Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities

Manuel Kleiner[a,b,1], Xiaoli Dong[a], Tjorven Hinzke[a,c,d], Juliane Wippler[e], Erin Thorson[a], Bernhard Mayer[a], and Marc Strous[a,1]

[a]Department of Geoscience, University of Calgary, Calgary, AB, Canada T2N 1N4; [b]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695 ; [c]Department of Pharmaceutical Biotechnology, Institute of Pharmacy, University of Greifswald, 17489 Greifswald, Germany; [d]Institute of Marine Biotechnology, 17489 Greifswald, Germany; and [e]Symbiosis Department, Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany

Edited by Margaret J. McFall-Ngai, University of Hawaii at Manoa, Honolulu, HI, and approved May 2, 2018 (received for review December 21, 2017)

Measurements of stable carbon isotope ratios ($\delta^{13}C$) are widely used in biology to address questions regarding food sources and metabolic pathways used by organisms. The analysis of these so-called stable isotope fingerprints (SIFs) for microbes involved in biogeochemical cycling and microbiota of plants and animals has led to major discoveries in environmental microbiology. Currently, obtaining SIFs for microbial communities is challenging as the available methods either only provide low taxonomic resolution, such as the use of lipid biomarkers, or are limited in throughput, such as nanoscale secondary ion MS imaging of single cells. Here we present "direct protein-SIF" and the Calis-p software package (https://sourceforge.net/projects/calis-p/), which enable high-throughput measurements of accurate $\delta^{13}C$ values for individual species within a microbial community. We benchmark the method using 20 pure culture microorganisms and show that the method reproducibly provides SIF values consistent with gold-standard bulk measurements performed with an isotope ratio mass spectrometer. Using mock community samples, we demonstrate that SIF values can also be obtained for individual species within a microbial community. Finally, a case study of an obligate bacteria–animal symbiosis shows that direct protein-SIF confirms previous physiological hypotheses and can provide unexpected insights into the symbionts' metabolism. This confirms the usefulness of this approach to accurately determine $\delta^{13}C$ values for different species in microbial community samples.

Protein-SIP | metaproteome | microbial ecology | microbiome | Q Exactive

Measurements of stable carbon isotope ratios ($^{13}C/^{12}C$, commonly called $\delta^{13}C$) are used in many different scientific fields, including atmospheric sciences, biology, paleoclimatology, oceanography, geology, environmental sciences, food and drug authentication, and forensic applications (1). In biology, stable isotope ratios (stable isotope fingerprints, SIFs) can be used to address at least two major questions. First, what is the food source of an organism? This question can be answered based on the principle that heterotrophic organisms usually have a SIF similar to their food source ("you are what you eat") (2). This has been used, for example, to assess the diet of animals (3) and to determine which microorganisms consume a specific carbon source (e.g., methane) in marine sediments (4). The second question, which can be addressed for those organisms that grow on C_1 carbon sources (bicarbonate, CO_2, or methane), is which metabolic pathway is used to assimilate the carbon source. This question can be answered based on the principle that most metabolic pathways/enzymes for C_1 assimilation discriminate against ^{13}C, which leads to characteristic carbon isotope fractionation effects. The extent of isotope fractionation of different C_1 assimilation pathways varies and thus the metabolic pathway used can be predicted based on the extent of the isotope fractionation (2, 5). This has, for example, been used in the past to distinguish plants with different carbon assimilation physiologies (6) and to predict differences in carbon fixation pathways used by symbionts of marine animals (7, 8).

Obtaining the SIFs of individual species in microbial communities is in theory a very promising tool to help unravel important abiotic and biotic interactions in global biogeochemical cycles as well as in microbiota of plants and animals. For example, if SIFs of individual species in the intestinal microbiota of humans were known, we could deduce which dietary components are used by different species in the intestine. However, there is currently no experimental approach to determine the specific SIFs of a large number of species in communities with reasonable effort and cost. The presently available approaches either have no or limited taxonomic resolution or are low-throughput. The most common approach for measuring $^{13}C/^{12}C$ ratios, isotope ratio mass spectrometry (IRMS), usually determines highly

Significance

To understand the roles that microorganisms play in diverse environments such as the open ocean or the human intestinal tract, we need an understanding of their metabolism and physiology. A variety of methods such as metagenomics and metaproteomics exist to assess the metabolism of environmental microorganisms based on gene content and gene expression. These methods often only provide indirect evidence for which substrates are used by a microorganism in a community. The direct protein stable isotope fingerprint (SIF) method that we developed allows linking microbial species in communities to the environmental carbon sources they consume by determining their stable carbon isotope signature. Direct protein-SIF also allows assessing which carbon fixation pathway is used by autotrophic microorganisms that directly assimilate CO_2.

Author contributions: M.K. and M.S. designed research; M.K., X.D., T.H., J.W., E.T., B.M., and M.S. performed research; M.K., X.D., and M.S. analyzed data; and M.K., B.M., and M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The MS proteomics data and the protein sequence databases have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository for the pure culture data (dataset identifier PXD006762), the mock community data (dataset identifier PXD006118), and the *O. algarvensis* case study data (dataset identifier PXD007510).

[1]To whom correspondence may be addressed. Email: manuel_kleiner@ncsu.edu or mstrous@ucalgary.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1722325115/-/DCSupplemental.

Published online May 29, 2018.

accurate C isotope ratios for bulk organic samples that have been converted to the measurement gas CO_2 via thermal decomposition. In IRMS the measured C isotope ratios are reported as $\delta^{13}C$ values, which give the per mille (‰) deviation of the measured ratio from the internationally accepted standard V-PDB (Vienna Pee Dee Belemnite). If lipid biomarkers are separated followed by their C isotope analysis using IRMS, high-level taxonomic groups can be resolved (9). Recently separation of proteins has also been combined with C isotope analysis using IRMS (P-SIF) (10). The P-SIF approach theoretically allows assigning $\delta^{13}C$ values to 5–10 taxa per sample; however, the approach has only been used on two bacterial pure cultures so far as it is extremely low-throughput because the mass spectrometer run time required for peptide identification alone amounts to around 2 wk per sample. A final approach is the combination of fluorescence in situ hybridization and nanoscale secondary ion MS (nanoSIMS), which enables measurement of $\delta^{13}C$ values of individual cells (4). The nanoSIMS approach, however, is currently difficult to use to measure natural abundance stable isotope ratios because the fixation and labeling procedures required for taxonomic cell identification lead to a large addition of reagent-derived carbon, nitrogen, hydrogen, and oxygen, thus diluting the true sample SIF (11, 12). Additionally, nanoSIMS has a low throughput because specific fluorescently labeled probes have to be applied for each individual species or higher-level taxonomic group.

Here we present "direct protein-SIF" and the Calis-p (The CALgary approach to ISotopes in Proteomics) software package, a method that enables high-throughput measurement of $\delta^{13}C$ (SIF) values for individual species within microbiota and environmental microbial communities using metaproteomics. We use the word "direct" to highlight the fact that the SIF data are directly extracted from a standard metaproteomic dataset (i.e., the same mass spectrometry data are used for both peptide identification and SIF estimation). The direct protein-SIF workflow (Fig. 1 and *SI Appendix*, Fig. S1) consists of a standard proteomics sample preparation to produce peptide mixtures of the samples and a reference material, followed by acquisition of 1D or 2D liquid chromatography tandem MS (LC-MS/MS) data of the peptide mixture using a high-resolution Orbitrap mass spectrometer (for details see *Methods*). The MS/MS data are used as input for a standard proteomic database search to produce a table with scored peptide spectrum matches (PSMs). The PSM tables for the samples and the reference material plus the raw MS data (in mzML format) are used as input for the Calis-p software. In a first step the software finds the isotopic peaks for each PSM and sums their intensity across a retention time window. The isotope peak intensities are reported together with the sum formula of the identified peptides. Isotope peak intensity patterns with low intensities or low search engine scores are discarded. The remaining isotope patterns are used as input for the stable isotope fingerprinting. In this step, experimentally derived isotope peak distributions are fitted to theoretical isotope peak distributions computed for a range of $\delta^{13}C$ values with a fast Fourier transform method adapted from Rockwood et al. (13). Peptides with a poor goodness of fit are discarded. The remaining peptides are used to compute an average $\delta^{13}C$ value for each individual species and associated SEs in per mille. In a final step the species $\delta^{13}C$ values are corrected for instrument isotope fractionation by applying the offset determined using the reference material (see details in *Results*). Just like IRMS, our approach provides a single, robust, averaged $\delta^{13}C$ value, one for each species. It is well known that ^{13}C contents can vary between biomolecules and even between positions within a single biomolecule (5, 14–16). We show that given sufficient data the ^{13}C content of individual amino acids might be resolved from average per-peptide values by multivariate regression.

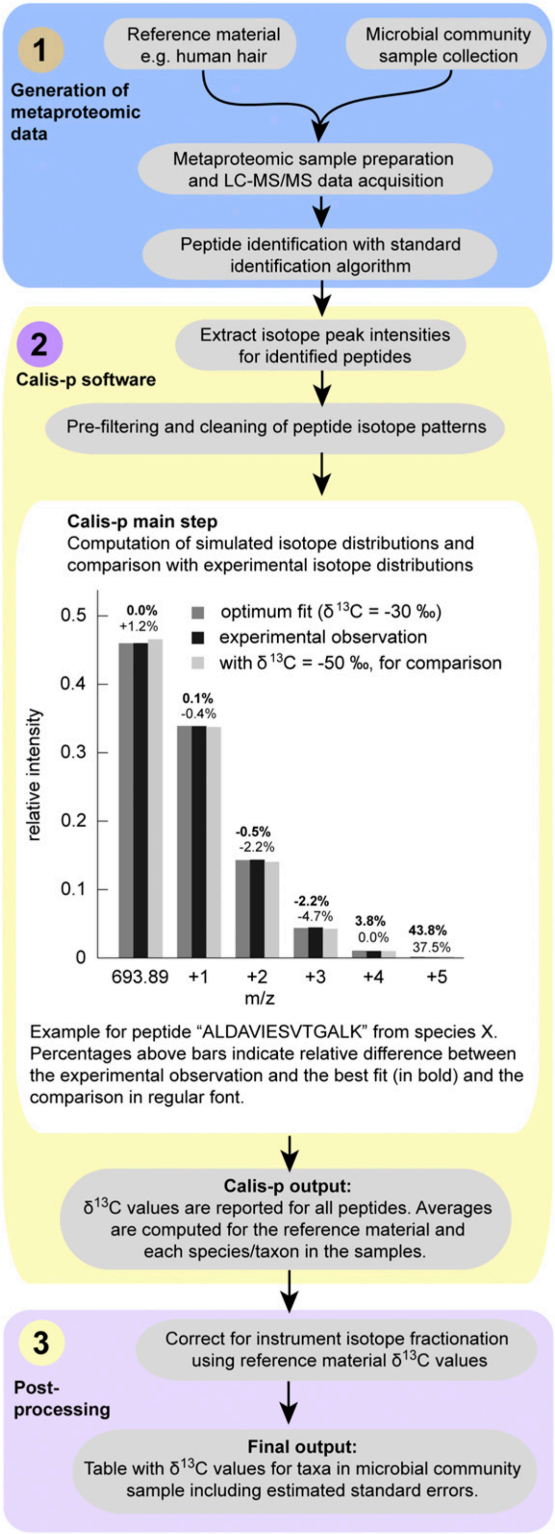
1
Generation of metaproteomic data
Reference material e.g. human hair
Microbial community sample collection
Metaproteomic sample preparation and LC-MS/MS data acquisition
Peptide identification with standard identification algorithm
2
Calis-p software
Extract isotope peak intensities for identified peptides
Pre-filtering and cleaning of peptide isotope patterns
Calis-p main step
Computation of simulated isotope distributions and comparison with experimental isotope distributions
optimum fit (δ 13C = -30 ‰)
experimental observation
with δ 13C = -50 ‰, for comparison
relative intensity
0.5
0.4
0.3
0.2
0.1
0
0.0%
+1.2%
0.1%
-0.4%
-0.5%
-2.2%
-2.2%
-4.7%
3.8%
0.0%
43.8%
37.5%
693.89
+1
+2
+3
+4
+5
m/z
Example for peptide "ALDAVIESVTGALK" from species X. Percentages above bars indicate relative difference between the experimental observation and the best fit (in bold) and the comparison in regular font.
Calis-p output:
δ13C values are reported for all peptides. Averages are computed for the reference material and each species/taxon in the samples.
3
Post-processing
Correct for instrument isotope fractionation using reference material δ13C values
Final output:
Table with δ13C values for taxa in microbial community sample including estimated standard errors.

Fig. 1. Direct protein-SIF workflow. In the main step of the data analysis with the Calis-p software, the experimentally derived isotope distributions for peptides are compared with theoretical isotope peak distributions computed based on peptide molecular formulae with a fast Fourier transform method. The comparison with theoretical distributions is done for a specified range of $\delta^{13}C$ values in increments. Goodness of fit is calculated for all comparisons and the $\delta^{13}C$ value for the best fit is reported if a predetermined goodness of fit threshold is passed. A more detailed workflow can be found in *SI Appendix*, Fig. S1.

Results

Benchmarking with Pure Culture Data. For benchmarking, we measured the stable carbon isotope ratios of 20 pure culture species using both direct protein-SIF and continuous flow elemental analysis IRMS (CF-EA-IRMS). CF-EA-IRMS is the most commonly used method to determine highly accurate carbon isotopic compositions of organic bulk samples with measurement uncertainties of ±0.15‰ or less and can be considered the gold standard. The 20 pure cultures represented 18 bacterial, 1 archaeal, and 1 eukaryotic species (Datasets S1 and S2). For seven of the species we obtained technical replicate measurements to determine the precision of the direct protein-SIF method.

The protein-SIF δ^{13}C values for the technical replicate measurements of individual species were highly consistent with each other and generally deviated by less than ±1‰ from the mean (Dataset S1). The direct protein-SIF δ^{13}C values were linearly correlated to the CF-EA-IRMS values (R^2 = 0.94). The direct protein-SIF δ^{13}C values showed a systematic offset from the CF-EA-IRMS values of −15.4‰ (SD = 2.55) (Fig. 2*A*). This systematic offset is likely caused by isotope fractionation in the Orbitrap mass spectrometer, which has been recently described (17). The large range of δ^{13}C values measured with both direct protein-SIF and CF-EA-IRMS corresponds to the isotope ratios of the different carbon sources used to grow the pure cultures (Dataset S3) and the heterotrophic versus autotrophic lifestyle of the different microorganisms.

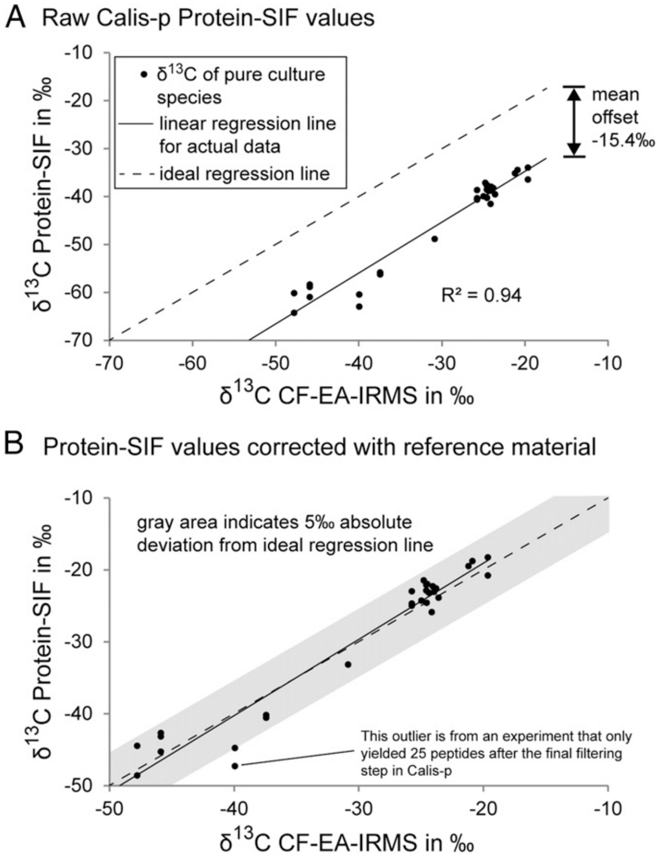

Fig. 2. Comparison of δ^{13}C measurements of pure cultures with protein-SIF and CF-EA-IRMS. Twenty pure cultures representing 18 bacterial, 1 archaeal, and 1 eukaryotic species were measured with both methods (detailed data in Datasets S1 and S2). For seven of the species technical replicate measurements were obtained. (*A*) The raw δ^{13}C values from the Calis-p software plotted against the IRMS-derived values. The average offset of protein-SIF values from the IRMS values is indicated. (*B*) Protein-SIF values after offset correction using the offset determined with reference material (human hair).

Correction of Instrument Isotope Fractionation with Reference Material. To correct for instrument isotope fractionation when using the direct protein-SIF method, we implemented the use of a reference material which is prepared, measured, and analyzed alongside the samples. With CF-EA-IRMS, reference materials with known isotopic compositions are also needed to correct for instrument isotope fractionation (18). We chose human hair as the reference material for the following reasons. (*i*) The bulk of its dry mass consists of protein, which allows directly correlating protein-SIF and IRMS δ^{13}C values. Interference from other biomolecules with potentially different isotope composition, such as lipids, is limited. (*ii*) It contains a large diversity of proteins providing hundreds to thousands of different peptides that can be measured for protein-SIF. (*iii*) The protein sequences required for peptide identification are known. (*iv*) It is easy to obtain in large batches (a few grams) to serve as a reference material for many years. Human hair with known δ^{13}C values measured by IRMS is also available from major chemical suppliers and the US Geological Survey (https://isotopes.usgs.gov/lab/referencematerials.html).

The offset between δ^{13}C values generated with direct protein-SIF and CF-EA-IRMS for the human hair reference material was −15.7‰ and thus almost identical to the average offset for the benchmarking pure culture δ^{13}C values (Fig. 2*A*). We corrected the pure culture protein-SIF δ^{13}C values with the reference material offset. After this, the absolute deviation of the protein-SIF δ^{13}C values from the CF-EA-IRMS δ^{13}C values was on average ±2.1‰ (SD = 1.5) (Fig. 2*B* and Dataset S2).

To explore whether the approach could also be used to estimate per-amino acid ^{13}C contents, we used multivariate regression to estimate the per-amino acid ^{13}C content for each of the 18 amino acids included in our analysis. *SI Appendix*, Fig. S2 shows consistent trends in ^{13}C content, with some amino acids strongly depleted in ^{13}C (δ^{13}C −78‰ for arginine and −64‰ for lysine) and others strongly enriched in ^{13}C (δ^{13}C +58‰ for asparagine and +50‰ for lysine) relative to an organism's average δ^{13}C value. *SI Appendix*, Fig. S2 also shows that these estimates are not very precise. Differences of 50‰ between replicates were not exceptional.

Detection Limit and Accuracy of Protein-SIF in Mock Communities. To determine the detection limit of our approach and to test how accurately we can measure SIFs in complex community samples, we used mock communities. These mock communities were created by mixing the 20 pure culture species used for benchmarking with 12 additional microbial strains and species so that the final community contained a total of 32 strains and species. We generated a total of 12 biological replicates of the community with differing species abundances and analyzed each replicate with two different LC-MS/MS methods (19). To evaluate the mock community direct protein-SIF data we only considered the 20 benchmarking species in the community for which IRMS and direct protein-SIF δ^{13}C values were known based on the results presented above.

The accuracy and detection limit of direct protein-SIF depends heavily on the available amount of data. We found that the number of peptides for each species that pass the final Calis-p filtering step has a large influence on the accuracy of the determined SIF values (Fig. 3). If fewer than 60 peptides were available for SIF value calculation, the protein-SIF δ^{13}C values differed from the IRMS-derived δ^{13}C values by more than 10‰ for more than half of the 20 species. If between 60 and 100 peptides were available, protein-SIF values differed less than 10‰ from IRMS values for six out of nine species, with only one large outlier. With more than 100 peptides available, over 95% of protein-SIF values differed less than 10‰ from δ^{13}C values determined by IRMS (mean deviation 4.7‰). As expected, the detection limit for direct protein-SIF depended on the amount of mass spectrometric data available, because to reach the necessary

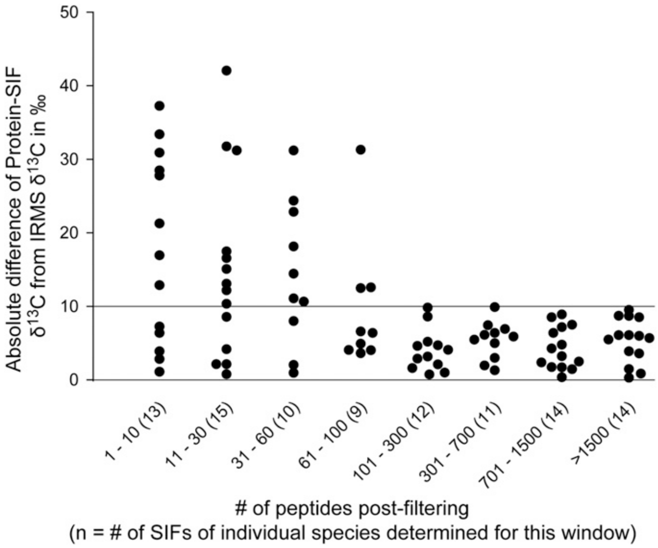

Fig. 3. Absolute difference between $\delta^{13}C$ values determined with direct protein-SIF of individual species in mock communities and IRMS of the corresponding pure cultures. Five mock community datasets with a total of 32 species and strains were analyzed. For 20 species, the $\delta^{13}C$ values were known from IRMS performed on pure cultures. For these species, the $\delta^{13}C$ values were determined with direct protein-SIF. Each dataset contained different amounts of data (Table 1). Different numbers of peptides were identified and passed the final Calis-p peptide filter for each species in each dataset. The absolute difference between $\delta^{13}C$ values obtained via protein-SIF and IRMS was calculated and sorted according to how many peptides were available for SIF calculation by Calis-p after filtering the peptides. The plot gives the absolute differences for different ranges of peptide numbers used for SIF calculation.

number of high-quality peptide isotopic patterns for low-abundant species more mass spectra are required (Table 1 and Dataset S4). For example, one 4-h LC-MS/MS run enabled determination of SIF values for 25% of the species in the mock community. These species were relatively abundant (proteinaceous biomass >5.7% of the total community). In contrast, combined data of 12 4-h runs enabled determination of SIF values for 75% of species, including species with lower (<1% of the total proteinaceous biomass) abundance (Table 1). We also found that the accuracy of SIF estimates of individual peptides strongly depended on the mass spectrometric peak intensities of the peptide (*SI Appendix*, Fig. S3). This could be explained by the presence of background noise, most likely originating from small interfering peaks that were not completely resolved from some of the analyzed peaks for a peptide, thus slightly changing individual isotopic peak intensities for peptides.

Case Study. To demonstrate the power and application of direct protein-SIF we applied it to a well-studied bacteria–animal symbiosis, the gutless marine oligochaete *Olavius algarvensis* (Fig. 4*A*). *O. algarvensis* lacks a digestive system. Instead, the worm relies on at least five bacterial symbionts under its cuticle for nutrition (Fig. 4 *B* and *C*). The metabolism and physiology of these symbionts and their interactions with the host have been extensively studied using metagenomics, metaproteomics, and physiological incubation experiments (20–22). Based on these previous studies, the carbon sources of the symbionts and host were thought to be as follows (Fig. 4*C*). Two sulfur oxidizers (γ1 and γ3) fix seawater-derived inorganic carbon using the Calvin–Benson–Bassham cycle with a Form IA RubisCO enzyme (20). Two sulfate reducers (δ1 and δ4) consume the host's organic waste products. For a spirochete, no data on metabolism and physiology were available. The host itself derives its carbon directly from the symbionts by digestion through endocytosis (23). Thus, the $\delta^{13}C$ values of the γ- and δ-symbionts as well as the host were expected to be similar, since all carbon is derived from the initial carbon fixation by the γ-symbionts. In a previous study the $\delta^{13}C$ value of complete worms was determined by IRMS to be −30.6‰ (21). Our expectation was therefore that all direct protein-SIF–derived $\delta^{13}C$ values would be somewhere in the range of −25 to −35‰. We assumed this range of possible $\delta^{13}C$ values because of the measurement uncertainty of direct protein-SIF, as well as the fact that different trophic levels in a heterotrophic food chain can vary in their $\delta^{13}C$ value by 0.5–2‰ from their food source (24, 25).

We put this model of carbon flow in *O. algarvensis* to the test by applying direct protein-SIF using a metaproteomic dataset obtained from multiple individual worms. To obtain sufficient peptides for measurement of $\delta^{13}C$ values of all of the symbionts, we combined 18 LC-MS/MS runs from a total of 14 individual worms. The $\delta^{13}C$ values obtained by direct protein-SIF of the host (−25.6‰), the two γ-symbionts (−32.9 and −31.8‰) and the δ1-symbiont (−34.8‰) were in the expected range (Fig. 4*D* and Dataset S5). However, the δ4-symbiont had a much higher $\delta^{13}C$ value (−17.9‰, SE ±1.8‰) than the δ1-symbiont. This was unexpected because the two δ-symbionts appear to be almost identical in terms of expressed carbon uptake and catabolic pathways. Both are characterized by many abundantly expressed high-affinity uptake transporters for sugars, amino acids, peptides, and organic acids, as well as pathways for the use of host-derived

Table 1. Detection limit of direct protein-SIF for species in mock communities depending on the amount of LC-MS/MS data available

	Total LC-MS/MS run time, h				
Experimental parameters and outcomes	92	52	31	8	4
Biological replicates	12	12	4	1	1
Gradient length, min	460	260	460	460	260
MS/MS spectra (in millions)	∼2.04	∼1.2	∼0.68	∼0.17	∼0.1
Species (out of 20) with protein-SIF*	15	15	10	6	5
Lower species abundance limit for SIF determination†, %	0.82	0.82	0.92	5.65	5.79
Mean deviation of protein-SIF $\delta^{13}C$ from IRMS $\delta^{13}C$, ‰	3.6	5.4	4.7	4.4	5.8
Minimum deviation, ‰	0.2	0.6	0.9	1.5	2.8
Maximum deviation, ‰	9.4	9.7	8.8	8.4	9.8

Mock community samples with 32 species and strains were used (14). For 20 of these species the IRMS $\delta^{13}C$ values were known. The detailed data can be found in *SI Appendix*, Table S4.

*Number of species with a sufficient number of peptides for protein-SIF (>100) after final Calis-p filtering.

†Abundance (percentage of total community protein) of lowest abundance species for which determining protein-SIF value was possible (i.e., >100 peptides after final Calis-p filtering).

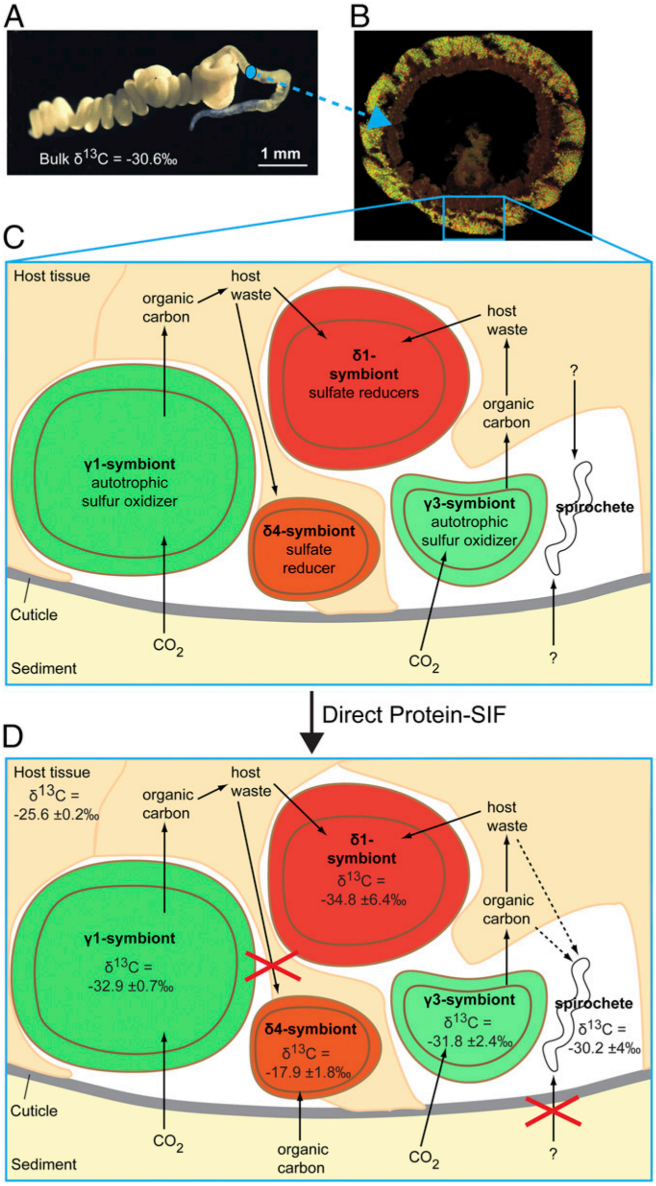
A
B
Bulk δ13C = -30.6‰
1 mm
C
Host tissue
organic carbon
host waste
host waste
δ1-symbiont sulfate reducers
organic carbon
?
γ1-symbiont autotrophic sulfur oxidizer
δ4-symbiont sulfate reducer
γ3-symbiont autotrophic sulfur oxidizer
spirochete
Cuticle
CO2
CO2
?
Sediment
Direct Protein-SIF
D
Host tissue δ13C = -25.6 ±0.2‰
organic carbon
host waste
host waste
δ1-symbiont δ13C = -34.8 ±6.4‰
organic carbon
γ1-symbiont δ13C = -32.9 ±0.7‰
δ4-symbiont δ13C = -17.9 ±1.8‰
γ3-symbiont δ13C = -31.8 ±2.4‰
spirochete δ13C = -30.2 ±4‰
Cuticle
CO2
organic carbon
CO2
?
Sediment

Fig. 4. Testing the model of physiological interactions in the *O. algarvensis* symbiosis using direct protein-SIF. (*A*) Live *O. algarvensis* specimen. The $\delta^{13}C$ value of bulk worms was determined by IRMS on six biological replicates in Kleiner et al. (21). Image courtesy of Christian Lott (photographer). (*B*) Cross-section through the worm. The bacterial symbionts right below the worm's cuticle are stained with specific fluorescence in situ hybridization probes (γ-symbionts in green, δ-symbionts in red). (*C*) Simplified model of carbon flow in the symbiosis based on previous metagenomic (22) and metaproteomic (26) studies. For the δ1- and δ4-symbionts the metaproteomic data suggested that these two symbionts are highly similar in terms of metabolism and physiology. (*D*) Adjusted model of carbon flow based on carbon sources predicted using direct protein-SIF–derived $\delta^{13}C$ values. Detailed protein-SIF data in Dataset S5. The ± value for each $\delta^{13}C$ value indicates the SE.

acetate and propionate (26). The direct protein-SIF data now point toward a functional difference between the two δ-symbionts. Apparently, the δ4-symbiont derives part or all of its carbon from external sources that have a different carbon isotopic composition. Currently we do not have an explanation of how the δ4-symbiont obtains external carbon and what the main physiological difference between the two δ-symbionts is. One potential mechanism for how the δ4-symbiont could obtain symbiosis-external carbon is that it lives in the sediment pore water and from there continuously infects *O. algarvensis*; however, continuous infection of the worm is unlikely as the worm does not have any orifices that would allow for such infection to occur and the δ4-symbiont genome does not encode for any known pathways that would allow it to penetrate the worm's cuticle. It is also possible that the two symbionts have differing affinities for specific substrates, which cannot be deduced from the transporter annotations. Future work using isotopically labeled substrates could help to address this hypothesis.

We were also able to measure a sufficient number of peptides to estimate a $\delta^{13}C$ value for the spirochetal symbiont, despite its low abundance. The spirochete $\delta^{13}C$ value of −30‰ was in the range of symbiosis-internal carbon, suggesting that this symbiont uses a symbiosis-internal carbon source. Currently, we are not aware of any symbiosis-external carbon sources that have a $\delta^{13}C$ value in the range of −30‰; however, we cannot exclude that the spirochete has access to a symbiosis-external source with such a negative $\delta^{13}C$ value. In summary, by applying direct protein-SIF to the *O. algarvensis* symbiosis we derived insights into carbon flow that were not suggested by any of the previous metaomics-based studies.

Discussion

The developed direct protein-SIF approach provides a means to directly and simultaneously access the SIFs (i.e., $\delta^{13}C$ values) of many individual species in microbial communities. As little as 4 h of LC-MS/MS time can be sufficient to estimate SIFs for the most abundant species in a sample. For this type of analysis only very small sample amounts are needed; recent advances in sample preparation allow for the production of metaproteomic data of sufficient quality from as little as 1 mg of wet weight cell mass. If the determination of SIFs for lower-abundant species is desired, more LC-MS/MS run time can provide the required metaproteomic depth. For longer or additional LC-MS/MS runs proportionally larger amounts of sample are needed (e.g., 2 mg for two 4-h runs). Additionally, enrichment of specific cell populations by filtration or centrifugation methods can be used to obtain better metaproteomic coverage (26, 27).

Differences Between IRMS and Direct Protein-SIF Measurements. We observed a range of differences between the $\delta^{13}C$ values measured by IRMS and direct protein-SIF when using both on single-species biomass (absolute deviation of the protein-SIF $\delta^{13}C$ values from the CF-EA-IRMS $\delta^{13}C$ values was on average ±2.1‰). Part of the observed variation between the two methods is likely due to a lower accuracy of direct protein-SIF using the highly complex metaproteomic mass spectrometric data. However, there are at least two other factors that might contribute to this variation in the observed $\delta^{13}C$ values.

First, direct protein-SIF measures the $\delta^{13}C$ value of proteins, while bulk IRMS measures all cell components such as protein, lipids, DNA, and metabolites, providing a weighted average. It has been shown that the $\delta^{13}C$ values of different cell components can differ. For example, lipids have a 1.6‰ lower $\delta^{13}C$ value than protein in *Escherichia coli* (28) and the lipids of Calvin–Benson–Basham cycle autotrophs have around 6‰ lower $\delta^{13}C$ values compared with the $\delta^{13}C$ values of the total biomass (5, 7). The possible difference between protein $\delta^{13}C$ values and bulk organic matter $\delta^{13}C$ values should thus be considered when interpreting direct protein-SIF results. In addition, different amino acids have different ^{13}C contents, and even carbon positions within amino acids differ in their $^{13}C/^{12}C$ ratios. These differences might also lead to a detectable bias if the amino acid composition is highly skewed or if only few peptides are available. Since protein usually makes up the majority of a cell in terms of mass, for example, 55% of *E. coli* dry weight (BioNumbers ID 104954) (29), protein $\delta^{13}C$ values will be a good approximation of bulk $\delta^{13}C$ values in most cases.

Second, the variation in the intensities of the peptide isotopic peaks used for direct protein-SIF is mostly due to variation in the ratio of ^{13}C to ^{12}C, because ^{13}C is, with a natural abundance of ~1.1%, the most abundant heavy stable isotope in the considered peptides. However, very large isotope fractionation of the three other elements (hydrogen, oxygen, and nitrogen) in the peptides that we consider (sulfur-containing peptides are excluded) could change the measured $\delta^{13}C$ values by several per mille. For example, the hydrogen isotope fractionation in photosynthate compared with the water used by the photosynthetic organism is $\varepsilon = -171‰$ (5). The $^2H/^1H$ ratio in the Vienna Standard Mean Ocean Water (VSMOW) reference is 0.00015576, which means that the $^2H/^1H$ ratio of photosynthate is 0.000129 if ocean water is the hydrogen source (i.e., a change in the fifth decimal of the fraction of heavy atoms). For comparison, carbon isotope fractionation influences the third and fourth decimal. As the $^{13}C/^{12}C$ ratio of V-PDB is 0.0111802, a $\delta^{13}C$ value of −47.2‰ corresponds to a $^{13}C/^{12}C$ ratio of 0.01065. Carbon isotope fractionation by only −3‰ would already yield a $^{13}C/^{12}C$ ratio of 0.01115, a shift in the fraction of heavy atoms similar to that produced by the fractionating hydrogen isotopes by −171‰. Or, to put it differently, a δ^2H value of −171‰ would change the estimated $\delta^{13}C$ value of −47.2‰ to −50.1‰ in direct protein-SIF if the reference material used had a δ^2H value similar to that of VSMOW (0‰). However, protein reference materials used for direct protein-SIF will have a much more negative δ^2H value compared with VSMOW [e.g., the δ^2H values of human hair range typically between −130 and −80‰ (30)], thus removing the most common hydrogen isotope fractionation effects when applying the reference material-based offset correction to the direct protein-SIF $\delta^{13}C$ values. Therefore, while nitrogen, hydrogen and oxygen isotope fractionation will usually not have a major effect on direct protein-SIF $\delta^{13}C$ values, it should always be considered during interpretation of direct protein-SIF results. Measurement of δ values for these three elements by IRMS bulk analyses of samples and reference material would provide insights into any major deviations that should be considered to achieve even better accuracy of the results.

Can This Approach Be Used for Protein-SIP? In recent years, metaproteomics has been successfully combined with stable isotope probing (protein-SIP) to follow the incorporation of isotopically labeled substrates by individual members in microbial communities (31–34). Several different algorithms have been used to compute isotope incorporation levels in protein-SIP and some of these algorithms employ approaches similar to our direct protein-SIF approach. All current protein-SIP approaches, however, can compute isotope incorporation levels only with low precision (i.e., several atom percent changes in isotope ratios are needed to obtain a clear readout). Naturally, the question arises if the direct protein-SIF approach could be also employed for protein-SIP. Answering this question will require further method development, but we predict that the direct protein-SIF approach should be applicable for protein-SIP approaches that use small percentages of labeled substrate or short labeling pulses. If we assume that the protein identification algorithm used correctly identifies peptides if the intensity of the monoisotopic peak is at least 33% of its original (unlabeled) intensity, the maximum amount of label assimilated that would still enable correct identification of >90% of all peptides would be ~1%. Since the Calis-p software resolves differences between $\delta^{13}C$ values of <10‰ units, the resulting resolution would be 0.01%. The Calis-p approach would thus not be a replacement for other protein-SIP approaches but would complement them by adding the capability to detect incorporation of very low amounts of heavy atoms into proteins.

Potential Limitations. There are two potential limitations of the direct protein-SIF approach when it comes to resolving $\delta^{13}C$ values for species-level microbial populations. First, strains within one species often share a high protein sequence identity, which makes it difficult to uniquely assign most peptides to individual strains within one species. This means that if we sample a microbial community that has multiple strains of the same species and these strains use different carbon substrates, the direct protein-SIF method will only report an "averaged" SIF for the species, which would hamper deductions about which carbon substrates were used. This strain resolution challenge could be addressed if a strain-resolved metagenomic database were available for the sample in question. With such a database, SIFs for individual strains could be resolved by filtering for strain-unique peptides before calculating the SIFs. Additionally, a careful evaluation of "expressed" metabolism and physiology of the species-level population using the metaproteomic data would provide insights into which and how many substrates might be used by a population and thus aid in the interpretation of the SIF data. Second, it is conceivable that individual cells of a population consisting of a single strain use different carbon substrates depending on their position in an environmental matrix or the same cell uses multiple substrates at once. This again would lead to an "averaged" SIF for the strain, confounding the signals from the carbon substrates used. In this second case, the use of the metaproteomic data to analyze the "expressed" metabolism would be essential for the SIF data interpretation.

Conclusions

The direct protein-SIF approach provides us with a key ability that no other method can provide at the moment. By measuring $\delta^{13}C$ values of individual species in microbial communities we can now make inferences about the food sources for these species as well as the metabolic pathways used for carbon assimilation. Ongoing technological development will further improve the accuracy, detection limits, and capabilities of the direct protein-SIF approach in the future in at least three ways. First, the fractionation of isotopic species in the mass spectrometer could be reduced with specialized methods and potentially with improvements of the instrumentation. Current approaches for reducing isotope fractionation in Orbitrap mass spectrometers are very promising even though they are not yet usable for metaproteomics (17), as these new approaches significantly reduce the number of MS/MS spectra acquired and thus reduce the number of identified peptides. Second, mass spectrometers with increased resolving power at high scan rates will make it possible to separate the isotopologues of peptides based on which element provides the heavy isotope (hydrogen, nitrogen, carbon, or oxygen). For example, the exchange of one ^{14}N atom with one ^{15}N atom in a peptide changes its mass by 0.99703489 Da, while the exchange of one ^{12}C with one ^{13}C changes the mass by 1.0033548378 Da. Once isotopic species can be resolved on a per-element basis, it should become feasible to not only determine carbon isotope ratios but also isotope ratios for other elements in peptides such as nitrogen, oxygen, and hydrogen. The necessary mathematical approach to calculate the required fine-structure peptide isotope patterns via a multidimensional Fourier transform has been recently demonstrated by Ipsen (35). Third, if sufficient numbers of peptides are measured for a species it is in theory possible to estimate amino acid-specific $\delta^{13}C$ values. $\delta^{13}C$ values of individual amino acids can provide additional information about carbon sources or biosynthetic pathways for a species (5, 14, 15).

Methods

Sample Preparation. The cultivation of the pure cultures and the creation of the mock communities are described in Kleiner et al. (19). For the case study, 14 *O. algarvensis* specimens were collected off the coast of Sant' Andrea Bay, Elba, Italy (42°48′26″N, 010°08′28″E) in August 2015 from shallow-water (6- to 8-m water depth) sediments next to seagrass beds. Live worms were transported in native Elba sediment and seawater to the Max Planck Institute for Marine Microbiology in Bremen, Germany. The sand which was used for worm transport and storage was washed three times with clean freshwater followed by three washes with clean seawater to remove life and dead meiofauna and other potential sources of organic substrates. The worms were kept for 1 mo in the dark and at Elba marine sediment temperature. Worms were then carefully removed from the sediment and frozen at −80 °C until further processing.

The human hair used as a standard for correction of instrument isotope fractionation was obtained from M.K.

Peptide samples for proteomics were prepared and quantified as described by Kleiner et al. (19) following the filter-aided sample preparation protocol described by Wiśniewski et al. (36). The only modification for the *Olavius* samples compared with the pure cultures and the mock communities was that no bead beating step was used. The bead beating step was used for the human hair reference material.

One-Dimensional LC-MS/MS. Samples were analyzed by 1D LC-MS/MS as described in Kleiner et al. (19). One or two wash runs and one blank run were performed between samples to reduce carryover. For the 1D LC-MS/MS runs of pure culture samples, 2 μg of peptide were loaded onto a 5-mm, 300-μm i.d. C18 Acclaim PepMap 100 precolumn (Thermo Fisher Scientific) using an UltiMate 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific). After loading, the precolumn was switched in line with a 50-cm × 75-μm analytical EASY-Spray column packed with PepMap RSLC C18, 2 μm material (Thermo Fisher Scientific). For the 1D LC-MS/MS runs of the *Olavius* samples, 0.8–4 μg of peptide were loaded onto a 2-cm, 75-μm i.d. C18 Acclaim PepMap 100 precolumn (Thermo Fisher Scientific) using an EASY-nLC 1000 Liquid Chromatograph (Thermo Fisher Scientific) set up in two-column mode. The precolumn was also connected to a 50-cm × 75-μm analytical EASY-Spray column packed with PepMap RSLC C18, 2 μm material. In both cases the analytical column was connected via an Easy-Spray source to a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Peptides were separated on the analytical column using 260- or 460-min gradients and mass spectra were acquired in the Orbitrap as described by Petersen et al. (37). For the mock communities, the existing 1D LC-MS/MS data from Kleiner et al. (19) were used.

Input Data Generation, Algorithm, and Software Development for Protein-SIF. ***Peptide identification.*** For peptide and protein identification of pure culture samples, protein sequence databases were created using the reference protein sequences for each species separately. The databases are available from the PRIDE repository (PXD006762). For the mock community samples, the database from the PRIDE repository (PXD006118) was used (19). For the Olavius samples, an existing *O. algarvensis* host and symbiont protein database from project PXD003626 (ftp://massive.ucsd.edu/MSV000079512/sequence/) was used. To this *Olavius* database we added additional symbiont protein sequences from recently sequenced metagenomes. The new *Olavius* database is available from the PRIDE repository (PXD007510). For the human hair standard, the human reference protein sequences from UniProt (UP000005640) were used. CD-HIT was used to remove redundant sequences from the multimember databases using an identity threshold of 95% (38). The cRAP protein sequence database (https://www.thegpm.org/crap/) containing protein sequences of common laboratory contaminants was appended to each database. One important consideration for creating protein sequence databases for use with the Calis-p software is that Calis-p will calculate taxon/population-specific $\delta^{13}C$ values based on accession number prefixes that indicate to which taxon a sequence belongs. The prefix should be separated from the accession number by an underscore (e.g., >TAX_00000). MS/MS spectra were searched against the databases using the Sequest HT node in Proteome Discoverer version 2.0.0.802 (Thermo Fisher Scientific) and peptide spectral matches were filtered using the Percolator node as described by Petersen et al. (37). The FidoCT node in Proteome Discoverer was used for protein inference and to restrict the protein-level false discovery rate to below 5% (FidoCT q-value < 0.05).

Input files. Examples for all input files are provided in PRIDE project PXD006762 alongside the raw data. The LC-MS/MS-produced raw files were converted into mzML format using MSConvertGUI via ProteoWizard (39) with the following options set: Output format: mzML, Binary encoding precision: 64-bit, Write index: checked, TPP compatibility: checked, Filter: Peak Picking, Algorithm: Vendor, MS Levels: 1 (The MS/MS scans are not needed for isotope pattern extraction). The peptide-spectrum match (PSM) files generated by Proteome Discoverer were exported in tab-delimited text format. The mzML files and the PSM files were used to extract isotopic patterns for all identified peptides.

Isotopic pattern extraction. The steps described here are carried out by the Calis-p software in a fully automated fashion upon provision of correctly formatted input files. The entries in the PSM files were excluded if they (*i*) had any identified posttranslational modifications, (*ii*) contained "M" or "C" in the peptide sequence, (*iii*) had a peptide confidence score not equal to "High," or (*iv*) had a PSM Ambiguity value equal to "Ambiguous." The remaining entries were subjected to the isotopic pattern extraction from the mzML mass spectrum input files. In a first step, the *m/z* value of the monoisotopic peak (A) of each peptide was searched for in the full scans in the mass spectrum file in a defined retention time window (peptide scan start time ±0.5). All peaks with the same *m/z* value in the retention time window were considered to be from the same peptide, because most peptides are analyzed multiple times in subsequent MS^1 scans in the defined RT window size. In a second step, each MS^1 scan of the same peptide was searched for the isotopic peaks of the peptide produced by replacement of single or multiple atoms in the peptide with heavier isotopes (i.e., A+1, A+2…). The isotopic peaks were defined as all of the peaks following the monoisotopic peak in the full scan mass spectrum with *m/z* values being n*(1/charge) distance removed from the monoisotopic *m/z* value (n is the peak count i.e., A+1, A+2…). Since the number of isotopic peaks found for a peptide in different MS^1 scans is variable due to decreasing signal intensity for the higher-number isotopic peaks, we chose to report the peak intensities only for the isotopic peaks detected in the majority of scans. To clarify, if we found the A, A+1, A+2 peaks for five MS^1 scans, A, A+1, A+2, A+3 peaks for three MS^1 scans, A, A+1, A+2, A+3, A+4 for seven MS^1 scans, and A, A+1, A+2, A+3, A+4, A+5 for three MS^1 scans, we only reported peak intensities for A to A+4. Scans that did not have all isotopic peaks to be reported were excluded (i.e., for the preceding example this would mean a scan that only had peaks A to A+3 would be excluded). To maximize carbon ratio calculation accuracy the intensities of each isotopic peak from all of the included scans of the same peptide were summed up. The filtered PSM entries with the identified isotopic patterns and aggregated intensities were written to an isotopicPattern text file in a tab delimited format. Additional information needed for SIF computation such as peptide sequence, peptide sum formula, charge, and peptide identification quality score (Xcorr) was written into the isotopicPattern file. For examples of isotopic pattern files see PRIDE project PXD006762.

SIF computation. Peptide isotope patterns were filtered to remove exact duplicates (peptides associated with identical spectra). Also, the following default parameters were used: peptides associated with more than three proteins or that were identified with low confidence (Xcorr < 2.0), or had fewer than three isotopic peaks in their spectra, or had a total intensity (sum of all peaks) of less than 0.5×10^8, were not considered further.

To calculate the $\delta^{13}C$ value for peptides, we compared the experimentally derived isotope peak intensities with simulated isotope peak intensities. For the peptides remaining after filtering, the absolute intensity of each peak (monoisotopic mass, A+1, A+2, etc.) was divided by the summed intensity of all peaks to obtain relative peak intensities. The $^{13}C/^{12}C$ ratio for each peptide was computed as follows. For $^{13}C/^{12}C$ ratios between 0 and 0.1 and fixed (natural) relative abundances of ^{15}N, ^{18}O, ^{17}O, and ^{2}H, the theoretical spectrum (relative intensity of each peak) was computed using a fast Fourier transform modified after Rockwood et al. (13) based on the peptide's molecular formula. In contrast to the approach used by Rockwood et al., we did not model the peaks as Gaussians because we used centroided data. Instead, the theoretical relative peak intensities were computed as described by Rockwood et al. (13), with the Gaussian peak shape function $s(\mu) = 1$:

$$\text{Peptide spectrum} = \bar{F}(m) = \mathbf{IFT}\left(\prod_{E}\left[\mathbf{FT}\left(\overline{RA_E}\right)\right]^{N_E}\right),$$

with **IFT** inverse Fourier transform, **Π** product (for each element C, N, O, and H), **FT** Fourier transform, RA_E relative abundances of the isotopes (e.g., ^{12}C, ^{13}C, and ^{14}C) for each element, and N_E number of atoms in the peptide for each element.

Because not all predicted peaks were detected experimentally, the predicted peak spectrum was truncated by discarding all peaks that were not detected experimentally and the relative peak intensity of the remaining peaks was renormalized, so that the predicted intensities of the remaining peaks was equal to one, as it was for the experimentally detected spectrum. Finally, the goodness of fit between the experimental and theoretical spectrum was calculated as the sum of squares of the difference between observed intensity and predicted intensity for each peak. The goodness of fit was optimized (lowest sum of squares) by decreasing the step size of the $^{13}C/^{12}C$ value from 0.01 to 0.0001 in four iterations. For each species (population) in the sampled community (identified by an accession number prefix for the taxon), the average $^{13}C/^{12}C$ ratio was calculated as the intensity-weighted average of the $^{13}C/^{12}C$ ratio of each peptide attributed to that population. The intensity-weighted average was used because there was a clear correlation between spectral intensity and accuracy of the SIF estimate for individual peptides (*SI Appendix*, Fig. S3). These data were also used to estimate the SE and the $\delta^{13}C$ for each population. To eliminate results from peptides suffering from interference from other peptides (due to overlapping spectra), only results with a sum of squares lower than 0.00005 or those results with a goodness of fit in the upper 33% of the data for that population were used in the calculation of the average $\delta^{13}C$ values.

Correction of direct protein-SIF values with reference standard. To correct for isotope fractionation introduced during sample processing and analysis, mostly by isotope fractionation in the mass spectrometer (17) (see also *Results*), we used human hair as a standard. $\delta^{13}C$ values for the human hair standard were obtained both with direct protein-SIF and CF-EA-IRMS (discussed below) and the offset between the two measurement methods was calculated. The offset value was then used to correct the direct protein-SIF values obtained for the pure culture and for the community samples.

CF-EA-IRMS. Stable isotope ratios of carbon and nitrogen of bacterial strains, human hair, and culture media were determined (Datasets S2 and S3) by CF-EA-IRMS. Frozen cell pellets of bacterial strains were dried at 105 °C and stored at 4 °C. Dried cell pellets as well as carbon and nitrogen sources used for cultivation were weighed (~1 mg per sample) into tin capsules and stored in a desiccator. $\delta^{13}C$ and $\delta^{15}N$ values were measured using CF-EA-IRMS on a Delta V Plus (Thermo) mass spectrometer coupled to an ECS 4010 elemental analyzer (Costech). A Zero Blank autosampler (Costech) was used for dropping the samples onto a quartz tube combustion column heated to 1,000 °C. Simultaneously with sample dropping, an O_2 pulse was injected to flash-combust the samples. Ultra-high-purity helium was used as a carrier gas to transport the resulting gases through a water trap and a reduction furnace (600 °C) in which NOx species were reduced to N_2. The generated N_2 and CO_2 were separated on a GC column. A Finnigan ConFlo III (Thermo) interface was used to deliver the gases to the ion source of the mass spectrometer. Peak areas for the isotopic species of sample CO_2 and N_2 were compared with those of reference gas peaks to determine the $\delta^{13}C$ and $\delta^{15}N$ values of the samples. International standards (Dataset S3) were used at the beginning and end of the measurement sequence to normalize the measurements to the internationally accepted delta (δ) scale with respect to V-PDB for $\delta^{13}C$ and atmospheric N_2 for $\delta^{15}N$. Additionally, caffeine (C-0750; Sigma-Aldrich) was included at the beginning and gelatin (G-9382; Sigma-Aldrich) after each fifth sample as internal laboratory standards. The measurement uncertainty was equal or better than ±0.15‰ for $\delta^{13}C$ and $\delta^{15}N$ values determined by IRMS.

Data and Software Availability. The Calis-p software was implemented in Java and is freely available for download, use, and modification at https://sourceforge.net/projects/calis-p/. The MS proteomics data and the protein sequence databases have been deposited to the ProteomeXchange Consortium (40) via the PRIDE partner repository for the pure culture data with the dataset identifier PXD006762, dataset identifier PXD006118 for the mock community data published previously in Kleiner et al. (19), and the *O. algarvensis* case study data with dataset identifier PXD007510.

ACKNOWLEDGMENTS. We thank Jianwei Chen, Emmo Hamann, Marc Mussmann, Jessica Kozlowski, Sean Booth, Jessica Duong, Johanna Voordouw, Kenneth Sanderson, JoongJae Kim, Joenel Alcantara, Anupama P. Halmillawewa, Michael F. Hynes, and Heidi Gibson for donations of cultures for the mock communities; Christine E. Sharp for help with bacterial stock collection; Jackie Zorz, Maryam Ataeian, and Angela Kouris for help with proteomics sample preparation; Angela Kouris and Eric Miller for grammar and spell-checking of the manuscript; Christian Lott for providing the *O. algarvensis* image used in Fig. 4; Steve Taylor and his team at the University of Calgary Stable Isotope Laboratory for measurements using isotope ratio MS; and two anonymous reviewers for thoughtful comments and feedback. Sequencing of new *O. algarvensis* metagenomes was conducted by the US Department of Energy Joint Genome Institute and is supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231. This study was supported by the Campus Alberta Innovation Chair Program (M.S. and X.D.), the Canadian Foundation for Innovation (M.S.), the North Carolina State Chancellor's Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities (M.K.), a DAAD fellowship by the German Academic Exchange Service (to T.H.), and the Natural Sciences and Engineering Research Council (NSERC) of Canada through a Banting fellowship (to M.K.), an NSERC undergraduate student research award (to E.T.), and a NSERC Discovery grant (to M.S. and B.M.). We acknowledge the support of the International Microbiome Center as well as funding support from the Government of Canada through Genome Canada, the Government of Alberta through Genome Alberta, Genome Prairie, Research Manitoba, and Genome Quebec. This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund.

1. Coplen TB, et al. (2006) New guidelines for $\delta^{13}C$ measurements. *Anal Chem* 78: 2439–2441.
2. Pearson A (2010) Pathways of carbon assimilation and their impact on organic matter values $\delta^{13}C$. *Handbook of Hydrocarbon and Lipid Microbiology*, ed Timmis KN (Springer, Berlin), pp 143–156.
3. DeNiro MJ, Epstein S (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochim Cosmochim Acta* 42:495–506.
4. Orphan VJ, House CH, Hinrichs K-U, McKeegan KD, DeLong EF (2001) Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* 293:484–487.
5. Hayes JM (2001) Fractionation of carbon and hydrogen isotopes in biosynthetic processes. *Rev Mineral Geochem* 43:225–277.
6. O'Leary MH, Madhavan S, Paneth P (1992) Physical and chemical basis of carbon isotope fractionation in plants. *Plant Cell Environ* 15:1099–1104.
7. Suzuki Y, et al. (2006) Host-symbiont relationships in hydrothermal vent gastropods of the genus *Alviniconcha* from the Southwest Pacific. *Appl Environ Microbiol* 72:1388–1393.
8. Beinart RA, et al. (2012) Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proc Natl Acad Sci USA* 109:E3241–E3250.
9. Zhang CL (2002) Stable carbon isotopes of lipid biomarkers: Analysis of metabolites and metabolic fates of environmental microorganisms. *Curr Opin Biotechnol* 13:25–30.
10. Mohr W, Tang T, Sattin SR, Bovee RJ, Pearson A (2014) Protein stable isotope fingerprinting: Multidimensional protein chromatography coupled to stable isotope-ratio mass spectrometry. *Anal Chem* 86:8514–8520.
11. Woebken D, et al. (2015) Revisiting N_2 fixation in Guerrero Negro intertidal microbial mats with a functional single-cell approach. *ISME J* 9:485–496.
12. Musat N, et al. (2014) The effect of FISH and CARD-FISH on the isotopic composition of $^{(13)}$C- and $^{(15)}$N-labeled Pseudomonas putida cells measured by nanoSIMS. *Syst Appl Microbiol* 37:267–276.
13. Rockwood AL, Van Orden SL, Smith RD (1995) Rapid calculation of isotope distributions. *Anal Chem* 67:2699–2704.
14. Macko SA, Fogel ML, Hare PE, Hoering TC (1987) Isotopic fractionation of nitrogen and carbon in the synthesis of amino acids by microorganisms. *Chem Geol Isot Geosci Sect* 65:79–92.
15. McMahon KW, Fogel ML, Elsdon TS, Thorrold SR (2010) Carbon isotope fractionation of amino acids in fish muscle reflects biosynthesis and isotopic routing from dietary protein. *J Anim Ecol* 79:1132–1141.
16. Wolyniak CJ, Sacks GL, Pan BS, Brenna JT (2005) Carbon position-specific isotope analysis of alanine and phenylalanine analogues exhibiting nonideal pyrolytic fragmentation. *Anal Chem* 77:1746–1752.
17. Su X, Lu W, Rabinowitz JD (2017) Metabolite spectral accuracy on orbitraps. *Anal Chem* 89:5940–5948.
18. Bay LJ, Chan SH, Walczyk T (2015) Isotope ratio analysis of carbon and nitrogen by elemental analyser continuous flow isotope ratio mass spectrometry (EA-CF-IRMS) without the use of a reference gas. *J Anal At Spectrom* 30:310–314.
19. Kleiner M, et al. (2017) Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun* 8:1558.
20. Kleiner M, Petersen JM, Dubilier N (2012) Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. *Curr Opin Microbiol* 15:621–631.
21. Kleiner M, et al. (2015) Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. *Environ Microbiol* 17:5023–5035.
22. Woyke T, et al. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443:950–955.
23. Wippler J, et al. (2016) Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*. *BMC Genomics* 17:942.
24. McCutchan JH, Lewis WM, Kendall C, McGrath CC (2003) Variation in trophic shift for stable isotope ratios of carbon, nitrogen, and sulfur. *Oikos* 102:378–390.
25. Tiunov AV (2007) [Stable isotopes of carbon and nitrogen in soil ecological studies]. *Izv Akad Nauk Ser Biol* 34:475–489. Russian.
26. Kleiner M, et al. (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci USA* 109:E1173–E1182.
27. Ponnudurai R, et al. (2017) Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *ISME J* 11:463–477.
28. Blair N, et al. (1985) Carbon isotopic fractionation in heterotrophic microbial metabolism. *Appl Environ Microbiol* 50:996–1001.
29. Milo R, Jorgensen P, Moran U, Weber G, Springer M (2010) BioNumbers–The database of key numbers in molecular and cell biology. *Nucleic Acids Res* 38:D750–D753.
30. Ehleringer JR, et al. (2008) Hydrogen and oxygen isotope ratios in human hair are related to geography. *Proc Natl Acad Sci USA* 105:2788–2793.
31. Jehmlich N, Vogt C, Lünsmann V, Richnow HH, von Bergen M (2016) Protein-SIP in environmental studies. *Curr Opin Biotechnol* 41:26–33.

32. Jehmlich N, Schmidt F, von Bergen M, Richnow H-H, Vogt C (2008) Protein-based stable isotope probing (Protein-SIP) reveals active species within anoxic mixed cultures. *ISME J* 2:1122–1133.
33. Pan C, et al. (2011) Quantitative tracking of isotope flows in proteomes of microbial communities. *Mol Cell Proteomics* 10:M110.006049.
34. Fischer CR, Bowen BP, Pan C, Northen TR, Banfield JF (2013) Stable-isotope probing reveals that hydrogen isotope fractionation in proteins and lipids in a microbial community are different and species-specific. *ACS Chem Biol* 8:1755–1763.
35. Ipsen A (2014) Efficient calculation of exact fine structure isotope patterns via the multidimensional Fourier transform. *Anal Chem* 86:5316–5322.
36. Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6:359–362.
37. Petersen JM, et al. (2016) Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat Microbiol* 2:16195.
38. Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.
39. Chambers MC, et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 30:918–920.
40. Vizcaíno JA, et al. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 32:223–226.