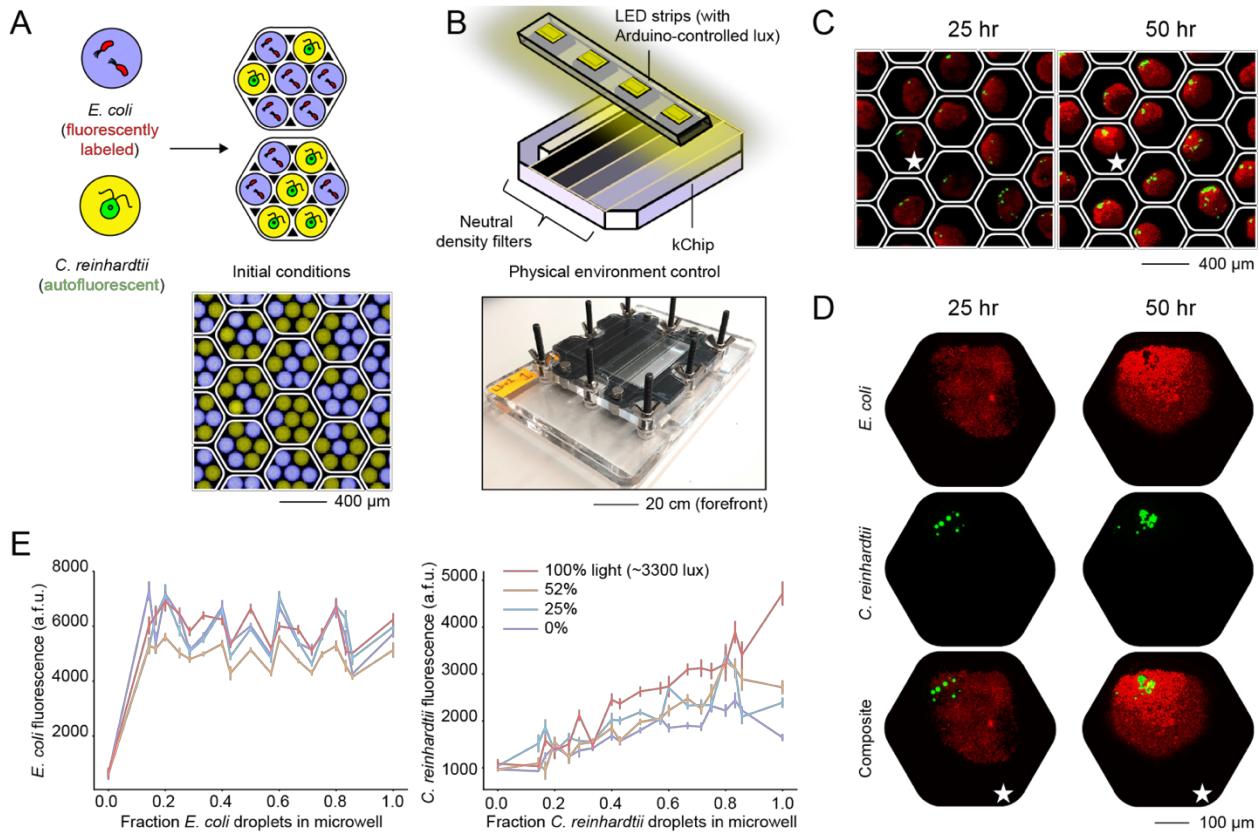
described in **Fig. 3D**). **C.** (Top) Compositions that contain [BaL + Ra] consistency imparted high yield and robustness to Hf-GFP. Compositions were identified that contained [BaL + Ra] (*i.e.* [BaL + Ra + isolate X]) and imparted a facilitative effect size and robustness to Hf-GFP that was greater than that imparted by the composition [BaL + Ra] alone, *e.g.* the composition [BaL + Ra + Ps] (labeled). (Middle, Bottom) The composition BaL and the composition Ra did not impart high robustness to Hf-GFP, and compositions containing BaL or Ra often did not impart robustness. Gray distribution = All compositions (represented  $\geq 30$  times). Red distribution = Compositions that contain at least isolate given in plot title. Black-encircled point = Composition given in title.



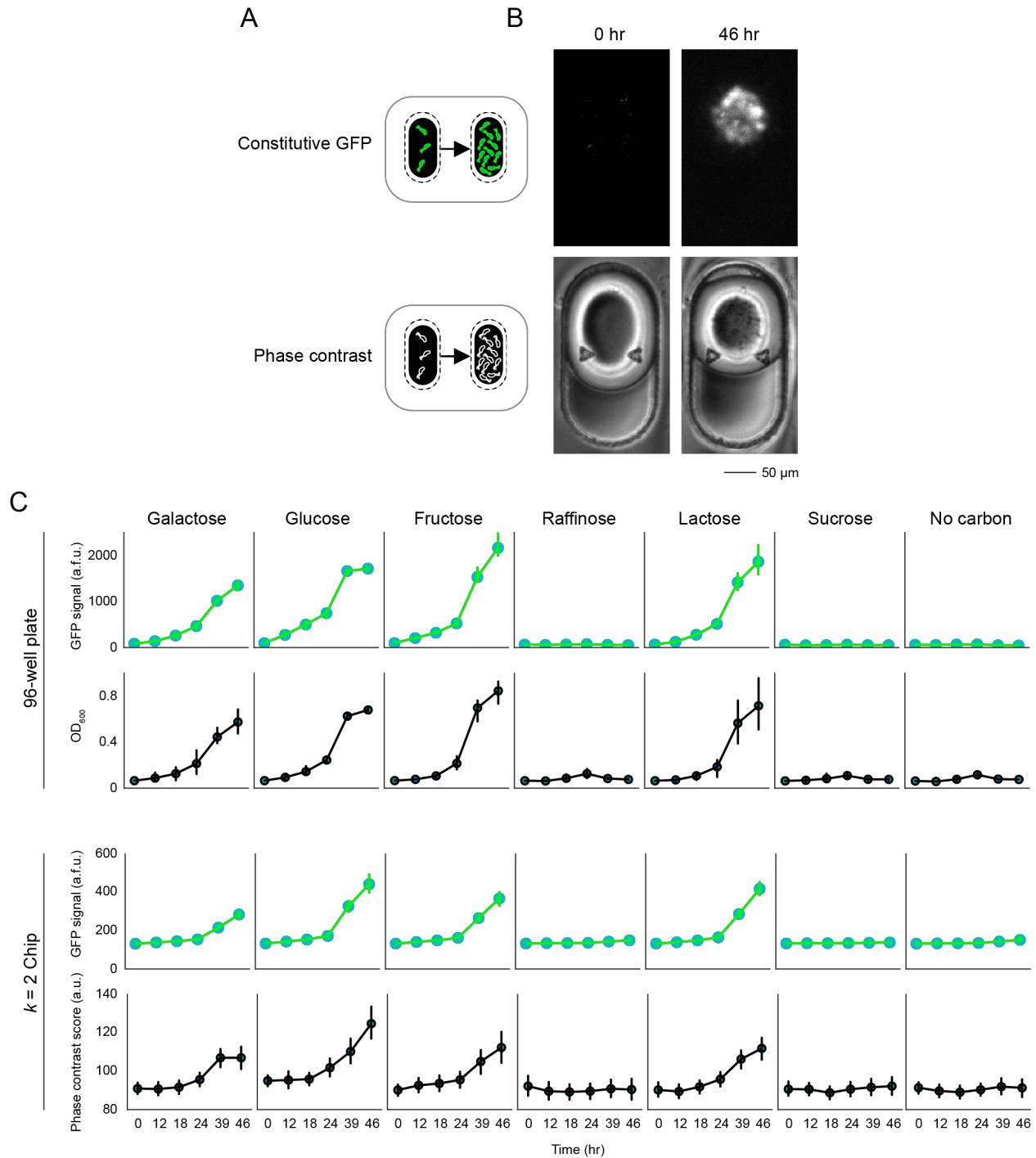
**Figure S24. BSA does not impact microbial growth but affects droplet size.** **A.** We measured growth of a panel of fluorescently labeled strains on different carbon sources with BSA. We observed almost no difference in growth dynamics for 0.05% w/v BSA (1X, the working concentration for droplet generation) or 0.5% BSA (10X) compared to the MM-only control. Full names of labeled strains are listed in **Table S1**. **B.** Different medium formulations and fluorosurfactant (FS) concentrations produced droplets of different sizes, which affects the optimal microwell geometries (**Fig. S2**). We constructed an 8-channel 300- $\mu$ L flow cell, into which droplets of fluorescently dyed media (Alexa Fluor 647) could be loaded and imaged to measure droplet size via the circular Hough transform (implemented in a Matlab script). **C.** The mean and distribution of droplet size were affected by the medium, the inclusion of BSA (which is included to improve the retention of hydrophobic small molecules within the droplets), and FS concentration. We worked with droplets of MM + 0.05% w/v BSA made with 2% w/w FS (to promote droplet stability and inhibit surface wetting). These droplets (distribution highlighted in yellow) had a mean diameter of ~135  $\mu$ m (~1.3 nL). For this medium formulation, droplets became larger as FS concentration increased. Droplet size was notably larger when BSA was not added irrespective of FS concentration.



**Figure S25. Droplet evaporation was typically <10% between 24 and 74 hours.** **A.** For each kChip used in the Hf-GFP facilitation screen (as described in Fig. 3), the area of the merged droplet was measured at each time point (24, 48, and 72 hr). The distributions represent all droplets passing quality filters (with the total number of droplets per distribution listed in **Dataset S2**). Because the height of each droplet is assumed to be flat (in contact with the kChip microwell on one side and glass on the other), and this height is approximately uniform for all microwells, we assume that volume scales linearly with area for this analysis. Lines in each plot connect the mean values of the distributions. Distributions for each kChip at a given time point have been offset to aid in visualization. **B.** The area at 72 hr (final time point) relative to 24 hr (first time point) indicates <10% evaporation for  $k \geq 2$ , and slightly more for  $k = 1$ , in the timespan of the time points.



**Figure S26. Additional kChip screening functionalities include co-culture monitoring and control of physical abiotic environment conditions.** **A.** Experimental setup whereby droplets containing either *E. coli* ( $OD_{600} = 0.01$ , or ~10 cells/droplet) or the phototrophic alga *Chlamydomonas reinhardtii* CC-503 ( $OD_{600} = 0.125$ , or ~1-2 cells/droplet) were randomly grouped on a  $k = 7$  Chip. Each microwell contained a random fraction of each droplet type, such that a wide range of relative starting densities of the two organisms was achieved upon droplet merging. **B.** The kChip was incubated at 30°C with slow shaking (25 RPM) beneath Arduino-controlled LED strips. An array of neutral density filters was constructed to fit on the surface of the loading apparatus (Fig. S1), dividing the kChip into sections exposed to different amounts of illumination (100%, 52%, 25%, and 0%). **C.** Co-cultures were monitored (2X imaging) for 50 hours, enabling the tracking of the abundance of each organism in each microwell. **D.** 10X images of microwells (stars corresponding to microwells in panel c) showing that the spatial arrangement of two co-cultured organisms can also be tracked. **E.** At 50 hours, the yield of each organism was measured for each initial starting fraction and light condition. For each organism in each microwell, a mean fluorescence intensity over the microwell area was measured. Data points represent a median of all replicates and error bars represent standard error. Noise in the curves may be attributable to stochasticity in initial cell count per droplet.



**Figure S27. Phase contrast microscopy enables a label-less and reagent-free growth readout for unlabeled microbes.** **A.** While an assay like constitutive GFP expression requires labeling microbes of interest, phase contrast microscopy allows for the visualization of unlabeled microbes by converting changes in lightpaths caused by the presence of cellular components to brightness changes in the image. Microbial growth in droplets can be inferred through scale-dependent contrast changes over time. **B.** Micrographs of *E. coli* cultures in the same  $k = 2$  microwell (post-merge, 2X magnification) at two timepoints used for the GFP assay (top) and phase contrast assay (bottom). **C.** *E. coli* cultures constitutively expressing

GFP were cultured on a  $k = 2$  Chip (and monitored using GFP expression and phase contrast) and 200  $\mu\text{L}$  cultures in 96-well plates (and monitored using GFP expression and optical density ( $\text{OD}_{600}$ )) in MM containing different single carbon sources (0.5% w/v). The growth metric in phase (“Phase contrast score”) is estimated by measuring local variation in contrast: (1) For each pixel in a given droplet, the range of pixel intensities in a local 10x10 pixel neighborhood centered on each pixel is measured (via the rangefit function implemented in a Matlab script); (2) the mean of these values is taken to measure the contrast of a single droplet; (3) a Phase contrast score is calculated as the median of all replicate measurements (reported here). Based on broad between-assay agreement (GFP signal on kChip and phase contrast score on kChip) and between-platform agreement (phase contrast score on kChip  $\text{OD}_{600}$  in well plates), we conclude that phase contrast is a suitable assay for measuring the growth of a microbe on the kChip. Cultures on plates were conducted in triplicate; the number of replicates on the kChip for each carbon source was {82, 77, 66, 69, 88, 75, 79} (left to right). Error bars represent 95% confidence interval.

Abbreviation	Genus	Species	Strain	Fluorescent protein	Integration/Plasmid	Original source
Ec	<i>Escherichia</i>	<i>coli</i>	K-12 MG1655	GFP	Plasmid (pZE21, Kan resistant)	Jim Collins
Pae	<i>Pseudomonas</i>	<i>aeruginosa</i>	PAO1	GFP	Plasmid	Katharina Ribbeck
Pf	<i>Pseudomonas</i>	<i>fluorescens</i>	A506	YFP	Integration	Nadav Kashtan
Pci	<i>Pseudomonas</i>	<i>citronellolis</i>	ATCC#13674	YFP	Integration	ATCC
Pch	<i>Pseudomonas</i>	<i>chlororaphis</i>	ATCC#9446	YFP	Integration	ATCC
Pv	<i>Pseudomonas</i>	<i>veronii</i>	ATCC#700474	YFP	Integration	ATCC
Pau	<i>Pseudomonas</i>	<i>aurantiaca</i>	ATCC#33663	YFP	Integration	ATCC
Pp	<i>Pseudomonas</i>	<i>putida</i>	ATCC#12633	GFP	Integration	ATCC
Ps	<i>Pseudomonas</i>	<i>syringae</i>	B728a	GFP	Plasmid	Steve E. Lindow
Hf	<i>Herbaspirillum</i>	<i>frisingense</i>	GSF30(T)	GFP	Plasmid (pJBA28, Kan resistant)	Mike Rothballer

**Table S1. Fluorescently labeled strains.** Labeling strains via constitutive fluorescent protein expression, e.g. GFP or YFP, can be used to monitor growth a microbial culture in a droplet. The growth of this panel of labeled strains was used to assess how growth in droplets compared to growth in 96-well plates as well as to compare growth across different carbon sources (**Fig. 2, S5, S6**).

	<b>Compound</b>	<b>Type</b>	<b>Chemical formula</b>	<b>Carbon atoms</b>	<b>Total weight</b>	<b>Component weight</b>	<b>4X concentration (%w/v)</b>
1	D-Ribose	monosaccharide	C5H10O5	5	150.13	150.13	2
2	D-Fructose	monosaccharide	C6H12O6	6	180.16	180.16	2
3	D-Galactose	monosaccharide	C6H12O6	6	180.16	180.16	2
4	D-Glucose (Dextrose)	monosaccharide	C6H12O6	6	180.16	180.16	2
5	N-Acetyl-D-glucosamine (GlcNAc)	monosaccharide	C8H15NO6	8	221.21	221.21	2
6	Sodium citrate dihydrate	carboxylate ion	Na3C6H5O7•2H2O	6	294.1	189.1	2
7	Sodium fumarate dibasic	carboxylate ion	C4H2Na2O4	4	160.04	114.06	2
8	D-Mannitol	sugar alcohol	C6H14O6	6	182.17	182.17	2
9	D-Sorbitol	sugar alcohol	C6H14O6	6	182.17	182.17	2
10	L-Alanine (Ala, A)	amino acid (hydrophobic)	C3H7NO2	3	89.09	89.09	2
11	L-Serine (Ser, S)	amino acid (polar)	C3H7NO3	3	105.09	105.09	2
12	D-Cellobiose	oligosaccharide (glucose + glucose)	C12H22O11	12	342.3	342.3	2
13	D-Sucrose	oligosaccharide (glucose + fructose)	C12H22O11	12	342.3	342.3	2
14	D-Lactose monohydrate	oligosaccharide (glucose + galactose)	C12H22O11•H2O	12	360.31	342.3	2
15	D-Raffinose pentahydrate	oligosaccharide (glucose + galactose + fructose)	C18H32O16•5H2O	18	594.5	504.42	2
16	Arabinogalactan	polysaccharide chain (arabinose + galactose)					2

**Table S2. Carbon compounds used on kChip.** These carbon compounds have been tested for solubility and compatibility with fluorescent dyes used to make up the color codes. They appear in experiments associated assay validation (**Fig. 2**) and/or the Hf-GFP facilitation screen (**Fig. 3, 4**).

### **Additional data Dataset S1.** (attached separately)

**Dataset S1. Soil isolate panel.** These isolates were used in the Hf-GFP facilitation screen (**Fig. 3, 4**). A phylogenetic tree is shown in Fig. S9.

### **Additional data Dataset S2.** (attached separately)

**Dataset S2. Hf-GFP facilitation screen breakdown.** A per-kChip breakdown of the number of individual assays in the Hf-GFP facilitation screen (**Fig. 3, 4**) and the number of instances where constructed communities were composed of all unique strains. An "average" kChip was also calculated and is shown as the first entry in the table. N/A indicates that the number cannot be calculated, *e.g.* because no combinations of 19 unique inputs can be constructed with a library <19 inputs in size. In support of data in **Fig. 4**, where the two control droplets (not containing a bacterial isolate) are not included in the construction of distributions, we include a breakdown of the numbers for the relevant 14 (non-control) inputs.

## References

1. Skhiri Y, et al. (2012) Dynamics of molecular transport by surfactants in emulsions. *Soft Matter* 8(41):10618.
2. Pan M, Lyu F, Tang SKY (2015) Fluorinated Pickering Emulsions with Nonadsorbing Interfaces for Droplet-based Enzymatic Assays. *Anal Chem* 87(15):7938–7943.
3. Gruner P, et al. (2016) Controlling molecular transport in minimal emulsions. *Nat Commun* 7:10392.
4. Kulesa A, Kehe J, Hurtado JE, Tawde P, Blainey PC (2018) Combinatorial drug discovery in nanoliter droplets. *Proc Natl Acad Sci U S A* 115(26):6685–6690.
5. González-Pinzón R, Haggerty R, Myrold DD (2012) Measuring aerobic respiration in stream ecosystems using the resazurin-resorufin system. *Journal of Geophysical Research: Biogeosciences* 117(G3). doi:10.1029/2012jg001965.
6. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947–2948.
7. Lefort V, Longueville J-E, Gascuel O (2017) SMS: Smart Model Selection in PhyML. *Mol Biol Evol* 34(9):2422–2424.
8. Cole JR, et al. (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42(Database issue):D633–42.
9. Celiker H, Gore J (2013) Cellular cooperation: insights from microbes. *Trends Cell Biol* 23(1):9–15.
10. Kowarsky MA, et al. (2017) Humans are colonized by many uncharacterized and highly divergent microbes. doi:10.1101/113746.
11. Hu J, et al. (2016) Probiotic Diversity Enhances Rhizosphere Microbiome Function and Plant Disease Suppression. *MBio* 7(6). doi:10.1128/mBio.01790-16.
12. Borody TJ, Khoruts A (2011) Fecal microbiota transplantation and emerging applications. *Nat Rev Gastroenterol Hepatol* 9(2):88–96.
13. Bhattacharyya PN, Jha DK (2012) Plant growth-promoting rhizobacteria (PGPR): emergence in agriculture. *World J Microbiol Biotechnol* 28(4):1327–1350.
14. Benizri E, Baudoin E, Guckert A (2001) Root Colonization by Inoculated Plant Growth-Promoting Rhizobacteria. *Biocontrol Sci Technol* 11(5):557–574.
15. Caballero S, et al. (2017) Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. *Cell Host Microbe* 21(5):592–602.e4.
16. Lawley TD, et al. (2012) Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice. *PLoS Pathog* 8(10):e1002995.
17. Brugiroux S, et al. (2016) Genome-guided design of a defined mouse microbiota that confers

- colonization resistance against *Salmonella enterica* serovar Typhimurium. *Nat Microbiol* 2:16215.
18. Claußell-Tormos J, et al. (2008) Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms. *Chem Biol* 15(5):427–437.