



# Measurement Error and Resolution in Quantitative Stable Isotope Probing: Implications for Experimental Design

 Ella T. Sieradzki,<sup>a</sup> Benjamin J. Koch,<sup>b,f</sup> Alex Greenlon,<sup>a</sup> Rohan Sachdeva,<sup>c</sup> Rex R. Malmstrom,<sup>d</sup> Rebecca L. Mau,<sup>b</sup> Steven J. Blazewicz,<sup>e</sup> Mary K. Firestone,<sup>a</sup> Kirsten S. Hofmockel,<sup>g,h</sup> Egbert Schwartz,<sup>b,f</sup>  Bruce A. Hungate,<sup>b,f</sup>  Jennifer Pett-Ridge<sup>e</sup>

<sup>a</sup>University of California Berkeley, Environmental Science and Policy Management, Berkeley, California, USA

<sup>b</sup>Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona, USA

<sup>c</sup>University of California Berkeley, Earth and Planetary Sciences, Berkeley, California, USA

<sup>d</sup>Department of Energy Joint Genome Institute, Berkeley, California, USA

<sup>e</sup>Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California, USA

<sup>f</sup>Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona, USA

<sup>g</sup>Pacific Northwest National Laboratory, Richland, Washington, USA

<sup>h</sup>Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa, USA


**ABSTRACT** Quantitative stable isotope probing (qSIP) estimates isotope tracer incorporation into DNA of individual microbes and can link microbial biodiversity and biogeochemistry in complex communities. As with any quantitative estimation technique, qSIP involves measurement error, and a fuller understanding of error, precision, and statistical power benefits qSIP experimental design and data interpretation. We used several qSIP data sets—from soil and seawater microbiomes—to evaluate how variance in isotope incorporation estimates depends on organism abundance and resolution of the density fractionation scheme. We assessed statistical power for replicated qSIP studies, plus sensitivity and specificity for unreplicated designs. As a taxon's abundance increases, the variance of its weighted mean density declines. Nine fractions appear to be a reasonable trade-off between cost and precision for most qSIP applications. Increasing the number of density fractions beyond that reduces variance, although the magnitude of this benefit declines with additional fractions. Our analysis suggests that, if a taxon has an isotope enrichment of 10 atom% excess, there is a 60% chance that this will be detected as significantly different from zero (with alpha 0.1). With five replicates, isotope enrichment of 5 atom% could be detected with power (0.6) and alpha (0.1). Finally, we illustrate the importance of internal standards, which can help to calibrate per sample conversions of %GC to mean weighted density. These results should benefit researchers designing future SIP experiments and provide a useful reference for metagenomic SIP applications where both financial and computational limitations constrain experimental scope.

**IMPORTANCE** One of the biggest challenges in microbial ecology is correlating the identity of microorganisms with the roles they fulfill in natural environmental systems. Studies of microbes in pure culture reveal much about their genomic content and potential functions but may not reflect an organism's activity within its natural community. Culture-independent studies supply a community-wide view of composition and function in the context of community interactions but often fail to link the two. Quantitative stable isotope probing (qSIP) is a method that can link the identity and functional activity of specific microbes within a naturally occurring community. Here, we explore how the resolution of density gradient fractionation affects the error and precision of qSIP results, how they may be improved via additional experi-

**Citation** Sieradzki ET, Koch BJ, Greenlon A, Sachdeva R, Malmstrom RR, Mau RL, Blazewicz SJ, Firestone MK, Hofmockel KS, Schwartz E, Hungate BA, Pett-Ridge J. 2020. Measurement error and resolution in quantitative stable isotope probing: implications for experimental design. *mSystems* 5:e00151-20. <https://doi.org/10.1128/mSystems.00151-20>.

**Editor** William Orsi, Woods Hole Oceanographic Institution

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply. Address correspondence to Bruce A. Hungate, [Bruce.Hungate@nau.edu](mailto:Bruce.Hungate@nau.edu), or Jennifer Pett-Ridge, [pettridge2@llnl.gov](mailto:pettridge2@llnl.gov).

 Exploring resolution, reproducibility and replication for cost-benefit stable isotope probing experimental design @SieraVirus @Jeffnerca @BruceHungate @archaeologist @KHofmockel @AGreenlon @Firestone\_lab

**Received** 24 February 2020

**Accepted** 29 June 2020

**Published** 21 July 2020

mental replication, and discuss cost-benefit balanced scenarios for SIP experimental design.

**KEYWORDS** stable isotope probing, environmental microbiology, experimental design, metagenomics, microbial communities, microbial ecology, qSIP, statistical power

Stable isotope probing (SIP) of nucleic acids is one of the few non-culture-dependent methods that can identify the functionality of microorganisms in their native communities, making it one of the most powerful techniques in microbial ecology (1–6). In SIP, a substrate labeled with a heavy isotope is added to an environmental sample. Following an incubation period ranging from hours to weeks (depending on the substrate uptake rate), the DNA (or RNA) of growing microbial populations that have consumed the isotope-enriched substrate becomes more dense due to their incorporation of the heavy isotope. Community nucleic acids can then be extracted and separated along a density gradient using ultracentrifugation. DNA/RNA from organisms that incorporated the labeled substrate will appear in denser fractions of the gradient compared to where it would be with addition of an unlabeled substrate (4, 7). While previous studies include some consideration of best practices for designing SIP experiments and handling SIP data (4, 6, 8–12), here we address several additional issues regarding reproducibility, sensitivity, cost, and the minimum detectable effect size of experiments with different designs.

Quantitative SIP (qSIP) is a recently developed adaptation of SIP that makes substrate uptake measurements possible at the scale of individual taxa (9, 13). In qSIP, isopycnic separation of nucleic acids in cesium chloride is combined with a mathematical model to quantify isotope enrichment. This approach allows a user to measure growth and mortality rates of individual taxa in complex communities, particularly when using  $^{18}\text{O}$ -labeled “heavy water” ( $\text{H}_2^{18}\text{O}$ ) as a universal substrate, since cells incorporate oxygen from water during nucleic acid synthesis, quantitatively reflecting cell division (DNA synthesis) and metabolism (RNA synthesis) (14, 15). Similarly, cell mortality rates may be quantitatively related to the degradation of unlabeled nucleic acids. In a similar manner, qSIP can be used to calculate enrichment of specific taxa due to incorporation of substrate  $^{15}\text{N}$  or  $^{13}\text{C}$  (16, 17). By normalizing relative abundance to the total number of organisms per fraction (estimated by quantitative PCR [qPCR] of the 16S-rRNA gene), the sensitivity of qSIP has been shown to be less susceptible to taxon abundance, the fraction of the community that incorporates isotope, and level of enrichment, compared to other SIP methods (10).

Designing modern SIP and qSIP experiments involves a tension between collecting many density fractions per sample (to increase precision) versus the costs of labor and sequencing (18). While early SIP studies inspected only the “heaviest” fractions—considered to host the most isotopically enriched DNA—these fractions may contain unlabeled high-GC-content DNA. Conversely, enriched low-GC-content DNA may not reach the heaviest fractions. The current practice is to examine many density fractions and perform statistical analyses comparing isotope-labeled versus unlabeled control to indicate the extent to which organisms have “shifted” within a density gradient in response to the isotope treatment (9, 10). Density shifts can be used to calculate substrate assimilation rate per taxon (atom% excess), and with the universal substrate  $\text{H}_2^{18}\text{O}$ , they can be used to infer specific growth rates (13, 15, 19–21). However, even the most basic SIP experiment (e.g., one type of substrate with labeled and unlabeled versions, 2 treatments, 2 time points, 3 replicates, 10 density fractions per sample) can easily generate nearly 250 samples for processing and sequencing. Thus, it is critical to plan an experimental design that will ensure high-quality data at sustainable costs. Doing so will become even more important as the field of environmental microbiology transitions to more ambitious applications, such as metagenomics qSIP (MG-qSIP) and SIP-metatranscriptomics (22), since shotgun sequencing adds substantially higher costs

**TABLE 1** Data sets used in this study, including source, number of replicates, and analyses performed

Data set	Source	No. of replicates	No. of density fractions
1	[ <sup>13</sup> C]naphthalene-enriched seawater	1	50
2	SPRUCÉ peatland H <sub>2</sub> <sup>18</sup> O addition	5 <sup>a</sup>	18
3	Grassland soil H <sub>2</sub> <sup>18</sup> O addition	3	9
4	<i>E. coli</i> and <i>P. putida</i>	4–6	30
5	Mock community	9	48

<sup>a</sup>The number of replicates is per preincubation temperature.

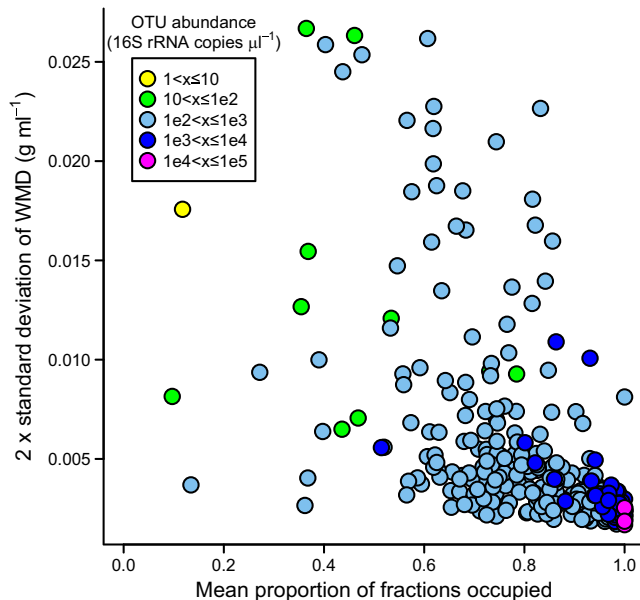
relative to amplicon sequencing and the amount of data per sample may become a computational limitation.

The majority of SIP studies published in the past 20 years targeted marker genes to identify substrate assimilators, usually with 16S rRNA gene-based analysis. However, focusing only on 16S rRNA misses a wealth of genetic information. A handful of more recent studies have combined SIP with metagenomics or metatranscriptomics to investigate genomic potential and actively expressed genes by functional guild (23–25), but the combination of SIP with shotgun sequence analysis quickly becomes limiting both financially and computationally. Some investigators have tried to downsize these burdens by sequencing only highly labeled fractions (12), pooling density fractions, or sequencing unfractionated DNA and matching assembled genomes to SIP-identified substrate assimilators (26–30). One additional way to address the cost issues inherent to metagenomic analysis (e.g., larger amounts of DNA required, higher sequencing costs, and higher computational complexity) is to use only the minimum number of density fractions that are needed to yield a desired level of sensitivity. In addition, adding more sample replicates but collecting a reduced number of fractions (gradient resolution) could lead to higher accuracy with similar effort.

Using multiple SIP data sets (Table 1), we tested the effects of fraction size and sample replication on the robustness and sensitivity of qSIP. We combined (*in silico*) density fractions and measured the effects of lower gradient resolution on per-taxon density shifts and unlabeled weighted mean density. We also calculated the repercussions of reducing the number of density fractions in replicated and unreplicated data sets from both marine and terrestrial microbial communities. Our results show that reducing the gradient resolution from an average density fraction size of 0.002 g ml<sup>-1</sup> (~50 fractions) down to 0.011 g ml<sup>-1</sup> (~9 fractions in a 5.1-ml tube) yields comparable shift detection and a detection limit of 0.005 g ml<sup>-1</sup> (equal to 9% enrichment with <sup>13</sup>C). We discuss using the small inherent variability between replicates as a way to define a shift detection limit and show that this inherent variability is comparable between replicates centrifuged together (within spin) and between replicates centrifuged separately (between spins). Finally, we stress the need for internal standards that can be spiked into each sample, and evaluate the statistical and financial benefits of different experimental design options.

## RESULTS

**Variability of taxon density shifts is greatest for rare OTUs.** In SIP, taxon density shifts, or changes in weighted mean density (WMD) due to incorporation of a stable-isotope-labeled substrate, are the basis for calculating isotope enrichment. *In silico* qSIP analyses have shown that these shifts are detectable in moderately to highly abundant operational taxonomic units (OTUs) (>0.1% relative abundance) (10). Using unlabeled sample experimental data from data set 2, we first examined variation in qSIP-derived density estimates for both rare and common OTUs. We found that OTU abundance (log<sub>10</sub> transformed) was positively correlated with the proportion of density fractions in which that OTU was detected (Pearson's  $r = 0.704$ ,  $P < 0.0001$ ). In other words, the most abundant OTUs tended to occur in all (or nearly all) fractions of the density gradient, whereas the least abundant OTUs tended to occur in fewer density fractions. Variability of the unlabeled WMD was lowest for common OTUs and greatest for rare



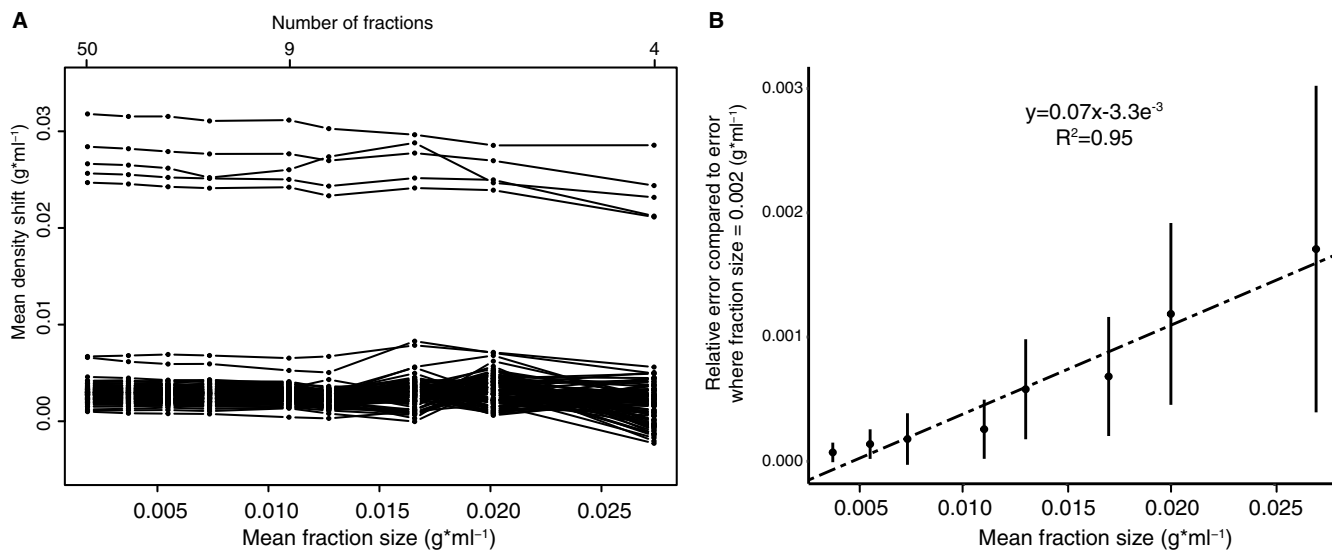
**FIG 1** Effect of OTU abundance on the standard deviation of density measured using qSIP. Points represent individual OTUs from a qSIP experiment ( $n = 320$ , data set 2), where  $\text{H}_2^{18}\text{O}$  was added to a forested peat bog soil. Only unlabeled ( $\text{H}_2^{16}\text{O}$ ) experimental replicates incubated at an intermediate temperature ( $n = 5$ ) are plotted here. We calculated the mean proportion of density fractions in which each OTU occurred (x axis) as well as  $2\times$  standard deviation of WMD for each OTU (y axis). OTUs are colored by their mean unfractionated abundance across all unlabeled replicates (units, 16S rRNA gene copies  $\mu\text{l}^{-1}$ ). The variability of qSIP-calculated unlabeled weighted mean density (WMD) was lowest for common OTUs and greatest for rare OTUs. Highly abundant OTUs tended to occur in the majority of density fractions.

OTUs (Fig. 1). WMD variation (expressed as  $2\times$  standard deviation of the WMD for a given OTU) correlated negatively with the proportion of density fractions in which that OTU was detected (Fig. 1; Pearson's  $r = -0.560$ ,  $P < 0.0001$ ). Similarly, WMD variation also correlated negatively with  $\log_{10}$ -transformed OTU abundance (Pearson's  $r = -0.217$ ,  $P < 0.0001$ ).

**Density shifts are consistent across medium-to-high gradient fraction resolution.** While it is widely assumed that the detection limit in a SIP experiment will depend on the number of fractions collected from each density gradient (18), to our knowledge, this has not been comprehensively tested. We first examined how fraction resolution affects density shift, using the 100 most abundant OTUs (from all fractions combined) in an unreplicated data set of OTUs from naphthalene-enriched seawater (data set 1) where the DNA had been divided into 50 fractions, of which 45 had quantifiable DNA. Consecutive density fractions were consolidated *in silico* (every 2, 3, 4 fractions, etc.) to represent a range of fraction sizes spanning 0.002 to 0.028  $\text{g ml}^{-1}$  (4 to 50 fractions, assuming a standard 5.1-ml tube). We found that the estimated magnitude of OTU density shifts remained consistent at fraction sizes from 0.002 up to 0.011  $\text{g ml}^{-1}$ , expanding previous results that demonstrated this trend with fractions of 0.003 to 0.008  $\text{g ml}^{-1}$  (Fig. 2A) (10). This indicates that fractionating the sample into more than nine fractions had negligible additional effects on OTU density shifts.

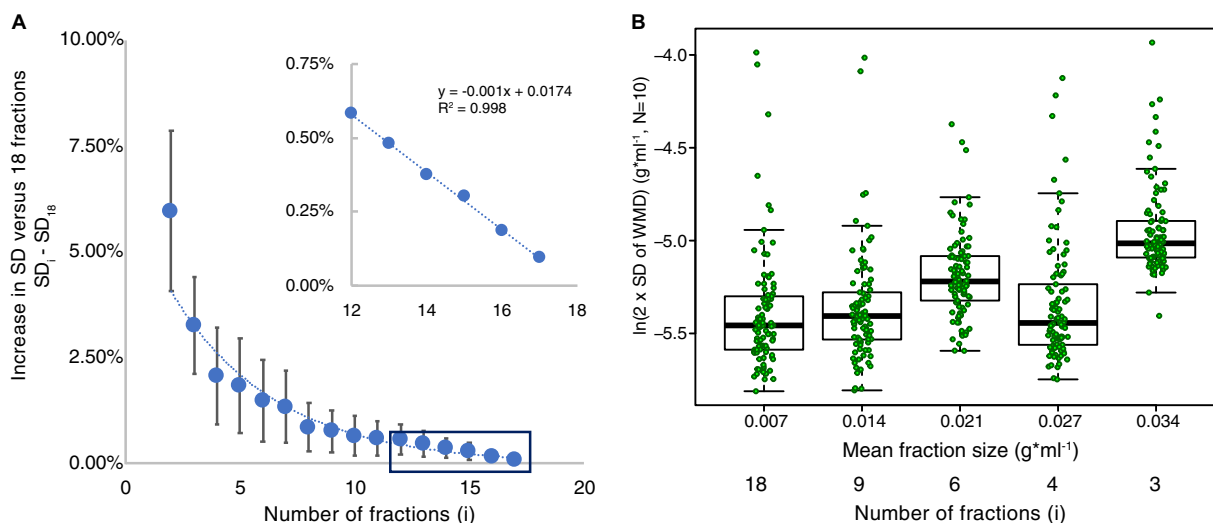
The magnitude of density shift data can also be represented as a relative error, which we define as the density shift in resolution  $r$  ( $r <$  original number of fractions) compared to the density shift realized with maximum resolution. As relative error increases, power declines, so the likelihood that taxa are misassigned as nonisotope incorporators increases. We found a positive linear correlation between fraction size and mean relative error ( $R^2 = 0.95$ ; Fig. 2B). Additionally, the increase in the mean relative error is accompanied by an increase in its variation.

We next assessed the correlation between the number of fractions and the relative standard deviation of the WMD in a study with replication (data set 2,  $n = 5$ , preincu-



**FIG 2** Relationship between fraction size and the estimate of density shift using qSIP (A) and the relative error in that estimate (B). Data from an unreplicated SIP study of [<sup>13</sup>C]naphthalene-enriched seawater (data set 1) where DNA was partitioned into 50 density fractions. To assess the effects of lower fraction resolution, consecutive density fractions were consolidated *in silico*. (A) Density shift (y axis) of the 100 most abundant OTUs (each line represents one OTU), plotted at a range of fraction resolutions (x axis). Mean fraction sizes smaller than 0.011 g ml<sup>-1</sup> (corresponding to nine fractions in a 5.1-ml ultracentrifuge tube) do not affect the density shift of OTUs. While highly enriched taxa are easily discerned even at a fraction size of 0.02 g ml<sup>-1</sup> (four fractions), when fewer than nine fractions are used, the shifts of less strongly enriched OTUs are inconsistent and negatively affect percent atom excess downstream analyses. (B) Relative error was calculated as the absolute difference between density shift of an OTU in a simulated fraction size *r* (e.g., *r* = 0.011 g ml<sup>-1</sup>) and the density shift of the same OTU with the original fraction size (0.0018 g ml<sup>-1</sup>). The points represent the mean relative error of the OTUs plotted in panel A; the error bars represent the standard deviations.

bation temperature of 15°C); here we found the relationship is not linear. Instead, between 2 and 11 fractions, every additional fraction reduced the standard deviation exponentially, but for 12 fractions or more, the proportional differences are much smaller and linearly correlated with the number of fractions (Fig. 3A).



**FIG 3** Variability in enrichment is negatively correlated with the number of density fractions in replicated data. Data plotted are from data set 2 (SPRUCE, *n* = 5, 100 permutations). (A) Relative error (compared to the original 18 fractions) decreases as the number of fractions (*i*) increases. The inset shows a linear decrease in relative error in the 12- to 18-fraction range. (B) Two standard deviations of the WMD in the 100 most abundant OTUs from unlabeled soil at time zero (data set 2, time zero, preincubation temperatures of 15°C and 45°C, *n* = 10) as a function of fraction size. Higher fraction size corresponds to lower gradient resolution. The horizontal line within the box represents the median, the box represents the 25 and 75 percentiles, and whiskers represent the 10 and 90 percentiles for 100 OTUs in each fraction size. The raw data are plotted on top of the boxes.

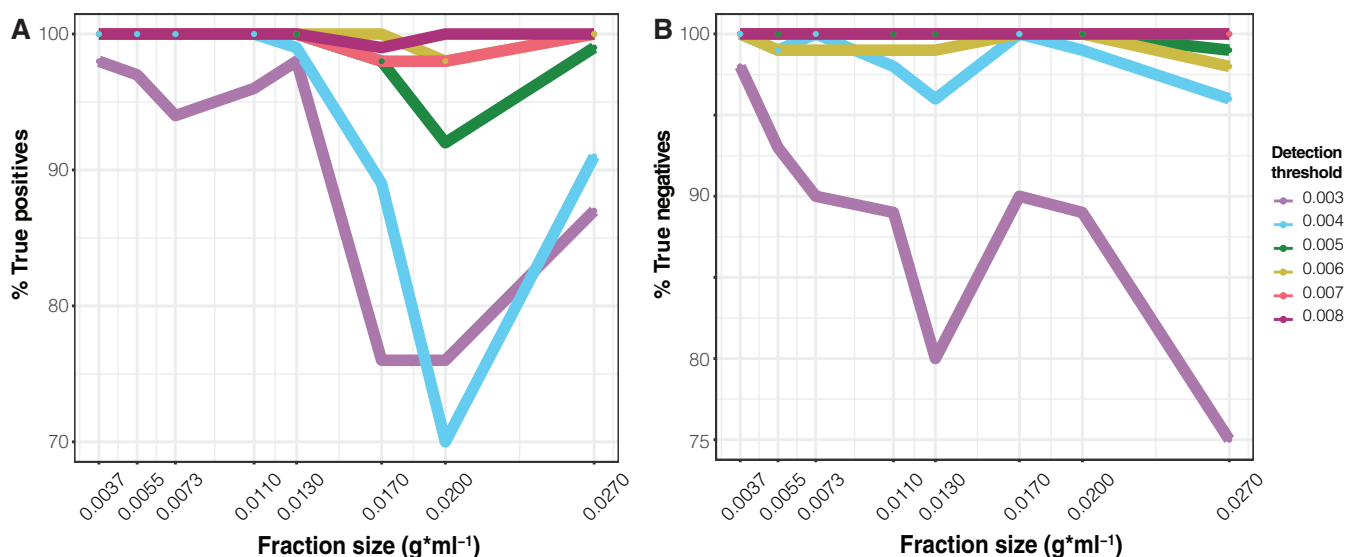
**Low variance influences minimum detectable density shift.** In a SIP experiment, it is important to identify statistically significant density shifts between samples treated with a labeled substrate versus control samples (unlabeled substrate). To define a minimum detectable effect size, we calculated the inherent variability of WMD of unlabeled DNA from taxa in data set 2—a well-replicated SIP experiment ( $n = 10$ , time zero). Reducing the resolution of the density gradient may impact this variability. We evaluated this by merging and averaging data from an increasing number of consecutive fractions. Our initial analysis with medium resolution (fraction size  $0.007 \text{ g ml}^{-1}$ ; e.g., 18 fractions in a 4.7-ml tube) revealed that the WMD of abundant taxa varied little between replicates (Fig. 3B). The WMD per OTU was calculated over 10 replicates. The resulting WMDs varied around a median of  $0.004$  to  $0.007 \text{ g ml}^{-1}$  at gradient resolutions varying from  $0.007$  to  $0.034 \text{ g ml}^{-1}$  (3 to 18 fractions). With progressively fewer fractions (decreasing gradient resolution), variation in WMD remained statistically indistinguishable for fraction sizes between  $0.007$  and  $0.027 \text{ g ml}^{-1}$  (4 to 18 fractions) but increased significantly (analysis of variance [ANOVA], Tukey 95% confidence level) with a fraction size of  $0.034 \text{ g ml}^{-1}$  (3 fractions in a 4.7-ml tube) (Fig. 3B). This variability of the WMD as a function of fraction size can be used to choose an acceptable minimum shift in density which a researcher would like to detect with a desired confidence level.

We found the range of relative error increased with larger fraction sizes. To explore how this variation affects the detection of substrate incorporators, we calculated the putative sensitivity (proportion of true positives) and specificity (proportion of true negatives) as a function of the shift detection threshold for all gradient resolutions in data set 1. The shift detection threshold can be thought of as the smallest difference between labeled and unlabeled WMD that would be considered a significant density shift. We performed these calculations using the assumption that a density shift higher than a specific threshold in the original experimental setup (data set 1, 50 fractions) represented significant enrichment. Therefore, all our comparisons are relative to the originally measured results, which likely contain some inherent error. As the sensitivity and specificity of qSIP have been shown to be high, we assume this error is very small. Both sensitivity and specificity parameters were stable down to a fraction resolution of  $0.011 \text{ g ml}^{-1}$  (9 fractions), using a shift detection threshold of  $0.005 \text{ g ml}^{-1}$  or higher. Specificity was  $>95\%$  for all gradient resolutions at a threshold of  $>0.003 \text{ g ml}^{-1}$ , but sensitivity was more impacted by a gradient resolution of  $>0.013 \text{ g ml}^{-1}$  (Fig. 4).

We also tested how fraction number affects error in the estimates of isotope incorporation using a replicated experiment (data set 2, SPRUCE [Spruce and Peatland Responses Under Changing Environments] soil). Our analyses show that confidence intervals widen as the number of density fractions declines (see Fig. S2 in the supplemental material). In other words, as the number of density fractions declines, the likelihood of detecting a given level of isotope assimilation declines as well.

**Replication, statistical power, and detection thresholds in SIP experiments.** In experiments with replicate samples, the number of replicates and desired statistical power influence the detection limit. When designing an experiment, it is valuable to use the desired statistical power and desired enrichment detection threshold in advance, to decide on the number of replicates needed. Both parameters depend on the scientific purpose of the study. The higher the power and threshold, the fewer replicates are necessary (Fig. 5).

**Variation in mean weighted density is comparable within and between spins.** To assess the impact of spin-to-spin and tube-to-tube variability on WMD, we tested the variability of WMD using pure culture DNA (data set 4). DNA extracted from pure cultures of unlabeled *Escherichia coli* and unlabeled or 100%  $^{13}\text{C}$ -labeled *Pseudomonas putida* was aliquoted into replicates, ultracentrifuged in  $\text{CsCl}$ , and fractionated in an automated pipeline. The known genome differences in %GC of these organisms (i.e., distance between their peak densities) allowed us to calculate their WMD, even when both were unlabeled. When we compared the WMD mean of *E. coli* replicates between

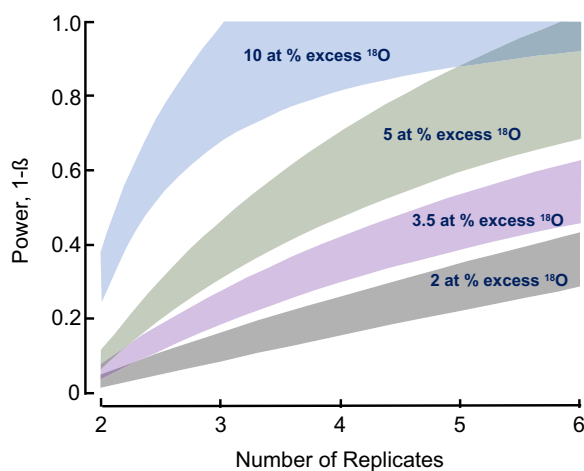


**FIG 4** Rate of false discoveries increases at low gradient resolution and with lower shift detection threshold. Sensitivity (rate of true positives) (A) and specificity (rate of true negatives) (B) calculated for the 100 most abundant OTUs from a naphthalene-enriched seawater SIP study (data set 1) and plotted versus fraction size. The colors represent detection limit thresholds ( $\text{g ml}^{-1}$ ). Here, we use the phrase “true positive” in the statistical sense, where the power of a test refers to its statistical sensitivity, its true positive rate, or its probability of detection. “True negative” refers to the proportion of actual negatives that are correctly identified.

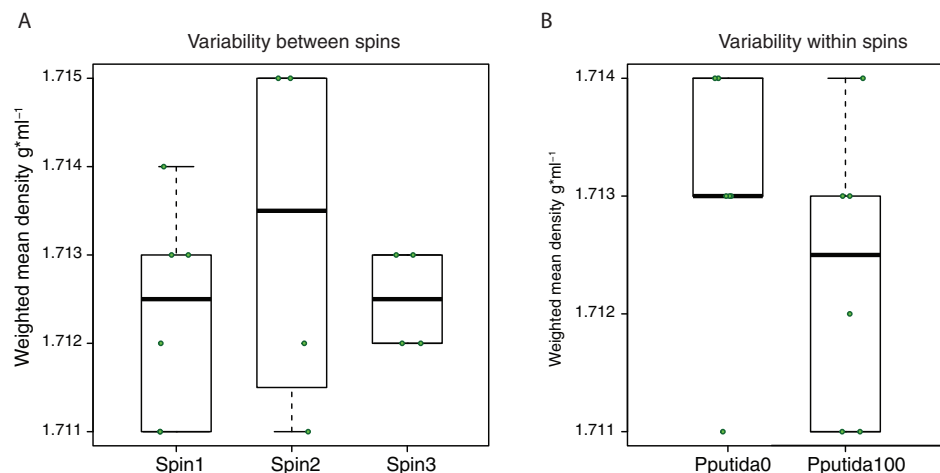
spins, we found the variation was comparable to the range of WMDs measured within a spin (up to  $0.004 \text{ g ml}^{-1}$ ) (Fig. 6).

We also tested within- and between-spin effects with a genomic mock community (data set 5) by centrifuging triplicates of a DNA mixture comprised of four organisms with distinguishable genomic %GC in equal quantities. In this test, between-spin variation was  $0.0013$  to  $0.0025 \text{ g ml}^{-1}$ , whereas within-spin variation ranged up to  $0.0056 \text{ g ml}^{-1}$  (see Data Set S1 in the supplemental material [raw data]).

**Using genomic mock communities to explore density-to-GC content conversion.** One potential strategy for decreasing the costs and labor needed for meta-



**FIG 5** Statistical power of enrichment detection as a function of number of sample replicates used in a qSIP experiment. To determine the necessary number of replicates in a qSIP experiment, users can choose their desired statistical power and desired detection threshold (represented by different colors). For example, the minimum number of replicates necessary to detect a density shift corresponding to 5 atom% excess  $^{18}\text{O}$  and a statistical power of 0.60 would be 4. This plot is calculated for  $^{18}\text{O}$  based on data set 2 (SPRUCE) but can easily be converted for carbon or nitrogen isotopes. Ten atom% enrichment (at percent excess [APE]) due to incorporation of  $0.0065 \text{ g ml}^{-1}$   $^{18}\text{O}$ -labeled substrates corresponds to 12 APE with the same incorporation of  $^{13}\text{C}$  substrates, or 31 APE with  $^{15}\text{N}$  substrates (50% GC).



**FIG 6** Tube-to-tube variability and effect of other taxa. Variability in the weighted mean density (WMD) of pure culture DNA. (A) WMD of unlabeled *Escherichia coli* DNA in three separate centrifuge spins ( $n = 6$ ,  $n = 4$ , and  $n = 4$ ) mixed with unlabeled or labeled *Pseudomonas putida* grown with 99 atom% [<sup>13</sup>C]glucose at a 1:1 ratio, and (B) within the same spin in the presence of unlabeled (Pputida0) versus 100 atom% <sup>13</sup>C-labeled *P. putida* (Pputida100) ( $n = 6$ ). The horizontal line within the box represents the mean, the box represents the 25 and 75 percentiles, and the whiskers represent the 10 and 90 percentiles for 100 OTUs in each fraction size. The raw data are plotted on top of the boxes.

genomic-SIP experiments would be to sequence only the labeled samples and calculate an approximate density shift using the %GC of the genomic bins. The conversion of density to the GC content of a genome is a linear function of the unlabeled (<sup>12</sup>C) weighted mean. There is a canonical equation describing this function (31–33), but it has also been determined empirically in the past with a ladder of three bacterial taxa with different GC contents (9). If this equation were identical for each gradient in a SIP experiment (as is usually assumed), unlabeled replicates of the same organism should have the same weighted mean in every run. However, as shown here and previously, this is not the case (9, 16). To address this variation, we hypothesized that an external standard could help to calibrate the conversion of %GC to the mean weighted density for each sample. To test this, we ran nine replicates of a mock community (data set 5) with a wide range of known GC content, generated a calibration curve from each, and fitted a linear equation to it. In this mock community, which consisted of high-molecular-weight genomic DNA from four bacterial taxa, we found mean weighted density and %GC were highly correlated with a linear relationship ( $n = 9$ ,  $R^2 > 0.94$ ) but that the slopes and intercepts for each individual replicate varied (Table 1, Fig. S3). We found similar results when using the peak (highest) buoyant density per genome rather than the WMD. There was a significant difference between the WMD as calculated according to the canonical equation  $\rho = (0.098[G + C]) + 1.66$  (32) and the observed mean of WMDs per genome over all replicates ( $n = 9$ , paired  $t$  test,  $P = 0.02$ ). For example, using the canonical equation, the GC content of the four mock community members in replicate 1 are 38%, 46.5%, 62%, and 72.1%, respectively. These values deviate by several percent from the known GC content of these genomes (Table 2).

## DISCUSSION

SIP is a powerful tool for investigating taxon-specific microbial functions in complex assemblages. Like any method, SIP-derived measurements have some inherent variability which can be managed—within limits—to address research questions of interest. Our results show how experimental design (number of replicates, number of fractions, use of standards) can be customized to achieve the goals of a particular research question and the level of sensitivity/detection required. Despite the wide use of SIP, there has been relatively little benchmarking of interpretation of its results. Here, we attempt to shed light on some practical aspects of the method and discuss how to adjust it to maximize results within the limits of labor and cost.



**TABLE 2** Relationship between observed versus expected GC content for known genomes varies between mock community replicates<sup>a</sup>

Parameter or replicate	GC value (%)				Slope	Intercept	R <sup>2</sup>
	<i>T. pseudethanolicus</i>	<i>B. licheniformis</i>	<i>B. longum</i>	<i>S. violaceoruber</i>			
Known value	34.5	45.9	60.1	72.7			
Replicate							
1	37.8	46.9	62.2	72.4	0.090	1.665	0.995
2	37.8	52	64.3	NA	0.099	1.664	0.984
3	40.8	48	64.3	75.5	0.093	1.666	0.993
4	39.8	48	64.3	76.5	0.098	1.664	0.995
5	40.8	50	66.3	78.6	0.099	1.665	0.997
6	NA	45.9	53.1	68.4	0.082	1.666	0.943
7	39.8	50	63.3	74.5	0.090	1.668	0.999
8	40.8	51	63.3	75.5	0.089	1.669	0.999
9	42.9	52	65.3	NA	0.087	1.671	0.999
Mean	39.8	49	63.3	74.5	0.090	1.668	0.999

<sup>a</sup>Calculated %GC (32) for nine replicates of a four-member genomic mock community, and slope, intercept, and R<sup>2</sup> for linear plots of observed versus expected values. The top row shows the known %GC in each genome. NA, not available.

Within a replicated qSIP experiment, it is important to evaluate statistical power—the probability of detecting a given level of isotopic enrichment. Yet, power is evaluated infrequently, because avoiding type I errors is typically prioritized above avoiding type II errors, reflecting the common practice of identifying thresholds for alpha (0.01, 0.05), while beta ( $\beta$ ) thresholds remain unreported. This perspective makes sense in some contexts, e.g., incorrectly inferring that a medical treatment is effective is often viewed as more hazardous than concluding it is not effective, when in fact it is. However, in contrast, applying the precautionary principle to the impact of a potential pollutant or toxin, it is in some cases wiser to avoid type II errors over type I (i.e., erring on the side of environmental caution [34]). For a SIP experiment, power analysis involves evaluating the trade-offs among several parameters: (i) the effect size of interest, which for qSIP studies is the density shift (or amount of isotope incorporation) that the researcher wishes to detect (this can be thought of as the minimum detectable difference); (ii) the acceptable  $\alpha$  value, or acceptable probability of type I error (for qSIP, a type I error occurs when the researcher concludes that there is isotope incorporation, when in fact none occurred); (iii) the acceptable  $\beta$  value, or acceptable probability of type II error (for qSIP, a type II error occurs when the researcher infers “no isotope incorporation,” when in fact some isotope incorporation actually occurred); and (iv) the number of true, independent replicates (sample size) used in the experiment. Power is defined as  $1 - \beta$  and is the probability that a true difference will be detected in a given experimental design. Applied to qSIP, power analysis can show how increasing the number of replicates increases the probability of detecting a given level of isotope incorporation. Power analysis can also show, at a constant level of power, how increasing the number of replicates decreases the threshold level of isotope incorporation that can be detected—making it more likely to detect taxa that are slower growing, or perhaps are secondary consumers in a substrate trophic chain. Last, power analysis can clarify the trade-offs between type I and type II errors and gives useful context for interpreting results from qSIP experiments.

In any SIP experiment, the detection of density shifts is a critical goal. However, variability in a number of factors, including fraction resolution and replication, can limit this ability. In addition, variation of the weighted mean density (WMD) of the same unlabeled taxon over numerous replicates, observed even between samples processed simultaneously and identically, implies that there are physical and/or chemical factors unrelated to genomic %GC that can affect SIP metrics. The results of the unlabeled WMD variation and interpretation error analyses we conducted imply that when using qSIP, researchers can expect that a minimum detectable effect size of 0.005 g ml<sup>-1</sup> in unreplicated <sup>13</sup>C experiments would result in low type I and type II errors. This cutoff applies when a DNA density gradient has been divided into at least four density

fractions, but we note that precision will increase if additional fractions are utilized. Adding additional replication would lead to increased statistical power using the same detection limit. However, we found that density shift estimates of taxon-specific isotope incorporation are broadly robust across a wide range of fraction sizes. For example, the relative error of high incorporators only varied by an average of 0.02% in shift from fraction size 0.002 to 0.011 g ml<sup>-1</sup> ( $n = 5, 50$  to 9 fractions in a 5.1-ml tube).

Variable WMD values mean that the same organism may peak at a density anywhere within a specific range. Moreover, the unlabeled mean and labeled mean of a single replicate can deviate in different directions, increasing the observed density shift and leading to a type I error. We show that the potential for such deviation increases at low gradient resolution. This may explain why variation in the WMD relative error increases as resolution decreases.

The physics of DNA behavior within a density gradient during centrifugation and processing will affect the number of fractions where OTUs can be detected. For example, the long tails of DNA concentration versus density distributions (e.g., see Fig. S1 in the supplemental material) have been attributed to DNA smearing along the centrifuge tube walls (10). It stands to reason that the more abundant an OTU, the higher its representation in this smear will be. In addition, the detection limit of an OTU affects the number of fractions where it can be detected. We found that when inspecting presence/absence of OTUs in all fractions, OTU abundance was positively correlated with the proportion of fractions in which it was present. Abundant OTUs appear in almost all fractions, whereas rare OTUs appear in relatively few fractions, and in some cases only one fraction. These rare OTUs also tended to have highly variable WMD values.

Low gradient resolution in combination with higher variability can lead to false classification of borderline taxa as isotope incorporators when in truth, they are not (type I error). For example, a recent qSIP model simulation showed that the rate of true negatives (specificity) and true positives (sensitivity) is 88% and >90%, respectively, with no effect of fraction size (in the range of 0.003 to 0.008 g ml<sup>-1</sup>) (10). When we examined the relative specificity and sensitivity of qSIP using real unreplicated data over an even wider range of fraction sizes, we found that gradient resolution and shift detection limit strongly affect specificity and sensitivity. However, the reliability of qSIP remained extremely high as long as the detection limit was 0.005 g ml<sup>-1</sup> or higher (sensitivity > 90% and specificity > 95%) regardless of gradient resolution. This detection limit is comparable to 2 standard deviations of unreplicated unlabeled WMD, further supporting this threshold. This means that if one decides to add experimental replication, even as few as 3 replicates, one can increase the power of this analysis to have very low type II error (assuming a 0.005 g ml<sup>-1</sup> detection threshold). Thus, our analysis can be used for experimental design based on the desired statistical power.

SIP is an expensive method (18). All of the experimental design factors we assessed (number of replicates, number of fractions, distribution of fractions across tubes/spins, internal standards) directly affect the cost of a SIP experiment—in terms of supplies, enriched isotope labels, labor, and sequencing. For many researchers, processing only 9 to 12 density fractions (as our analysis suggests is sufficient) could significantly mitigate labor and sequencing costs. Cost savings will be considerable for amplicon-SIP studies, and even more so for metagenomic SIP (MG-SIP), given the high cost of metagenomic analyses.

The combination of metagenomics and SIP (26, 35) involves sequencing metagenomes instead of amplified marker genes from density fractions, and assembly of genomic bins from those metagenomes. But this approach, while powerful, remains rarely used. Early obstacles such as low DNA yield, multiple displacement amplification (MDA) biases, and low throughput (5) have now been resolved by improvements in sequencing platforms and library preparation kits. However, to date, most MG-SIP studies have sequenced only a few heavy fractions, or at best, two or three light fractions as well (27, 28, 36). While analyzing so few fractions saves money, doing so limits detection of substrate incorporators in several ways. (i) Choosing which fractions

to sequence relies on density shift of the entire community, which may be subtle. (ii) Low-GC genomes, even if highly enriched, may not become heavy enough to reach the heavy fractions. (iii) Low-GC genomic islands may not be well covered and therefore not assembled, leading to increased genome fragmentation (Fig. S4, data set 3). (iv) Depending on the density of the fractions sequenced, high-GC genomes may be highly represented regardless of enrichment. (v) As we and others have demonstrated, abundant organisms can be found in all density fractions (10) and may be erroneously classified as incorporators. Sequencing all fractions in MG-SIP studies should overcome all of these obstacles, even if only a small number of fractions are used (37). While low gradient resolution (<4 fractions) may limit detection to only highly enriched taxa, medium gradient resolution (and the decreasing price of shotgun sequencing) should keep the financial and computational costs of MG-SIP manageable while maintaining acceptable detection limits. Specifically, our data suggest that with circa 10 density fractions (fraction size  $0.011 \text{ g ml}^{-1}$ ), the increase in error compared to higher resolution is minor.

Reducing the number of fractions from  $\sim 40$  (typical for many high-resolution SIP studies [16, 17, 21, 38]) to only 10 will substantially impact the feasibility of a SIP study, particularly for MG-SIP. As an example, we compared the resources and yield of a simplified experiment. For a high-resolution MG-SIP study (40 fractions  $\times$  labeled/unlabeled = 80 metagenomes) with minimal replication (3), one could easily need to sequence 240 metagenomes. Assuming shallow sequencing of 2 Gbp per metagenome, this translates to 480 Gbp of data that would need to be stored, manipulated, and assembled and total costs (sequencing, library prep) approaching \$50,000. This highly simplified experiment could easily balloon if experimental treatments or time points were added. In comparison, our findings suggest one may reduce the number of fractions to only  $\sim 10$  ( $0.011 \text{ g ml}^{-1}$ ) and keep to a relative error lower than  $0.005 \text{ g ml}^{-1}$ . We show that for fraction sizes below  $0.011 \text{ g ml}^{-1}$ , the mean relative error increases, as does its variability (in both replicated and unreplicated data sets). However, this increase in variability can be mitigated by reallocating effort toward replication. Indeed, reducing the number of fractions by just a factor of 2 allows for doubling the number of replicates without additional costs, while increasing the statistical power of downstream analyses.

All commonly used SIP protocols rely on a linear conversion between the peak density (as a measure of the central tendency of the distribution) of an unlabeled genome and its GC content (4, 8, 29, 32, 33). However, the inherent variations that our results illustrate suggest that GC content cannot be accurately converted to density with the canonical equation (32), since the relationships between calculated %GC and known GC content have a very high  $R^2$  but various slopes and intercepts. Instead, researchers may need to create a calibration curve of %GC/WMD (or peak density) per tube, using an internal standard. Accurate conversion between WMD and %GC is particularly important for MG-SIP experiments where only the heavy fractions of labeled samples are sequenced. Once genomic bins are assembled, their %GC can be converted into a theoretical unlabeled WMD which can be used to calculate the density shift, and thus the enrichment level of those bins. It is conceivable that a reliable calculation like this could obviate the need to analyze most of the unlabeled controls and is thus another experimental design customization that can save substantial labor and costs.

If one were to combine MG-SIP with an internal %GC/density ladder, one could significantly decrease the number of unlabeled controls sequenced and use the calibration curve from the labeled tubes with the GC content of genome bins to calculate unlabeled WMD and density shift values. The internal standard should be informatically separable from the sample. This could be done by creating a mock community of organisms which are highly unlikely to appear in the sample (e.g., in a soil sample, use genomes of strictly marine organisms) and could be customized per experiment. Due to the variation within spin, an external ladder (a mock community in a separate ultracentrifuge tube) would be insufficient. However, it could be argued that finding a

suitable set of nonindigenous genomes distinguishable from a highly diverse environment such as soil may prove difficult. Alternatively, if highly complete genomic bins can be assembled, then their %GC would be more reliable, and their WMD can be calculated from the gradient. Such bins could be used as an internal standard for generating a WMD-to-GC formula. As the generation of high-completion bins could only be assessed *post hoc*, we would still recommend the use of an internal standard.

**Conclusions.** The inherent variability in qSIP stems from its many steps: replicate variation, bottle effects during incubation, extraction efficiencies, tube-to-tube variation in the gradient, amplification bias, strain heterogeneity, among-treatment shifts in community composition, and OTU clustering errors (for marker genes). Quantifying the sensitivity of qSIP to these factors will improve existing amplicon-based qSIP techniques and facilitate efficient ways of extending SIP to more ambitious applications, such as metagenome-assembled genome-based SIP. The analyses we present here illustrate how customizing experimental design factors can allow researchers to achieve ideal levels of sensitivity and specificity while keeping costs in check.

## MATERIALS AND METHODS

We used five unpublished data sets from different ecosystems for our *in silico* analyses: a high-resolution unreplicated SIP study of [<sup>13</sup>C]naphthalene in seawater, two medium-resolution replicated SIP experiments where H<sub>2</sub><sup>18</sup>O was added to soils (SPRUCE [Spruce and Peatland Responses Under Changing Environments] peatland from northern Minnesota, Mediterranean grassland from northern California), replicated genomic DNA from pure cultures of *Escherichia coli* and *Pseudomonas putida*, and a replicated genomic mock community comprised of high-molecular-weight genomic DNA of *Thermoanaerobacter pseudethanolicus*, *Bacillus licheniformis*, *Bifidobacterium longum* subsp. *infantis*, and *Streptomyces violaceoruber*, purchased from ATCC. Table 1 contains the number of density fractions and replicates per data set. As these experiments were performed by different laboratories, using slightly different protocols, we describe their SIP pipelines separately (see Text S1 in the supplemental material). All postsequencing steps were performed identically for all 16S rRNA operational taxonomic units (OTUs).

**Data set 1. [<sup>13</sup>C]naphthalene-enriched seawater.** Data set 1 contains bacterial 16S rRNA OTUs from [<sup>13</sup>C]naphthalene-enriched seawater collected from the port of Los Angeles, California (30). For full details, see Text S1.

**Data set 2. SPRUCE peatland H<sub>2</sub><sup>18</sup>O addition.** Data set 2 contains 16S-rRNA OTUs from a forested peat bog soil, part of the Spruce and Peatland Responses Under Changing Environments (SPRUCE) experiment in the Marcell Experimental Forest in Minnesota incubated with H<sub>2</sub><sup>18</sup>O. For full details, see Text S1.

**Data set 3. Grassland soil H<sub>2</sub><sup>18</sup>O addition.** Data set 3 contains metagenomes from grassland soil (California) incubated with H<sub>2</sub><sup>18</sup>O. For full details, see Text S1.

**Data set 4. *E. coli* and *P. putida* cultures.** A 1:1 mix of DNA from unlabeled *E. coli* and 0.5 μg DNA from unlabeled ( $n = 8$ ) or labeled ( $n = 12$ ) *P. putida* was processed on an automated SIP pipeline at the Joint Genome Institute. For full details, see Text S1.

**Data set 5. Genomic mock communities.** DNA for a genomic mock community was purchased from ATCC, resuspended in Tris-EDTA buffer, mixed in equal proportions, and aliquoted into replicates. The mock community was composed of high-molecular-weight DNA of *Thermoanaerobacter pseudethanolicus*, *Bacillus licheniformis*, *Bifidobacterium longum* subsp. *infantis*, and *Streptomyces violaceoruber* (see Table S1 for accession numbers), and these genomes were selected for their distinct %GC content (34.5%, 46%, 60% and 73%, respectively). For full details, see Text S1.

**16S rRNA gene-based microbial community composition in individual fractions.** Each data set was processed separately. For data sets 1 and 2, reads were quality trimmed using Trimmomatic version 0.33 (40) with parameters set at LEADING:20 TRAILING:20 SLIDINGWINDOW:15:25. The resulting reads were merged using Usearch version 7 (41), clustered in mothur following the MiSeq standard operating procedure (SOP) (42–44). To track individual operational taxonomic units (OTUs) in separate density fractions, the relative abundance of the OTU in each fraction was multiplied by either the concentration of DNA in the same fraction (seawater, data set 1; as in reference 20) or by the total 16S rRNA gene copy number (SPRUCE soil, data set 2; as in reference 9).

**Density shifts.** The weighted mean density of each OTU in labeled and control samples was calculated by multiplying density by OTU abundance (amount of DNA/16S rRNA gene copies × OTU relative abundance) within each fraction, summing the products and dividing them by the sum of abundances of the OTU across all fractions. The density shift was calculated by subtracting the weighted mean density of the OTU in the unlabeled control from the its weighted mean density in the labeled treatment (9). Density shifts were then plotted in R (45) for the 100 most abundant OTUs in each sample. The relative error was calculated as the difference between the density shift per OTU in resolution  $r$  minus the shift per OTU in the original high-resolution  $r_{\max}$ .

**Sensitivity analysis.** We used the variation in the weighted mean density (WMD) of the same OTU in replicated samples to assess the minimum density shift that can be detected. The leading principle here was that a shift smaller than the natural variation could not be statistically significant.

Using unlabeled replicated ( $n = 10$ ) samples from data set 2 time zero (SPRUCE soil, with five subsamples from each of two preincubation temperatures [15°C and 45°C]), we calculated the WMD for each OTU in each replicate and the standard deviation of the 10 resulting means for each of the 100 most abundant OTUs. We also assessed the effect of OTU abundance on WMD variability using all 320 OTUs from the same data set (Text S1). The use of standard deviation has an underlying assumption of normality. Consistent with this assumption, we observed only slight leptokurtosis in the distribution of taxon relative abundance as a function of density that is not thought to strongly influence inference (Fig. S1).

**Sensitivity to number of fractions.** We used data set 1 (seawater) and data set 2 (SPRUCE soil) to estimate how precision in the estimate of taxon isotope incorporation varies with the number of density fractions collected in a qSIP experiment. We combined fractions *in silico* to simulate the results of these experiments had they been run with fewer fractions ( $f$ ), with the following principles. (i) Only adjacent fractions were combined. (ii) Fraction combinations were conducted to create new, combined fractions that were approximately equal in size and sequencing depth (i.e., with minimal variation in the range of densities represented by each simulated fraction). For example, to simulate an experiment where only two density fractions were collected instead of 18, we ran three possible scenarios using the original empirical data: combining the lightest 8, 9, or 10 fractions into a simulated fraction and combining the remaining heaviest fractions into a second simulated fraction. Similarly, to simulate an experiment with 10 density fractions instead of 18, we ran all 45 possible scenarios of combining eight pairs of adjacent density fractions with two uncombined single density fractions. In this way, we simulated data sets across the full range of possible fraction numbers, from 2 to 18. We did this to simulate typical approaches to SIP experiments, where fractions that span similar density ranges are usually combined. We computed the weighted-average density for each taxon in each replicate of each simulated data set using the midpoint of each density fraction (combined or not) and weighting each of those density values by the area under the curve for that fraction, where the curve is delineated by the plot of 16S rRNA gene copies versus density. The sum of those weighted densities, divided by the total area under the curve, yielded the weighted-average density metric used in subsequent analyses. For nonpermuted analyses, fractions were combined from preincubation samples (temperature, 15°C and 45°C [ $n = 10$ ]). For each permuted combination (using the replicated data set 2), we ran the qSIP code (13, 46) and estimated the atom percent excess  $^{18}\text{O}$  for each replicate sample and then calculated the standard deviation of that estimate across all replicates ( $n = 5$ , preincubation temperature of 15°C). Finally, we calculated the relative standard deviation as a function of increasing number of simulated fractions compared to the original number of fractions.

**Power analysis.** We evaluated statistical power using the SPRUCE data set (data set 2). These data came from a SIP experiment where peat bog soils were incubated for 10 days at an intermediate temperature (15°C) and then harvested after 5 or 10 days of exposure to  $\text{H}_2^{18}\text{O}$ . An unlabeled control sample was sampled at day 0. For the power analysis, we omitted 27 taxa that did not occur in all 15 samples ( $n = 5$  for control,  $\text{H}_2^{18}\text{O}$  at day 5, and  $\text{H}_2^{18}\text{O}$  at day 10), 5 rare taxa (relative abundance of  $<0.026\%$ ), and 7 taxa identified as outliers for variance of the estimate of weighted average density ( $P < 0.05$ , Grubb's test). Excluding taxa with high variance will inflate power, but the effect is likely negligible given the small number of taxa excluded for this reason. Of the remaining 236 taxa, 211 were included in the power analysis. We used the observed variation among taxa in weighted average density shift, which ranged from  $-0.003$  to  $0.033 \text{ g ml}^{-1}$ . This captures a wide range of possible values of isotope uptake, from  $\sim 0$  to  $\sim 60$  atom% excess  $^{18}\text{O}$ . We used *in silico* resampling with replacement to estimate power. For each taxon at each sample date,  $n$  random samples were drawn (with replacement) from each of the  $^{18}\text{O}$ -labeled and unlabeled data sets, a  $t$  test was performed, and the  $P$  value was recorded. This was repeated 1,000 times, and power was estimated as the frequency of significant  $t$  tests among the 1,000 simulations (47, 48).  $N$  (number of sample replicates) was varied between 2 and 6 to simulate experiments with different replication by pruning or duplicating replicates from the original data set. Average power was calculated across all taxa. The upper 10th percentile of power was also calculated to estimate power typical for more dominant taxa.

**Data availability.** Data set 1 sequence data are available in ENA under project number [PRJEB26952](#) and accession numbers [ERS2507530](#) to [ERS2507619](#). Data set 2 sequence data were submitted to NCBI Sequence Read Archive (SRA) under project ID [PRJNA641177](#) and accession numbers [SRR12071929](#) to [SRR12072082](#).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**TEXT S1**, DOCX file, 0.02 MB.

**FIG S1**, PDF file, 0.1 MB.

**FIG S2**, PDF file, 0.3 MB.

**FIG S3**, PDF file, 0.1 MB.

**FIG S4**, PDF file, 1.7 MB.

**TABLE S1**, PDF file, 0.02 MB.

**DATA SET S1**, XLSX file, 0.02 MB.

## ACKNOWLEDGMENTS

We thank the SPRUCE team, Michaela Hayer, Paul Dijkstra, Sheryl Bell, and Ember Morrissey for generously providing data and Andy Tomatsu at the DOE Joint Genome Institute for operating the high-throughput SIP pipeline for the mock communities.

Support for analyses and data integration was provided by the U.S. Department of Energy, Office of Biological and Environmental Research, Genomic Science Program LLNL “Microbes Persist” Soil Microbiome Scientific Focus Area (award SCW1632), and awards DE-SC0016207 and DE-SC0020172 at Northern Arizona University. Work conducted at LLNL was contributed under the auspices of the U.S. Department of Energy under contract DE-AC52-07NA27344. The work conducted by the U.S. Department of Energy Joint Genome Institute was supported under contract DE-AC02-05CH11231. E.T.S. was supported by U.S. Department of Energy Office of Science award DE-SC0014079 to UC Berkeley.

## REFERENCES

- Radajewski S, Ineson P, Parekh NR, Murrell JC. 2000. Stable-isotope probing as a tool in microbial ecology. *Nature* 403:646–649. <https://doi.org/10.1038/35001054>.
- Manefield M, Whiteley AS, Griffiths RI, Bailey MJ. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl Environ Microbiol* 68:5367–5373. <https://doi.org/10.1128/aem.68.11.5367-5373.2002>.
- Radajewski S, McDonald IR, Murrell JC. 2003. Stable-isotope probing of nucleic acids: a window to the function of uncultured microorganisms. *Curr Opin Biotechnol* 14:296–302. [https://doi.org/10.1016/s0958-1669\(03\)00064-8](https://doi.org/10.1016/s0958-1669(03)00064-8).
- Neufeld JD, Vohra J, Dumont MG, Lueders T, Manefield M, Friedrich MW, Murrell JC. 2007. DNA stable-isotope probing. *Nat Protoc* 2:860–866. <https://doi.org/10.1038/nprot.2007.109>.
- Chen Y, Murrell JC. 2010. When metagenomics meets stable-isotope probing: progress and perspectives. *Trends Microbiol* 18:157–163. <https://doi.org/10.1016/j.tim.2010.02.002>.
- Dumont MG, García MH (ed). 2019. Stable isotope probing: methods and protocols. Springer, New York, NY.
- Lueders T, Manefield M, Friedrich MW. 2004. Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environ Microbiol* 6:73–78. <https://doi.org/10.1046/j.1462-2920.2003.00536.x>.
- Lueders T. 2010. Stable isotope probing of hydrocarbon-degraders, p 4011–4026. *In* Timmis KN (ed), *Handbook of hydrocarbon and lipid microbiology*. Springer, New York, NY.
- Hungate BA, Mau RL, Schwartz E, Caporaso JG, Dijkstra P, van Gestel N, Koch BJ, Liu CM, McHugh TA, Marks JC, Morrissey EM, Price LB. 2015. Quantitative microbial ecology through stable isotope probing. *Appl Environ Microbiol* 81:7570–7581. <https://doi.org/10.1128/AEM.02280-15>.
- Youngblut ND, Barnett SE, Buckley DH. 2018. SIPSim: a modeling toolkit to predict accuracy and aid design of DNA-SIP experiments. *Front Microbiol* 9:570. <https://doi.org/10.3389/fmicb.2018.00570>.
- Barnett SE, Youngblut N, Buckley DH. 2019. Data analysis for DNA stable isotope probing experiments using multiple window high-resolution SIP, p 109–128. *In* Dumont MG, García MH (ed), *Stable isotope probing: methods and protocols*. Springer, New York, NY.
- Barnett SE, Buckley DH. 2020. Simulating metagenomic stable isotope probing datasets with MetaSIPSim. *BMC Bioinformatics* 21:37. <https://doi.org/10.1186/s12859-020-3372-6>.
- Koch BJ, McHugh TA, Hayer M, Schwartz E, Blazewicz SJ, Dijkstra P, Gestel N, Marks JC, Mau RL, Morrissey EM, Pett-Ridge J, Hungate BA. 2018. Estimating taxon-specific population dynamics in diverse microbial communities. *Ecosphere* 9:e02090. <https://doi.org/10.1002/ecs2.2090>.
- Schwartz E. 2007. Characterization of growing microorganisms in soil by stable isotope probing with H<sub>2</sub><sup>18</sup>O. *Appl Environ Microbiol* 73:2541–2546. <https://doi.org/10.1128/AEM.02021-06>.
- Blazewicz SJ, Schwartz E. 2011. Dynamics of <sup>18</sup>O incorporation from H<sub>2</sub><sup>18</sup>O into soil microbial DNA. *Microb Ecol* 61:911–916. <https://doi.org/10.1007/s00248-011-9826-7>.
- Morando M, Capone DG. 2016. Intraclade heterogeneity in nitrogen utilization by marine prokaryotes revealed using stable isotope probing coupled with tag sequencing (Tag-SIP). *Front Microbiol* 7:1932. <https://doi.org/10.3389/fmicb.2016.01932>.
- Morrissey EM, Mau RL, Schwartz E, Koch BJ, Hayer M, Hungate BA. 2018. Taxonomic patterns in the nitrogen assimilation of soil prokaryotes. *Environ Microbiol* 20:1112–1119. <https://doi.org/10.1111/1462-2920.14051>.
- Angel R. 2019. Experimental setup and data analysis considerations for DNA- and RNA-SIP experiments in the omics era. *In* Dumont MG, García MH (ed), *Stable isotope probing: methods and protocols*. Springer, New York, NY.
- Blazewicz SJ, Schwartz E, Firestone MK. 2014. Growth and death of bacteria and fungi underlie rainfall-induced carbon dioxide pulses from seasonally dried soil. *Ecology* 95:1162–1172. <https://doi.org/10.1890/13-1031.1>.
- Papp K, Mau RL, Hayer M, Koch BJ, Hungate BA, Schwartz E. 2018. Quantitative stable isotope probing with H<sub>2</sub><sup>18</sup>O reveals that most bacterial taxa in soil synthesize new ribosomal RNA. *ISME J* 12:3043–3045. <https://doi.org/10.1038/s41396-018-0233-7>.
- Blazewicz SJ, Hungate BA, Koch BJ, Nuccio EE, Morrissey E, Brodie EL, Schwartz E, Pett-Ridge J, Firestone MK. 2020. Taxon-specific microbial growth and mortality patterns reveal distinct temporal population responses to rewetting in a California grassland soil. *ISME J* 14:1520–1532. <https://doi.org/10.1038/s41396-020-0617-3>.
- Hu A, Lu Y, Hernández García M, Dumont MG. 2019. Targeted metatranscriptomics of soil microbial communities with stable isotope probing. *Methods Mol Biol* 2046:163–174. [https://doi.org/10.1007/978-1-4939-9721-3\\_13](https://doi.org/10.1007/978-1-4939-9721-3_13).
- Sul WJ, Park J, Quensen JF, III, Rodrigues JLM, Seliger L, Tsoi TV, Zylstra GJ, Tiedje JM. 2009. DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Appl Environ Microbiol* 75:5501–5506. <https://doi.org/10.1128/AEM.00121-09>.
- Wilhelm RC, Singh R, Eltis LD, Mohn WW. 2019. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J* 13:413–429. <https://doi.org/10.1038/s41396-018-0279-6>.
- Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, Szeto E, Salamov A, Grigoriev IV, Suci D, Levine SR, Markowitz VM, Rigoutsos I, Tringe SG, Bruce DC, Richardson PM, Lidstrom ME, Chistoserdova L. 2008. High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol* 26:1029–1034. <https://doi.org/10.1038/nbt.1488>.
- Dumont MG, Radajewski SM, Miguez CB, McDonald IR, Murrell JC. 2006. Identification of a complete methane monooxygenase operon from soil by combining stable isotope probing and metagenomic analysis. *Environ Microbiol* 8:1240–1250. <https://doi.org/10.1111/j.1462-2920.2006.01018.x>.
- Dombrowski N, Donaho JA, Gutierrez T, Seitz KW, Teske AP, Baker BJ. 2016. Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. *Nat Microbiol* 1:16057. <https://doi.org/10.1038/nmicrobiol.2016.57>.
- Thomas F, Corre E, Cébron A. 2019. Stable isotope probing and metagenomics highlight the effect of plants on uncultured phenanthrene-

- degrading bacterial consortium in polluted soil. *ISME J* 13:1814–1830. <https://doi.org/10.1038/s41396-019-0394-z>.
29. Murrell JC, Whiteley AS (ed). 2011. Stable isotope probing and related technologies. ASM Press, Washington, DC.
  30. Sieradzki ET, Morando M, Fuhrman JA. 2019. Metagenomics and stable isotope probing offer insights into metabolism of polycyclic aromatic hydrocarbon degraders in chronically polluted seawater. *bioRxiv* <https://doi.org/10.1101/777730>.
  31. Rickwood D, Birnie GD. 1978. Introduction: Principles and practices of centrifugation, p 1–6. *In* Birnie GD, Rickwood D (ed), *Centrifugal separations in molecular and cell biology*. Butterworth-Heinemann, Oxford, United Kingdom.
  32. Schildkraut CL, Marmur J, Doty P. 1962. Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. *J Mol Biol* 4:430–443. [https://doi.org/10.1016/s0022-2836\(62\)80100-4](https://doi.org/10.1016/s0022-2836(62)80100-4).
  33. Buckley DH, Huangyutham V, Hsu S-F, Nelson TA. 2007. Stable isotope probing with <sup>15</sup>N achieved by disentangling the effects of genome G+C content and isotope enrichment on DNA density. *Appl Environ Microbiol* 73:3189–3195. <https://doi.org/10.1128/AEM.02609-06>.
  34. Underwood AJ. 1997. Environmental decision-making and the precautionary principle: what does this principle mean in environmental sampling practice? *Landscape Urban Plan* 37:137–146. [https://doi.org/10.1016/S0169-2046\(97\)80000-X](https://doi.org/10.1016/S0169-2046(97)80000-X).
  35. Schwarz S, Waschkowitz T, Daniel R. 2006. Enhancement of gene detection frequencies by combining DNA-based stable-isotope probing with the construction of metagenomic DNA libraries. *World J Microbiol Biotechnol* 22:363–368. <https://doi.org/10.1007/s11274-005-9042-z>.
  36. Fortunato CS, Huber JA. 2016. Coupled RNA-SIP and metatranscriptomics of active chemolithoautotrophic communities at a deep-sea hydrothermal vent. *ISME J* 10:1925–1938. <https://doi.org/10.1038/ismej.2015.258>.
  37. Starr EP, Shi S, Blazewicz SJ, Probst AJ, Herman DJ, Firestone MK, Banfield JF. 2018. Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize microbially-processed plant-derived carbon. *Microbiome* 6:122. <https://doi.org/10.1186/s40168-018-0499-z>.
  38. Pepe-Rannek C, Campbell AN, Koechli CN, Berthrong S, Buckley DH. 2016. Unearthing the ecology of soil microorganisms using a high resolution DNA-SIP approach to explore cellulose and xylose metabolism in soil. *Front Microbiol* 7:703. <https://doi.org/10.3389/fmicb.2016.00703>.
  39. Reference deleted.
  40. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
  41. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
  42. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
  43. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–1829. <https://doi.org/10.1093/bioinformatics/bts252>.
  44. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79:5112–5120. <https://doi.org/10.1128/AEM.01043-13>.
  45. R Core Team. 2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
  46. Finley BK, Hayer M, Mau RL, Purcell AM, Koch BJ, van Gestel NC, Schwartz E, Hungate BA. 2019. Microbial taxon-specific isotope incorporation with DNA quantitative stable isotope probing. *Methods Mol Biol* 2046:137–149. [https://doi.org/10.1007/978-1-4939-9721-3\\_11](https://doi.org/10.1007/978-1-4939-9721-3_11).
  47. Cohen J. 2013. *Statistical power analysis for the behavioral sciences*. Academic Press, Cambridge, MA.
  48. Crowley PH. 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. *Annu Rev Ecol Syst* 23:405–447. <https://doi.org/10.1146/annurev.es.23.110192.002201>.