# Evaluating the accuracy of DNA stable isotope probing

Nicholas D. Youngblut and Daniel H. Buckley*

School of Integrative Plant Science, Cornell University, 306 Tower Road, Ithaca, NY, 14853

*Corresponding author:
Daniel H. Buckley
Section of Soil and Crop Sciences
306 Tower Road
Ithaca, NY 14853
dbuckley@cornell.edu
(607) 255-1716

Running title: *Evaluating the accuracy of DNA-SIP*

Keywords: SIP, DNA, method, rRNA, microbial, bacterial, community, functional

## Originality-Significance Statement

By combining DNA Stable Isotope Probing (DNA-SIP) with multiplexed high throughput DNA sequencing (HTS-DNA-SIP), it is now possible to identify patterns of isotope incorporation for thousands of microbial taxa. HTS-DNA-SIP has enormous potential to reveal patterns of carbon and nitrogen exchange within microbial food webs. A current limitation is that, due to the expense of these experiments, it has been impossible to evaluate the accuracy of DNA-SIP methods. We have developed a model that simulates DNA-SIP data, and we use the model to systematically evaluate and validate the accuracy of DNA-SIP analyses. This model can determine the analytical accuracy of DNA-SIP experiments in a range of contexts. Furthermore, the ability to predict experimental outcomes, as a function of experimental design and community characteristics, should be of great use in the design and interpretation DNA-SIP experiments.

## Summary

DNA Stable isotope probing (DNA-SIP) is a powerful method that identifies *in situ* isotope assimilation by microbial taxa. Combining DNA-SIP with multiplexed high throughput DNA sequencing (HTS-DNA-SIP) creates the potential to map *in situ* assimilation dynamics for thousands of microbial taxonomic units. However, the accuracy of methods for analyzing DNA-SIP data has never been evaluated. We have developed a toolset (SIPSim) for simulating HTS-DNA-SIP datasets and evaluating the accuracy of methods for analyzing HTS-DNA-SIP data. We evaluated two different approaches to analyzing HTS-DNA-SIP data: "high resolution stable isotope probing" (HR-SIP) and "quantitative stable isotope probing" (q-SIP). HR-SIP was highly specific and moderately sensitive, with very few false positives

23 but potential for false negatives. In contrast, q-SIP had fewer false negatives but many false
24 positives. We also found HR-SIP more robust than q-SIP with respect to experimental
25 variance. Furthermore, we found that the detection sensitivity of HTS-DNA-SIP can be
26 increased without compromising specificity by evaluating evidence of isotope
27 incorporation over multiple windows of buoyant density (MW-HR-SIP). SIPSim provides a
28 platform for determining the accuracy of HTS-DNA-SIP methods across a range of
29 experimental parameters, which will be useful in the design, analysis, and validation of
30 DNA-SIP experiments.
31

32 **Introduction**

33 Stable isotope probing of nucleic acids (DNA-SIP and RNA-SIP) is a powerful culture-
34 independent method for linking microbial metabolic functioning to taxonomic identity
35 (Radajewski *et al.*, 2003). In particular, DNA-SIP has been used extensively to identify
36 microbial assimilation of various $^{13}C$- and $^{15}N$-labeled substrates in a multitude of
37 environments (Uhlík *et al.*, 2009). DNA-SIP identifies microbes that assimilate isotope into
38 their DNA ("incorporators") by exploiting the increased buoyant density (BD) of
39 isotopically labeled ("heavy") DNA relative to unlabeled ("light") DNA. For example, fully
40 $^{13}C$- and $^{15}N$-labeled DNA will increase in BD by 0.036 and 0.016 g $ml^{-1}$, respectively (Birnie
41 and Rickwood, 1978).
42

43 Ideally, isopycnic centrifugation could be used to completely separate labeled and
44 unlabeled DNA fragments based solely on this difference in BD. However, several factors
45 besides BD can impact the position of DNA in isopycnic gradients. For example, G + C

content variation across a single genome can produce unlabeled DNA fragments that vary in BD by up to 0.03 g $ml^{-1}$, while G + C content variation between microbial genomes can cause the average BD of unlabeled DNA fragments to vary by up to 0.05 g $ml^{-1}$ (Youngblut and Buckley, 2014). In addition, DNA in SIP experiments will often be partially labeled as a consequence of isotope dilution from unlabeled endogenous substrates. Therefore, it is unlikely that nucleic acid SIP experiments will ever achieve complete separation of labeled and unlabeled DNA.

In the absence of complete separation between labeled and unlabeled DNA, isotope incorporators must be identified using some statistical procedure suitable for comparing the BD distributions of DNA fragments from labeled and unlabeled samples (Pepe-Ranney *et al.*, 2016a). The use of multiplexed high throughput sequencing with DNA-SIP ("HTS-DNA-SIP") makes it possible to sequence SSU rRNA amplicons across many density gradient fractions and simultaneously determine the BD distributions for thousands of taxa. The problem then becomes one of identifying those taxa that have increased in BD in the isotopically labeled samples relative to the corresponding unlabeled controls.

Different analytical approaches have been applied to HTS-DNA-SIP datasets to identify changes in DNA BD in response to isotopic labeling. These include "high resolution stable isotope probing" (HR-SIP) and "quantitative stable isotope probing" (q-SIP), which both analyze SSU rRNA amplicons across numerous gradient fractions (Hungate *et al.*, 2015; Pepe-Ranney *et al.*, 2016a; Pepe-Ranney *et al.*, 2016b). However, these methods differ in the statistical procedures used to detect taxa that incorporate isotopic label. HR-SIP

identifies isotopically labeled taxa by evaluating the sequence composition of high density "heavy" fractions using a differential abundance quantification framework that evaluates sequence count data in isotopically labeled samples relative to their corresponding unlabeled controls. Differential abundance between the "heavy" fractions of labeled and control gradients is measured with DESeq2 (Love *et al.*, 2014), which uses sophisticated statistical methods to reduce technical error and increase analytical power for analysis of microbiome data (McMurdie and Holmes, 2014). In a very different approach, q-SIP transforms SSU rRNA relative abundance values by using qPCR estimates of total SSU rRNA gene copies present within gradient fractions. These normalized data are used to estimate average BD for each taxon across density gradients for both isotopically labeled samples and corresponding unlabeled controls (Hungate *et al.*, 2015). Incorporators are then determined by using a permutation procedure to identify those taxa whose BD shifts are unlikely to occur as a result of chance.

While DNA-SIP is a powerful method for the discovery and characterization of microorganisms *in situ*, systematic assessment of the specificity or sensitivity of this method has not been performed. Empirical validations of DNA-SIP methods typically include only one or a few organisms (Lueders *et al.*, 2004; Buckley *et al.*, 2007; Cupples *et al.*, 2007; Wawrik *et al.*, 2009; Andeer *et al.*, 2012), and such approaches do not adequately replicate the complexity of the DNA fragment BD distributions expected in a typical DNA-SIP experiment (Youngblut and Buckley, 2014). DNA-SIP experiments vary in the diversity of the target community, DNA G + C content distribution, the number of incorporators, incorporator relative abundance, and the atom % excess of labeled DNA. Systematic

92 evaluation of method accuracy should address the effects that all of these variables have on
93 the sensitivity and specificity of detecting isotope incorporators. Since DNA-SIP
94 experiments are costly, technically difficult, and laborious, it is not practical to perform
95 empirical assessment across this full range of variables.
96
97 Fortunately, the physics of isopycnic centrifugation have been well characterized
98 mathematically, and the behavior of individual DNA fragments in CsCl gradients is highly
99 reproducible and predictable from first principles (Meselson *et al.*, 1957; Fritsch, 1975;
100 Birnie and Rickwood, 1978). In addition, genome sequences are available for thousands of
101 diverse microorganisms, and these genomes can be used to generate DNA fragments
102 representative of community DNA (Youngblut and Buckley 2014). Hence, we can simulate
103 realistic HTS-DNA-SIP data for *in silico* microbial communities that differ in diversity
104 (richness, evenness, and composition), where the relative abundance, genome G + C
105 content, and atom % excess isotope are defined for discrete DNA fragments from every
106 genome. We have developed a computational toolset for simulating HTS-DNA-SIP data
107 (SIPSim) and used this simulation framework to systematically and objectively evaluate
108 how changes in key SIP experimental parameters affect HTS-DNA-SIP accuracy.
109
110 **Results**
111 *Model validation and parameter estimation*
112 The SIPSim model starts with a set of user-designated genomes and user-designated
113 experimental parameters (*e.g.* number of gradient fractions, desired community
114 characteristics, desired isotopic labeling characteristics) as described (see *Experimental*

115 *Procedures* and *Supporting Information*). Briefly, the genomes are fragmented as would
116 occur during DNA extraction, isotopic labeling is applied to some number of genomes as
117 specified by the user, the BD distributions are determined for each DNA fragment and
118 fragment collections are then binned into gradient fractions, fragments are sampled from
119 each fraction as would occur during amplification and DNA sequencing of SSU rRNA genes,
120 and then the relative abundance is calculated for each OTU (Figure 1). The model produces
121 results that are highly similar to those observed in empirical experiments, including the
122 ability to detect DNA fragments throughout the density gradient (Figure 2).
123
124 The development of the simulation model was guided by established centrifugal theory and
125 by comparison of simulated results to empirical data (as in *Experimental Procedures* in
126 *Supporting Information*). First, we performed a simple evaluation of model performance by
127 recreating results from a prior DNA-SIP experiment with *Methanosarcina barkeri* MS and
128 *Methylobacterium extorquens* AM1 (Lueders *et al.*, 2004) (Figure S1). Simulated DNA
129 distributions (both in terms of total DNA and SSU rRNA gene amplicon copies) significantly
130 and strongly correlated with the empirical data for both taxa ($p < 0.003$ for all
131 comparisons; see Table S1). In addition, the simulated SSU rRNA gene amplicon-fragment
132 BD distributions were shifted 0.007 g $ml^{-1}$ toward the middle of the BD gradient relative to
133 the shotgun-fragments ("total DNA"), a phenomenon also observed in the empirical data.
134 This central tendency for SSU rRNA amplicon-fragments reflects G + C conservation of the
135 *rrn* operon, as previously described (Youngblut and Buckley, 2014).
136

Next, we determined whether the simulation accurately modeled variation in BD within complex mixtures of unlabeled DNA by comparing simulation results to empirical results obtained with unlabeled DNA from soil. For this purpose we used empirical data from an experiment (Youngblut *et al*., in prep.) in which DNA was extracted from soil microcosms at 1, 3, 6, 14, 30, and 48 days following the addition of an unlabeled carbon source mixture. These six DNA samples were equilibrated in CsCl gradients, fractionated by BD, and SSU rRNA gene amplicons were sequenced for ~24 fractions from each gradient. Simulation input included 1147 microbial genomes (see *Experimental Procedures*), hence the soil data was resampled to 1147 OTUs in order to standardize the richness of the simulated and empirical data. Ideally, we could map SSU rRNA sequences from soil to all bacterial genomes available in public databases, but genome composition can vary dramatically across taxa that have identical SSU rRNA gene sequences. Since the genome sequences of taxa in the empirical HTS-DNA-SIP dataset cannot be confidently assigned to genomes in existing databases, a direct mapping of taxa (and their genomes) between the empirical and simulated datasets was not possible. We therefore employed metrics that capture variation in DNA fragment BD distributions within density gradients, and which thereby allow for gradient to gradient comparison of DNA BD distributions (see *Supporting Information*).

The empirical BD distributions (Figure S2) show that temporal change in soil mesocosm community composition caused dramatic shifts in the Shannon diversity of 'heavy fractions' even in the absence of isotopic labeling (Figure S2B), with heavy fraction diversity increasing at later time points. Moreover, taxonomic similarity within a gradient

is auto-correlated across the BD gradient (Figure S2C). Lastly, variance in amplicon fragment BD is positively correlated with OTU relative abundance in the community (Figure S2D), with highly abundant OTUs found throughout the CsCl gradient. We found that the simulation model was able to recapitulate these results across a wide range of parameter space, and that the variance between simulated and empirical results was less than that observed between replicate empirical samples (Figure S3). We used these comparisons to determine model parameter values (Table S2), which provided the best fit to the actual behavior of DNA fragments in CsCl gradients (as described in *Supporting Information*).

*The influence of isotope incorporation on HTS-DNA-SIP accuracy*

We hypothesized that both the number of taxa that incorporate isotope and the atom % isotope incorporation per taxon would substantially affect the accuracy of HTS-DNA-SIP methods. To test these predictions, we simulated HTS-DNA-SIP datasets for both $^{13}C$-labeled samples and unlabeled controls (3 replicates of each), while varying both the number of incorporators (1, 5, 10, 25, or 50 % of taxa) and the atom % isotope incorporation for each taxon (0, 15, 25, 50, or 100 atom % excess $^{13}C$). Taxa in the control were always set to 0 % isotope incorporation. Each simulation was replicated 10 times, with differing taxa randomly designated as incorporators in each replicate. We evaluated 4 methods used to analyze HTS-DNA-SIP data: Heavy-SIP, q-SIP, HR-SIP, and MW-HR-SIP. Heavy-SIP involved simply identifying as incorporators all taxa observed in "heavy" gradient fractions of labeled gradients, which provided a baseline of accuracy for the more complex HTS-DNA-SIP analyses. q-SIP and HR-SIP were performed as described in Hungate

183 *et al.*, (2015) and Pepe-Ranney *et al.*, (2016a), respectively. MW-HR-SIP was performed
184 similarly to HR-SIP, but with multiple overlapping "heavy" buoyant density windows (see
185 *Experimental Procedures*).
186
187 As expected, both the number of incorporators and the amount of isotope incorporated
188 affected accuracy (Figure 3). However, the effect of these parameters on specificity and
189 sensitivity varied depending on the analytical method (Figure 3). Specificity is the
190 proportion of true negatives observed out of all true negatives expected, and so specificity
191 declines in direct relation to an increase in the number of false positives. For example, a
192 specificity of 0.8 would generate 200 false positives in a sample of 1000 unlabeled taxa.
193 Specificity, as measured across a wide range in parameters, was highest for MW-HR-SIP (1
194 ± 0; ave. ± s.d.) and HR-SIP (1 ± 0), substantially lower for q-SIP (0.88 ± 0.06), and very low
195 for Heavy-SIP (0.28 ± 0.16) (Figure 3).
196
197 Sensitivity is the fraction of true positives observed out of all true positives expected. For
198 example, a sensitivity of 0.7 means that a method failed to detect 30 % of the incorporators
199 present. Both q-SIP and Heavy-SIP had relatively high sensitivity (median values of 0.91
200 and 0.93, respectively), and the sensitivity of these methods was largely insensitive to the
201 atom % excess of DNA and the number of incorporators (Figure 3). In contrast, the
202 sensitivities of both HR-SIP and MW-HR-SIP were highly responsive to the atom % excess
203 of DNA, and the number of incorporators (Figure 3). For these methods, sensitivity
204 declined in proportion to the atom % excess $^{13}C$ label in DNA.
205

206 Balanced accuracy is calculated as the mean of specificity and sensitivity. We observed a
207 tradeoff in balanced accuracy in relation to the atom % excess $^{13}C$ of DNA. MW-HR-SIP had
208 the highest accuracy when % atom excess $^{13}C$ exceeded 50 %, but q-SIP had higher
209 accuracy at lower levels of isotope incorporation (Figure 3). This tradeoff in balanced
210 accuracy resulted from a difference in the tolerance for false positives. For example, MW-
211 HR-SIP produced nearly zero false positives but as a result of its high specificity, it lost
212 sensitivity at lower levels of isotope incorporation. In contrast, q-SIP detected labeled taxa
213 across a wider range of isotope incorporation, but it did so at the cost of a large number of
214 false positives.
215
216 *The influence of community variation on HTS-DNA-SIP accuracy*
217 All HTS-DNA-SIP analyses rely upon comparisons made between isotopically enriched
218 experimental treatments and their corresponding unlabeled controls. In real SIP
219 experiments the composition of replicate post incubation communities are likely to vary
220 somewhat as a result of sample heterogeneity and incubation effects. However, the
221 simulations described above assume random sampling from identical pre-fractionation
222 (post-incubation) community structures. We hypothesized that an increase in variation in
223 community composition between treatment and control samples would decrease the
224 accuracy of HTS-DNA-SIP analyses. To test this hypothesis, we generated simulations in
225 which isotope incorporation was held constant (100 atom % excess $^{13}C$; 10 % of OTUs are
226 incorporators) but beta-diversity was varied among 3 replicate treatment and 3 replicate
227 control samples. We varied beta-diversity in two ways: *i*) using permutation to vary the
228 rank abundance of a fixed proportion of community members and *ii*) varying the

proportion of taxa shared between communities. For each simulation scenario, we calculated the mean Bray-Curtis distance among communities in order to provide a real-world metric for gauging the potential accuracy of actual DNA-SIP experiments.

As hypothesized, increased beta-diversity among samples had a substantial impact on the accuracy of HTS-DNA-SIP methods (Figure 4). Accuracy was impacted more by the number of taxa shared between samples than by differences in taxon abundance (Figure S4). The sensitivity of q-SIP declined as beta-diversity increased, falling to 0.64 ± 0.04 (ave. ± s.d.) when samples shared 80 % of their OTUs (Figure S4). In contrast, the sensitivities of MW-HR-SIP and Heavy-SIP were least affected by changes in beta-diversity and these methods had the highest sensitivity overall (0.81 ± 0.04 and 0.82 ± 0.03 at 80 % shared OTUs, respectively; ave. ± s.d.). Increasing the beta-diversity of samples had little effect on the specificity of q-SIP but diminished slightly the specificity of HR-SIP and MW-HR-SIP (Figure 4). Despite these declines, HR-SIP and MW-HR-SIP maintained specificity that was greater than or equal q-SIP and Heavy-SIP throughout most parameter space (Figure 4).

Of the methods evaluated, MW-HR-SIP had the highest balanced accuracy across the widest range of parameters tested (Figure 4). Regardless, the balanced accuracy for MW-HR-SIP was negatively affected by an increase in beta-diversity, falling from 0.98 ± 0.02 to 0.86 ± 0.02 (ave. ± s.d.) when the Bray-Curtis dissimilarity between samples was increased beyond 0.5 (Figure 4). These results highlight the overall negative impact that sample-to-sample variation has on HTS-DNA-SIP accuracy, and the importance of minimizing

experimental variation between unlabeled controls and labeled treatments in SIP experiments.

*Using HTS-DNA-SIP data to quantify atom % excess*

So far we have focused on the accuracy of HTS-DNA-SIP methods with respect to the identification of taxa that incorporate isotope into their DNA. However, changes in DNA BD can also be used to quantify the isotope enrichment of DNA from particular taxa. Two approaches have been used to evaluate isotope enrichment from HTS-SIP data: q-SIP and ΔBD, with the latter being a complementary analysis to HR-SIP (Pepe-Ranney *et al.*, 2016a). Both ΔBD and q-SIP derive quantitative estimates from measuring taxon BD shifts (and thus atom % excess) in the labeled treatment gradient(s) versus their unlabeled counterparts. The ΔBD method attempts to measure the extent of the BD shift directly from the compositional sequence data, while q-SIP utilizes relative abundances transformed by qPCR counts of total SSU rRNA copies. Therefore, ΔBD accuracy likely suffers from compositional effects inherent to HTS datasets, while q-SIP accuracy is dependent on qPCR accuracy and variation.

We assessed the quantification accuracy of both methods using the simulations described previously, where either the amount of isotope incorporation or sample beta-diversity was varied. We found that ΔBD produced estimates of isotope incorporation that were closer on average to the true value compared to q-SIP, but ΔBD values had much higher variance than q-SIP estimates (Figure 5). Furthermore, the variance in ΔBD atom % excess $^{13}C$ estimates increased substantially with even moderate increases in beta-diversity between

274 samples, while the q-SIP estimations were largely invariant across the simulation
275 parameter space (Figure S5A). However, mean q-SIP values consistently underestimated
276 the true $^{13}C$ atom % excess by 30.2-39.2% (Figure 5B). Overall, quantitative estimates of
277 isotope incorporation for individual taxa were less variable with q-SIP, though q-SIP
278 consistently miss-estimated actual levels of isotope enrichment.
279
280 **Discussion**
281 Our simulation framework (SIPSim) provides a tractable platform for evaluating the
282 accuracy of DNA-SIP methods and for developing new methods to analyze DNA-SIP data.
283 Given the laborious nature of DNA-SIP experiments, it is impossible to use empirical
284 analyses with mock communities to evaluate the range of parameter values that can be
285 investigated through simulation. In addition, both the physics of density gradient
286 centrifugation and the physical properties of genomic DNA are well established, making the
287 simulation of DNA-SIP data both tractable and reliable. Without rigorous assessment of
288 DNA-SIP methods, it is difficult to determine the likelihood of false negatives (Type II error)
289 and false positives (Type I error) across the wide range of experimental conditions in
290 which DNA-SIP has been employed in the literature. Issues of Type I and Type II statistical
291 error are compounded in the analysis of HTS-DNA-SIP by the nature of HTS data, where it
292 is necessary to make many thousands of comparisons to identify those OTUs that change in
293 response to treatment. This multiple comparison problem has major implications for
294 statistical power and the likelihood of false detection (Paulson *et al.*, 2013). We have used
295 SIPSim to test the effects of multiple parameters on the accuracy of current methods for
296 analyzing HTS-DNA-SIP data. Furthermore we used observations from the model to

297 develop MW-HR-SIP, an analytical approach with balanced accuracy higher than any other
298 current method, with a higher sensitivity than HR-SIP, a higher specificity than q-SIP and
299 Heavy-SIP, and a higher robustness to inter-sample beta-diversity than all other currently
300 available methods.
301
302 Although both HR-SIP and q-SIP use high throughput SSU rRNA amplicon sequencing of
303 many gradient fractions, their different approaches for detecting isotope incorporators
304 result in substantial differences in sensitivity and specificity. In q-SIP, taxon relative
305 abundance is transformed using qPCR data to estimate counts of SSU rRNA gene copies
306 across gradient fractions; however, this approach resulted in a large number of false
307 positives (8 ± 0.3 to 15 ± 0.7 % of the unlabeled taxa evaluated were misidentified as
308 labeled; Figure 3 and Figure 4). Moreover, the number of false positives detected by q-SIP
309 increased dramatically in response to variation in community structure between samples
310 (Figure 4). In contrast, HR-SIP had negligible false positives under a wide range of
311 parameters (Figure 3 and Figure 4), but had lower sensitivity (more false negatives) than
312 other methods. However, we found that the sensitivity of HR-SIP was improved without
313 compromising specificity by using a multi-window analysis (MW-HR-SIP) in place of the
314 single window analysis used in HR-SIP. MW-HR-SIP has high sensitivity and specificity
315 across a range of experimental parameters, provided that the atom % excess $^{13}C$ of DNA is
316 in the range of 50-100% (Figure 3). At lower levels of isotope incorporation, q-SIP has
317 better sensitivity, but this sensitivity comes at the cost of detecting a large number of false
318 positives (*i.e.* low specificity). This tradeoff between specificity and sensitivity can be
319 contextualized by considering a community that contains 1045 unlabeled taxa and 55

labeled taxa. If these 55 taxa are labeled at 50% atom excess $^{13}C$, both methods do a good job detecting truly labeled taxa (MW-HR-SIP: 51 ± 2; q-SIP: 50 ± 2), but q-SIP detects far more false positives (MW-HR-SIP: 0 ± 1; q-SIP: 126 ± 8). If these 55 taxa are instead labeled at 25% atom excess $^{13}C$ then MW-HR-SIP detects fewer labeled taxa than q-SIP (MW-HR-SIP: 33 ± 3; q-SIP: 50 ± 2), but q-SIP still detects far more falsely labeled taxa (MW-HR-SIP: 1 ± 0; q-SIP: 122 ± 8). In these examples, about 71% of the taxa identified by q-SIP as labeled are actually unlabeled.

When considering the relative importance of sensitivity versus specificity for DNA-SIP experiments, the ability to detect taxa that incorporate isotope is only useful if those identifications can be made with high confidence (*i.e.* with a low number of false positives). Therefore, based on our results, MW-HR-SIP is the most robust method for identifying isotope incorporators from HTS-DNA-SIP data. In addition to its high specificity and better ability to handle variance between replicate samples, MW-HR-SIP has the added advantage of not requiring qPCR to be performed on each gradient fraction. It should be noted, however, that the primary objective for which MW-HR-SIP was designed is the accurate detection of labeled taxa, regardless of level of isotopic enrichment, while a major goal of q-SIP is to quantify the atom % excess of individual taxa.

In regards to methods used to quantify the atom % excess of individual taxa from HTS-DNA-SIP data, we found that the utility of q-SIP or ΔBD varied depending on the hypothesis being evaluated. ΔBD produced more accurate estimates of mean $^{13}C$ atom % excess than q-SIP (Figure 5 and Figure S5), and so this approach may be suitable when seeking to make

343 relative comparisons in the degree of labeling between large groups of taxa (as described in
344 Pepe-Ranney, *et al.*, 2016). However, the high variability of this approach causes ΔBD to be
345 unreliable in determining differences in atom % excess $^{13}C$ at the scale of individual OTUs.
346 Alternatively, q-SIP produced much more stable estimates of atom % excess $^{13}C$ among
347 individual taxa, but the method resulted in systematic underestimates of isotope
348 incorporation.
349
350 The SIPSim framework makes it possible to both evaluate hypothetical outcomes of DNA-
351 SIP experiments before they are performed and to evaluate the accuracy of HTS-DNA-SIP
352 data analysis methods. For brevity, we have only focused on a few key variables that could
353 affect the accuracy of HTS-DNA-SIP methods. However, SIPSim can also be used to assess
354 the accuracy of DNA-SIP methods across a range of possible real-world scenarios. For
355 instance, spatial or population-level heterogeneity could result in taxa that are not
356 homogeneously labeled (Lennon and Jones, 2011). Such systematic heterogeneity in
357 labeling would manifest as "split" (bimodal or multimodal) distributions of DNA fragments
358 in an isopycnic gradient. It would be challenging to evaluate such scenarios empirically, but
359 SIPSim can be readily used to evaluate a range of such scenarios. SIPSim can also be used to
360 evaluate the effect of sequencing depth on the statistical power needed to resolve isotope
361 incorporation in rare taxa. Such information should be useful in planning HTS-DNA-SIP
362 experiments, to ensure that the experiment has a reasonable chance of success before it is
363 performed. Finally, SIPSim provides a toolkit for developing and improving analytical
364 methods used in DNA-SIP experiments. For example, a hybrid method that combines

365 aspects of MW-HR-SIP and q-SIP may be able to produce robust incorporator identification
366 while also providing accurate estimates of the atom % excess of individual taxa.
367
368 **Conclusion**
369 With our newly developed simulation toolset, we determined that MW-HR-SIP has the
370 highest accuracy of currently available methods for identifying taxa that have incorporated
371 isotope in HTS-DNA-SIP experiments. The use of MW-HR-SIP resulted in a negligible
372 number of false positives and its ability to detect true positives varied in relation to the
373 isotopic enrichment of DNA. Generally, we found that the specificity of all HTS-DNA-SIP
374 methods declined with increased beta-diversity among replicate samples. Thus, given that
375 accuracy declined most rapidly between a mean Bray-Curtis distance of 0 and 0.2 for all
376 methods evaluated (Figure 4), we recommend that researchers strive for mean Bray-Curtis
377 distances of <0.2 among replicate samples used in SIP experiments (*i.e.* between
378 treatments and their corresponding controls).
379
380 **Experimental Procedures**
381 *Theory underlying the simulation framework*
382 DNA stable isotope probing employs isopycnic centrifugation to separate isotopically
383 enriched ("heavy") DNA molecules from unlabeled ("light") DNA based on their differences
384 in buoyant density (BD). Isopycnic centrifugation is distinguished from other
385 centrifugation methods in that centrifugation is carried out long enough to both generate a
386 density gradient (typically using CsCl for DNA-SIP) and have all macromolecules of interest
387 reach sedimentation equilibrium, which is the point at which sedimentation rates equal

388 rates of diffusion (Hearst and Schmid, 1973; Birnie and Rickwood, 1978). Empirical studies
389 have shown that the average BD (ρ) of a mixture of DNA molecules is linearly related to the
390 average G + C content for that collection of molecules:
391
392 $$\rho = 0.098[G + C] + 1.66 \quad (1)$$
393
394 where [G + C] is the mole fraction of G+C content (Schildkraut *et al.*, 1962; Birnie and
395 Rickwood, 1978). In addition, empirical studies have also shown that homogeneous
396 mixtures of DNA molecules form a Gaussian distribution in an isopycnic gradient when at
397 sedimentation equilibrium (Meselson *et al.*, 1957; Fritsch, 1975). Therefore, in order to
398 model the BD distribution of a heterogeneous set of genomic DNA fragments, a Gaussian
399 distribution must be estimated for each homogeneous subset of molecules rather than
400 using discrete BD values (as described in *Supporting Information*). Based on the work of
401 Meselson and colleagues (Meselson *et al.*, 1957), Fritsch (1975) derived an equation
402 describing time to reach sedimentation equilibrium, which can be reworked to calculate
403 the standard deviation (σ) of the Gaussian distribution:
404
405 $$\sigma = \frac{L}{e^{(\gamma - 1.26)}} \quad (2)$$
406
407 $$\gamma = \frac{t\omega^4 r_p^2 s}{\beta^{\circ}(p_p - p_m)} \quad (2.1)$$
408

where $L$ is the effective length of the gradient (cm), $t$ is time in seconds, $\omega$ is the angular velocity (radians $sec^{-1}$), $r_p$ is the distance of the particle from the axis of rotation (cm), $s$ is the sedimentation coefficient of the particle, $\beta°$ is the coefficient specific to the density gradient medium (*e.g.* CsCl); $p_p$ and $p_m$ are the maximum and minimum distances between the gradient and axis of rotation (cm) (Fritsch, 1975). By assuming that sedimentation equilibrium has been reached for all macromolecules of interest, Clay and colleagues derived a simplified equation for determining σ from the calculations in (Schmid and Hearst, 1972):

$$\sigma = \sqrt{\frac{\rho RT}{\beta^2 G M_C l}} \qquad (3)$$

where $\rho$ is the BD of the particle, $R$ is the universal gas constant, $T$ is the temperature in Kelvins, $\beta$ is a proportionality constant for aqueous salts of specific densities, $G$ is a buoyancy factor as described in (Clay *et al.*, 2003), $M_C$ is the molecular weight per base pair of DNA, and $l$ is the fragment length (bp). For most DNA-SIP experiments, the assumption of sedimentation equilibrium for all DNA fragments is likely to be unrealistic for relatively short DNA fragments (*e.g.* < 4 kb), given that the time to equilibrium rises dramatically with decreasing fragment length (Meselson *et al.*, 1957; Birnie and Rickwood, 1978; Youngblut and Buckley, 2014). However, the ultracentrifugation durations used in typical DNA-SIP experiments should still generally produce small σ values for short DNA fragments according to Eq. 2 (Neufeld *et al.*, 2007). Therefore, equation Eq. 3 provides a

good approximation for modeling the BD distribution of DNA in density gradients generated in typical DNA-SIP experiments.

The distribution of a heterogeneous mixture of DNA fragments in an isopycnic gradient can thus be modeled by integrating the Gaussian distributions of each homogeneous subset of DNA fragments, where the mean of each Gaussian is determined by Eq. 1 and the standard deviation derived from Eq. 3. In this way, the BD distribution for a given genome in an isopycnic gradient can be modeled by the following steps: simulate genome fragmentation resulting from DNA extraction, bin gDNA fragments with respect to length and G + C content, model Gaussian distribution for each fragment bin, and then integrate these distributions to describe the cumulative DNA distribution in the gradient.

We found that empirical DNA fragment distributions differed from the expectations of a strictly Gaussian model (Figure S2), and we determined that these differences could be reconciled on the basis of established principles of fluid mechanics (as described below and in *Supporting Information*). Based on empirical measurements, we found that most taxa with relative abundances > 0.1% are detected in all gradient fractions when unlabeled DNA is subjected to CsCl gradient centrifugation and SSU rRNA amplicon sequencing is performed across a wide range of density gradient fractions (Figure S2). This observation is in general congruence with observations in the literature (Birnie and Rickwood, 1978; Lueders *et al.*, 2004; Leigh *et al.*, 2007), but it does not match the expectation that DNA fragment distributions are strictly Gaussian, since the Gaussian model predicts that DNA fragments should be undetectable (i.e. probability density < $1e^{-7}$) at either end of the

density gradient (Figure S6). We explain the difference between observed and expected DNA distributions as a function of fluid mechanics during gradient reorientation.

During isopycnic centrifugation, the buoyant density gradient forms perpendicular to the axis of rotation (Figure S7), and gradient reorientation during centrifuge deceleration is dramatic, especially for vertical rotors (Flamm *et al.*, 1966). While the distortion of the BD gradient during reorientation has been shown to be minimal in the aggregate (Fisher *et al.*, 1964; Flamm *et al.*, 1966), the inevitable presence of a diffusive boundary layer along the tube wall is sufficient to entrain quantities of DNA, which are small but should be readily detectable by high throughput sequencing methods. The flow field that occurs during gradient reorientation entrains along the tube wall a volume with a dimension proportional to flow velocity, fluid viscosity, and surface topography (Tritton, 1977; Cohen and Dowling, 2012). Following gradient reorientation, DNA from the entrained volume will combine with DNA from the reoriented volume, thereby introducing a small amount of non-BD-equilibrium DNA into each gradient fraction (Figure S7). The ability of the diffusive boundary to introduce non-BD-equilibrium DNA into gradient fractions can be modeled as a function of rotor geometry (Figure S7). Assuming sedimentation equilibrium, BD (ρ) can be directly related to the distance from axis of rotation (Birnie and Rickwood, 1978):

$$x = \sqrt{(p - p_m)\frac{2\beta^{\circ}}{\omega^2} + r_c^2} \qquad (4)$$

From this calculation, the location of DNA molecules in the centrifuge tube, both during centrifugation and fractionation, can be ascertained by using simple trigonometry along with knowledge of centrifuge tube dimensions and angle to the axis of rotation. A full description of the calculations along with an example can be found at https://github.com/nick-youngblut/SIPSim. The fraction of a taxon's DNA fragments that are in the boundary layer ($D_{ti}$) is modeled as:

$$D_{ti} = A_{ti}\gamma + \alpha \quad (5)$$

where $A_{ti}$ is the pre-fractionation community relative abundance of taxon $t$ in gradient $i$, $\gamma$ is a weight parameter determining the contribution of $A_{ti}$ to $A_b$, and $\alpha$ is the baseline fraction DNA in $A_b$.

Assimilation of the commonly used isotopes $^{13}C$ and $^{15}N$ into genomic DNA produces linear shifts in BD, with a maximum shift of 0.036 and 0.016 g $ml^{-1}$, respectively (Birnie and Rickwood, 1978). Thus the shift in BD (ρ) can be modeled as:

$$\rho_{13C} = I_{i,max}A_i + \rho_{12C} \quad (6)$$

where $I_{i,max}$ is the maximum possible BD shift if 100% atom excess for isotope $i$, $A$ is the atom % excess of isotope $i$, and $\rho_{12C}$ is the buoyant density at 0% atom excess.

*SIP data simulation framework overview*

Based on the theory described above, our SIP data simulation framework simulates the distribution of gDNA fragments in isopycnic gradients at sedimentation equilibrium. Furthermore, it generates the HTS-DNA-SIP datasets obtained from fractionating isopycnic gradient(s) and performing high throughput sequencing on many of the gradient fractions. Our framework also implements all of the HTS-DNA-SIP analysis methods assessed in this study (Heavy-SIP, HR-SIP, MW-HR-SIP, q-SIP, and ΔBD) and evaluates their accuracy of identifying incorporators or quantifying BD shifts. An overview of our simulation framework is shown in Figure 1.

Our simulation framework is a modular collection of steps that can be grouped in workflow stages that are further broken down into steps (Figure 1). The input is a set of reference genomes in fasta format and a text file designating the experimental design, which includes the number of gradients for labeled treatments and unlabeled controls.

Stage 1 involves generating a BD distribution of gDNA fragments for each genome. Step 1a involves simulating the pool of gDNA fragments that is extracted from SIP incubation samples and then loaded into the isopycnic gradients. If amplicon sequence data (*e.g.* SSU rRNA) is to be generated, amplicons from only the fragments containing the PCR template ("amplicon-fragments") are sequenced, while shotgun metagenomic sequencing can target all gDNA fragments ("shotgun-fragments"). If ≥ 1 PCR primer set is provided, amplicon-fragments are generated from genomic regions fully encompassing genome locations that produced amplicons by *in silico* PCR. Alternatively, shotgun-fragments are randomly

519 generated from all possible genomic locations. The fragment size distribution is user-
520 defined (Table S2).
521
522 As described in Eq. 1 & 3, the length and G + C content of a DNA fragment can be used to
523 calculate a probability distribution of its location in the gradient, assuming sedimentation
524 equilibrium. Step 1b uses the fragments simulated in Step 1a to generate a 2-dimensional
525 Gaussian kernel density estimation (KDE) for each taxon, which describes the joint
526 probability of obtaining fragments with a certain length and G + C content from that taxon.
527 From this 2D-KDE, a large number of [length, G + C] vectors can be simulated efficiently for
528 more precise estimations of the fragment BD distributions. Fragment BD distributions are
529 calculated for each taxon in Step 1c by sampling [length, G + C] vectors from the 2D-KDE
530 and calculating Gaussian distribution from each, where the mean is based on Eq. 1 and the
531 standard deviation based on Eq. 3. The collection of Gaussian distributions for all fragments
532 for each taxon is integrated into a BD distribution for all fragments of a taxon with Monte
533 Carlo error estimation, which involves sampling BD values from the collection of Gaussian
534 distributions and estimating a probability density function (PDF) of the fragment BD
535 distribution as a one-dimensional Gaussian KDE. The result is a list of KDEs, with each
536 describing the probability of detecting the gDNA fragments of a taxon at any point along the
537 isopycnic gradient. These fragment BD distributions are modified in steps 1d and 1e by
538 adding diffusive boundary layer (DBL) effects (see Theory) and isotope incorporation,
539 respectively. The “smearing” due to DBL effects is modeled as a uniform distribution
540 describing the increased fragment BD uncertainty, and this uncertainty is integrated into
541 the fragment BD distributions by Monte Carlo error estimation as in Step 3b. The BD shift

542 due to isotope incorporation is modeled in a similar manner, except BD uncertainty is a
543 result of inter- and intra-population variation in the amount of isotope incorporated.
544 Variation of isotope incorporation is modeled as a hierarchical set of mixture models
545 (weighted sets of standard distributions; such as two Gaussians), where the parameters for
546 intra-population mixture models that describe the amount of isotope incorporated by each
547 individual are themselves defined by inter-population mixture models that describe how
548 isotope incorporation varies among taxa.
549
550 Stage 2 involves simulating the isopycnic gradients for a particular experimental design.
551 Step 2a involves simulating the BD range size of each fraction of each gradient. Sizes are
552 drawn from a user-defined distribution. Step 2b involves simulating the relative abundance
553 distribution of taxa in the gDNA pools loaded into each gradient ("pre-fractionation
554 communities"). The abundance distribution of each pre-fractionation community is user-
555 defined and can vary among gradients. Furthermore, the amount of taxa shared or rank-
556 abundances permutated among communities (*i.e.* the beta-diversity) is user-defined.
557
558 Stage 3 involves generating a HTS-DNA-SIP dataset based on the fragment BD distributions
559 simulated in Stage 1 along with the isopycnic gradient data generated in Stage 2. In Step 3a,
560 an OTU (taxon) abundance table is generated by sampling from the fragment BD
561 distributions of each taxon generated in Stage 1, with sampling depth determined by pre-
562 fractionation community abundances simulated in Step 2b. The subsampled fragments are
563 then binned into gradient fractions simulated in Step 2a. The resulting OTU table lists the
564 number of gDNA fragments of each taxon in each gradient fraction in each gradient. If the

565 simulated fragments are amplicons, then PCR amplification efficiency biases are simulated
566 in Step 3b based on the PCR kinetic model described in Suzuki and Giovannoni (1996). The
567 model assumes that efficiencies decrease as the product concentration increases due to an
568 increased propensity of single stranded products to re-anneal to their homologous
569 complements. Sequence data is simulated in Step 3c by subsampling from the table of
570 fragment counts (the DNA fragment pool), which produces a final table ("HTS-DNA-SIP
571 dataset") of taxon relative abundances in each gradient fraction in each gradient.
572
573 *SIP data simulation framework parameters*
574 Unless stated otherwise, we made the following assumptions for all simulations in this
575 study. Community abundance distributions were simulated as lognormal distributions with
576 a mean of 10 and a standard deviation of 2. All taxa were shared among communities, and
577 no rank-abundances were permuted (unless otherwise stated as for when evaluating beta-
578 diversity effects). The total number of fragments in each gradient was $1e^9$. Gradient
579 fragment BD range sizes were sampled from a normal distribution, with a mean of 0.004
580 and a standard deviation of 0.0015. SSU rRNA amplicon-fragments were simulated using
581 the V4-targeting 16S rRNA primers: 515F and 927R (5'-GTGYCAGCMGCMGCGGTRA-3'; 5'-
582 CCGYC AATTYMTTTRAGTTT-3'), as used by Pepe-Ranney and colleagues (Pepe-Ranney, *et*
583 *al.*, 2016a). The amplicon-fragment size distribution was a left-skewed normal distribution
584 with a mean of ~12 kb, which is similar to size distributions produced from common bead
585 beating cell lysis methods (Kauffmann *et al.*, 2004; Roh *et al.*, 2006; Thakuria *et al.*, 2008). A
586 total of $1e^4$ amplicon-fragments were simulated per genome, which equated to > 100X
587 coverage for the genomic region of interest. Monte Carlo error estimation was conducted

588 with $1e^5$ sampling replicates. Ultracentrifugation conditions were set as in Pepe-Ranney
589 and colleagues (Pepe-Ranney *et al.*, 2016a), with a Beckman TLA-110 rotor spun at $5.5e^4$
590 rpm for 66 hours at 20°C and an average density gradient 1.7 g $ml^{-1}$. Inter-population
591 variation in isotope incorporation was binary (either 0 % or X % atom excess), and intra-
592 population variation was set to zero. Two key parameters were estimated from empirical
593 HTS-DNA-SIP data: the bandwidth (smoothing factor) for kernel density estimation, and
594 the gamma parameter in Eq. 5. See Table S2 for a full listing of simulation parameters.
595
596 *Implementing HTS-DNA-SIP analyses*
597 The HR-SIP method was performed as described in (Pepe-Ranney *et al.*, 2016a; Pepe-
598 Ranney *et al.*, 2016b). Briefly, we used a "heavy" BD window of 1.71-1.75 g $ml^{-1}$, a sparsity
599 cutoff of 0.25 (*i.e.* OTUs must be present in >25% of samples), a $\log_2$ fold change null
600 threshold of 0.25, and a false discovery rate cutoff of 10 %. ΔBD was determined as
601 described by Pepe-Ranney and colleagues (Pepe-Ranney *et al.*, 2016a), with OTU
602 abundances linearly interpolated across 20 evenly spaced values across the gradient BD
603 range.
604
605 We hypothesized that HR-SIP sensitivity could be improved by altering the "heavy" BD
606 window (1.71-1.75 g $ml^{-1}$) in which sequence composition is compared between treatment
607 and control. We evaluated different approaches and found that the analysis of multiple
608 windows (hereby called "MW-HR-SIP") resulted in a significant improvement in sensitivity
609 relative to HR-SIP. MW-HR-SIP evaluated sequence composition within BD windows of:
610 1.70-1.73, 1.72-1.75, 1.74-1.77 g $ml^{-1}$ (Figure S8) while adjusting for multiple comparisons.

611

612 q-SIP was conducted as described in Hungate and colleagues (Hungate *et al.*, 2015), with 90
613 % confidence intervals calculated from 1000 bootstrap replicates. The variance among
614 qPCR replicates was modeled based on the qPCR data provided in Table S2 of Hungate et
615 al., (2015). Specifically, we found the qPCR count variance ($\sigma^2$) to increase as a function of
616 the mean ($\mu$). The following polynomial regression was found to best describe this
617 relationship and was used for simulating all qPCR count values:

618

619 $$\sigma^2 = 5889 + \mu + 0.714\mu^2 \quad (7)$$

620

621 where $\mu$ was set as the total number of simulated DNA fragments in the gradient fraction
622 (designated in the OTU table from Step 4a).

623

624 *Datasets*

625 The genome dataset used to simulate genomic DNA fragments was obtained from Genbank
626 (Benson *et al.*, 2008). From a list of all bacterial genomes designated as "complete", one
627 representative was chosen per species in order to reduce the bias toward highly
628 represented species. We found the dataset to contain a rather high proportion (~12 %) of
629 low G + C organisms (< 30 % G + C); most of which were obligate endosymbionts. We
630 randomly sampled a subset of these low G + C genomes in order to reduce the proportion
631 of low G + C organisms to just 1 % of the genome dataset. The resulting dataset consisted of
632 1147 bacterial genomes.

633

In order to simulate empirical data from Lueders and colleagues (Lueders *et al.*, 2004), the genome sequences of *Methanosarcina barkeri* MS and *Methylobacterium extorquens* AM1 were downloaded from Genbank. Amplicon-fragments were simulated with the primers Ar109f (5'-ACKGCTCAGTAACACGT-3'), Ar915r (5'-GTGCTCCCCCGCCAATTCCT-3'), Ba519f (5'-CAGCMGCCGCGGTAANWC-3'), and Ba907r (5'-CCGTCAATTCMTTTRAGTT-3'). Atom % excess was assumed to be 100 %, and isopycnic centrifugation conditions were simulated as specified in Lueders et al., (2004).

For model evaluation (see *Supporting Information* - *Results*), we downloaded the genomes *Clostridium ljungdahlii* DSM 13528, *Escherichia coli* 1303, and *Streptomyces pratensis* ATCC 33331 from Genbank.

The HTS-DNA-SIP dataset from Youngblut and colleagues consisted of SSU rRNA MiSeq sequences (V4 region) of ~24 fractions per gradient from 6 gradients of unlabeled controls (Youngblut et al., in prep.). These data were subsampled to obtain a total richness equal to the 1147 OTUs in our reference genome dataset. The sequence data is available from the NCBI under BioProject PRJNA382302.

*Software implementation*

The SIP simulation framework was mostly written in Python v2.7.11, with some accompanying code written in C++ v4.9.2 and R v3.2.3 (R Core Team, 2016). MFEprimer v2.0 was used to perform *in silico* PCR (Qu *et al.*, 2009). The software, along with documentation and examples, can be found at https://github.com/nick-youngblut/SIPSim.

All genomes were downloaded from Genbank with the R package *genomes* v2.12.0 (Stubben, 2014), and all data analysis was conducted in R with the following packages: ggplot2 v2.1.0, dplyr v0.4.3, tidyr v 0.4.1, and cowplot v0.6.2.

Further methodological details are provided in the *Supporting Information*.

**Acknowledgements**

We thank Samuel Barnett for consultation on modeling the diffusive boundary layer effects and Chuck Pepe-Ranney for helpful discussions on the modeling approach used in this work. This material is based upon work supported by the Department of Energy, Office of Biological & Environmental Research Genomic Science Program under Award Numbers DE-SC0010558 and DE-SC0004486.

**References**

Andeer, P., Strand, S.E., and Stahl, D.A. (2012) High-sensitivity stable-isotope probing by a quantitative terminal restriction fragment length polymorphism protocol. *Appl. Environ. Microbiol.* **78**: 163–169.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2008) GenBank. *Nucleic Acids Res.* **36**: D25–D30.

Birnie, G.D. and Rickwood, D. (1978) Centrifugal separations in molecular and cell biology Butterworths, Boston, MA.

Buckley, D.H., Huangyutitham, V., Hsu, S.-F., and Nelson, T.A. (2007) Stable isotope probing with $^{15}N$ achieved by disentangling the effects of genome G+C content and isotope enrichment on DNA density. *Appl. Environ. Microbiol.* **73**: 3189–3195.

Clay, O., Douady, C.J., Carels, N., Hughes, S., Bucciarelli, G., and Bernardi, G. (2003) Using analytical ultracentrifugation to study compositional variation in vertebrate genomes. *Eur. Biophys. J.* **32**: 418–426.

Cohen, I.M. and Dowling, D.R. eds. (2012) Boundary layers and related topics. In Fluid Mechanics (Fifth Edition). Kundu, Pijush K. (ed). Boston, USA: Academic Press, pp. 361–419.

Cupples, A.M., Shaffer, E.A., Chee-Sanford, J.C., and Sims, G.K. (2007) DNA buoyant density shifts during $^{15}N$-DNA stable isotope probing. *Microbiol. Res.* **162**: 328–334.

Fisher, W.D., Cline, G.B., and Anderson, N.G. (1964) Density gradient centrifugation in angle-head rotors. *Anal. Biochem.* **9**: 477–482.

Flamm, W.G., Bond, H.E., and Burr, H.E. (1966) Density-Gradient centrifugation of DNA in a fixed-angle rotor. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **129**: 310–317.

703 Fritsch, A. (1975) Preparative density gradient centrifugations. Beckman Instrument
704 International, S.A. Geneva.
705
706 Hearst, J.E. and Schmid, C.W. (1973) Density gradient sedimentation equilibrium. In
707 Methods in Enzymology. Hirs, S.N.T. (ed). Boston, USA: Academic Press, pp. 111–127.
708
709 Hungate, B.A., Mau, R.L., Schwartz, E., Caporaso, J.G., Dijkstra, P., van Gestel, N., et al. (2015)
710 Quantitative microbial ecology through stable isotope probing. *Appl. Environ. Microbiol.* **81**:
711 7570–7581.
712
713 Kauffmann, I.M., Schmitt, J., and Schmid, R.D. (2004) DNA isolation from soil samples for
714 cloning in different hosts. *Appl. Microbiol. Biotechnol.* **64**: 665–670.
715
716 Leigh, M.B., Pellizari, V.H., Uhlík, O., Sutka, R., Rodrigues, J., Ostrom, N.E., et al. (2007)
717 Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated
718 with polychlorinated biphenyls (PCBs). *ISME J.* **1**: 134–148.
719
720 Lennon, J.T. and Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary
721 implications of dormancy. *Nat. Rev. Microbiol.* **9**: 119–130.
722
723 Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and
724 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**: 550.
725

726 Lueders, T., Manefield, M., and Friedrich, M.W. (2004) Enhanced sensitivity of DNA- and
727 rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic
728 centrifugation gradients. *Environ. Microbiol.* **6**: 73–78.
729
730 mcmurdie, p.j. and holmes, s. (2014) Waste not, want not: why rarefying microbiome data
731 is inadmissible. *PLoS Comput. Biol.* **10.**:
732
733 Meselson, M., Stahl, F.W., and Vinograd, J. (1957) Equilibrium sedimentation of
734 macromolecules in density gradients. *Proceedings of the National Academy of Sciences* **43**:
735 581–588.
736
737 Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., and
738 Murrell, J.C. (2007) DNA stable-isotope probing. *Nat. Protoc.* **2**: 860–866.
739
740 Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013) Differential abundance analysis for
741 microbial marker-gene surveys. *Nat. Methods* **10**: 1200–1202.
742
743 Pepe-Ranney, C., Campbell, A.N., Koechli, C.N., Berthrong, S., and Buckley, D.H. (2016)
744 Unearthing the ecology of soil microorganisms using a high resolution DNA-SIP approach
745 to explore cellulose and xylose metabolism in soil. *Front. Microbiol.* **7**: 703.
746
747 Pepe-Ranney, C., Koechli, C., Potrafka, R., Andam, C., Eggleston, E., Garcia-Pichel, F., and
748 Buckley, D.H. (2016) Non-cyanobacterial diazotrophs mediate dinitrogen fixation in

749 biological soil crusts during early crust formation. *ISME J.* **10**: 287–298.

750

751 Qu, W., Shen, Z., Zhao, D., Yang, Y., and Zhang, C. (2009) MFEprimer: multiple factor

752 evaluation of the specificity of PCR primers. *Bioinformatics* **25**: 276–278.

753

754 Radajewski, S., McDonald, I.R., and Murrell, J.C. (2003) Stable-isotope probing of nucleic

755 acids: a window to the function of uncultured microorganisms. *Curr. Opin. Biotechnol.* **14**:

756 296–302.

757

758 R Core Team (2016) R: A Language and Environment for Statistical Computing R

759 Foundation for Statistical Computing, Vienna, Austria.

760

761 Roh, C., Villatte, F., Kim, B.-G., and Schmid, R.D. (2006) Comparative study of methods for

762 extraction and purification of environmental DNA from soil and sludge samples. *Appl.*

763 *Biochem. Biotechnol.* **134**: 97–112.

764

765 Schildkraut, C.L., Marmur, J., and Doty, P. (1962) Determination of the base composition of

766 deoxyribonucleic acid from its buoyant density in CsCl. *J. Mol. Biol.* **4**: 430–443.

767

768 Schmid, C.W. and Hearst, J.E. (1972) Sedimentation equilibrium of DNA samples

769 heterogeneous in density. *Biopolymers* **11**: 1913–1918.

770

771 Stubben, C. (2014) genomes: Genome sequencing project metadata. R Package version

772 3.6.0.

773

774 Suzuki, M.T., Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification

775 of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62**: 625-630.

776

777 Thakuria, D., Schmidt, O., Mac Siúrtáin, M., Egan, D., and Doohan, F.M. (2008) Importance of

778 DNA quality in comparative soil microbial community structure analyses. *Soil Biol.*

779 *Biochem.* **40**: 1390–1403.

780

781 Tritton, D.J. (1977) Boundary Layers and Related Topics. In Physical Fluid Dynamics.

782 Netherlands: Springer, pp. 101–118.

783

784 Uhlík, O., Jecná, K., Leigh, M.B., Macková, M., and Macek, T. (2009) DNA-based stable isotope

785 probing: a link between community structure and function. *Sci. Total Environ.* **407**: 3611–

786 3619.

787

788 Wawrik, B., Callaghan, A.V., and Bronk, D.A. (2009) Use of inorganic and organic nitrogen

789 by *Synechococcus* spp. and diatoms on the west florida shelf as measured using stable

790 isotope probing. *Appl. Environ. Microbiol.* **75**: 6662–6670.

791

792 Youngblut, N.D. and Buckley, D.H. (2014) Intra-genomic variation in G + C content and its

793 implications for DNA stable isotope probing. *Environ. Microbiol. Rep.* **6**: 767–775.

794

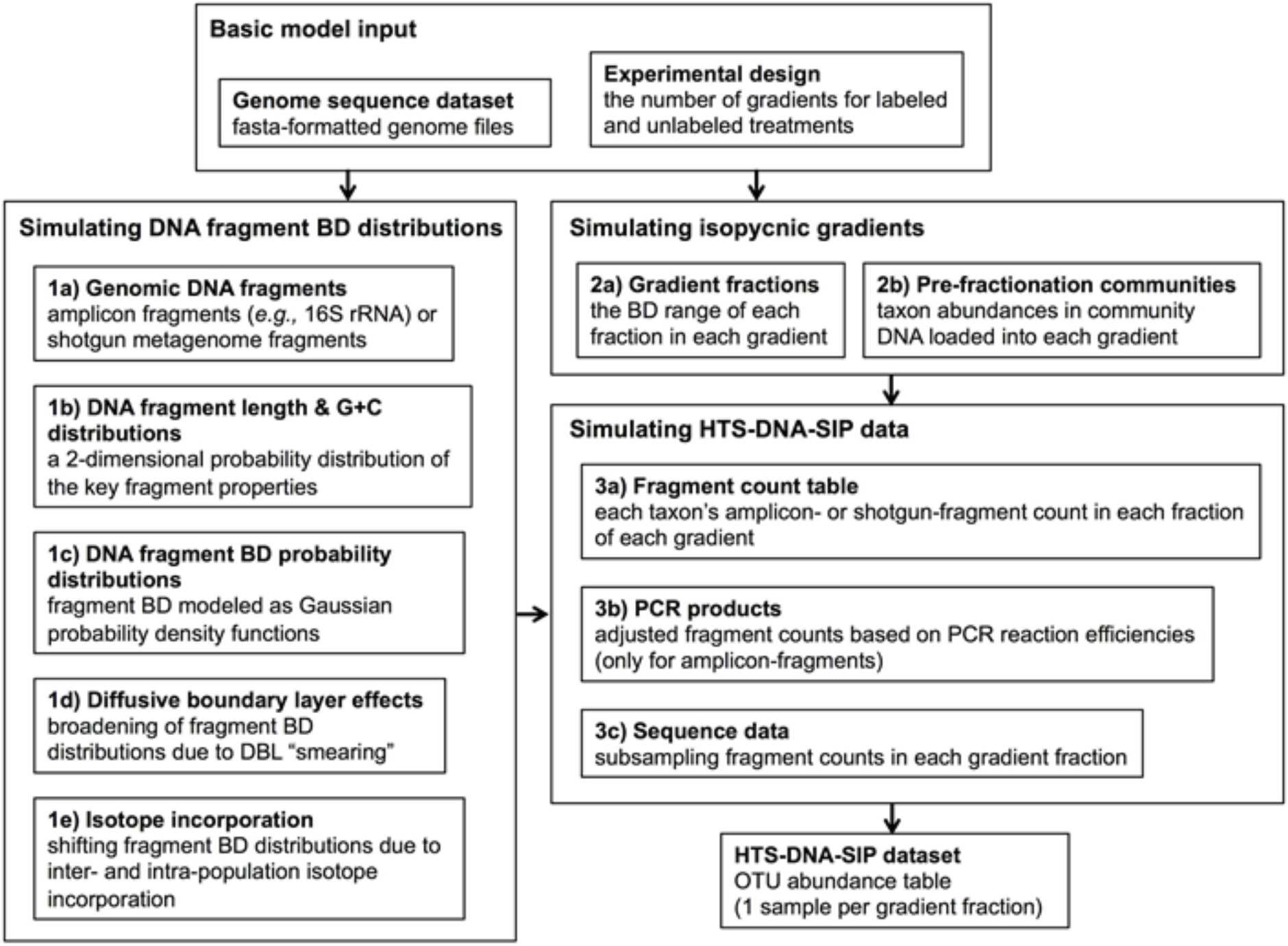
## Figure Legends

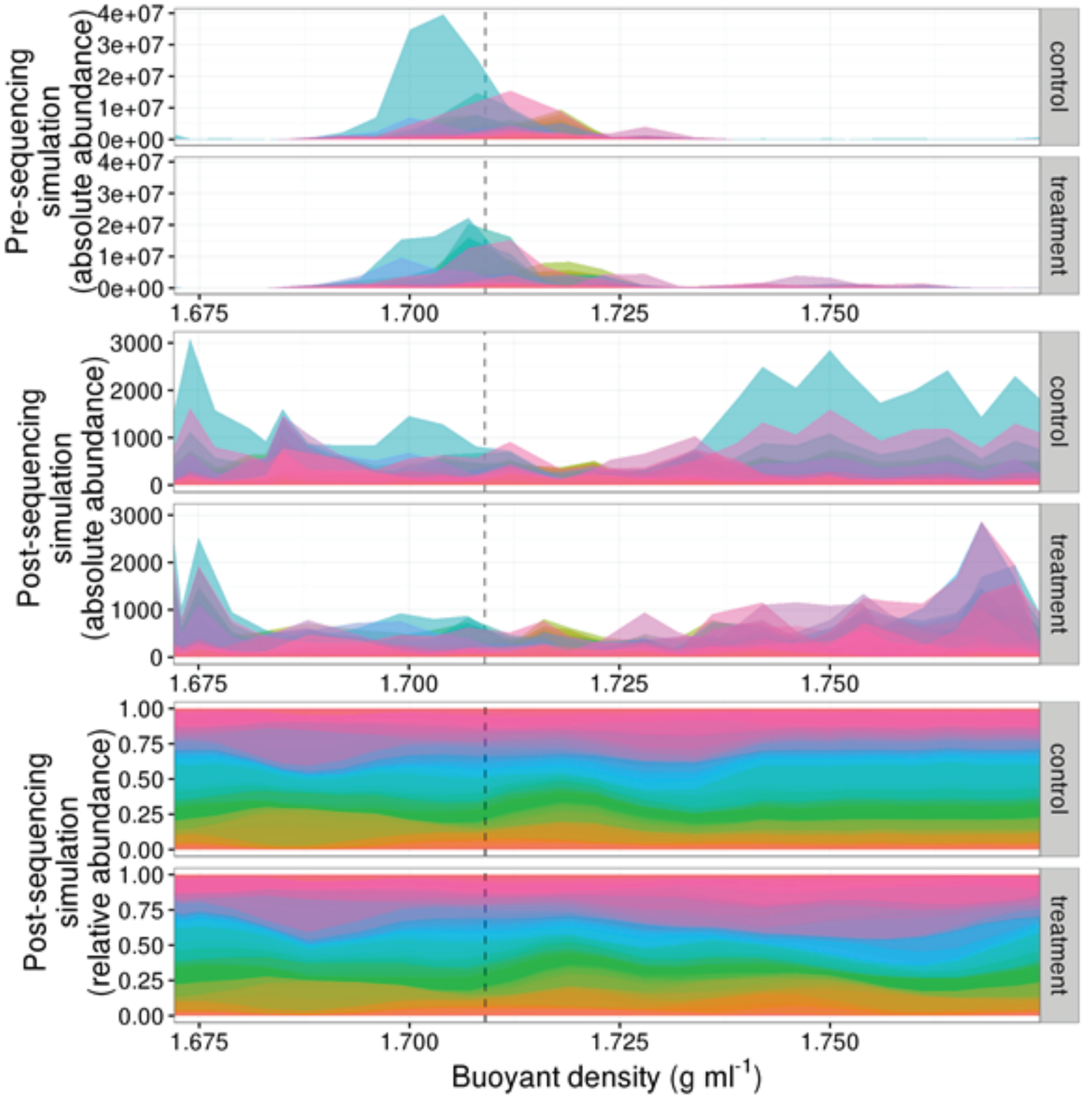
**Figure 1.** The SIPSim simulation workflow involves three major stages, which are broken down into multiple steps. Stage 1 involves generating a buoyant density distribution of gDNA fragments for each genome. Stage 2 involves simulating the isopycnic gradients for a particular experimental design. Stage 3 involves generating a HTS-DNA-SIP dataset based on the fragment BD value distributions simulated in Stage 1 along with the isopycnic gradient data generated in Stage 2. The output is a table ("HTS-DNA-SIP dataset") of taxon relative abundances in each gradient fraction in each gradient. See *Experimental Procedures* for a more detailed description of the simulation workflow.

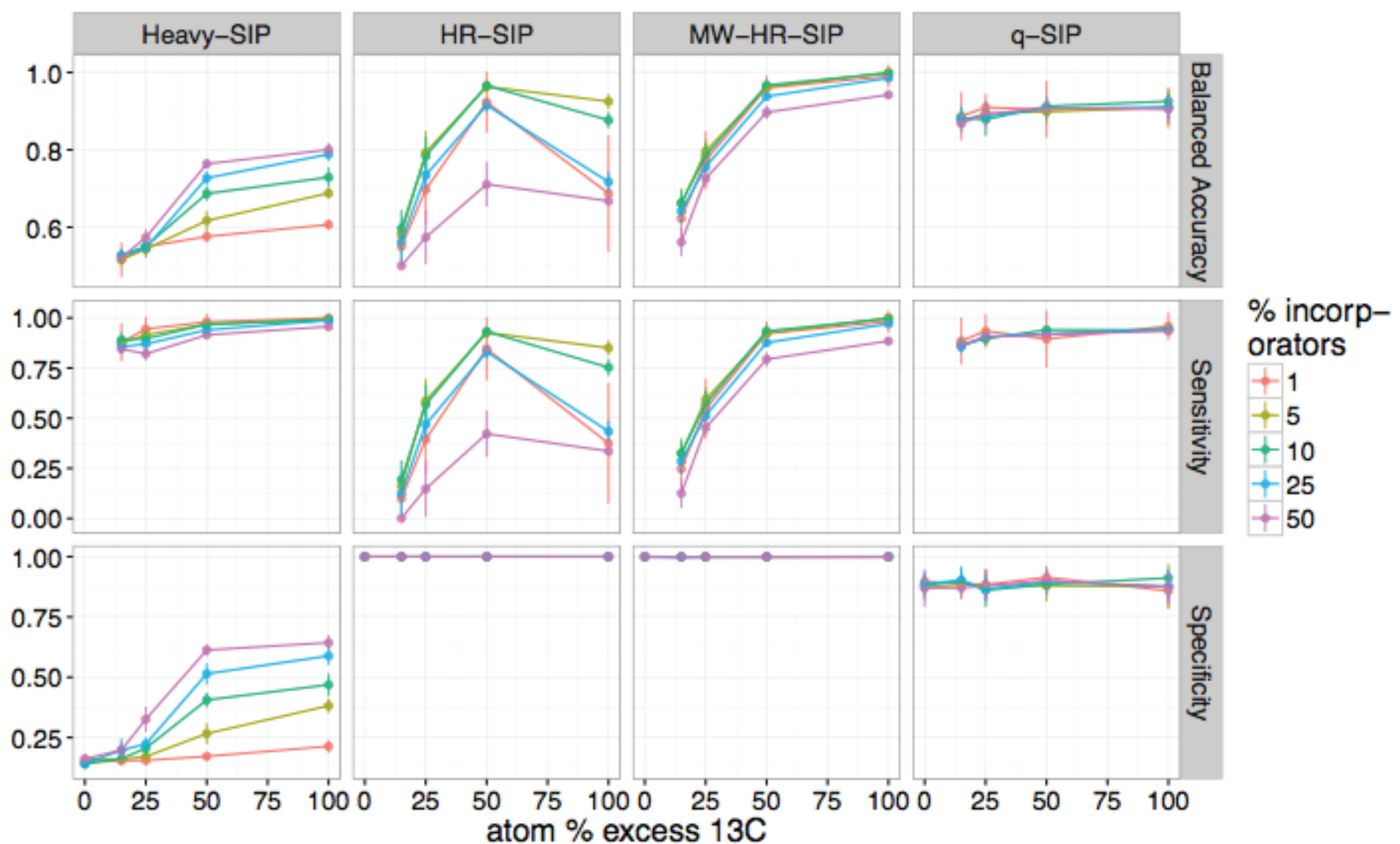
**Figure 2.** *Compositional effects can distort and obscure BD shifts resulting from $^{13}C$ isotope incorporation*. The plots show DNA fragment distributions resulting from simulation of 1147 taxa (one color per taxon) within CsCl gradient pairs consisting of: a $^{12}C$-control ("control") and a $^{13}C$-treatment ("treatment") gradient. For this simulation, all taxa in the control gradient had 0% atom excess $^{13}C$, while 10% of taxa in the treatment gradient were randomly assigned 100% atom excess $^{13}C$. "Pre-sequencing simulation" (top) and "Post-sequencing simulation" (middle, and bottom) show fragment BD distributions before and after simulating the effect of uniform random sampling which results from high throughput sequencing of all gradient fractions at an equal number of sequences. The "absolute abundance" (top and middle) indicates the number of DNA fragments from each taxon in each gradient fraction, while "relative abundance" (bottom) indicates the relative abundance of each taxon. Note that the top plot represents the actual amplicon-fragment distributions in an isopycnic gradient at equilibrium, while the bottom plot represents the

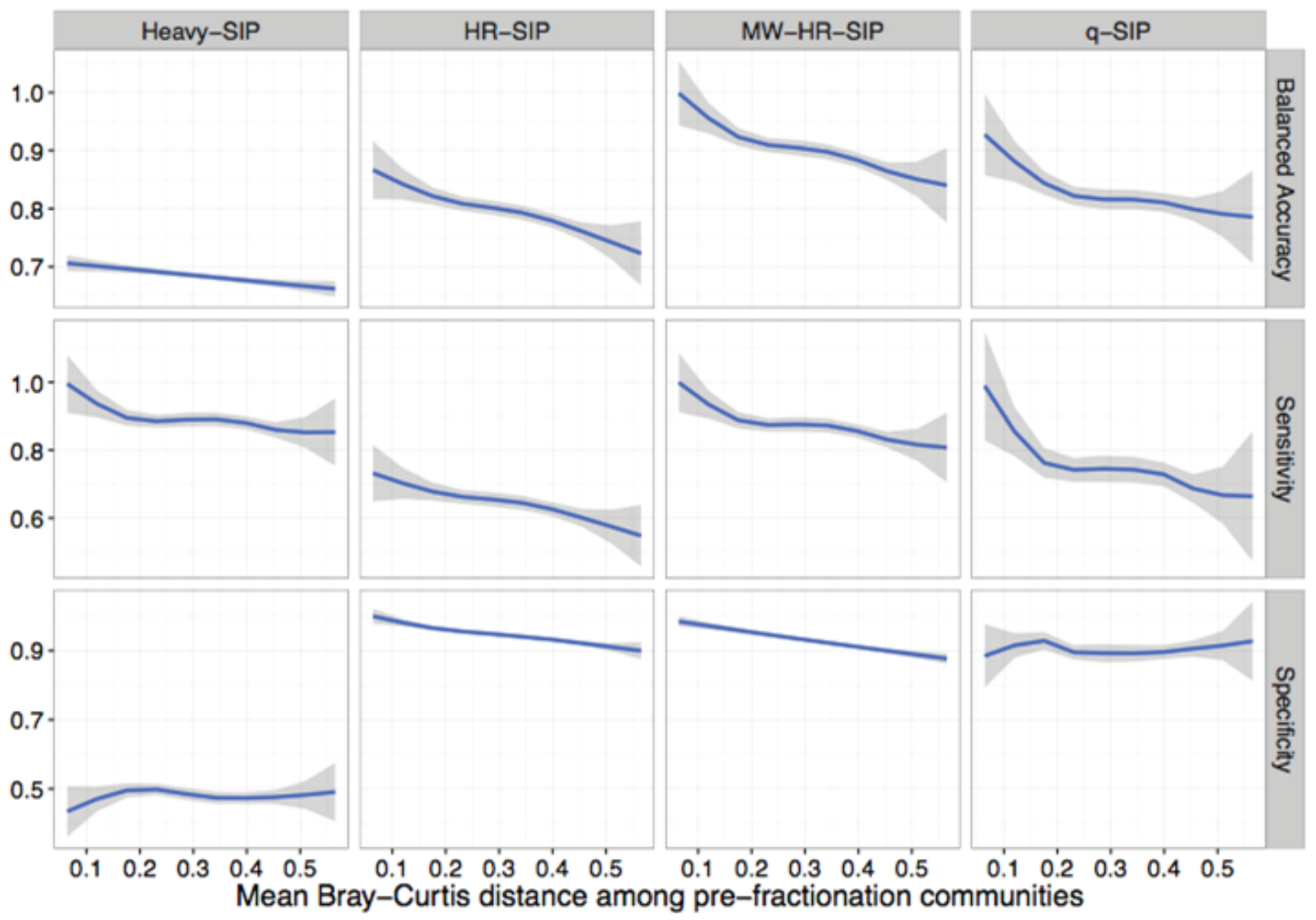
818 sampled fragment distributions obtained after high throughput sequencing. The dashed
819 vertical line is provided as a point of reference and designates the theoretical buoyant
820 density of an unlabeled DNA fragment with 50 % G + C (as modeled in Eq. 1).
821
822 **Figure 3**. HTS-DNA-SIP methods vary in accuracy depending on the $^{13}C$ atom % excess of
823 DNA and the number of taxa that incorporate isotope. Points and bars represent means and
824 standard deviations, respectively (n = 10 simulations). Specificity indicates the proportion
825 of true negatives that are identified correctly and it is used to quantify false positives.
826 Sensitivity indicates the proportion of labeled taxa (true positives) identified correctly.
827 Balanced accuracy is a function of both specificity and sensitivity. The *x*-axis indicates the
828 amount of $^{13}C$ isotope present in taxa that are labeled, and different colors are used to
829 indicate the percentage of taxa that have incorporated $^{13}C$ as indicated by the legend.
830
831 **Figure 4.** *HTS-DNA-SIP methods differ in their sensitivity to community dissimilarity between*
832 *replicate samples*. Beta diversity, expressed as Bray-Curtis dissimilarity, was varied
833 between simulated replicates (3 replicates each for $^{12}C$-control and $^{13}C$-treatment
834 gradients) to determine the effect that community dissimilarity between replicates has on
835 method accuracy. Variation in beta diversity was simulated by systematically varying two
836 parameters: the percent of taxa shared between replicate samples (80, 85, 90, 95, or 100
837 %) and the percent of taxa whose rank abundances that were permuted (0, 5, 10, 15, or 20
838 %), with 10 simulation replicates for each parameter set. The blue lines are LOESS curves
839 fit to accuracy values for all simulations ($n$ = 250), and the grey regions represent 99%

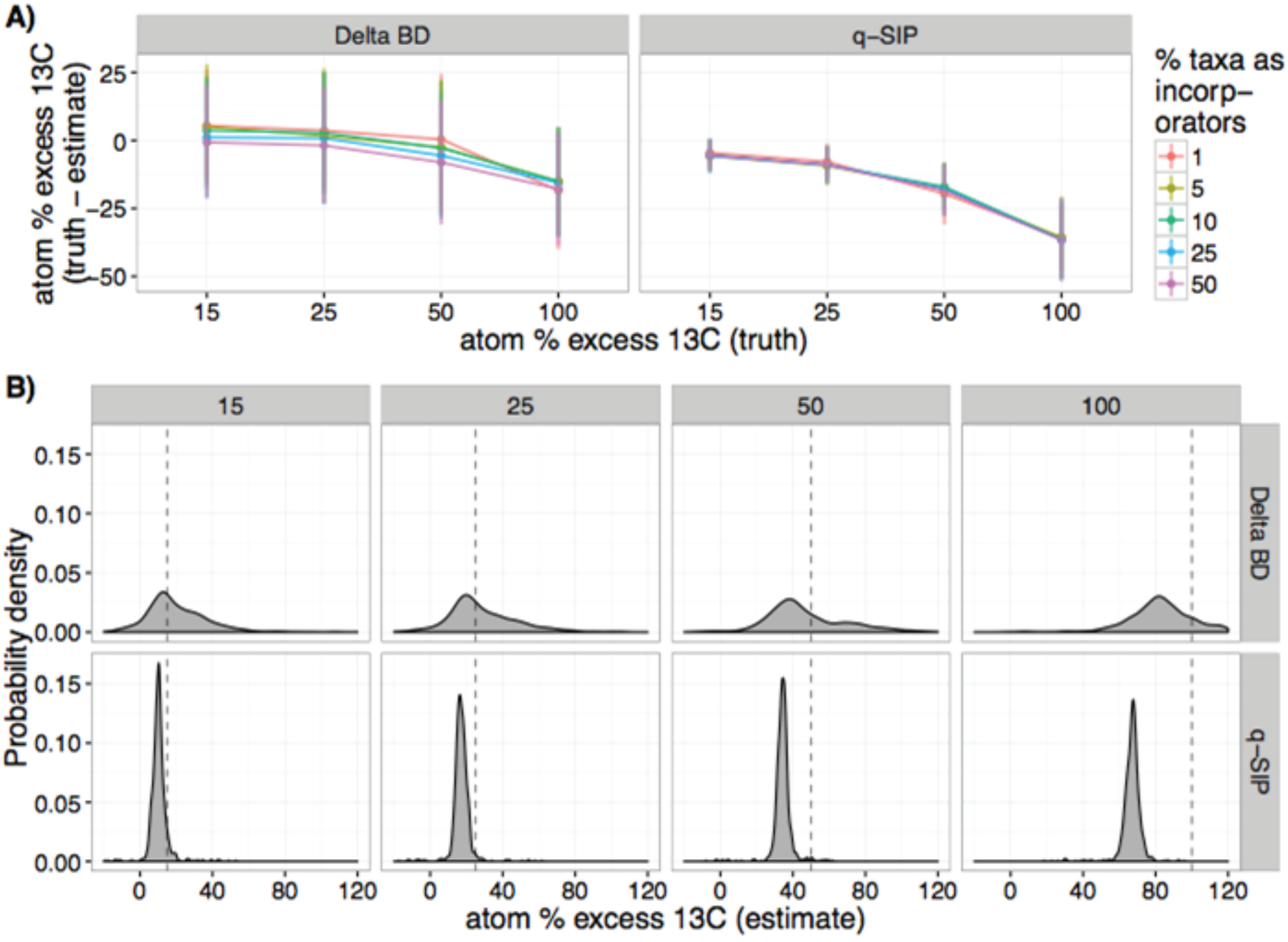
840 confidence intervals. For all simulations, 10% of the community were incorporators (100
841 % atom excess $^{13}C$).
842
843 **Figure 5.** *ΔBD and qSIP vary in their accuracy at estimating $^{13}C$ atom % excess of labeled DNA*
844 *fragments.* (A) The accuracy of both methods declines as the amount of $^{13}C$ in DNA increases,
845 but accuracy is not affected by the percent of taxa that are labeled; values indicate the mean and
846 standard deviation (n = 10 simulations). (B) Probability density plots indicate that estimates of
847 $^{13}C$ atom % excess made using *ΔBD* have greater variance than those made using *qSIP*, but both
848 estimates systematically underestimate levels of isotope incorporation. Each vertical pair of
849 panels indicates the probability density for estimates made across different levels of isotope
850 incorporation (15, 25, 50, and 100 atom % excess), and the dashed line indicates the actual level
851 of isotopic enrichment. For the calculation of probability density, 10 % of taxa were labeled
852 using the level of enrichment indicated in each panel.


Basic model input
Genome sequence dataset
fasta-formatted genome files
Experimental design
the number of gradients for labeled and unlabeled treatments
Simulating DNA fragment BD distributions
1a) Genomic DNA fragments
amplicon fragments (e.g., 16S rRNA) or shotgun metagenome fragments
1b) DNA fragment length & G+C distributions
a 2-dimensional probability distribution of the key fragment properties
1c) DNA fragment BD probability distributions
fragment BD modeled as Gaussian probability density functions
1d) Diffusive boundary layer effects
broadening of fragment BD distributions due to DBL "smearing"
1e) Isotope incorporation
shifting fragment BD distributions due to inter- and intra-population isotope incorporation
Simulating isopycnic gradients
2a) Gradient fractions
the BD range of each fraction in each gradient
2b) Pre-fractionation communities
taxon abundances in community DNA loaded into each gradient
Simulating HTS-DNA-SIP data
3a) Fragment count table
each taxon's amplicon- or shotgun-fragment count in each fraction of each gradient
3b) PCR products
adjusted fragment counts based on PCR reaction efficiencies (only for amplicon-fragments)
3c) Sequence data
subsampling fragment counts in each gradient fraction
HTS-DNA-SIP dataset
OTU abundance table
(1 sample per gradient fraction)


Pre-sequencing simulation (absolute abundance)
4e+07
3e+07
2e+07
1e+07
0e+00
control
treatment
1.675
1.700
1.725
1.750
Post-sequencing simulation (absolute abundance)
3000
2000
1000
0
Post-sequencing simulation (relative abundance)
1.00
0.75
0.50
0.25
0.00
Buoyant density (g ml-1)


Heavy−SIP
HR−SIP
MW−HR−SIP
q−SIP
Balanced Accuracy
Sensitivity
Specificity
1.0
0.8
0.6
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
atom % excess 13C
% incorp−orators
1
5
10
25
50


Heavy−SIP
HR−SIP
MW−HR−SIP
q−SIP
Balanced Accuracy
Sensitivity
Specificity
1.0
0.9
0.8
0.7
0.6
0.5
0.1
0.2
0.3
0.4
0.5
Mean Bray−Curtis distance among pre−fractionation communities


A)
Delta BD
q−SIP
atom % excess 13C (truth − estimate)
25
0
−25
−50
15
25
50
100
atom % excess 13C (truth)
% taxa as incorp−orators
1
5
10
25
50
B)
15
25
50
100
Delta BD
q−SIP
Probability density
0.15
0.10
0.05
0.00
0
40
80
120
atom % excess 13C (estimate)