Full factorial construction of synthetic microbial communities

Juan Diaz-Colunga[1,2] *, Pablo Catalan[3] *, Magdalena San Roman[1,2] *, Andrea Arrabal[1,2] *, and Alvaro Sanchez[1,2] * ✉

[1]Instituto de Biología Funcional y Genómica (IBFG), CSIC & Universidad de Salamanca, 37007 Salamanca, Spain
[2]Centro Nacional de Biotecnologia (CNB), CSIC, 28049 Madrid, Spain
[3]Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganés, Spain
*All authors contributed equally.
✉alvaro.sanchez@usal.es

Abstract

Constructing combinatorially complete species assemblages is often necessary to dissect the complexity of microbial interactions and to find optimal microbial consortia. At the moment, this is accomplished through either painstaking, labor intensive liquid handling procedures, or through the use of state-of-the-art microfluidic devices. Here we present a simple, rapid, low-cost, and highly accessible liquid handling methodology for assembling all possible combinations of a library of microbial strains, which can be implemented with basic laboratory equipment. To demonstrate the usefulness of this methodology, we construct a combinatorially complete set of consortia from a library of eight *Pseudomonas aeruginosa* strains, and empirically measure the community-function landscape of biomass productivity, identify the highest yield community, and dissect the interactions that lead to its optimal function. This easy to implement, inexpensive methodology will make the assembly of combinatorially complete microbial consortia easily accessible for all laboratories.

Introduction

Microbial consortia are playing a growingly important role in biotechnology. They possess substantial potential advantages over monocultures, including larger metabolic capabilities, division of labor, and a potentially higher ecological and evolutionary stability.[1] Synthetic microbial communities are being engineered for goals as diverse as degrading pollutants,[2] producing high-value molecules such as biofuels,[3,4] vitamins,[5,6] and flavonoids,[7,8] and preventing the invasion of pathogens.[9]

To fully realize the promise of microbial consortia in biotechnology we must develop tools for their optimization. Given a library of candidate strains, which consortia should we form if we wish to maximize a function of interest?[10] One way to address this question is purely empirical, following a full factorial design in which we form every possible assemblage and simply select whichever one is best.[11–15] Even for candidate libraries of small size (less than 10 species) the combinatorial nature of this strategy makes the assembly of all possible communities painstaking and tedious, limiting the scale and replicability of full factorial design. The number of unique liquid handling events required to form all possible combinations of m species scales as $m2^{m-1}$, as each of the m species needs to be added to each of the 2^{m-1} consortia where it is present. Manually making all combinations is therefore draining, slow, and prone to human error. Because of the long time required to complete such a task, the risk of contamination is high, and replication is constrained. Perhaps for this reason, studies reporting full factorial construction of microbial consortia are very rare,[11,13,14,16] and the field has largely relied on fractional factorial design, where only a subset of representative species combinations are constructed.[10,17–19]

The ability to form full factorial combinations of a library of species would have benefits beyond biotechnology. In ecology, forming every possible species assemblage is required if we wish to investigate the complex, combinatorial nature of microbial interactions[16,20–25] and the role that high-order interactions play in the function and dynamics of microbial communities.[19,24,26–30] In sum, the quantitative understanding of high-order microbial interactions, the empirical mapping of community assembly networks, the quantitative exploration of community-function landscapes, and the development of predictive models of microbial community assembly and function all require us to be able to reliably assemble communities of growing complexity.[15,18,30]

Robotic liquid handlers can facilitate the task of assembling a full combinatorial set of microbial communities.[11] While this alleviates some of the challenges associated to combinatorial liquid handling, particularly those associated with human error, others remain: the execution of hundreds or thousands of pipetting events is still inevitably slow, and as a result additional equipment (i.e., a HEPA filter) is required to avoid contamination. Robotic liquid handlers are also expensive, technically sophisticated equipment and they are not routinely available. As an alternative to deal with the combinatorial complexity of community assembly, Blainey et al. have introduced a droplet-based microfluidic system (kChip) that is capable of forming hundreds of thousands of species assemblages,[27,31] facilitating their full factorial construction. Notwithstanding their power and unparalleled high-throughput, microdroplet-based approaches such as kChip and others[32] are still state-of-the-art experimental methods, requiring specialized equipment and training that are not yet available to the vast majority of research groups worldwide.

In this paper, we describe a simple, rapid, inexpensive, and highly accessible liquid handling methodology for the full factorial design of microbial consortia and inocula, which requires only basic laboratory equipment and which can be easily implemented by the vast majority of research laboratories. Using a standard multichannel pipette, a single user can manually assemble all possible combinations of up to

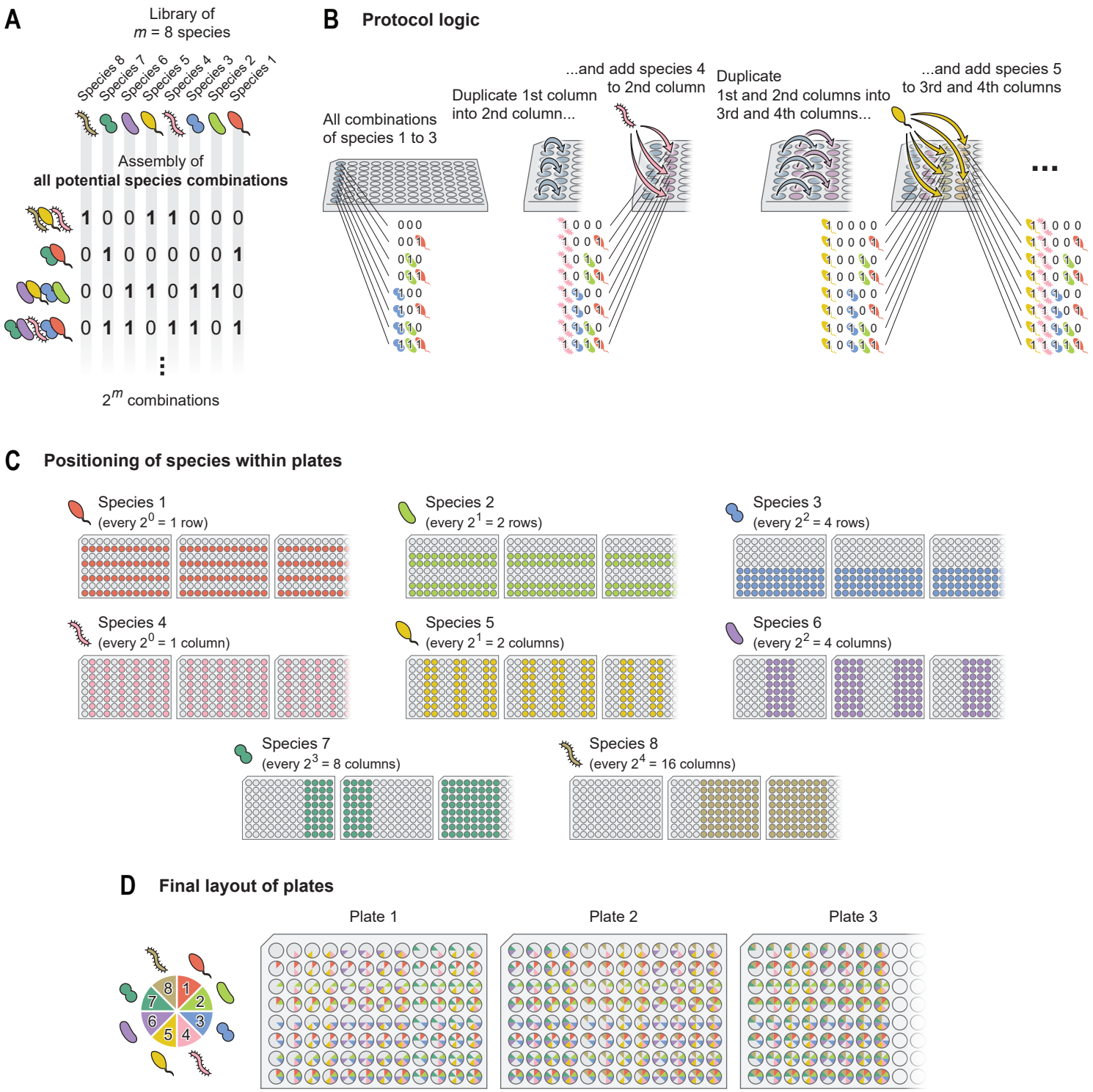

Figure 1. Logic of the assembly protocol. (**A**) We consider a set of m species (in this illustration, $m = 8$), and we identify each of the 2^m consortia that can be formed with them by the binary number $c = x_m x_{m-1} \cdots x_2 x_1$ (with $x_k = 1, 0$ representing the presence/absence of species k, respectively; see equation 1). (**B**) Our protocol is based on the idea that one could assemble each of the $2^3 = 8$ combinations of species 1 to 3 in the first column of a 96-well plate. Iteratively duplicating these consortia and adding species 4 to m to them would eventually result in all 2^m combinations of species, as described in the main text. (**C**) Each of the m species would be present in the indicated wells (see equation 2). (**D**) Positioning of the 2^m consortia within the 96-well plates used in the protocol.

10 species in less than one hour, a timescale that is shorter than the replication time of most bacteria in minimal media. Higher-dimensional designs are possible too, with the only limitation being the plasticware that is required. In addition to providing detailed protocols for a specific set of conditions (e.g., volume of the culture, density of the bacteria, etc.) we provide an R script to help the user tailor the protocol to their needs.

To illustrate the usefulness of this approach, we apply it to empirically characterize the full, combinatorially complete community-function landscape of a model microbial consortium consisting of eight *Pseudomonas aeruginosa* strains. We characterize the full absorption spectrum of all consortia and, as a model function, we determine their total biomass. We show that the implementation of our protocol, enabling the full factorial assembly of all strain combinations, allows

us to quantitatively determine the relationship between community diversity and function, identify optimal strain combinations, and characterize all pairwise and higher-order interactions among all members of the consortia. We expect this methodology will rapidly expand the number of factorially constructed microbial consortia in the literature.

Methods

In what follows, we describe the reasoning behind our protocol in full detail. Readers interested in the practical implementation of this logic are referred to the "Full protocol description" section of the Supplementary Methods, where we provide a step-by-step guide for the assembly of all consortia from a particular example, consisting of a library of eight microbial isolates. In addition, we provide an R script which can be used to generate similar step-by-step guides for different numbers of species and pipetting volumes.

Logic of the assembly protocol

The mathematical basis of our method lies in identifying each microbial consortia by a unique binary number. For a set of m species, we can represent any consortium (which we generically call c) as

$$c = x_m x_{m-1} x_{m-2} \cdots x_2 x_1 \tag{1}$$

where $x_k = 0, 1$ represents the absence (0) or presence (1) of species k in the consortium (Figure 1A). For example, for $m = 6$ species, the consortium where only species 1, 2, and 3 are present would be represented as 000111.

Binary numbers are the natural way to denote combinations in mathematics, and indeed the correspondence between the two sets is often used in combinatorics. For instance, there are $\binom{m}{n}$ combinations of n species from a set of m species. The same number counts how many binary strings of length m contain n ones (equivalently, n zeros).

Now, the most important aspect of this notation is that merging two disjoint consortia becomes a simple addition: if we combine consortium 110000 with consortium 000011, the resulting consortium is represented by the binary number 110011, which is the sum of the binary representations of the initial consortia: 110000 + 000011 = 110011. Note, however, that this operation only works for disjoint consortia. If we add two consortia sharing one or more species, binary addition is no longer a good representation: combining consortium 100001 with consortium 000001 should result in the first consortium again, 100001 (we are for now ignoring potential differences in species abundances/densities, but will address this issue later in the text). Yet, the addition of the two binary representations leads to 100010, which does not represent the appropriate consortium (remember that in binary addition 01 + 01 = 10). As we will see in what follows, our protocol minimizes the number of liquid handling events by making extensive use of this addition property, but only for disjoint consortia.

The second property that we will use to our advantage is that 96-well plates have 8 rows, which is a power of 2 ($2^3 = 8$). This allows us to form all combinations from a 3-species set (say, species 1, 2, and 3) in one column of the plate. It is convenient to arrange these 8 consortia following the order of their binary representation: the empty consortium (000) in the first well, followed by 001, 010, 011, 100, 101, 110, and 111; corresponding to the decimal numbers from 0 to 7 in increasing order.

We can now duplicate these eight consortia, pipetting them into the second column of the 96-well plate, and then pipette species 4 (1000) into all wells of this second column using a multichannel pipette (Figure 1B). This operation is equivalent to the binary addition of consortium 1000 (species 4 alone) with each of the starting eight consortia, and results in all $2^4 = 16$ consortia which can be assembled with species 1 to 4: in the first column, as explained before, we have 0000, 0001, 0010, 0011, 0100, 0101, 0110, and 0111 (decimal numbers 0 to 7); while in the second column we now have 1000, 1001, 1010, 1011, 1100, 1101, 1110, and 1111 (decimal numbers 8 to 15, Figure 1B). We can next duplicate each of these 16 consortia, pipetting the first and second columns of the plate into the third and fourth columns, respectively. If we then add species 5 (10000) to each of the duplicated consortia, we generate all $2^5 = 32$ combinations of species 1 to 5 (Figure 1B).

In what follows, we will describe how this logic can be implemented most efficiently in practice, but hopefully the reader can already intuit how to generalize this algorithm for consortia with more species. Note that for more than 6 species, multiple 96-well plates will be required: for instance, with $m = 7$ one could form $2^7 = 128$ potential consortia, more than what a single 96-well plate can fit. Although we will not make use of them here, the algorithm is also generalizable for 384-well plates, as these have 16 rows. In this case, we would use the first column to assemble all consortia made up by species 1 to 4, and then proceed as described above for all subsequent species.

Plate arrangement

Following the logic we just discussed, it is straightforward to notice that species 4 will be present in all wells of columns 2, 4, 6, 8... of the 96-well plate. In turn, species 5 will be present in alternating pairs of columns: 3, 4, 7, 8, 11, 12... Species 6 will be present in alternating sets of 4 columns: 5, 6, 7, 8, 13, 14, 15, 16... (Figure 1C). Note that 96-well plates only have 12 columns: what we here call columns 13 to 24 would in practice be columns 1 to 12 of a second 96-well plate, columns 25 to 36 would correspond to a third plate, and so on. By a similar reasoning, species 1 to 3 will be present in alternating sets of 1, 2, and 4 rows (instead of columns) respectively (Figure 1C). In summary, we have that species k will be present in:

Species k present in:

$$\begin{cases} \text{alternating sets of } 2^{k-1} \text{ rows} & \text{if } k = 1, 2, 3 \\ \text{alternating sets of } 2^{k-4} \text{ columns} & \text{if } k = 4, 5, \cdots, m \end{cases} \tag{2}$$

Conveniently, this arrangement makes it so consortia are located in order from top to bottom and from left to right within the plate(s): the first column contains the consortia corresponding to the binary representations of decimal integers 0 (00000) to 7 (00111), the second column corresponds to decimal integers 8 (01000) to 15 (01111), the third column to decimal integers 16 (10000) to 23 (10111), and so on.

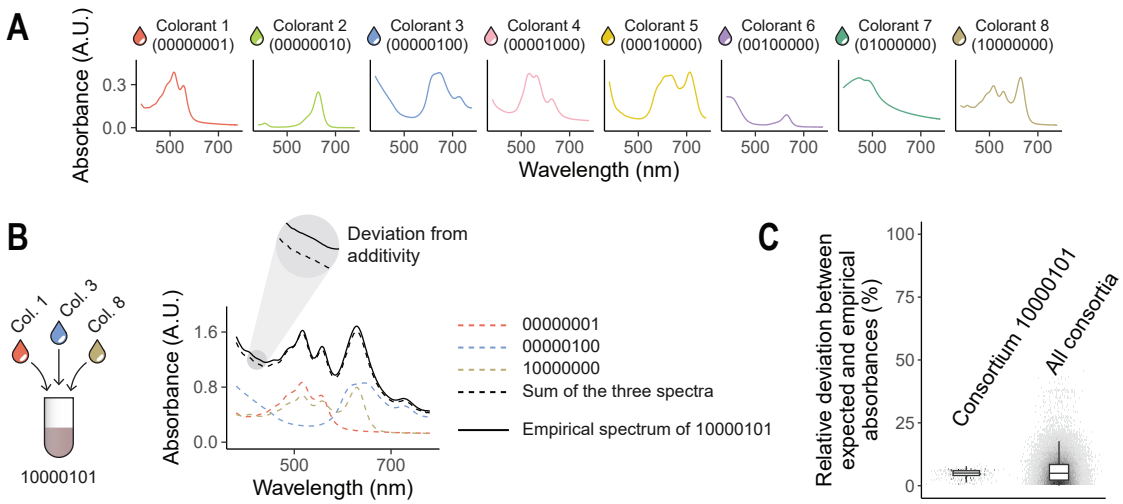

Figure 2. Full factorial construction of synthetic colorant combinations. (**A**) We consider a set of eight synthetic colorants, each of which exhibits a different absorbance spectrum in the range 380-780 nm. (**B**) As an example, we show the empirically measured absorbance spectrum (solid black line) for the combination of colorants 1, 3, and 8 (10000101). The dashed black line represents the additive expectation, i.e., the sum of the spectra of the three constituent colorants (colored dashed lines). (**C**) Relative deviation between the empirical spectrum and the additive expectation (see equation 6) for the example combination 10000101 (left) and for all mixtures of 2 or more colorants (right). Relative errors were only computed when the true absorbance was above 0.1 A.U..

Homogenizing species' densities across consortia

As we just saw, it would be possible to form all combinations of m species simply by pipetting each one of them row-wise (for species 1 to 3) or column-wise (for species 4 to m) in their corresponding positions (equation 2) with the aid of multichannel pipettes. There is, however, one important caveat: if we pipette equal volumes of each species' monocultures, we will end up with variable volumes across wells. As an example, if we pipette a volume v_0 of each monoculture, we will have a volume of $2v_0$ in well D01 of the 96-well plate (where consortium 011 is located), but a volume of $3v_0$ in well H01 (where consortium 111 is located). Species 1 is present in both of these example consortia at the same total population size, but at different population densities differing by a factor of 2/3.

Often, we may want the density of each species to be consistent across consortia. To achieve this, we can pipette additional liquid (sterile water, growth medium, saline buffer...) into each well in order to compensate for the differences in volume. Note that, prior to this density homogenization step, the maximum volume in a well would be mv_0, corresponding to the consortium with all m species. This naturally also corresponds to the consortium where each of the individual species is present at the lowest population density. In every other consortium (say, consortium c), the volume of buffer (which we denote $v_B(c)$) that we would need to add in order to reach this minimum population density is given by

$$v_B(c) = v_0\left(m - H(c)\right) \tag{3}$$

where $H(c)$ represents the Hamming weight of the binary representation of consortium c. The Hamming weight of a binary number is the number of ones in its representation, for instance, $H(0000) = 0$, $H(0010) = 1$, $H(1111) = 4$, etc. Thus, $H(c)$ simply counts the number of species present in consortium c.

In practice, we do not need to manually pipette the buffer well-by-well. The convenient arrangement of the plate allows us to streamline this process. In what follows, we use i and j as the row and column indexes of the plate (from top to bottom and from left to right, respectively). For instance, well C02 (containing consortium 1010) would correspond to $i = 3$ and $j = 2$. As a reminder, we are considering that $j = 13$ to 24 correspond to columns 1-12 of a second 96-well plate, $j = 25$ to 36 correspond to a third plate, etc.

In order to streamline the pipetting of buffer, it would be convenient to find an expression for $H(c)$ in terms of the positional indexes i and j. Now, due to the particular arrangement of consortia in the plate, we can divide the binary representation of a consortium c into two. First, the rightmost three bits of c are determined by the row in which it is located. For instance, all consortia in row 7 will always have a binary representation ending in 110, independently of the column in which they appear. In general, the three rightmost digits of c will be equal to $B(i-1)$, the binary representation of $i-1$ with i being the index of the row where consortium c is located. Similarly, the leftmost $m-3$ bits of c will depend only on the column, in the following way: take a column j and ignore the three rightmost digits; then the binary representation of all consortia in that column will be the same, and equal to $B(j-1)$. For instance, with $m = 8$, the binary representation of all consortia in column 4 is $00011x_3x_2x_1$ — the last three bits may be 0 or 1 depending on the specific well, but the first five bits are common to all wells in the column. Therefore, in order to calculate the Hamming weight of a given consortium, we just need to sum the Hamming weight of the row and the Hamming weight of the column:

$$H(c) = H\Big(B(i-1)\Big) + H\Big(B(j-1)\Big) \tag{4}$$

We can use equation 4 to rewrite equation 3 in terms of the positional indexes i and j. After some algebra, we arrive at

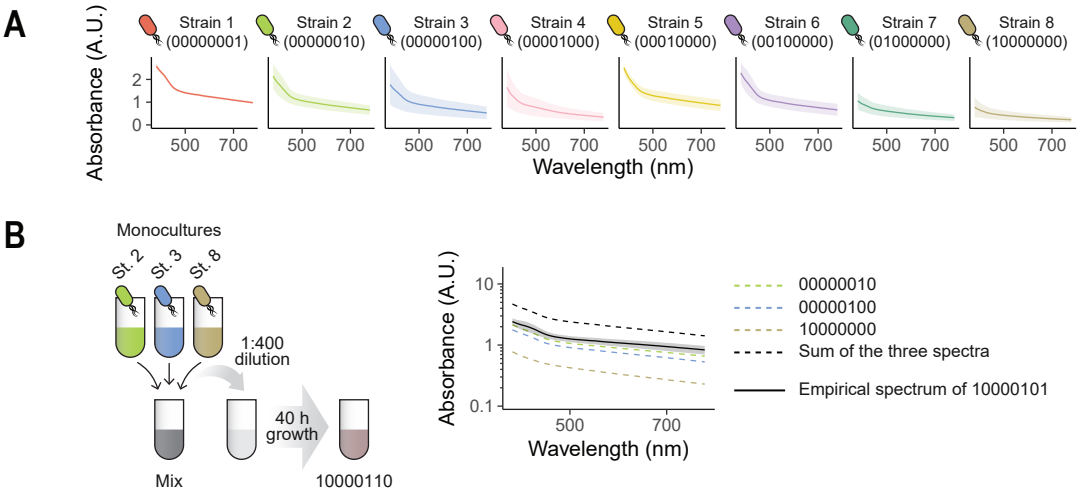

Figure 3. Co-culturing combinations of bacterial isolates. (**A**) We consider eight *P. aeruginosa* strains obtained from a previous experiment.[33–35] Here we represent the absorbance spectra of the eight monocultures of these strains. Solid lines and shaded areas represent means and standard deviations across three independent biological replicates. (**B**) We followed our protocol to assemble different combinations of strains (in this example, the consortium formed by strains 2, 3, and 8, represented as 10000110). Each consortium was diluted into fresh medium by a factor of 1:400 and allowed to re-grow for 40 h, after which the absorbance spectrum of the culture was measured. Solid black line: empirical spectrum of consortium 10000110 (mean of three independent biological replicates, shaded region represents the standard deviation across them). Dashed black line: additive expectation if the three strains did not interact. Dashed colored lines: monoculture spectra of the three constituent strains.

$$\begin{aligned} v_B(i,j) = & v_0 \left[3 - H\Big(B(i-1)\Big)\right] \\ & + v_0 \left[m - 3 - H\Big(B(j-1)\Big)\right] \end{aligned} \tag{5}$$

While terms in this equation could be arranged in a more compact form, equation 5 as it is allows us to implement the buffer pipetting in a very efficient way: We can start by pipetting liquid buffer column-wise (a volume $v_0\,[m-3-H(B(j-1))]$ into every well of column j). We can then pipette the buffer row-wise, pipetting a volume $v_0\,[3-H(B(i-1))]$ into every well of row i. Furthermore, the row-wise buffer pipetting can be bypassed if we homogenize the densities of the starting consortia formed with species 1 to 3, as we explain in detail in the Supplementary Methods (section "Full protocol description for $m = 8$ species").

Results

Proof of concept: construction of colorant combinations

As a first demonstration of the method, and to establish its feasibility and accuracy, we used commercial food colorants and temperas to build all 256 combinations made up from eight colors. We diluted these colorants in water such that all of them had comparable maximum absorbances in the range from 380 nm to 780 nm. Each color exhibits a different absorbance spectrum in this range (Figure 2A).

Using synthetic colorants to test our protocol has various advantages. First, color differences are clearly visible to the naked eye, making it easy to keep track of the location of each colorant. Second, we do not expect these colorants to interact with one another, or to do so very weakly. This means that the absorbance spectrum of any combination of colorants may be reasonably expected to match the sum of the individual spectra of each colorant. Therefore, the magnitude of the deviations between an empirical spectrum and this additive expectation (Figure 2B) can be used to estimate an upper bound for the pipetting error of the protocol. We can obtain an estimation of the accuracy of our protocol by examining these deviations.

We followed the protocol described above, setting $m = 8$ and $v_0 = 25$ µL (a detailed step-by-step protocol for these parameters is given in the Supplementary File `protocol_8species_25uL.txt`), to assemble all 256 color combinations. The absorbance spectrum of each combination was quantified using a MultiSkan SkyHigh plate reader (Thermo Scientific). The implementation of the protocol required three 96-well plates, 8 Falcon tubes, an 8-channel pipette and a 12-channel pipette. Completing the whole protocol took a single experimentalist less than an hour. Figure S1 shows the three 96-well plates at six different steps of the protocol.

Figure 2B shows the absorbance spectrum for an example colorant mixture (10000101), comparing it to the additive expectation that its three constituent colorants (1, 3, and 8) do not interact. We estimated the pipetting error of the protocol as the relative deviation (denoted δ) between the empirical spectra and the additive expectations at all wavelengths (Figure 2B), that is:

$$\delta(c,\lambda) = \frac{\left|Abs_\lambda(c) - Abs_\lambda^{(\mathrm{add})}(c)\right|}{Abs_\lambda(c)} \tag{6}$$

where we have denoted $Abs_\lambda(c)$ the absorbance at a wavelength λ (between 380 and 780 nm) of the colorant mixture c, and $Abs_\lambda^{(\mathrm{add})}(c)$ the additive expectation (i.e., the sum of the absorbances of each individual constituent colorant).

In Figure 2C we show the distribution of relative deviations for the example mixture 10000101 at all wavelengths (left), as well as the distribution corresponding to all mixtures at all wavelengths (right). Figure S2 shows the ab-

Figure 4. Full factorial design of synthetic microbial consortia. We represent the absorbance spectra of all consortia (of 2 or more strains) that can be formed with our library of eight *P. aeruginosa* isolates. As in Figure 3B, solid black lines and shaded areas represent the mean and standard deviation of the co-culture spectrum across three biological replicates; the dashed colored lines represent the individual monoculture spectra; and dashed black lines represent the sum of these individual spectra.

solute deviations ($|Abs_\lambda(c) - Abs_\lambda^{(\mathrm{add})}(c)|$) for all mixtures at all wavelengths (median absolute deviation ~ 0.025 A.U.). We found that relative deviations were generally small (Figure 2C: mean ~ 5.8 %, median ~ 4.9 %). We also found that the magnitude of these deviations did not increase with the number of colorants in a mixture (Figure S3), suggesting that the precision of our protocol is not compromised by the size of the set from which mixtures are assembled. Deviations were also roughly similar in magnitude across all wavelengths (Figure S4).

Full factorial design of synthetic bacterial consortia

To illustrate the potential applications of our protocol for finding optimal microbial consortia and quantitatively studying the complexity of microbial interactions, we adopted a collection of eight *Pseudomonas aeruginosa* strains obtained from a previous experiment.[33–35] Monocultures of each of these strains grown in liquid LB medium exhibit differences in their absorbance spectra (Figure 3A). Unlike the synthetic colorants, bacterial strains can be expected to interact with one another through a variety of potential mechanisms, including competition for limiting resources or facilitation which may lead to changes in the population sizes of different species.[15,19]

We first cultured each strain separately by resuspending a single colony into 15 mL of LB medium in 50 mL conical Falcon tubes. Strains were allowed to grow to carrying capacity for 24 h at 37 °C with no agitation. Then, monocultures were fully homogenized and used to assemble all 256 strain combinations following the same protocol described before (detailed steps can be found in the Supplementary Methods, section "Full protocol description for m = 8 species"). Phosphate buffer saline (1×PBS) was used to adjust volumes/densities as indicated in the protocol. After completing our protocol (which took a single experimentalist less than one hour), each consortium was inoculated into fresh LB medium (0.5 µL of inoculum into 200 µL of medium for a dilution factor of 1:400; note that our protocol design makes it so this is equivalent to inoculating 0.5 µL/8 = 0.0625 µL of each strain's monoculture into the fresh medium) in 96-well plates. Samples were incubated still at 37 °C for 40 h, after which the absorbance spectrum of each sample was quantified in the same plate reader we used before.

Figure 3B shows the absorbance spectrum for one example consortium after the 40 h incubation period. Consistent with our reasoning, this spectrum deviates strongly from the additive expectation (sum of each monoculture spectrum), indicating a prominent effect of interactions between strains. Figure 4 shows the spectra for all consortia with two or more members, where deviations from the additive expectation are similarly large.

Because we have constructed all possible assemblages, we can systematically investigate the interactions between species and how they may result in different community-level properties and functions. As an illustration, we focus on the absorbance at 600 nm (Abs_{600}), a metric of total biomass in our cultures. We found a non-monotonic relationship between community biomass and community richness (number of strains), peaking at 3 strains (Figure 5A). Our protocol allows us not only to characterize this type of broad-scale relationships between diversity and function, but also to identify specific high-performing consortia (Figure 5B). In our experiment, the consortium with the highest biomass contains only three strains (consortium 00001101, containing strains 1, 3, and 4). In Figure 5C we represent the full map between the composition and function (biomass) of all 256 consortia in our experiment (i.e., the *community structure-function landscape*[15,16,19,22]).

From this type of data, one can readily quantify the sign and magnitude of strain interactions with respect to the community-level function of interest, in our case total biomass. In Figure 5D we show that the inclusion of strain 1 in a given consortium (which we refer to as an *ecological background*) results in an increase in biomass, and the same is true for the inclusion of strain 2. We call the difference in function between two consortia with and without a given focal strain as the *functional effect* of that strain in that ecological background.[10] If two strains did not interact, we may expect that including both of them in a same ecological background would result in a change in function equal to the sum of their two individual functional effects (see Figure 5D, dashed line represents this additive expectation). If the empirical function deviates from the additive expectation, the two strains engage in a *functional interaction* (denoted as ϵ in Figure 5D) in that particular ecological background.[10,19]

Following this definition, one could dissect the interaction structure of all eight *P. aeruginosa* strains in our system. As an illustration, in Figure 5E we show the community-function landscape corresponding to all consortia that could be formed solely with strains 1, 3, and 4, i.e., the trio that maximizes biomass yield (see Figure 5B-C indicating that the maximum Abs_{600} is achieved by consortium 00001101). Although pairwise interactions are negative within this consortium, the three strains engage in a positive third-order functional interaction (see inset in Figure 5E) which rescues the function of the trio. This observation highlights the potential of full factorial design to interrogate and generate hypotheses with respect to the mechanisms that govern the emergence of community-level functions, particularly (but not only) in high-performing consortia.

Beyond this particular consortium, in Figure 5F we represent the distributions of functional effects of all eight strains in our library across all of the 128 ecological backgrounds where they may be included. We found that the functional effects of all eight *P. aeruginosa* strains exhibit large variation in sign and magnitude, indicating that the contribution of a strain to the total community biomass largely depends on the presence or absence of additional community members in the ecological background. This indicates a prominent role of inter-strain interactions in this system.

In Figure 5G we represent the magnitude of the functional interactions between all pairs of strains i and j (denoted ϵ_{ij}) across all the ecological backgrounds where the two may be included. We found that these pairwise interactions are highly variable across backgrounds. Pairwise interactions being sensitive to the presence or absence of additional community members can be interpreted as a signature of higher-order ecological interactions (HOIs).[19,36,37] By this definition, HOIs do exist in our particular experimental system (as evidenced by the variation in the pairwise interaction terms ϵ_{ij} shown in Figure 5G); however, most of the variance in community biomass can be explained by low-order interactions (Figure 5H, the fraction of functional variance attributed to pairwise and higher-order interactions was computed as described elsewhere;[30] see Supplementary Meth-

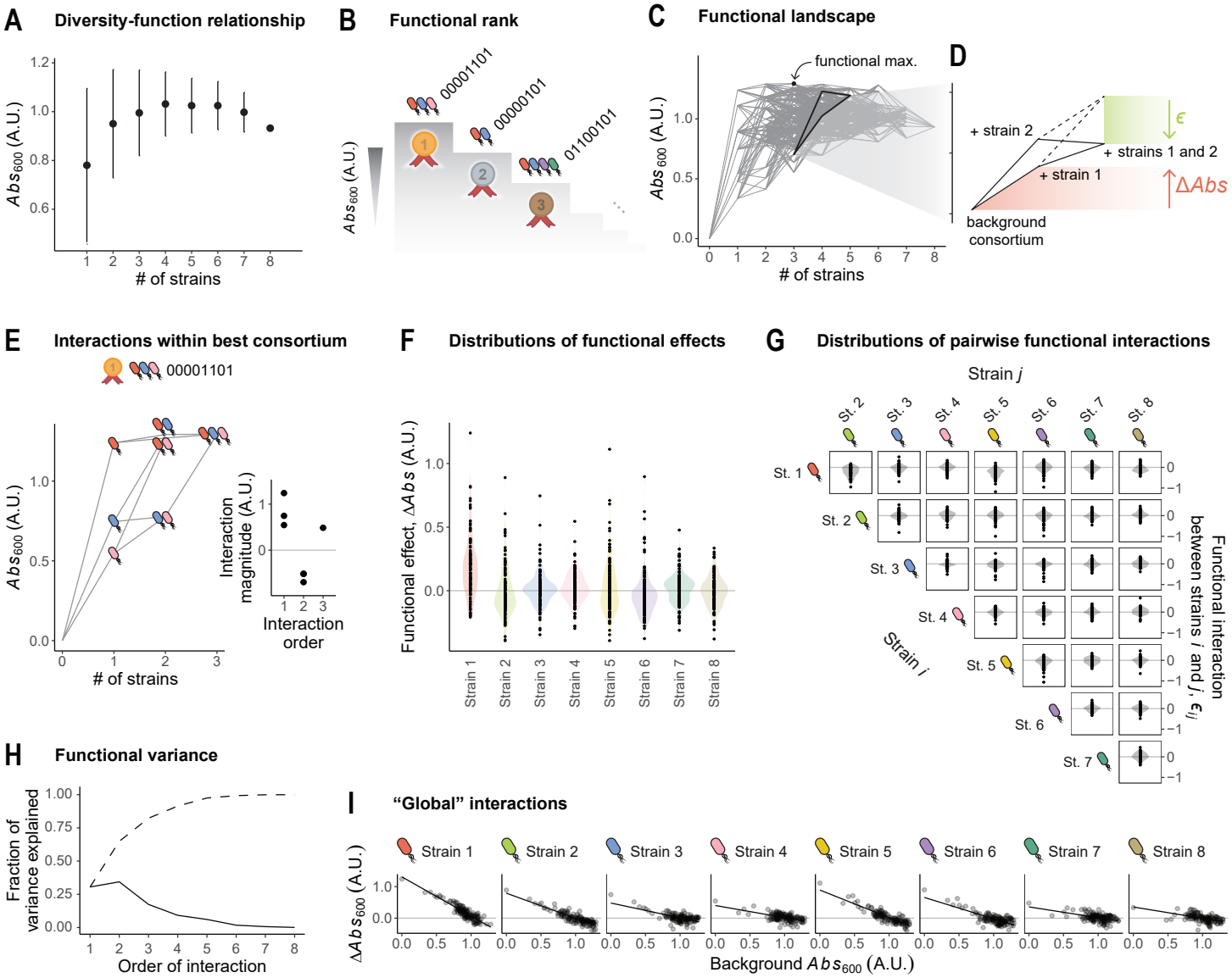

Figure 5. Full factorial design of microbial consortia enables the dissection of microbial interactions. We focus on the absorbance at 600 nm as a proxy for the biomass of a consortium. (**A**) Consortium function (biomass quantified as Abs_{600}) against the number of strains in the consortium. Dots and error bars represent the average and standard deviation across all consortia with a same number of strains. (**B**) Top three consortia exhibiting the highest biomass across all 2^8 combinations. (**C**) We represent the function (biomass; values shown are averages across three biological replicates) for the consortia assembled using our protocol. Nodes correspond to the different consortia, edges connect consortia that differ in the presence of a single particular strain. (**D**) Detail showing the effect of including strain 1, strain 2, and both, in a background consortium (in this example, the background corresponds to consortium 11001000). Red: the *functional effect* of strain 1 is quantified as the difference in function between the background consortium and the consortium resulting from the inclusion of that strain. Green: the *functional interaction* between strains 1 and 2 in this particular background is quantified as the difference between the additive expectation (dashed line) and the empirical value for the function of the consortium containing both strains, as explained in the main text. (**E**) Detail of the functional landscape corresponding to the three strains that are part of the overall highest-biomass consortium (00001101). Inset: functional interactions within this 3-strain consortium. Order 1 corresponds to the monoculture functions. Order 2 corresponds to pairwise species-by-species interactions as defined in panel D of this figure. Order 3 corresponds to the third-order interaction between these strains, defined as the deviation between the empirical function of the trio with respect to the null expectation from the additive and pairwise interactions.[19,29] [19,29] (**F**) Distributions of functional effects of the eight strains used in our experiments across all of the 128 backgrounds where each may be included. (**G**) Distributions of functional interactions between all pairs of strains across all backgrounds where both may be included. (**H**) Fraction of functional variance due to interactions of different orders (order 1 corresponds to the additive effects of each strain). The fractions were computed as explained in the Supplementary Methods.[30] Dashed line is the cumulative sum. (**I**) We have recently shown that the functional effect of a strain is often predictable from the function of the background consortium through simple, linear regressions.[10] For the eight *P. aeruginosa* strains in our experiment, linear regressions do an excellent job at linking these functional effects with the function of the background consortia.

ods and Figure S5 for details). This has also been shown to be the case for other datasets of microbial community function.[30]

We have recently shown that the aggregate effect of pairwise and higher-order microbial interactions often results in emergent statistical patterns, which make it so the functional effect of a strain may be predictable from the function of the ecological background where that strain is included.[10] The relationships between the functional effect of a strain and the function of the background community can often be cap-

tured through simple linear models, which mirror the patterns of *global epistasis* commonly reported in genetics. This phenomenon is also observed in the particular set of *P. aeruginosa* strains that we analyzed here (Figure 5I, R^2 averages 0.54), each of which exhibits a different relationship between its functional effect and the background function. Adopting the same analytical approach we used in previous work,[10] we find that such strong global epistasis patterns deviate from the null expectation of a random community-function landscape (Figure S6).

Together, these analyses illustrate how a full factorial construction of microbial consortia may allow us both to identify optimal assemblages and to dissect how the interplay of pairwise and higher-order interactions shape community-level properties and functions. To our knowledge, the experiments we have just reported represent one of barely a handful of datasets consisting of every potential combination from a library of strains. The fact that it was done in under an hour per replicate gives a sense of the potential our design has to dramatically expand the number of empirical community-function landscapes in the literature.

Discussion

We have introduced a rapid, simple, and inexpensive protocol for assembling all possible combinations of a given set of species using basic laboratory equipment such as a multichannel pipette and 96-well plates. Our protocol enables a single individual to construct all 256 possible consortia from a pool of eight species within an hour, and can be easily expanded to even higher-diversity libraries. Even without the aid of specialized equipment such as robotic liquid handlers, this procedure will make it possible to efficiently generate all communities with up to 10-12 species.

For laboratories equipped with robotic liquid handlers, the protocol can be easily expanded to accommodate larger libraries, with the main limitation being the amount of plasticware that will be used. The rationale underlying our assembly protocol allows for flexible implementation depending on the specific needs of each laboratory (e.g., it is straightforward to generalize for 384-well plates). Our implementation also has the advantage of keeping pipetting error minimal and independent of consortium size, as we showed by creating all 256 color combinations from eight synthetic colorants. To facilitate automation and adaptation to the needs of other groups, we provide an accompanying R script to facilitate the generation of comprehensive step-by-step protocols for any number of species and pipetting volumes.

We have demonstrated the utility of this protocol in synthetic microbial ecology by generating all possible communities from a collection of eight *P. aeruginosa* strains and quantifying the strength of pairwise and higher-order functional interactions between them. Our experiment illustrates how the full factorial design of microbial consortia enables for the full characterization of ecological community-function landscapes, which can provide valuable insights into the emergence of community-level functions (such as, but not limited to, biomass productivity, as we studied here). The entire experiment, from spreading colonies to measuring community function, could be completed from beginning to end by a single experimentalist in less than one working week, and most of that time was taken by waiting for cell growth.

Although initially designed for building combinations of microbial species, our protocol can be readily applied for creating combinations of any soluble (or at least homogeneously distributed) compounds beyond microorganisms. For instance, it could be employed to generate all possible media compositions from a set of resources, all possible combinations of a set of antibiotics, toxins, or even bacteriophages, as well as mixed combinations thereof. We therefore anticipate that our methodology will find extensive applications and utility across diverse fields such as microbiology, ecology, evolution, and biotechnology.

Acknowledgements

We thank Sara Hernando-Amado for providing the *P. aeruginosa* strains used in our experiment. A.S. and J.D.-C. acknowledge support from grant PID2021-125478NAI00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF: A way of making Europe." P.C. acknowledges support from grants PID2022-142185NB-C21 and PID2022-142185NB-C22 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF: A way of making Europe." A.S. acknowledges funding by the European Union (ERC, *ECOPROSPECTOR*, 101088469). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Data and code availability

All original data and code for analysis is available at github.com/jdiazc9/full_factorial_design. This repository also contains the R script for the automatic generation of assembly protocols with custom parameters.

Competing interests

The authors declare no competing interests.

References

1. Brenner K, You L and Arnold FH (2008). Engineering microbial consortia: a new frontier in synthetic biology. *Trends in Biotechnology* **26(9)**:483–489

2. Arias-Sanchez FI, Vessman B, Haym A, Alberti G and Mitri S (2023). Artificial selection optimizes pollutant-degrading bacterial communities. *bioRxiv* pp. 2023–07

3. Minty JJ, Singer ME, Scholz SA, Bae CH, Ahn JH, Foster CE, Liao JC and Lin XN (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proceedings of the National Academy of Sciences* **110(36)**:14592–14597

4. Senne de Oliveira Lino F, Bajic D, Vila JCC, Sánchez A and Sommer MOA (2021). Complex yeast–bacteria interactions affect the yield of industrial ethanol fermentation. *Nature Communications* **12(1)**:1498

5. Wang EX, Ding MZ, Ma Q, Dong XT and Yuan YJ (2016). Reorganization of a synthetic microbial consortium for one-step vitamin C fermentation. *Microbial Cell Factories* **15**:1–11

6. Wang Y, Liu L, Jin Z and Zhang D (2021). Microbial cell factories for green production of vitamins. *Frontiers in Bioengineering and Biotechnology* **9**:661562

7. Du Y, Yang B, Yi Z, Hu L and Li M (2020). Engineering Saccharomyces cerevisiae coculture platform for the production of flavonoids. *Journal of Agricultural and Food Chemistry* **68(7)**:2146–2154

8. Xu P, Marsafari M, Zha J and Koffas M (2020). Microbial coculture for flavonoid synthesis. *Trends in Biotechnology* **38(7)**:686–688

9. Oliveira R, Pandey B, Lee K, Yousef M, Chen RY, Triebold C, McSpadden E, Haro F, Aksianiuk V et al. (2024). Statistical design of a synthetic microbiome that clears a multi-drug resistant gut pathogen. *bioRxiv* pp. 2024–02

10. Diaz-Colunga J, Skwara A, Vila J, Bajic D and Sanchez A (2022). Global epistasis and the emergence of ecological function. *bioRxiv* pp. 2022–06

11. Angeles-de Paz G, Ledezma-Villanueva A, Robledo-Mahón T, Pozo C, Calvo C, Aranda E and Purswani J (2023). Assembled mixed co-cultures for emerging pollutant removal using native microorganisms from sewage sludge. *Chemosphere* **313**:137472

12. Kuebbing SE, Classen AT, Sanders NJ and Simberloff D (2015). Above-and below-ground effects of plant diversity depend on species origin: an experimental test with multiple invaders. *New Phytologist* **208(3)**:727–735

13. Langenheder S, Bulling MT, Solan M and Prosser JI (2010). Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity. *PLOS One* **5(5)**:e10834

14. Purswani J, Romero-Zaliz RC, Martín-Platero AM, Guisado IM, González-López J and Pozo C (2017). BSocial: deciphering social behaviors within mixed microbial populations. *Frontiers in Microbiology* **8**:919

15. Sanchez A, Bajic D, Diaz-Colunga J, Skwara A, Vila JC and Kuehn S (2023). The community-function landscape of microbial consortia. *Cell Systems* **14(2)**:122–134

16. Gould AL, Zhang V, Lamberti L, Jones EW, Obadia B, Korasidis N, Gavryushkin A, Carlson JM, Beerenwinkel N et al. (2018). Microbiome interactions shape host fitness. *Proceedings of the National Academy of Sciences* **115(51)**:E11951–E11960

17. Clark RL, Connors BM, Stevenson DM, Hromada SE, Hamilton JJ, Amador-Noguez D and Venturelli OS (2021). Design of synthetic human gut microbiome assembly and butyrate production. *Nature Communications* **12(1)**:3254

18. Eng A and Borenstein E (2019). Microbial community design: methods, applications, and opportunities. *Current Opinion in Biotechnology* **58**:117–128

19. Sanchez-Gorostiaga A, Bajić D, Osborne ML, Poyatos JF and Sanchez A (2019). High-order interactions distort the functional landscape of microbial consortia. *PLOS Biology* **17(12)**:e3000550

20. Baichman-Kass A, Song T and Friedman J (2023). Competitive interactions between culturable bacteria are highly nonadditive. *eLife* **12**:e83398

21. Bergelson J, Kreitman M, Petrov DA, Sanchez A and Tikhonov M (2021). Functional biology in its natural context: A search for emergent simplicity. *eLife* **10**:e67646

22. Eble H, Joswig M, Lamberti L and Ludington WB (2023). Master regulators of biological systems in higher dimensions. *Proceedings of the National Academy of Sciences* **120(51)**:e2300634120

23. Friedman J, Higgins LM and Gore J (2017). Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution* **1(5)**:0109

24. Ludington WB (2022). Higher-order microbiome interactions and how to find them. *Trends in Microbiology* **30(7)**:618–621

25. Picot A, Shibasaki S, Meacock OJ and Mitri S (2023). Microbial interactions in theory and practice: when are measurements compatible with models? *Current Opinion in Microbiology* **75**:102354

26. Ansari AF, Acharya NI, Kumaran S, Ravindra K, Reddy YB, Dixit NM and Raut J (2019). 110th anniversary: high-order interactions can eclipse pairwise interactions in shaping the structure of microbial communities. *Industrial & Engineering Chemistry Research* **58(51)**:23508–23518

27. Kehe J, Kulesa A, Ortiz A, Ackerman CM, Thakku SG, Sellers D, Kuehn S, Gore J, Friedman J et al. (2019). Massively parallel screening of synthetic microbial communities. *Proceedings of the National Academy of Sciences* **116(26)**:12804–12809

28. Morin MA, Morrison AJ, Harms MJ and Dutton RJ (2022). Higher-order interactions shape microbial interactions as microbial community complexity increases. *Scientific Reports* **12(1)**:22640

29. Sanchez A (2019). Defining higher-order interactions in synthetic ecology: lessons from physics and quantitative genetics. *Cell Systems* **9(6)**:519–520

30. Skwara A, Gowda K, Yousef M, Diaz-Colunga J, Raman AS, Sanchez A, Tikhonov M and Kuehn S (2023). Statistically learning the functional landscape of microbial communities. *Nature Ecology & Evolution* **7(11)**:1823–1833

31. Kehe J, Ortiz A, Kulesa A, Gore J, Blainey PC and Friedman J (2021). Positive interactions are common among culturable bacteria. *Science Advances* **7(45)**:eabi7159

32. Hsu RH, Clark RL, Tan JW, Ahn JC, Gupta S, Romero PA and Venturelli OS (2019). Microbial interaction network inference in microfluidic droplets. *Cell Systems* **9(3)**:229–242

33. Hernando-Amado S, Sanz-García F and Martínez JL (2020). Rapid and robust evolution of collateral sensitivity in Pseudomonas aeruginosa antibiotic-resistant mutants. *Science Advances* **6(32)**:eaba5493

34. Hernando-Amado S, Laborda P, Valverde JR and Martínez JL (2022). Mutational background influences P. aeruginosa ciprofloxacin resistance evolution but preserves collateral sensitivity robustness. *Proceedings of the National Academy of Sciences* **119(15)**:e2109370119

35. Hernando-Amado S, Laborda P, Valverde JR and Martínez JL (2022). Rapid decline of ceftazidime resistance in antibiotic-free and sublethal environments is contingent on genetic background. *Molecular Biology and Evolution* **39(3)**:msac049

36. Billick I and Case TJ (1994). Higher order interactions in ecological communities: what are they and how can they be detected? *Ecology* **75(6)**:1529–1543

37. Letten AD and Stouffer DB (2019). The mechanistic basis for higher-order interactions and non-additivity in competitive communities. *Ecology Letters* **22(3)**:423–436

Supplementary Methods

Full protocol description for $m = 8$ species

In this section, we describe our protocol for the assembly of all consortia from a $m = 8$ species library in full detail. The reader will hopefully find that generalizing the protocol for larger (or smaller) library sizes is straightforward. We also provide an R script which generates complete step-by-step protocols with customizable parameters (e.g., different values of m).

Initial considerations. As before, we will call v_0 the minimum pipetting volume throughout the protocol. It is important to keep in mind that each well of the final plates will carry a volume of $m\,v_0$. For $m = 8$, we can choose $v_0 = 25$ µL, a volume which can be easily handled by standard pipettes, such that every well of the plates will carry a total volume of 200 µL by the end of the assembly protocol. In what follows, we will assume that at the start of the protocol we have enough volume of each of the $m = 8$ individual species in monoculture. For some applications, it might be desirable to standardize the density of the monocultures, typically by diluting all cultures to the same target optical density. One should keep in mind that each species will be present in half of the 2^m consortia, that is, we will need a minimum volume of $2^{m-1}v_0$ (3.2 mL with our choices of m and v_0) of each monoculture. It is of course recommended to have some extra volume available.

Step 1: Buffer pipetting. We start the protocol by pipetting the necessary buffer column-wise. Making this the first step of the protocol has the advantage that all liquid buffer handling can be done using the same set of pipette tips with no risk of cross-contamination. Using an 8-channel pipette, we pipette a volume of $v_0\ [m - 3 - H(B(j-1))]$ (note that this corresponds to the second term of equation 5) into all wells of column j. We do this for every column from the first ($j = 1$) to the final one ($j = 2^{m-3}$, in our case $j = 32$ since we have $m = 8$ species). For 8 species, and with $v_0 = 25$ µL, we have:

Column index, j	$B(j-1)$	$H(B(j-1))$	**Buffer volume,** $v_0\ [m-3-H(B(j-1))]$
1	00000	0	125 µL
2	00001	1	100 µL
3	00010	1	100 µL
4	00011	2	75 µL
5	00100	1	100 µL
6	00101	2	75 µL
7	00110	2	75 µL
8	00111	3	50 µL
9	01000	1	100 µL
10	01001	2	75 µL
11	01010	2	75 µL
12	01011	3	50 µL
13	01100	2	75 µL
14	01101	3	50 µL
15	01110	3	50 µL
16	01111	4	25 µL
17	10000	1	100 µL
18	10001	2	75 µL
19	10010	2	75 µL
20	10011	3	50 µL
21	10100	2	75 µL
22	10101	3	50 µL
23	10110	3	50 µL
24	10111	4	25 µL
25	11000	2	75 µL
26	11001	3	50 µL
27	11010	3	50 µL
28	11011	4	25 µL
29	11100	3	50 µL
30	11101	4	25 µL
31	11110	4	25 µL
32	11111	5	0 µL

Step 2: Species 4 to m. Using an 8-channel pipette, we now pipette a volume $v_0 = 25$ µL of the monocultures of species 4 to m into all wells of their corresponding columns (see equation 2 and Figure 1C):

Species	Volume	Columns, j
4	25 µL	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 (alternating columns)
5	25 µL	3, 4, 7, 8, 11, 12, 15, 16, 19, 20, 23, 24, 27, 28, 31, 32 (alternating sets of 2 columns)
6	25 µL	5, 6, 7, 8, 13, 14, 15, 16, 21, 22, 23, 24, 29, 30, 31, 32 (alternating sets of 4 columns)
7	25 µL	9, 10, 11, 12, 13, 14, 15, 16, 25, 26, 27, 28, 29, 30, 31, 32 (alternating sets of 8 columns)
8	25 µL	17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32 (alternating sets of 16 columns)

We recommend using different pipette tips for each pipetting event to avoid cross-contamination — but note that e.g. all pipetting of species 4 could be performed with the same set of tips with no risk of cross-contamination. Similarly, species 5 could be first pipetted into all necessary odd columns, and then into all necessary even columns, with the same set of tips.

Step 3: Combinations of species 1 to 3. We next assemble all $2^3 = 8$ combinations of species 1, 2, and 3. We recommend first assembling these consortia in large vessels, typically 15 mL or 50 mL Falcon tubes which we label as T1 to T8. We then pipette the following volumes of species 1, 2, 3 monocultures and buffer into each of the tubes:

Tube	Consortium	Buffer volume	Species 3 volume	Species 2 volume	Species 1 volume
T1	000	3 mL	0	0	0
T2	001	2 mL	0	0	1 mL
T3	010	2 mL	0	1 mL	0
T4	011	1 mL	0	1 mL	1 mL
T5	100	2 mL	1 mL	0	0
T6	101	1 mL	1 mL	0	1 mL
T7	110	1 mL	1 mL	1 mL	0
T8	111	0	1 mL	1 mL	1 mL

Lastly, we pipette a volume of $3v_0 = 75$ µL per well from tubes T1 to T8 into rows $i = 1$ to 8 of the plates respectively, which can be easily done with the aid of a 12-channel pipette. Note that this not only places species 1 to 3 in their corresponding positions (Figure 1C), but also adds the necessary amount of buffer per row (see first term of equation 5). Also note that, since a volume of $3v_0$ per well is added from each tube to every row of the plates, a minimum volume of $3v_0 \cdot 2^{m-3} = 2.4$ mL per tube is required — but again, some spare volume is desirable, so here we have rounded that up to 3 mL per tube.

Reagents and strains

For our experiment with synthetic colorants, we used a mix of food colorants (Dr. Oetker) and temperas, which were diluted in water until they all showed similar maximum absorbance values in the range 380-780 nm (Figure 2A).

All eight *P. aeruginosa* strains were kindly donated by Sara Hernando-Amado, and are described elsewhere.[33–35] Strains 1 to 8 (as labeled in the main text) correspond, in order, to mutants pAOV, PA14, NfxB, ParR-CAZ, mexZ, OrfN, NfxB-CAZ, and MDR6-CAZ in the original publications. For our experiment, strains were streaked on LB agar (Condalab) plates from frozen stocks, and a single colony of each strain was resuspended in 15 mL of LB medium (Condalab) (contained in 50 mL conical tubes, Falcon). Monocultures were allowed to grow for 24 h at 37 °C, after which we used them to assemble all 2^8 strain combinations as described in the main text. After assembling every combination, samples were diluted into fresh LB medium by a factor of 1:400 (0.5 µL of inoculum into 200 µL of fresh medium) in 96-well plates (Thermo Scientific). Plates were covered with transpirable seals (Excel Scientific) and cultures were incubated for 40 h at 37 °C with no agitation. After incubation, plates were transferred to a MultiSkan SkyHigh plate reader (Thermo Scientific; software version: SkanIt 7.0) to obtain the absorbance spectra.

Quantification of pairwise functional interactions between species

Following recent work in microbial ecology,[10,19,29] we define the pairwise functional interaction between two species i and j in a background consortium B as

$$\epsilon_{ij}(B) = F(B+i+j) - \left[F(B) + \Delta F_i(B) + \Delta F_j(B)\right] \tag{S1}$$

where we have called $B+i$, $B+j$, and $B+i+j$ the consortia resulting from including species i, species j, or both, in the ecological background B. Here, F generically denotes the function of a consortium, which in our case is the absorbance at 600 nm (Abs_{600}). The terms in brackets represent the additive expectation for the function of consortium $B+i+j$, that is, the background function $F(B)$ plus the two individual functional effects of species i and j in that background (denoted $\Delta F_i(B)$ and $\Delta F_j(B)$, respectively). The functional effect of species i in background B, as explained in the main text, is simply quantified as the difference in function between consortium $B+i$ and B:

$$\Delta F_i(B) = F(B+i) - F(B) \tag{S2}$$

Thus, equation S1 can be written as

$$\epsilon_{ij}(B) = F(B+i+j) - F(B+i) - F(B+j) + F(B) \tag{S3}$$

Figure 5F in the main text shows the values of the interaction coefficients ϵ_{ij} for all pairs of strains i and j across all potential backgrounds B, calculated using equation S3.

Analysis of functional variance by interaction order

Recent analyses of already published microbial consortia[30] used the following approach to quantifying interactions: instead of representing the presence/absence of a species i with a variable $x_i = 1, 0$, one can represent it using a variable $\sigma_i = +1, -1$. The function (in our case, biomass quantified as Abs_{600}) of a consortium c can thus be expressed as

$$Abs_{600}(c) = \alpha_0 + \sum_i \alpha_i \sigma_i + \sum_i \sum_j \alpha_{ij} \sigma_i \sigma_j + \sum_i \sum_j \sum_k \alpha_{ijk} \sigma_i \sigma_j \sigma_k + \cdots \tag{S4}$$

In this expression, the coefficients α_i represent the average difference in function between two consortia differing solely on the presence or absence of species i, that is, the average functional effect (as defined in the main text) of species i across all potential ecological backgrounds where this species may be included.[30] Therefore, the coefficients i are the averages of the distributions shown in Figure 5E. In turn, the coefficients α_{ij} are proportional to the average pairwise functional interaction between species i and j (in the sense that we define them in the main text, see Figure 5D and equation S3), i.e., the averages of the distributions in Figure 5F.

Using this representation has the advantage that the total functional variance (in our case, the total variance in Abs_{600} across all 2^8 consortia) can be expressed as:

$$\mathrm{var}(Abs_{600}) = \sum_i \alpha_i^2 + \sum_i \sum_j \alpha_{ij}^2 + \sum_i \sum_j \sum_k \alpha_{ijk}^2 + \cdots \tag{S5}$$

The variance explained by the coefficients of order k can be computed as $\sum \alpha_{(k)}^2$, e.g., the functional variance explained by the first-order coefficients is $\sum_i \alpha_i^2$, the variance explained by the second-order terms is $\sum_i \sum_j \alpha_{ij}^2$, and so on. This allows us to calculate the fraction of variance which can be attributed to each order, represented in Figure 5G in the main text.

Supplementary Figures

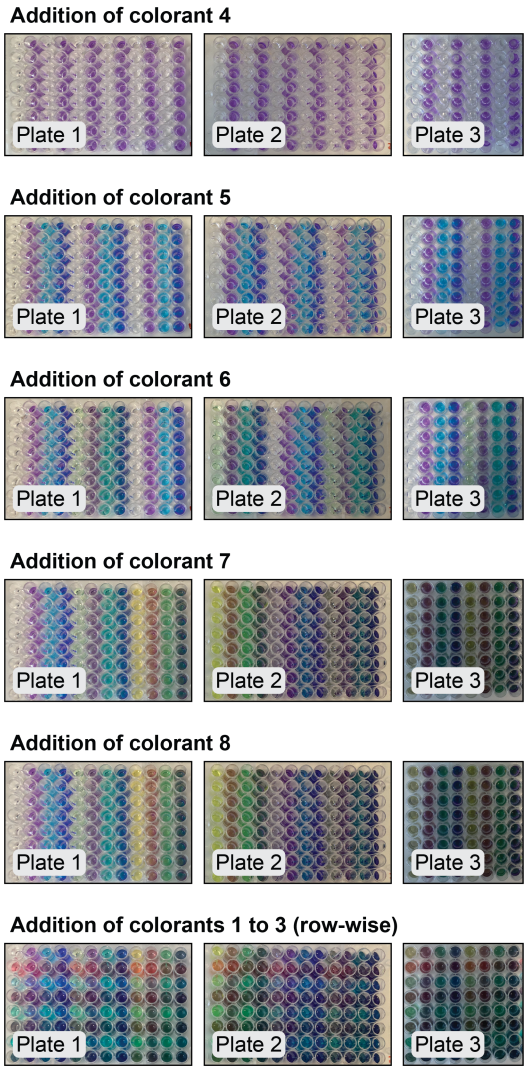

Figure S1. Plate layout at different stages of the assembly protocol for combinations of 8 colorants.

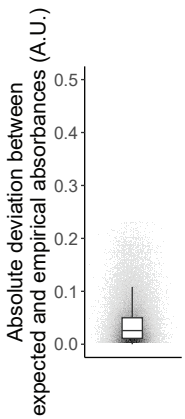

Figure S2. Absolute deviation, $|Abs(c) - Abs^{(add)}(c)|$, between the spectra of the 8-colorant consortia and the additive expectations, for all consortia c of 2 or more colorants and for all wavelengths λ between 380 nm and 780 nm. Median absolute error is ~ 0.025 A.U.

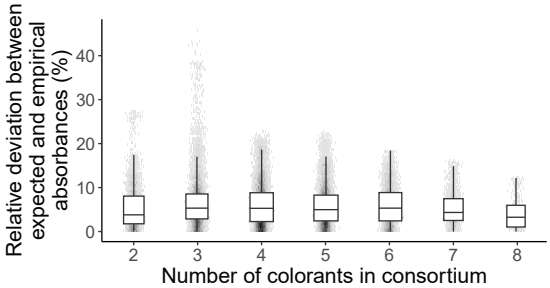

Figure S3. Relative deviation between the empirical spectra and the respective additive expectations (equation 6) versus the number of colorants in the consortium. Here we represent the deviations for all wavelengths in the 380-780 nm range.

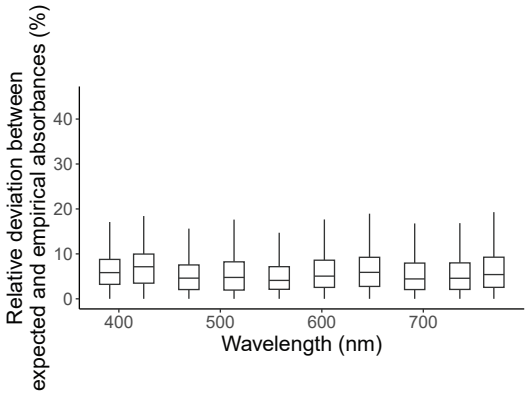

Figure S4. Relative deviation between the empirical spectra and the respective additive expectations (equation 6) as a function of the wavelength. Wavelengths in the 380-780 nm range were split into ten quantiles. Within each quantile, we represent the relative deviations (equation 6) for all combinations of two or more colorants.

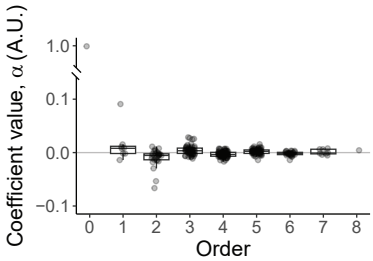

Figure S5. Previous work has expressed the community-level function of a consortium c as $F(c) = \alpha_0 + \sum_i \alpha_i \sigma_i + \sum_i \sum_j \alpha_{ij} \sigma_i \sigma_j + \sum_i \sum_j \sum_k \alpha_{ijk} \sigma_i \sigma_j \sigma_k + \cdots$.[30] In this expression, σ_i takes values +1 or -1 depending on the presence or absence of strain i in the consortium, respectively (note the difference with equation 1 in the main text, where presence/absence was denoted with 1s and 0s instead). The coefficients α_i, α_{ij}, etc. can be quantified exactly from empirical data if the full mapping between community compositions and functions is known, as is the case in our experiment with eight *P. aeruginosa* strains. Here we represent the magnitude of the coefficients α at all orders for that experiment. The equation above has the advantage that the total functional variance, var(F), can be computed as: var$(F) = \sum_i \alpha_i^2 + \sum_i \sum_j \alpha_{ij}^2 + \sum_i \sum_j \sum_k \alpha_{ijk}^2 + \cdots$, where each sum corresponds to the variance due to terms of order 1, 2, 3, and so on.

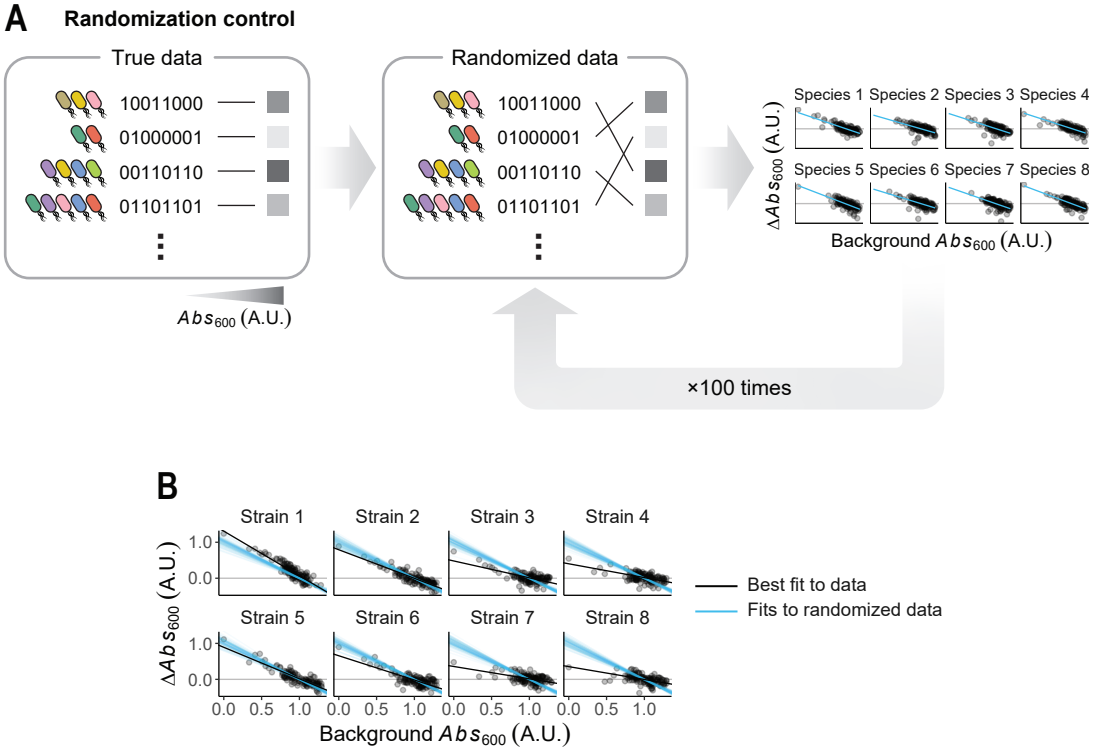

Figure S6. Correlations between the functional effect of a strain (Abs) and the function of the background consortium where it is included, such as those we report in Fig. 5I, can emerge when the mapping between community composition and function is entirely random.[10] (**A**) To test whether the empirical correlations we observe in Figure 5I could be attributed to this phenomenon, we randomized our empirical data 100 times. In each of these randomizations, we fit linear models linking the functional effect (ΔAbs) of each of the eight strains to the functions of their background consortia. (**B**) Here we compare the linear fits obtained in each of these 100 randomization controls (blue lines) with the true best fit to the empirical data (black lines). We find that the empirical fits are generally incompatible with the randomization controls.