Seasonal exometabolites are regulated by essential microbial metabolisms in the oligotrophic ocean

Erin L. McParland[1*], Fabian Wittmers[2,3], Luis M. Bolaños[4], Craig A. Carlson[5], Ruth Curry[6] Stephen J. Giovannoni[7], Michelle Michelsen[4], Rachel J. Parsons[6], Melissa C. Kido Soule[1], Gretchen J. Swarr[1], Ben Temperton[4], Kevin Vergin[8], Alexandra Z. Worden[3], Krista Longnecker[1], Elizabeth B. Kujawinski[1]

1 Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

2 Ocean EcoSystems Biology Unit, RD3, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

3 Marine Biological Laboratory, Woods Hole, MA, USA

4 School of Biosciences, University of Exeter, Exeter, United Kingdom

5 Department of Ecology, Evolution, and Marine Biology, Marine Science Institute, University of California, Santa Barbara, Santa Barbara, CA, USA

6 Bermuda Institute of Ocean Sciences, Saint George's, Bermuda

7 Department of Microbiology, Oregon State University, Corvallis, OR, 97331, USA

8 Microbial DNA Analytics, Phoenix, OR, USA

[*] Corresponding author email: *emcparland@whoi.edu*

Author contributions: E.L.M., C.A.C., K.L., E.B.K. designed research; L.M.B., S.J.G., M.M., B.T., K.L., E.B.K. contributed to field sampling of exometabolome, ASVs, or metagenomes; E.L.M., L.M.B., M.M., R.J.P., M.C.K., G.J.S., B.T. contributed to sample processing of exometabolome, ASVs, or metagenomes; E.L.M., F.W., L.M.B., C.A.C., R.C., S.J.G., R.J.P., M.C.K., B.T., K.V., A.Z.W., K.L. contributed to wavelet data analysis or metagenome queries; E.L.M. drafted the original manuscript and all authors contributed to manuscript editing and improvements.

This file includes: Main text, Figure captions, and Supplementary Materials. PDF contains Figures, Supplementary Figures, and Supplementary Tables

Abstract (max 250 words)

Predictions of how the biogeochemical reservoir of marine dissolved organic matter (DOM) will respond to future ocean changes are limited by bulk characterizations that mask a complex network of thousands of individual microbe-molecule interactions. The marine microbial community regulates the transformation and fate of organic carbon by utilizing DOM molecules as the products and reactants of their diverse enzymatic functions. Linking these molecules with the microbial taxa responsible for their utilization requires characterization of both microbial activity and the resulting metabolic footprint. Here we present a time series of the seasonal dynamics of both the exometabolome and the bacterioplankton community at the Bermuda Atlantic Time series Study (BATS) site and show that metabolic functions are greater predictors of DOM composition than microbial taxonomy. Putative exometabolite identifications (gonyol, glucose 6-sulfate, succinate, and trehalose) indicate that at least a portion of the exometabolome contains rapidly remineralized, labile molecules. We hypothesize that observations of seasonal accumulation of these labile molecules result from decoupled enrichments of microbial production and consumption enzymes. Critically, we found the composition of seasonal DOM features was more stable inter-annually than the microbial community structure. By estimating redundancy in the BATS metagenomes of metabolisms responsible for cycling these molecules, we propose a paradigm whereby microbial metabolisms that are essential, either to all or to a subset of marine microbes, determine DOM composition. The molecular-level characterization of DOM achieved herein greatly enhances possibilities for connecting the mechanisms behind the DOM-microbe network that cycle Earth's largest reservoir of organic carbon.

Significance statement (max 120 words)

Marine dissolved organic matter (DOM) is a major carbon reservoir that serves as a critical control on Earth's climate. Marine DOM production and removal is largely regulated by microbial activity, but a mechanistic understanding of these processes is lacking. By analyzing individual molecules of DOM across a three-year ocean time series with ultrahigh-resolution mass spectrometry, we found the composition of DOM molecules is more stable inter-annually than the microbial community. We suggest that while the identity of marine microbes may change, essential metabolisms within microbial communities are maintained and thus result in less variability of the DOM composition. Therefore, comprehensive and mechanistic knowledge of metabolic functions and pathways will be critical to predicting the response of marine DOM to future climate-driven changes.

Introduction

In the global ocean, thousands of chemically diverse organic molecules are cycled by a rich and diverse microbial community. These interactions regulate the flux and storage of carbon in the biogeochemical cycling of marine dissolved organic matter (DOM) and thus exert critical controls on Earth's climate (1). Both the chemical nature of DOM molecules (e.g., aromaticity, heteroatom content, size) and environmental conditions (e.g., community composition, microbial interactions, nutrient dynamics, temperature) have been proposed to control DOM flux (2–4). Disentangling the contributions of these different controlling mechanisms is key for predicting changes to DOM carbon flux in future oceans. To date, these mechanisms have most often been described by a framework of reactivity that condenses the thousands of DOM molecules into well-defined sub-pools. These range from labile and semi-labile DOM, which represent the most rapidly cycled substrates that sustain the microbial loop, to refractory DOM which represents molecules that evade microbial degradation and sequester carbon in the oceans for thousands of years (1). However, critical insights can be masked by bulk DOM characterization (e.g., Follett et al., 2014). This is especially true for biologically-labile molecules in DOM, which supply the metabolic demand of microbes and represent the greatest organic carbon flux in the ocean, but are found at concentrations below the detection limits of bulk methods (6).

Hidden within the details of bulk DOM are the results of hundreds of thousands of individual microbe-molecule interactions. These interactions form the fabric of the microbial web, supporting the relationships required to fix, exchange, metabolize, and ultimately remineralize carbon within the marine microbiome (7). These cryptic interactions inhibit our ability to predict when DOM composition will force a change in the microbial community or when microbial activity will alter the composition of DOM. Nevertheless, the tight link between marine microbes and DOM molecules would suggest that microbial taxonomy and metabolic functions should be, at least partially, predictive of DOM molecular composition.

Advances in ultrahigh-resolution mass spectrometry have now enabled marine metabolomics to directly detect individual molecules as metabolites, or the metabolic by-products of the microbial community (8–10). However, it is currently impossible to predict how marine microbial activity links to the metabolic imprint on the resulting DOM molecules. We know from previous culture experiments as well as biogeochemical models that in similar environmental settings, microbial community metabolic functions are often more predictable than taxonomic composition (11, 12). This work reflects the fundamental nature of many physiological processes and the genes that encode them, on which is overlayed species diversity created by adaptive traits not easily discerned in metabolic signatures. The influence of microbial assemblage processes on DOM biogeochemistry remains unknown. Recent marine metabolomics studies suggest that a limited number of marine metabolites are conserved across phylogenies as part of core metabolism, while others are taxonomically-specific, down to the strain level (13–15). Improved predictions of DOM cycling require experimental efforts that simultaneously probe both the marine microbial community and DOM molecules.

To address this core challenge within marine biogeochemistry, we analyzed dynamics across a three-year, depth-resolved time series of both dissolved metabolites (exometabolites) and free-living microbial prokaryotes (bacterioplankton) in the oligotrophic northwestern Sargasso Sea at the Bermuda Atlantic Time series Study (BATS) site. We present an unprecedented perspective into the seasonal variability of these two critical components of microbial-DOM interactions. Seasonal environmental changes represent recurring disturbances that induce shifts in microbial assemblage taxonomy and function (16), and thus create a natural perturbation ideal for testing the reproducibility of the resulting seasonal transformations in DOM composition. We assessed the seasonality and variability of both the exometabolome and bacterioplankton time series with wavelet analysis. We show that within the BATS microbial community, the inter-annual variability of bacterioplankton taxonomy is greater than that of exometabolite composition. This suggests that microbial assemblage mechanisms are metabolically redundant so that the resulting DOM biogeochemistry remains consistent. With a targeted investigation of surface BATS metagenomes, we find that there can be a wide range in the degree of metabolic redundancy for functions responsible for producing and consuming exometabolites and still produce consistent inter-annual patterns of exometabolites. Our work suggests that the identity and concentrations of exometabolites available for consumption in the oligotrophic ocean will be determined by the presence of essential metabolisms rather than by the presence of specific microbial taxa.

Results and Discussion

With more than three decades of sustained observations, the large-scale biogeochemical and physical fields of the water column at the BATS site are well-defined (17–20). The BATS site is seasonally oligotrophic with recurring annual patterns of temperature variability and mixing in the epipelagic and upper mesopelagic. In winter and early spring, the system experiences convective mixing, as deep as 200 – 300 m, where inorganic nutrients entrained from depth trigger an annual spring phytoplankton bloom (Steinberg et al. 2001; Lomas et al. 2013). Following the mixing period, a quiescent and stratified period develops in late spring and persists into mid-autumn in which the surface 100 m become highly oligotrophic. These physical dynamics also drive seasonal dynamics of dissolved organic carbon (DOC), which is used as a proxy for bulk DOM, in the top 300 m at BATS (21, 22). During the stratified season (typically May – October), DOM accumulates in the euphotic zone (0 - 120 m). A portion of the seasonally accumulated, residual DOM is redistributed throughout the mixed layer and exported to the upper mesopelagic during deep convective overturn during the mixing season (typically January – March). Following re-stratification, the exported DOC becomes trapped in the mesopelagic, where it is subsequently remineralized by the resident microbial community (Carlson et al., 1994; Steinberg et al., 2001; Hansell and Carlson, 2001b; Lomas et al., 2013; Liu et al., 2022). This physical framework guided our time series sampling to capture the major water column states for the epipelagic and upper mesopelagic zones during all four major seasons (summer stratified, fall transition, winter mixed, and spring transition) (Fig 1A).

Wavelet analysis detects seasonality in the exometabolome and bacterioplankton time series. With the physical framework in mind, we collected a depth-resolved time series of both the untargeted exometabolome, which represents a global overview of all DOM features detectable by solid phase extraction and liquid chromatography coupled to mass spectrometry, and the microbiome, using bacterioplankton V1-V2 16S rRNA gene amplicon sequence variants (ASVs). Parallel samples of DOM features and ASVs were collected bi-monthly for three-years from July 2016 to July 2019 at the surface (1 m), in the mixed layer (40 m), the bottom of the euphotic zone (120 m), and the upper mesopelagic zone (200 m) (Fig 1A) so that each DOM feature or ASV has an associated time series at every sampling depth (e.g. Fig 1B). We detected 7314 DOM molecules, defined by a unique mass-to-charge ratio and retention time, and 3158 bacterioplankton sub-species, defined as ASVs across the time series. Key seasonal taxonomic trends in bacterioplankton succession at the BATS site were previously described (23). Here we analyzed the temporal dynamics of the bacterioplankton dataset to compare with those of the untargeted exometabolome.

Due to the inherent challenges comparing different data types across a time series, we classified the temporal dynamics of DOM features and bacterioplankton ASVs using wavelet analysis to decompose each time series at every sampling depth into the frequency domain (24) (Fig S1). Previous work has used clustering and correlation networks to group unknown DOM features and microbial species by their surface temporal patterns (25). However, wavelet analysis allows us to extract greater temporal insights of individual molecules and ASVs that are concealed by correlation techniques, including the dominant periods (e.g., seasonal, 12 months), the magnitude of change (i.e., amplitude), and the timing of period peaks (e.g., winter or summer). In addition, unlike several other time series approaches, wavelet analysis can extract localized temporal information. This means that a period of interest does not have to occur globally across the entire time series in order to be detected (Martin-Platero et al. 2018), and thus wavelet analysis can be valuable for detecting inter-annual differences in patterns of plankton communities (27). Here we compare the highly resolved time series of the untargeted exometabolome and bacterioplankton using wavelet analysis to analyze the spatial and temporal dynamics resulting from the complex microbe-DOM interactions network.

The dominant period of a DOM feature or bacterioplankton ASV within the time series was assigned based on the highest median power, an estimate of best fit, across all calculated periods (2-12 months) (Fig S1B) that was statistically significant from a null hypothesis test of 'no periodicity' (median p-value ≤ 0.01) (Fig S2). Significant wavelets were found for 78% of DOM features (n = 5693) and 67% of ASVs (n = 2102) across the four sampling depths. The median powers ranged from 0.32 to 1.4, and dominant periods ranged from 5 to 12 months (Fig S2). Almost all significant wavelets exhibited periods greater than 6 months, indicating that DOM features and bacterioplankton with shorter frequency periods were more stochastic and too similar to random white noise to be significant (Fig S2). Higher-resolution sampling and a longer time series would better detect patterns with shorter frequencies. At almost every sampling depth, a

majority of DOM features and ASVs exhibited a dominant period of 12 months, underscoring the important influence of seasonal environmental conditions (Fig S3).

Exometabolites are differentiated by depth and season.

A seasonal period of 12 months emerged as the dominant period across our time series of thousands of unknown DOM features that comprise bulk DOM, and thousands of bacterioplankton ASVs that are, at least in part, responsible for the cycling of these molecules. A total of 2826 unique DOM features (39% of all) exhibited seasonality at one or more sampling depths (Fig 1C). Approximately two-fold more DOM features exhibited seasonality at 1 m and 40 m compared to DOM features at 120 m and 200 m (n = 1596, 1634, 742, and 723 seasonal DOM features, respectively). A total of 1385 unique ASVs (44% of all) exhibited seasonality at one or more sampling depths, and the greatest number of seasonal bacterioplankton was found at 200 m (n = 576, 291, 313, and 698 seasonal bacterioplankton at 1, 40, 120, and 200 m, respectively) (Fig 1C). To predict the season in which a seasonal DOM feature or bacterioplankton ASV time series reached a maximum, wavelets were reconstructed with a period of 12 months and the season was assigned based on the month in which the maximum occurred (Fig S1C). Most seasonal DOM features (~42-55%) and seasonal ASVs (83-95%) peaked in the summer stratified season at every sampling depth (Fig S4). The stratified and oligotrophic periods encompass a large portion of the annual physical/chemical regimes at BATS (Fig 1A), increasing the chances that our sampling would capture DOM features that exhibited a maximum during this period of elevated bulk DOM. However, we also found seasonal DOM features and bacterioplankton ASVs at each sampling depth in the time series that peaked in the other seasons (fall transition, winter mixed, and spring transition) (Fig S4). The fewest seasonal DOM features peaked in the spring, though this is likely because more frequent sampling is required to capture the short-lived spring transition (Fig 1A). Bulk DOC exhibits a consistent seasonal cycle at BATS (Hansell and Carlson, 2001; Carlson et al. 2004), and by examining individual molecules we find molecular patterns that reflect this bulk signal, but also those that result from independent mechanisms. For example, a large majority of seasonal DOM features at 1 m (Fig S4) peaked during the summer stratified season, but an almost equal number of DOM features at this sampling depth exhibited peaks during the winter mixed season (e.g., Fig 1B). These peaks in winter correspond with the lowest DOC concentrations and thus represent new seasonal molecular signatures not detected with bulk methods.

DOM export can contribute significantly to carbon export in the subtropical oligotrophic ocean where deep convective mixing or subduction occurs (Carlson et al. 1994, Hansell et al., 2009). For this reason, connectivity of DOM features that exhibited seasonality at more than one sampling depth was of interest, particularly features that exhibited seasonality in both the surface (1 m or 40 m) and the deeper (120 m or 200 m) sampling depths as these exometabolites could comprise a portion of DOC export. For example, the DOM feature FT01372 exhibited significant seasonality at both 40 m and 120 m (Fig 1B). However, the composition of the exometabolome was overall strongly vertically stratified and the transfer of seasonal DOM features between depths was limited (Fig 1C). The majority of the seasonal DOM features (n = 1501) exhibited seasonality at only one sampling depth (n = 502, 519, 215, and 265 seasonal DOM features were unique to the sampling

depths of 1, 40, 120, and 200 m, respectively; first four rows in Fig 1C). A smaller subset of seasonal DOM features exhibited connectivity across depths, where the same DOM feature exhibited seasonality at two (n = 893; rows 5-10 in Fig 1C), three (n = 370; rows 11-14 in Fig 1C), or all four (n = 81; row 15 in Fig 1C) sampling depths. The greatest connectivity was found in DOM features exhibiting seasonality at both 1 m and 40 m (n = 556; fifth row in Fig 1C). Connectivity between the surface and deeper depths was minimal. Similar to the exometabolome, 85% of seasonal bacterioplankton ASVs (n = 862) exhibited seasonality in distinct depth zones (Fig 1C). A small subset of seasonal ASVs exhibited seasonality at two (n = 144) or three (n = 6) major depth zones. In contrast to DOM features, a subset of bacterioplankton (n = 233; row 8 in Fig 1C) peaked in the stratified season at 1 m and 200 m, indicative of some connectivity between bacterioplankton in the surface and the upper mesopelagic. Although depth is known to structure the ocean's microbiome diversity (29), the influence on the resulting exometabolome's spatiotemporal dynamics was previously unknown. Our individual time series of thousands of DOM features detected temporal dynamics that complement trends in bulk DOC variability, but also independent dynamics that were unique to the exometabolome.

DOC is removed from the surface at BATS by physical export below the euphotic zone (~120 m) and into the upper mesopelagic (~200 m) during convective winter mixing. While this mechanism is apparent in bulk DOC concentrations and characterized polymers, i.e., total hydrolyzable amino acids (22), exometabolites were lacking in connectivity between our upper and lower sampling depths (Fig 1C). If a DOM feature persists in the surface and becomes entrained in the upper mesopelagic by physical export, its seasonal peaks would be expected to lag in timing, meaning the feature would peak in the surface during summer stratification and subsequently in the deep during winter convective mixing. However, of the few DOM features that exhibited connectivity, most reached maxima in the same stratified season at both depths (Fig S4). This would suggest that these seasonal DOM features were not persistent (recalcitrant) features redistributed by convective mixing, but rather introduced as by-products of common metabolisms across sampling depths or other mechanisms such as sinking particle solubilization or the vertical-migrating mesozooplankton shuttle (Steinberg et al, 2000, Maas et al., 2020). The limited connectivity may in part be explained by our Eulerian sampling approach and the inter-annual variability of the extent and duration of stratification and deep convection in our three-year time series. For example, the mixed layer extended deeper than 200 m in April 2017 and March 2019, but only to 174 m in March 2018 (Fig 1A). The different hydrographic conditions could have altered seasonal DOM compositions at 200 m inconsistently between years. This may partially explain why 200 m was the only sampling depth where wavelets of DOM features predominantly had a period of 8 months, rather than 12 months (Fig S3). However, the wavelet analysis is capable of detecting significant localized periods, and therefore if a DOM feature exhibited seasonality in only a portion of the time-series, the wavelet would have captured these trends. This suggests that the lack of connectivity is a true attribute of the exometabolome (Fig 1C). DOM export is likely best captured by bulk concentrations, while the concentrations of any individual exported DOM features may be diluted below detection during deep mixing. Instead of recapitulating bulk DOM dynamics, we

posit that the exometabolome reflects a unique subset of the DOM reservoir; specifically, the dissolved molecules that are cycled too quickly to be passively exported.

The stratified depth differentiation of seasonal DOM features and their unique seasonal patterns independent of bulk DOM dynamics support the hypothesis that the exometabolome captures the rapid metabolic rate processes of the microbial communities rather than the production of persistent molecules. Microbial taxa are known to be vertically stratified in the oligotrophic ocean and this structure presumably results from niche partitioning among microbial specialists along vertical nutrient and energy gradients (30–32). Culture-based work demonstrates that marine microbes release specific exometabolites that are distinct to specific phylogenetic groups and strains (15, 33). The exometabolome time series at BATS provides some of the first *in situ* evidence that vertical differentiation of the microbial community also promotes stratification of metabolic by-products.

Seasonal exometabolite patterns break away from bulk DOM.

In order to further differentiate the seasonality of distinct DOM features, we compared the magnitudes of temporal change by calculating the amplitudes of the predicted wavelets in each time series (Fig S1C). The amplitudes of seasonal DOM features ranged from 0.1 to 2.6 (0.7 ± 0.4, mean ± sd), where an amplitude of 1 reflects a seasonal peak intensity that is 2-fold higher than the seasonal mean. Seasonal DOM features with small, dampened amplitudes (≤ 0.5) were classified as static molecules. Based on the maximum observed amplitude across depths, most seasonal DOM features (n = 1234) exhibited a static amplitude (~42% of seasonal DOM features at each sampling depth). DOM features with large amplitudes (≥ 1) were classified as dynamic. A smaller, but still significant, number of seasonal DOM features (n = 613) were dynamic (~21% of seasonal DOM features at each sampling depth). In almost all cases, the dynamic DOM features were seasonally removed below the detection limits of our analytical method, and then re-emerged in the DOM pool, accumulating in their respective seasonal peak six months later (e.g., Fig 1C). The temporal patterns further differentiate the seasonal DOM features, but their identities remain unknown.

Untargeted exometabolomics techniques provide an opportunity to highlight important, but previously unrecognized, DOM molecules not detected with targeted techniques. Although identification of environmental metabolites is a notoriously challenging endeavor (34), we highlight four putatively identified (see Methods) seasonal DOM features of interest that are presumably labile molecules based on structure and potential availability for microbial metabolism: gonyol, glucose 6-sulfate (or the isomer galactose 6-sulfate), trehalose, and succinate (Fig 2, Table S*1*). All four exometabolites exhibited dynamic seasonality (amplitude ≥ 1) in the surface and were predicted to peak in the summer stratified season. Based on structures and existing literature, these four small exometabolites are expected to be rapidly metabolized. For example, gonyol has been shown to be rapidly consumed in 24 hours by the Alphaproteobacteria *Ruegeria pomeroyi* (35). Succinate is a widely used metabolite that acts as a crucial intermediate in the tricarboxylic acid cycle (see Supplementary Text for further discussion of potential microbial interactions). Thus, these exometabolites have the potential to play an important role in

the ocean's microbe-DOM network despite a paucity of *in situ* observations. In addition, previous work with structurally similar molecules would suggest that these four molecules have low extraction efficiencies (~ <1%) when using solid phase extraction (36) and, therefore must be present at high concentrations (potentially >100 nM) to be observable in this time series. Future efforts using novel extraction techniques (e.g., Widner et al. 2021; Sacks et al. 2022) will help elucidate these exometabolites' absolute concentrations. The composition of most seasonal DOM features observed here remains unknown. Indeed, solid phase extraction is known to select more recalcitrant-like properties in the context of bulk DOM (39), but extraction efficiencies of individual exometabolites are not easily predictable (Johnson et al. 2017). Therefore, exometabolomes collected using solid-phase extraction contain a vast range of molecular compositions and individual components have the potential to be remineralized quickly. This has also been demonstrated in previous research (40–43).

The mechanisms controlling bulk DOM recycling have traditionally been defined by a spectrum of turnover times and reactivities, from the most labile and reactive DOM to the most recalcitrant and long-lived DOM (1). In the context of the bulk DOM framework, seasonal DOM features in this time series would be classified as semi-labile or semi-refractory because of their apparent seasonal accumulation (44). However, our putative identifications indicate the presence of labile molecules within this seasonally-accumulating pool. This summer accumulation has been observed at BATS for other similarly small and easily metabolized molecules detected with targeted methods (43, 45), which can have rapid half-lives of ~24 hours (46). Signatures of these individual DOM molecules are likely lost within the μM resolution of bulk DOC methods. When tracking individual DOM molecules, it should be expected that these molecules will diverge from the bulk DOM framework and that observed patterns of exometabolites result from alternative mechanisms.

Based on the variable amplitudes, timing of seasonal peaks, and depth differentiation of seasonal DOM features, the BATS exometabolome captured molecules with abiotic and biotic control mechanisms that are likely unique to each molecule. Conservative dilution in the surface and physical export below the euphotic zone may produce the observed static and dynamic exometabolite patterns (21, 22), although this would require an input mechanism (either biotic or abiotic, e.g., photochemical) to return concentrations to their seasonal peaks. The persistence of these molecules may result from inherent recalcitrance or environmental conditions that do not support microbial communities capable of accessing the exometabolites. However, based on the strong vertical stratification of the exometabolome, we propose that many of the observed patterns are the result of biotic controls due to rapid microbial turnover rates that are altered by changes in seasonal environmental conditions and community composition. Static DOM features would result from synchronous microbial processes where production and consumption are at an equilibrium and maintain stable concentrations of these molecules with limited amplitudes of change. Alternatively, dynamic DOM features would result from asynchronous processes where production and consumption are decoupled allowing for the apparent seasonal accumulation. This

353 study provides a new framework for individual DOM-microbe interactions, which could be tested
354 in future work.

355 **Exometabolome composition is more stable inter-annually than microbial assemblages.**
356 During the three-year time series, we tested whether the observed temporal dynamics of the
357 exometabolome and bacterioplankton community remain consistent between years. We assessed
358 inter-annual variability using the wavelet's median power (Fig S1), which reflects its fit and thus
359 provides insight into the predictability of a given time series (Fig 3). A high median power
360 indicates the time series fits well to the seasonal wavelet. A low median power can be driven by a
361 poor fit and/or a signal that only appears in a portion of the three-year time series. Across all
362 sampling depths, the range of median powers of the seasonal DOM features and bacterioplankton
363 were similar (0.3 - 1.3). However, the distributions of median powers were significantly different
364 for DOM features than for bacterioplankton (Kolmogorov-Smirnov test $D = 0.2$, $p < 2e\text{-}16$) (Fig
365 3). Seasonal DOM features had an average median power of 0.6 ± 0.2 (mean ± std dev), which
366 was 0.1 units greater than the average median power of bacterioplankton (0.5 ± 0.1), indicating
367 that DOM features exhibit stronger recurring patterns between years at BATS.

368 The magnitude of richness between the DOM features and bacterioplankton ASVs being compared
369 are similar. However, we also compared the distributions of median powers with the time series
370 of ASVs condensed to node-resolved taxonomic resolution based on phylogenetic relationships
371 (Vergin et al. 2013a) to confirm that the inter-annual variability observed for the bacterioplankton
372 was not inflated by an overrepresentation of rare taxa in ASV microdiversity. Even after
373 condensing the highly resolved ASVs, the same difference in the distribution of median powers
374 was observed (Fig 3). This significant difference would suggest that distinct microbial taxa, not
375 just slightly different ASVs, are driving the inter-annual changes observed across the
376 bacterioplankton time-series.

377 The composition of seasonal DOM features remained statistically stable across the three years
378 despite changes in the bacterioplankton community (Fig 3). Other time series studies have found
379 that individual microbial taxa can vary between years (47, 48), but the resulting feedback on the
380 exometabolome composition was previously unknown. Here we show that taxonomic differences
381 of ASVs and their 16S rRNA gene are too highly-resolved to leave a discernible imprint on the
382 exometabolome composition.

383 **Essential metabolisms support a stable exometabolome.**
384 We hypothesized that recurring patterns of the same DOM features in the exometabolome result
385 from a bacterioplankton community that exhibits taxonomic variability across years but
386 convergent functionality for some key metabolisms. Marine microbial communities are extremely
387 diverse, but also share core groups of genes that confer the same functions (29). This metabolic
388 redundancy is thought to create a buffering capacity of microbial ecosystem functions when
389 communities change (49, 50). To test our hypothesis, we quantified the metabolic redundancy of
390 metabolisms related to the recycling of trehalose and succinate. Both succinate and trehalose
391 exhibited significant seasonality in the BATS exometabolome. However, the metabolisms

392 responsible for the production and consumption of these two exometabolites are expected to differ
393 across the microbial community. Succinate is broadly used as part of the tricarboxylic acid (TCA)
394 cycle, whereas trehalose is used more narrowly as a carbon substrate or for energy storage (51).

395 Metabolic transformations encoded by enzymes that require the identified exometabolites as
396 products or reactants were searched in metagenomes to assess the redundancy of succinate and
397 trehalose metabolisms (Table S*2*). Thus, here we are referring to metabolic redundancy as the
398 degree to which an enzyme, as detected by the same protein sequence, is present across a
399 metagenome. Proteins were surveyed using functional orthologs (KOs) in 22 years (1997 - 2019)
400 of all known publicly available surface ocean BATS metagenomes (n = 28 samples) (Table S3).
401 These samples were not uniformly collected but they capture all four seasons of the BATS physical
402 framework (n = 12 samples in summer stratified, 5 samples in fall transition, 9 samples in winter
403 mixed, and 2 samples in spring transition) (Table S3).

404 Most trehalose proteins (n = 6 transformations) and succinate proteins (n = 5 transformations) were
405 present in a majority of the metagenome samples (Fig 4A). In cases where more than one KO
406 exists for a given transformation (Table S3), we present the most common KO (Fig 4A).
407 Taxonomy of these functional proteins was also assessed, and, in general, was differentiated
408 between succinate and trehalose metabolisms (Fig 4A). Proteins encoding for enzymes required to
409 conduct the TCA cycle (succinate-semialdehyde dehydrogenases, succinyl CoA synthetases, and
410 succinate dehydrogenases) were present in all samples, and functional taxonomy reflected large-
411 scale seasonal community shifts. The most annotated taxonomic contributors of succinate proteins
412 were Pelagibacterales (Alphaproteobacteria) and SAR86 (Gammaproteobacteria). As these are
413 abundant groups in the surface ocean at BATS (52–54), it is not surprising that they dominate the
414 taxonomy of proteins required for the widely used TCA cycle. Proteins encoding for enzymes
415 required to metabolize trehalose are also present in a majority of metagenome samples, but the
416 functional taxonomy of these proteins reflected greater pathway-specificity. For example, many
417 different taxa, the most common annotation being Pelagibacterales, utilized the trehalose synthase
418 (K13057) to synthesize trehalose from ADP-glucose and glucose, whereas the phosphatase
419 (K01087) used to convert trehalose-6-phosphate to trehalose was limited to Flavobacteriales
420 UBA10066 and Rhodobacterales. While the presence and taxonomy of functional proteins differ
421 between succinate and trehalose, the resulting *in situ* seasonal patterns of the exometabolites are
422 similar.

423 We used the metric of contribution evenness (CE) to estimate metabolic redundancy in microbial
424 communities. Traditional metrics of functional redundancy, which are based on niche space, are
425 not easily translated for microbial communities. We determined CE by analyzing protein
426 abundances, which serve as proxies for the encoded metabolic functions (55). CE ranges from no
427 redundancy (CE = 0), indicating only one community member in the sample harbors the protein
428 of interest, to absolute redundancy (CE = 1), indicating all community members contribute equally
429 to the protein of interest. As would be expected based on the increase in sequencing power in the
430 last twenty years, sequencing depths varied by orders of magnitude across the different
431 metagenomes. CE accounts for these differences by normalizing KO abundances to total species

richness as estimated by the presence of universal single-copy marker genes, which are assumed to occur once in each genome (Fig S5). The median CE across all samples ranged from 0 to 0.1 for trehalose-related proteins and from 0 to 0.5 for succinate-related proteins (Fig 4B). CE of succinate-related proteins were overall significantly greater than those of trehalose-related proteins (Wilcoxon rank sum test $p \leq 0.01$), indicating greater redundancy in succinate metabolism. Thus, a similar pattern of stability for an exometabolite can result from either common or specialized metabolisms.

Altogether, our data indicate that only a minimum threshold of metabolic redundancy must be met to maintain the stable inter-annual composition of exometabolites. Trehalose proteins exhibited a specific-type redundancy, where trehalose utilization was limited to a narrow portion of the total community (56). In comparison, metabolic redundancy of succinate metabolism was observed globally within the microbial community, and thus displayed a broad-type redundancy which reflects its role in core microbial metabolism. While these differences supported our expectations based on previous literature, ecological theory would suggest that trehalose metabolism is more susceptible to community changes due to the lower metabolic redundancy (50, 55). And yet, despite these differences in metabolic redundancies, trehalose and succinate exhibited similarly stable inter-annual seasonal patterns in the surface ocean (Fig 2). Based on the time series and metagenomic analyses, it appears that the role of a metabolite in an essential metabolism is more important than the degree of metabolic redundancy. Even if the metabolite is essential for only a subset of the members, it will still be regularly available in the labile DOM pool.

In addition to inter-annual differences, both functional taxonomy and calculated CE exhibited seasonal shifts (Fig 4C) that reflect the asynchronous biotic controls of the dynamic DOM features (Fig 2, Fig S1) described above. In roughly half of the proteins assessed, no significant difference in CE was observed between the summer stratified and winter mixed seasons, but there was a general taxonomic shift in the major contributors. For example, CE of the most redundant KOs for succinate (K00135) was not significantly different between seasons, but the dominant functional taxonomy shifted from Pelagibacteraceae (Alphaproteobacteria) in the summer stratified season to Thalassarchaeaceae (Archaea) in the winter mixed season. In contrast, CE of the other proteins were significantly enhanced or suppressed between summer stratified and winter mixed seasons, meaning that the community's total potential to produce or consume the exometabolite changed seasonally (Fig 4C). Both observations could alter microbial production or consumption rates in an asynchronous manner. Changes in rates could result from differences in enzymatic efficiency associated with the taxonomic shifts, or from a change in the total number of taxa capable of interacting with the exometabolite. Through these targeted metagenomic analyses, we observe relationships that support this new framework of mechanisms responsible for the patterns of seasonal exometabolites at BATS.

Conclusions

In an environment that has already experienced 1.2°C of warming (57), it is critical to resolve the role of community functions in determining DOM biogeochemistry in order to predict future

471 changes in the ocean's organic carbon cycle. Understanding how the seed bank of marine microbial
472 diversity translates into a stable, yet similarly diverse, pool of DOM molecules is a key relationship
473 missing in our understanding of carbon cycling in the ocean. This knowledge gap is challenged by
474 the fact that metabolites do not share a common genetic code like that which encodes metabolic
475 functions. Moreover, many different microbes can produce and consume the same DOM molecule.
476 We demonstrate that the metabolic pathways regulating the exometabolome, which are made up
477 of a variety of microbial enzymes that share a common genetic code, could be leveraged by
478 oceanographers to link the exometabolome to microbial producers and consumers.

479 The untargeted exometabolome time series at BATS provided some of the first insights into the
480 variability of DOM molecules on seasonal and inter-annual timescales, in parallel with the
481 microbial community. Despite similar complexities with respect to composition, we found that the
482 mechanisms responsible for driving bacterioplankton taxonomy and exometabolites should be
483 expected to differ, and linking these two datasets requires careful consideration. The results of our
484 analyses are consistent with the perspective that many metabolic functions are shared across
485 diverse phylogenetically related taxa and ecological concepts such as Hubbell's neutral theory (58)
486 that predict variation in species with no corresponding functional variation. This work highlights
487 that in order to predict labile DOM flux, future models will need to represent essential microbial
488 metabolisms, or those pathways which are required for community function and utilized by either
489 all or a portion of the microbial community.

490 Significant taxonomic variability was detected in the bacterioplankton community at BATS during
491 the three-year time series, but the changes were not enough to influence the composition of the
492 resulting DOM biogeochemistry. How much can microbial taxonomy change inter-annually
493 before the presence of these essential metabolisms is altered? This work presents a major advance
494 in our understanding of variability across the individual molecules comprising DOM. However,
495 the seasonal patterns represent snapshots of standing stocks, and future studies that emphasize rate
496 measurements will be essential. Continuing to resolve the influences of metabolic redundancy and
497 essential metabolism of the microbial loop on DOM biogeochemistry will be critical for predicting
498 changes to the ecosystem function of heterotrophic carbon remineralization in future oceans.

499 **Materials and Methods**

500 **Exometabolome sample collection and extraction.**
501 Samples were collected aboard the R/V *Atlantic Explorer* monthly from fixed depths (1 m, 40 m,
502 120 m, 200 m) at or in the vicinity of the Bermuda Atlantic Time series Study (BATS) site from
503 July 2016 to July 2019. During July field campaigns, samples were collected roughly every 6 hours
504 for 72 hours. Samples were contextualized by a physical framework that defines the state of the
505 water column across the seasonal cycle at BATS (Fig 1A) (22, 43). All major seasons were
506 sampled every year, with the exception of the spring transition, which is a short-lived period and
507 was missed in 2017 when it likely occurred between our sampling in April and May 2017. 4L of
508 whole seawater was filtered through a 47mm 0.2µm Omnipore PTFE filter (Millipore, Burlington
509 MA, USA) using a peristaltic pump as described previously (59). The filtrate was acidified to a

510 pH of 2-3 with OmniTrace HCl (ThermoFisher Scientific, Waltham, MA, USA) and extracted via
511 solid phase extraction with styrene-divinylbenzene polymer columns (Bond Elut PPL, Agilent,
512 Santa Clara, CA, USA) as described previously (39, 60). Sample elutions were evaporated to near
513 dryness and reconstituted in Milli-Q water with 22 stable isotope labeled internal standards (SIL-
514 IS) (Table S*4*). A pooled sample was created with an aliquot of every sample.

515 **UHPLC-ESI-MSMS, exometabolite feature processing, and data filtering.**
516 The sample set (n = 374) was randomized across five batches. The pooled sample was used for
517 column conditioning and was also injected after every 5 samples and at the end of each sequence,
518 followed by process blanks and Milli-Q blanks. Batches were run in both positive and negative
519 ionization mode. Chromatography was performed as previously described (41, 59) using an
520 ultrahigh-performance liquid chromatography system (Vanquish UHPLC, Thermo Scientific)
521 coupled with an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific).
522 Detailed instrument parameters are provided in the Supplement.

523 Raw data files were converted to mzML format using msConvert (61) and transferred to a high-
524 performance computing cluster for processing with R (v 4.0.1). XCMS (v 3.10.2) was used for
525 peak-picking via CentWave, retention time alignment via Orbiwarp, and correspondence (62).
526 Parameter details are described in the Supplement. MS1 features were defined by a unique mass-
527 to-charge ratio and retention time. This analysis yielded a table of MS1 feature intensities (n =
528 12106 features in positive mode and n = 12678 features in negative mode) in each sample.
529 Presented intensities are unitless as this integration reflects an integration of all ion counts
530 associated with a given feature's mass-to-charge ratio bounded by the retention time window.
531 CAMERA was used to identify and filter isotopologues and adducts (63). Features were further
532 filtered using Milli-Q blanks, process blanks, and pooled samples as described in the Supplement.
533 Feature intensities were batch corrected using the BatchCorrMetabolomics package (v 0.1.14) with
534 a robust least-squares regression (64). The majority of SIL-IS exhibited an RSD <20% across all
535 injections after batch correction (Table S*4*). The four putatively identified exometabolites (Table
536 S*1*) were originally found via MS/MS matches using feature-based molecular networking in the
537 Global Natural Products Social Molecular Networking infrastructure (65), and are further
538 discussed in the Supplement.

539 **Microbial community, 16S rRNA amplicon sequencing, and data filtering**
540 Samples for 16S V1-V2 amplicon sequence variants (ASVs) were as described in Liu et al. (2022).
541 Only samples collected at 1 m, 40 m, 120 m, and 200 m were presented. Briefly, 4L of seawater
542 were filtered onto 0.2 µm Sterivex and stored at -80°C. DNA was extracted with a phenol-
543 chloroform protocol (Giovannoni et al., 1996). V1-V2 16S rRNA hypervariable region was
544 amplified with primers 27F (5′-AGAGTTTGATCNTGGCTCAG-3′) and 338RPL (5′-
545 GCWGCCWCCCGTAGGWGT-3′). Amplicon libraries were built using the Nextera XT Index
546 Kit (Illumina Inc.) and sequenced using the Illumina MiSeq platform (reagent kit v.2; 2×250 PE)
547 at the Center for Quantitative Life Sciences (CQLS), Oregon State University. Raw amplicon
548 datasets were processed as in Bolaños et al. (2022) using Dada2 v1.18 (Callahan et al., 2017) with
549 the following filtering parameters: maxEE=(2,2), truncQ=2, minLen=190, truncLen= 220,190,

maxN=0. Samples from the same sequencing run were processed together to accurately estimate the error frequency. Potential chimeras were removed with the removeBimeraDenovo command. Taxonomic assignment was performed with the assignTaxonomy command and the Silva non redundant database V.123 (Quast et al., 2013). Generated ASV and taxonomic tables were analyzed using phyloseq v1.34 (McMurdie et al., 2013). ASVs were presented as relative abundances, normalized to the total counts of all ASVs in a respective sample. ASVs were filtered to remove singletons and any ASV that was found in ≤ 5% of all samples. This yielded an ASV table with n = 3158 taxa. The terminal node collapse of the ASVs was conducted via PhyloAssigner with a global reference tree, resulting in a table of n = 1806 taxa (54).

Wavelet analysis

Wavelet analysis was used to decompose the exometabolome and ASV time series using the R package WaveletComp (24) (Fig S1). Most of the time series was sampled in odd months, except for samples collected in April. Wavelet analysis requires a uniform grid so we interpolated between months to create a monthly time series that allowed us to utilize the April data. This also avoided any distortion to the wavelet analysis which is sensitive to time series length. Significance was assessed with the null hypothesis of white noise and 1000 permutations were calculated for each time series. Similar trends were observed for the exometabolome in both ionization modes, and therefore only positive mode results are presented.

Metabolic redundancy and functional taxonomy

HMMER (v 3.3.1, hmmer.org) searches were conducted with HMM profiles previously created by KofamScan (66). KO numbers were collected based on analysis of KEGG Pathways (67) to find key enzymatic reactions required to conduct pathways related to the production or consumption of trehalose and succinate (Table S*2*). Multiple KOs can encode for the same metabolic transformation, and for brevity we present the most redundant KO only (Fig 4). In addition, single copy marker genes (SCMG) were also searched to estimate sample richness (K01409, K01869, K01873, K01875, K01883, K01887, K01889, K03106, K03110, K06942). Publicly available surface sample metagenomes collected at BATS were queried from 1997 – 2019, though with non-uniform sampling (Table S3). Metabolic KO HMM results were filtered with an e-value of 1×10^{-10}, and SCMG KOs were filtered based on threshold scores defined by KofamScan. Samples richness was calculated based on the number of contigs assigned to a SCMG. The taxonomy of metabolic KO genes was assigned using the contig level taxonomy annotations from MDMcleaner (v 0.8.2) 'clean' output with '—fast_run' settings (68). Presence was calculated as the number of contigs assigned to a metabolic KO. Contribution evenness was calculated as an estimate of metabolic redundancy as defined by (55).

Data presentation

All figures were created with ggplot2 (v 3.4.3) and curated with Inkscape (v 1.2.2).

<u>Data availability statement</u>

Metabolomics data, including raw files, mzML files, and feature tables, are deposited at MetaboLights under study accession number MTBLS5228. 16S amplicon sequences are

deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under project number PRJNA769790. Publicly available metagenomes were accessed from NCBI SRA project number PRJNA385855 (69) and newly deposited historical metagenomes from NCBI SRA project number PRJNA769790. CTD data are deposited in the Biological and Chemical Oceanography Data Management Office (BCO-DMO) at http://lod.bco-dmo.org/id/dataset/861266 for BIOS-SCOPE cruises, and at http://lod.bco-dmo.org/id/dataset/3782 for BATS cruises.

Code for processing the raw exometabolome data is available at https://github.com/KujawinskiLaboratory/UntargCode. Code for processing the raw amplicon data is available at https://github.com/lbolanos32/NAAMES_2020. Code for PhyloAssigner, analyzing the time series, querying the metagenomes, and calculating metabolic redundancy is available under git project https://github.com/BIOS-SCOPE/FunctionalRedundancy

Acknowledgements

We thank Rod Johnson, the BATS technical team, especially Julia Matheson and Paul Lethaby, the BIOS-SCOPE team for their efforts during sample collection, the MAGIC lab for hydrographic data, and the R/V *Atlantic Explorer* officers, technicians, and crew members. We greatly appreciate work by many others to develop and share the publicly available code and data utilized in this study. This project was funded by the Simons Foundation International's BIOS-SCOPE program. McParland was funded by the Woods Hole Oceanographic Institution Postdoctoral Scholar program and the Simons Postdoctoral Fellowship in Marine Microbial Ecology. This is the NSF Center for Chemical Currencies of a Microbial Planet (C-CoMP) publication #037.

832 **Figure Captions**

833 Fig 1: (A) Spatiotemporal coverage of samples collected in the three-year time series for the BATS
834 exometabolome and bacterioplankton community. Samples are colored by the physical framework
835 seasons. The black line reflects mixed layer depth. The blue color indicates bottom of euphotic
836 zone/transition to upper mesopelagic zone. (B) An example time series of an unknown
837 exometabolite (FT01372) at all four sampling depths. The black dots reflect the feature's intensity
838 in all samples. Diel sampling efforts are reflected in the subset of days with multiple points
839 displayed. At 40 m and 120 m, the grey line reflects the significant seasonal trend (12-month
840 period) detected with wavelet analysis at these two depths. (C) Upset plot of seasonal
841 bacterioplankton (ASVs) and exometabolites (DOM features). The top bar plot reflects the total
842 number of seasonal bacterioplankton and exometabolites at each depth. Side panels reflect
843 connectivity, where the middle portion indicates the depth at which seasonality was detected and
844 shading groups the different trends by the number of depths with connectivity. Rows 1-4 represent
845 bacterioplankton and exometabolites that were seasonal at only one sampling depth, rows 5-9
846 represent seasonality at two sampling depths, rows 10-14 represent seasonality at three sampling
847 depths, and row 15 represents seasonality at all four sampling depths. The bars reflect the sum of
848 bacterioplankton (left) or exometabolites (right) that meet these criteria. For example, FT01372
849 (from panel B) is part of the sum of exometabolites in row 9 as it exhibited seasonality at both 40
850 and 120 m.

851 Fig 2: Seasonal patterns across the three-year time series of four putatively identified
852 exometabolites: gonyol, glucose 6-sulfate (or the isomer galactose 6-sulfate), trehalose, and
853 succinate. Feature intensity units are arbitrary (see Methods).

854 Fig 3: Empirical cumulative distribution functions demonstrating the spread of median powers
855 calculated for seasonal exometabolites (green), seasonal bacterioplankton (defined as ASVs)
856 (yellow), and seasonal bacterioplankton nodes condensed by phylogenetic-relatedness (pink). The
857 insets demonstrate two examples of time series with different median powers. A higher median
858 power (green dots) reflects a seasonal wavelet fit that is predictable and exhibits a repeatable
859 dynamic pattern across the three-year time series. A lower median power (yellow dots) is driven
860 by a poor wavelet fit and absences across the three-years.

861 Fig 4: Functional taxonomy and redundancy of enzymes that encode metabolite transformations
862 that require succinate or trehalose as a product or reactant in surface metagenomes at BATS. Rows
863 represent individual metabolite transformations and their associated KO. + indicates production
864 (exometabolite is product). – indicates consumption (exometabolite is reactant). (A) Pie charts are
865 colored by the top three taxonomic contributors to total RPKM (reads per kilobase per million
866 mapped reads) in a given sample. Grey represents all other contributors. Samples are organized by
867 season (summer stratified, fall transition, winter mixed, and spring transition), and are arranged by
868 sample date. An absent pie chart indicates no KO was present in the sample. (B) Metabolic
869 redundancy was estimated based on the metric of contribution evenness (CE) of each gene in each
870 sample. The CE average ± standard deviation across samples is shown. Higher CE values reflects

871 greater metabolic redundancy. (C) CE was compared in summer stratified and winter mixed
872 samples (Wilcoxon rank sum test, $p \leq 0.1$). Sun indicates CE was significantly higher in summer
873 stratified samples, snowflake indicates CE was significantly higher in winter mixed samples, and
874 dash indicates no significant difference.

875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902

References

1. C. A. Carlson, D. A. Hansell, "DOM Sources, Sinks, Reactivity, and Budgets" in *Biogeochemistry of Marine Dissolved Organic Matter: Second Edition*, (2014), pp. 65–126.
2. Y. Wang, *et al.*, Linking Microbial Population Succession and DOM Molecular Changes in Synechococcus -Derived Organic Matter Addition Incubation . *Microbiol Spectr* **10** (2022).
3. E. K. Wear, C. A. Carlson, M. J. Church, Bacterioplankton metabolism of phytoplankton lysates across a cyclone-anticyclone eddy dipole impacts the cycling of semi-labile organic matter in the photic zone. *Limnol Oceanogr*, 1–15 (2020).
4. C. A. Carlson, *et al.*, Interactions among dissolved organic carbon, microbial processes, and community structure in the mesopelagic zone of the northwestern Sargasso Sea. *Limnol Oceanogr* **49**, 1073–1083 (2004).
5. C. L. Follett, D. J. Repeta, D. H. Rothman, L. Xu, C. Santinelli, Hidden cycle of dissolved organic carbon in the deep ocean. *Proceedings of the National Academy of Sciences* **111**, 16706–16711 (2014).
6. M. A. Moran, *et al.*, Microbial metabolites in the marine carbon cycle. *Nat Microbiol* **7**, 508–523 (2022).
7. E. B. Kujawinski, The Impact of Microbial Metabolism on Marine Dissolved Organic Matter. *Ann Rev Mar Sci* **3**, 567–599 (2011).
8. B. Pedler Sherwood, *et al.*, Metabolic characterization of a model heterotrophic bacterium capable of significant chemical alteration of marine dissolved organic matter. *Mar Chem* **177**, 357–365 (2015).
9. E. B. Kujawinski, *et al.*, Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. *Geochim Cosmochim Acta* **73**, 4384–4399 (2009).
10. A. D. Steen, *et al.*, Analytical and Computational Advances, Opportunities, and Challenges in Marine Organic Biogeochemistry in an Era of "Omics." *Front Mar Sci* **7** (2020).
11. J. E. Goldford, *et al.*, Emergent simplicity in microbial community assembly. *Science (1979)* **361**, 469–474 (2018).
12. V. J. Coles, *et al.*, Ocean biogeochemistry modeled with emergent trait-based genomics. *Science (1979)* **358**, 1149–1154 (2017).
13. B. P. Durham, *et al.*, Chemotaxonomic patterns in intracellular metabolites of marine microbial plankton. *Front Mar Sci* **9** (2022).
14. K. R. Heal, *et al.*, Marine Community Metabolomes Carry Fingerprints of Phytoplankton Community Composition. *mSystems* **6** (2021).

15. E. B. Kujawinski, *et al.*, Metabolite diversity among representatives of divergent Prochlorococcus ecotypes. *mSystems* **8** (2023).

16. M. Huelsmann, M. Ackermann, Community instability in the microbial world. *Science (1979)* **378**, 29–30 (2022).

17. A. F. Michaels, A. H. Knap, Overview of the U.S. JGOFS Bermuda Atlantic Time-series Study and the Hydrostation S program. *Deep Sea Research Part II: Topical Studies in Oceanography* **43**, 157–198 (1996).

18. N. R. Bates, R. J. Johnson, Acceleration of ocean warming, salinification, deoxygenation and acidification in the surface subtropical North Atlantic Ocean. *Communications Earth & Environment 2020 1:1* **1**, 1–12 (2020).

19. D. K. Steinberg, *et al.*, Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): A decade-scale look at ocean biology and biogeochemistry. *Deep Sea Res 2 Top Stud Oceanogr* **48**, 1405–1447 (2001).

20. M. W. Lomas, *et al.*, Two decades and counting: 24-years of sustained open ocean biogeochemical measurements in the Sargasso Sea. *Deep Sea Research Part II: Topical Studies in Oceanography* **93**, 16–32 (2013).

21. D. A. Hansell, C. A. Carlson, Biogeochemistry of total organic carbon and nitrogen in the Sargasso Sea: control by convective overturn. *Deep Sea Research Part II: Topical Studies in Oceanography* **48**, 1649–1667 (2001).

22. S. Liu, *et al.*, Linkages Among Dissolved Organic Matter Export, Dissolved Metabolites, and Associated Microbial Community Structure Response in the Northwestern Sargasso Sea on a Seasonal Scale. *Front Microbiol* **13**, 407 (2022).

23. S. J. Giovannoni, K. L. Vergin, Seasonality in ocean microbial communities. *Science (1979)* **335**, 671–676 (2012).

24. A. Roesch, H. Schmidbauer, WaveletComp: Computational Wavelet Analysis. *R package version 1.1.* (2018) (December 5, 2022).

25. J. Merder, *et al.*, Dissolved organic compounds with synchronous dynamics share chemical properties and origin. *Limnol Oceanogr* **66**, 4001–4016 (2021).

26. A. M. Martin-Platero, *et al.*, High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nature Communications 2018 9:1* **9**, 1–11 (2018).

27. L. M. Bolaños, *et al.*, Influence of short and long term processes on SAR11 communities in open ocean and coastal systems. *ISME Communications 2022 2:1* **2**, 1–11 (2022).

28. C. Carlson, H. Ducklow, A. Michaels, Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso Sea. *Nature Letters* **371** (1994).

29. S. Sunagawa, *et al.*, Structure and function of the global ocean microbiome. *Science (1979)* **348** (2015).

30. E. F. DeLong, *et al.*, Community genomics among stratified microbial assemblages in the ocean's interior. *Science (1979)* **311**, 496–503 (2006).

31. R. R. Malmstrom, *et al.*, Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific oceans. *The ISME Journal 2010 4:10* **4**, 1252–1264 (2010).

32. S. J. Giovannoni, M. S. Rappé, K. L. Vergin, N. L. Adair, 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the Green Non-Sulfur bacteria. *Proc Natl Acad Sci U S A* **93**, 7979–7984 (1996).

33. J. W. Becker, *et al.*, Closely related phytoplankton species produce similar suites of dissolved organic matter. *Front Microbiol* **5** (2014).

34. I. Koester, *et al.*, Illuminating the dark metabolome of Pseudo-nitzschia–microbiome associations. *Environ Microbiol* **24**, 5408–5424 (2022).

35. B. Gebser, K. Thume, M. Steinke, G. Pohnert, Phytoplankton-derived zwitterionic gonyol and dimethylsulfonioacetate interfere with microbial dimethylsulfoniopropionate sulfur cycling. *Microbiologyopen* **9**, e1014 (2020).

36. W. M. Johnson, M. C. Kido Soule, E. B. Kujawinski, Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics. *Limnol. Oceanogr.: Methods* **15**, 417–428 (2017).

37. B. Widner, M. C. Kido Soule, F. X. Ferrer-González, M. A. Moran, E. B. Kujawinski, Quantification of Amine- And Alcohol-Containing Metabolites in Saline Samples Using Pre-extraction Benzoyl Chloride Derivatization and Ultrahigh Performance Liquid Chromatography Tandem Mass Spectrometry (UHPLC MS/MS). *Anal Chem* **93**, 4809–4817 (2021).

38. J. S. Sacks, K. R. Heal, A. K. Boysen, L. T. Carlson, A. E. Ingalls, Quantification of dissolved metabolites in environmental samples through cation-exchange solid-phase extraction paired with liquid chromatography–mass spectrometry. *Limnol Oceanogr Methods* **20**, 683–700 (2022).

39. T. Dittmar, B. Koch, N. Hertkorn, G. Kattner, A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol Oceanogr Methods* **6**, 230–235 (2008).

40. C. L. Fiore, K. Longnecker, M. C. Kido Soule, E. B. Kujawinski, Release of Ecologically Relevant Metabolites by the Cyanobacterium, Synechococcus elongatus CCMP 1631. *Environ Microbiol* **17**, n/a-n/a (2015).

41. L. Weber, *et al.*, Extracellular Reef Metabolites Across the Protected Jardines de la Reina, Cuba Reef System. *Front Mar Sci* **7**, 1063 (2020).

42. A. Vorobev, *et al.*, Identifying labile DOM components in a coastal ocean through depleted bacterial transcripts and chemical signals. *Environ Microbiol* **20**, 1–19 (2018).

43. K. Longnecker, *et al.*, Seasonal and daily patterns in known dissolved metabolites in the northwestern Sargasso Sea. *Limnol Oceanogr* **9999**, 1–18 (2024).

44. D. A. Hansell, Recalcitrant dissolved organic carbon fractions. *Ann Rev Mar Sci* **5**, 421–445 (2013).

45. N. M. Levine, *et al.*, Revising upper-ocean sulfur dynamics near Bermuda: New lessons from 3 years of concentration and rate measurements. *Environmental Chemistry* **13**, 302–313 (2015).

46. R. Kiene, L. Linn, The fate of dissolved dimethylsulfoniopropionate (DMSP) in seawater: Tracer studies using 35 S-DMSP. *Geochim Cosmochim Acta* **64**, 2797–2810 (2000).

47. S. Lambert, J.-C. Lozano, F.-Y. Bouget, P. E. Galand, Seasonal marine microorganisms change neighbours under contrasting environmental conditions (2021) https:/doi.org/10.1111/1462-2920.15482 (November 24, 2022).

48. C. S. Ward, *et al.*, Annual community patterns are driven by seasonal switching between closely related marine bacteria. *The ISME Journal 2017 11:6* **11**, 1412–1422 (2017).

49. S. D. Allison, J. B. H. Martiny, Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A* **105 Suppl**, 11512–11519 (2008).

50. S. Louca, P. L, M. Doebeli, Decoupling function and taxonomy in the global ocean microbiome. *Science (1979)* **353**, 1272–1277 (2016).

51. E. L. McParland, H. Alexander, W. M. Johnson, The Osmolyte Ties That Bind: Genomic Insights Into Synthesis and Breakdown of Organic Osmolytes in Marine Microbes. *Front Mar Sci* **8**, 732 (2021).

52. A. H. Treusch, *et al.*, Seasonality and vertical structure of microbial communities in an ocean gyre. *The ISME Journal 2009 3:10* **3**, 1148–1163 (2009).

53. C. A. Carlson, *et al.*, Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* **3**, 283–295 (2009).

54. K. L. Vergin, *et al.*, High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *The ISME Journal 2013 7:7* **7**, 1322–1332 (2013).

55. T. M. Royalty, A. D. Steen, Contribution Evenness: A functional redundancy metric sensitive to trait stability in microbial communities. *bioRxiv* (2021) https:/doi.org/10.1101/2020.04.22.054593 (July 14, 2022).

56. W. H. Cheng, C. H. Hsieh, C. W. Chang, F. K. Shiah, T. Miki, New index of functional specificity to predict the redundancy of ecosystem functions in microbial communities. *FEMS Microbiol Ecol* **98** (2022).

57. N. R. Bates, R. J. Johnson, Acceleration of ocean warming, salinification, deoxygenation and acidification in the surface subtropical North Atlantic Ocean. *Communications Earth & Environment 2020 1:1* **1**, 1–12 (2020).

58. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, 2001).

59. M. C. Kido Soule, K. Longnecker, W. M. Johnson, E. B. Kujawinski, Environmental metabolomics: Analytical strategies. *Mar Chem* **177**, 374–387 (2015).

60. K. Longnecker, Dissolved organic matter in newly formed sea ice and surface seawater. *Geochim Cosmochim Acta* **171**, 39–49 (2015).

61. M. C. Chambers, *et al.*, A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology 2012 30:10* **30**, 918–920 (2012).

62. C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779–787 (2006).

63. C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, S. Neumann, CAMERA: An integrated strategy for compound spectra extraction and annotation of LC/MS data sets. *Anal Chem* **84**, 283 (2012).

64. R. Wehrens, *et al.*, Improved batch correction in untargeted MS-based metabolomics. *Metabolomics* **12** (2016).

65. L. F. Nothias, *et al.*, Feature-based molecular networking in the GNPS analysis environment. *Nature Methods 2020 17:9* **17**, 905–908 (2020).

66. T. Aramaki, *et al.*, KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).

67. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27–30 (2000).

68. J. Vollmers, S. Wiegand, F. Lenk, A. K. Kaster, How clear is our current view on microbial dark matter? (Re-)assessing public MAG & SAG datasets with MDMcleaner. *Nucleic Acids Res* **50**, e76–e76 (2022).

69. S. J. Biller, *et al.*, Data Descriptor : Marine microbial metagenomes sampled across space and time. *Nature Publishing Group*, 1–7 (2018).

Supplement

Supplemental Text 1

UHPLC-ESI-MS/MS performance

Large LC-MS/MS experiments are prone to retention time drift, contamination, and carry over between samples. To mitigate these factors LC-MS sequences were limited to 105 injections (~18 hrs), sample order was randomized, and the ESI probe was cleaned and the column was re-equilibrated between batches.

Separation was performed with a reverse phase Waters Acquity HSS T3 column (2.1 x 100 mm, 1.8 µm), equipped with a Vanguard pre-column. Column temperature was held at 40°C. The column was eluted at 0.5 ml/min with a combination of solvents: A) 0.1% formic acid in water and B) 0.1% formic acid in acetonitrile. The chromatographic gradient was as follows: 1% B (1 min), 15% B (1-3 min), 50% B (3-6 min), 95% B (6-9 min), 95% B (10 min). The column was washed and re-equilibrated with 1% B (2 min) between injections. The autosampler was set to 4°C and injection volumes were 5 µl. The electrospray voltage was set to 2600 V for negative mode and 3600 V for positive mode. The setting for source sheath gas was 55 and auxiliary gas was 20 (presented in arbitrary units). The heated capillary temperature was 350°C and the vaporizer temperature was 400°C. MS data were collected in the Orbitrap analyzer with a mass resolution of 120,000 FWHM at m/z 200. The automatic gain control (AGC) target was $4e^5$, the maximum injection time was 50 ms, and the scan range was 100 – 1000 m/z. Internal mass calibration of the Orbitrap analyzer was used to improve mass accuracy of the MS scan. Data-dependent MS/MS data were acquired in the Orbitrap analyzer using higher energy collisional dissociation (HCD) with a normalized collision energy of 35% and with mass resolution of 7500. The AGC target value for fragmentation spectra was $5e^4$ and the intensity threshold was $2e^4$. Cycle time was set at 0.6 s. Precursor selection was performed within the quadrupole with a 1 m/z isolation window. Dynamic exclusion was enabled, with 3s exclusion duration after n=1. All data were collected in profile mode. Raw data files were converted to mzML format using msConvert (1).

XCMS parameters

Peak-picking was performed using the CentWave algorithm with the following parameters: noise = 100, peak-width = 7-14, ppm = 15, prescan = 3, preintensity = 900, integrate = 2, mzdiff = -0.005, snthresh = 5, extendLengthMSW = TRUE. Peaks were filtered based on Gaussian fits with root mean square errors less than or equal to 0.125. Peaks were further filtered with refineChromPeaks. Any peaks with a maximum peakwidth > 40s were removed (CleanPeaksParam). Any peaks that were artificially split were merged (MergeNeighboringPeaksParam) by binning peaks with a similar m/z (± 10ppm) and retention time (± 4s), and subsequently merging if the intensity half-way between the two peaks was > 75% of the smallest of the two peak's intensities. Retention times were adjusted using Orbiwarp (binSize = 0.1). Pooled samples were used as the subset to define the retention time adjustments, and these alignments were subsequently interpolated and applied to all other samples. Correspondence (bw = 0.7, binSize = 0.025) between the peaks was conducted. Finally,

fillChromPeaks was used to fill in any peaks missed originally by searching each feature's retention time and m/z (± 25ppm). Feature values were integrated by the 'maxint' method.

Feature filtering

Features were first filtered based on results from CAMERA to remove isotopologues and adducts (2). The remaining features were then filtered using Milli-Q and process blanks (3). The mean log abundance across samples and blanks was calculated for each feature, and subsequently binned into 20, 40, 60, and 80th quantiles. For each bin, a threshold was calculated based on the 25th quartile of the difference between the mean log abundances of samples and blanks that were less than 0. The difference for all features in a given bin were required to be greater than the absolute value of this threshold. This data-adaptive method flagged 21% of features in negative mode and 17% of features in positive mode for blank contamination. Features were also filtered based on presence or absence in pooled samples. If a feature was detected in less than 80% of all pooled samples, it was removed. However, this filter surprisingly only identified 3% of features in negative mode and 1% of features in positive mode as low abundance. This results from the XCMS function fillChromPeaks, which fills in 'noise' when no peak can be integrated. The remaining features represent, to the best of our ability, unique molecules (defined by a m/z and retention time), but we note that these datasets will always contain some small portion of unidentifiable isomers, adducts, and isotopologues.

Metabolite identification

All four putative identifications had m/z matches to reference masses within ± 1ppm (Fig S6 - Fig S9). Based on confidence levels defined by the Metabolomics Standards Initiative (4), succinate and trehalose were identified to the highest level possible (Level 1) using authentic standards analyzed with the same analytical platform used to quantify the untargeted exometabolome, gonyol was identified with high confidence (Level 2) based on a match to a GNPS library standard, and glucose 6-sulfate was identified with medium confidence (Level 3) based on a match to a library standard of a similar compound, glucose 6-phosphate, as an authentic standard does not exist to our knowledge. We note that glucose 6-sulfate exhibited a significant cosine similarity with the library standard of glucose 6-phosphate, and the exact mass difference between the two different precursor masses (0.009 m/z) is equal to the expected mass difference between glucose 6-sulfate and glucose 6-phosphate (0.0095 m/z). Additionally, the ultrahigh-resolution of the Orbitrap MS/MS spectrum supports this identification. For example, a dominant MS2 fragment was m/z 96.959 (HSO_4^-), in comparison a phosphate containing fragment would have a mass of m/z 96.969 (H_2PO_4). However, we cannot rule out that the glucose 6-sulfate may instead be the isomer galactose 6-sulfate.

<u>Supplemental Text 2</u>

Here we further discuss the potential role of putatively identified exometabolites in the marine microbial loop. Gonyol is a reduced organic sulfur molecule similar in structure to the well-known metabolite dimethylsulfoniopropionate (DMSP) (5). Gonyol is produced with taxonomic-specificity and can be degraded by marine bacteria, but the genetic pathways responsible for

recycling this exometabolite are not yet known (6–8). As would be expected for a labile molecule, the first quantification of gonyol in the dissolved phase was found at low nM concentrations in the Pacific Ocean (9), and our time series suggests that these concentrations would likely change seasonally. Glucose 6-sulfate is similar to the core metabolite glucose 6-phosphate, where the phosphate group is substituted for an oxidized sulfate group. To our knowledge, this is the first detection of glucose 6-sulfate in the oligotrophic ocean, and its potential sources or sinks remain an open question. Glucose 6-sulfate (or its isomer galactose 6-sulfate) could be a degradation product of presumably abundant, but poorly characterized, sulfated polysaccharides that comprise algal cell walls (10). Little is known about these large biopolymers in DOM, and most knowledge is derived from studies of macroalgae, which would include *Sargassum* at BATS (11). Nevertheless, microalgae and bacteria can also produce sulfated polysaccharides (12). A diverse suite of known sulfatases could be utilized to remove the oxidized sulfate group and consume the remaining monosaccharide (12–14). Many questions remain to be answered about these two organic sulfur molecules, particularly considering past work in phosphorus-limited waters of the North Atlantic Ocean that has demonstrated the substitution of sulfur into core biomolecules (15).

Trehalose and succinate are both metabolites for which little is known about their presence in marine DOM, but their genetic pathways are well-characterized. Trehalose is a sugar that can be easily routed to glycolysis after breaking the disaccharide bond, but has also been shown to be synthesized or retained as an osmolyte (16–19). Succinate is a dicarboxylic acid produced as an intermediate metabolic product of the citric acid cycle and glyoxylate pathway making it a key part of core catabolism and anabolism pathways in marine microbes (20–22).

Supplementary Figure Captions

Fig S1: Example of the time series wavelet analysis for two different DOM features (left and right). (A) The original time series (black) is almost uniformly sampled and contains multiple samples from the same month in some cases. The time series was first transformed by averaging the samples and interpolating if needed across months to create a time series with one sample per month (purple) (n = 37 samples; see Methods). (B) Wavelet analysis is used to detrend the time series as reflected by the resulting power spectrum, where the calculated power (colorbar) for every sample (x-axis) is plotted as a function of every calculated period (1-12 months). The x-axis is the same as that in panel A. The side panels in (B) represent the median power for each period. A higher median power indicates a better wavelet fit. The highest median power was used to assign the dominant period of a time series. In both examples here, the median power is highest for a period of ≥ 11 months and was therefore considered to be a seasonal time series. (C) If the highest significant median power was ≥ 11 months, the wavelet was reconstructed (dashed black line) using a period of 12 months to predict the seasonal maximum and to calculate the amplitude. The month of the maximum value in the reconstructed time series was used to assign the peak season. The amplitude was calculated as the difference between the maximum reconstructed value and the mean of all reconstructed values (solid black line) normalized to the mean. The time series in the left panel represents a dynamic seasonal DOM feature (amplitude ≥ 1) that had a maximum peak

in the summer. The time series in the right represents a static seasonal DOM feature (amplitude ≤ 0.5) that had a maximum peak in the summer.

Fig S2: The median p-value and median power of all wavelets. Vertical grey line represents the median p-value cut-off of 0.01 and the horizontal grey lines represents the resulting minimum possible median power of 0.32. Color reflects the median period (1-12 months).

Fig S3: The percentage of all significant wavelets found in time series with a dominant period of 5-12 months for DOM features (green) and bacterioplankton (yellow) at each sampling depth (1, 40, 120, and 200 m).

Fig S4: Alluvial plots depicting the connectivity of seasonal (A) DOM features and (B) bacterioplankton across sampling depths. Horizontal boxes represent the total number of seasonal DOM features or bacterioplankton, while box width and color reflect the number of features that peaked in a given season at the respective sampling depth. Grey represents a feature that is not seasonal at that depth but becomes seasonal at another depth. The ribbon colors track the connectivity of seasonal DOM features or bacterioplankton at 1m through the water column.

Fig S5: Median richness of all single copy marker gene (SCMG) KOs versus median richness of all metabolic (trehalose and succinate production and consumption) KOs in all surface samples (n = 30) of each assembly queried. Color reflects the assembly name as defined in Table S3. The black line reflects a 1:1 relationship to demonstrate the linear relationship between the two groups of KOs queried, despite the order of magnitude differences in sequencing depth across the 20 years of metagenomic information.

Fig S6: (A) Succinate mirror plot of common MS2 fragments from samples (black) compared to the succinate reference spectrum in GNPS (green). (B) Succinate EIC of samples (blue) compared to standard (black).

Fig S7: (A) Trehalose mirror plot of common MS2 fragments from samples (black) compared to the trehalose reference spectrum in GNPS (green). (B) Trehalose EIC of samples (blue) compared to standard (black) reflecting chromatographic separation of trehalose and its isomer sucrose.

Fig S8: (A) Gonyol mirror plot of common MS2 fragments from samples (black) compared to the gonyol reference spectrum in GNPS (green).

Fig S9: (A) Glucose 6-sulfate (or galactose 6-sulfate) mirror plot of common MS2 fragments from samples (black) compared to the glucose 6-phosphate reference spectrum in GNPS (green). (B) Isotopic fine structure of glucose 6-sulfate highlighting the ^{34}S-containing m/z of 181.069.

<u>Supp Table Captions</u>

Table S1: Feature id, formula, identification level as defined by the Metabolomics Standards Initiative (4), ionization mode, detected m/z, adduct, retention time, and of putatively identified metabolites.

Table S2: List of metabolic KOs queried in all surface metagenomes for trehalose and succinate production or consumption.

Table S3: Metagenome assemblies queried for functional redundancy analyses.

Table S4: A mix of stable isotope labeled internal standards was added to every sample of the exometabolome. The label reflects which element was isotopically heavy. Ng/ml reflects the concentration added to each sample. Ion mode reflects which ionization mode the standard was detected in. The retention time reflects where the standard was detected in the chromatogram. The exact mass was calculated based on monoisotopic element composition. All standards were detected as either [M+H] or [M-H] adducts.

References

1. M. C. Chambers, *et al.*, A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology 2012 30:10* **30**, 918–920 (2012).

2. C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, S. Neumann, CAMERA: An integrated strategy for compound spectra extraction and annotation of LC/MS data sets. *Anal Chem* **84**, 283 (2012).

3. C. Schiffman, *et al.*, Filtering procedures for untargeted LC-MS metabolomics data. *BMC Bioinformatics* **20**, 1–10 (2019).

4. L. W. Sumner, *et al.*, Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).

5. A. Spielmeyer, G. Pohnert, Direct quantification of dimethylsulfoniopropionate (DMSP) with hydrophilic interaction liquid chromatography/mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **878**, 3238–3242 (2010).

6. B. Gebser, G. Pohnert, Synchronized regulation of different zwitterionic metabolites in the osmoadaption of phytoplankton. *Mar Drugs* **11**, 2168–2182 (2013).

7. B. Gebser, K. Thume, M. Steinke, G. Pohnert, Phytoplankton-derived zwitterionic gonyol and dimethylsulfonioacetate interfere with microbial dimethylsulfoniopropionate sulfur cycling. *Microbiologyopen* **9**, e1014 (2020).

8. K. R. Heal, *et al.*, Marine Community Metabolomes Carry Fingerprints of Phytoplankton Community Composition. *mSystems* **6** (2021).

9. J. S. Sacks, K. R. Heal, A. K. Boysen, L. T. Carlson, A. E. Ingalls, Quantification of dissolved metabolites in environmental samples through cation-exchange solid-phase extraction paired with liquid chromatography–mass spectrometry. *Limnol Oceanogr Methods* **20**, 683–700 (2022).

10. C. Arnosti, *et al.*, The Biogeochemistry of Marine Polysaccharides: Sources, Inventories, and Bacterial Drivers of the Carbohydrate Cycle (2021) https:/doi.org/10.1146/annurev-marine-032020 (November 25, 2022).

11. M. Wang, *et al.*, The great Atlantic Sargassum belt. *Science (1979)* **364**, 83–87 (2019).

12. W. Helbert, Marine polysaccharide sulfatases. *Front Mar Sci* **4**, 6 (2017).

13. A. G. Hettle, C. J. Vickers, A. B. Boraston, Sulfatases: Critical Enzymes for Algal Polysaccharide Processing. *Front Plant Sci* **13** (2022).

14. J. H. W. Saw, *et al.*, Pangenomics analysis reveals diversification of enzyme families and niche specialization in globally abundant SAR202 bacteria. *mBio* **11** (2020).

15. B. a S. Van Mooy, G. Rocap, H. F. Fredricks, C. T. Evans, A. H. Devol, Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proc Natl Acad Sci U S A* **103**, 8607–12 (2006).

16. J. C. Argüelles, Physiological roles of trehalose in bacteria and yeasts: a comparative analysis. *Archives of Microbiology 2000 174:4* **174**, 217–224 (2000).

17. N. Pade, J. Compaoré, S. Klähn, L. J. Stal, M. Hagemann, The marine cyanobacterium Crocosphaera watsonii WH8501 synthesizes the compatible solute trehalose by a laterally acquired OtsAB fusion protein. *Environ Microbiol* **14**, 1261–1271 (2012).

18. C. McLean, *et al.*, Harmful Algal Bloom-Forming Organism Responds to Nutrient Stress Distinctly From Model Phytoplankton. *bioRxiv*, 2021.02.08.430350 (2021).

19. A. K. Boysen, *et al.*, Particulate Metabolites and Transcripts Reflect Diel Oscillations of Microbial Activity in the Surface Ocean. *mSystems* **6** (2021).

20. O. Levitan, *et al.*, Remodeling of intermediate metabolism in the diatom Phaeodactylum tricornutum under nitrogen stress. **112** (2015).

21. S. E. Noell, *et al.*, Sar11 cells rely on enzyme multifunctionality to metabolize a range of polyamine compounds. *mBio* **12** (2021).

22. C. Koedooder, *et al.*, The role of the glyoxylate shunt in the acclimation to iron limitation in marine heterotrophic bacteria. *Front Mar Sci* **5**, 435 (2018).

Fig 1

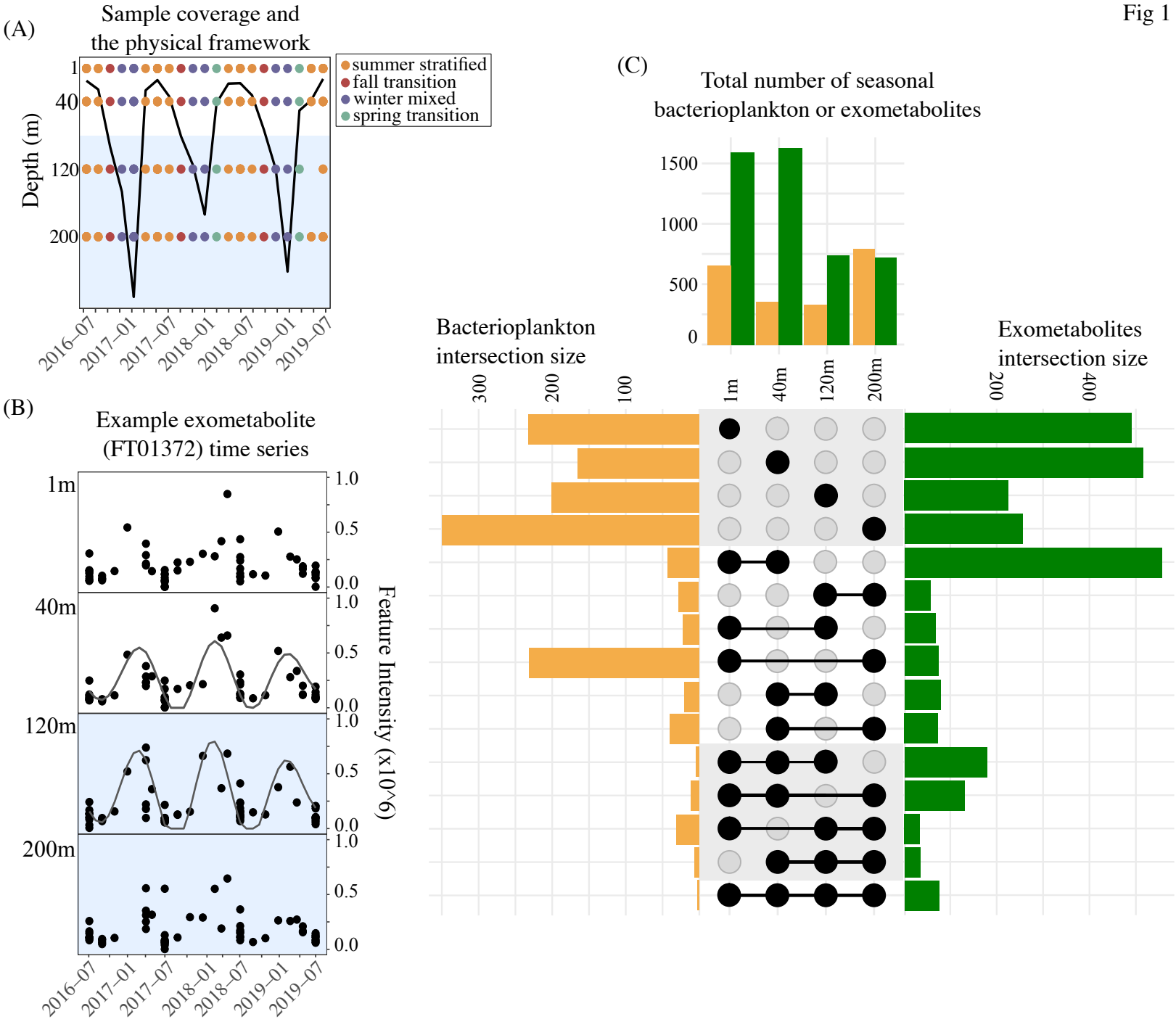

(A)
Sample coverage and the physical framework
summer stratified
fall transition
winter mixed
spring transition
Depth (m)
1
40
120
200
2016–07
2017–01
2017–07
2018–01
2018–07
2019–01
2019–07
(B)
Example exometabolite (FT01372) time series
1m
40m
120m
200m
Feature Intensity (x10^6)
1.0
0.5
0.0
(C)
Total number of seasonal bacterioplankton or exometabolites
1500
1000
500
0
Bacterioplankton intersection size
300
200
100
1m
40m
120m
200m
Exometabolites intersection size
200
400

Fig 2

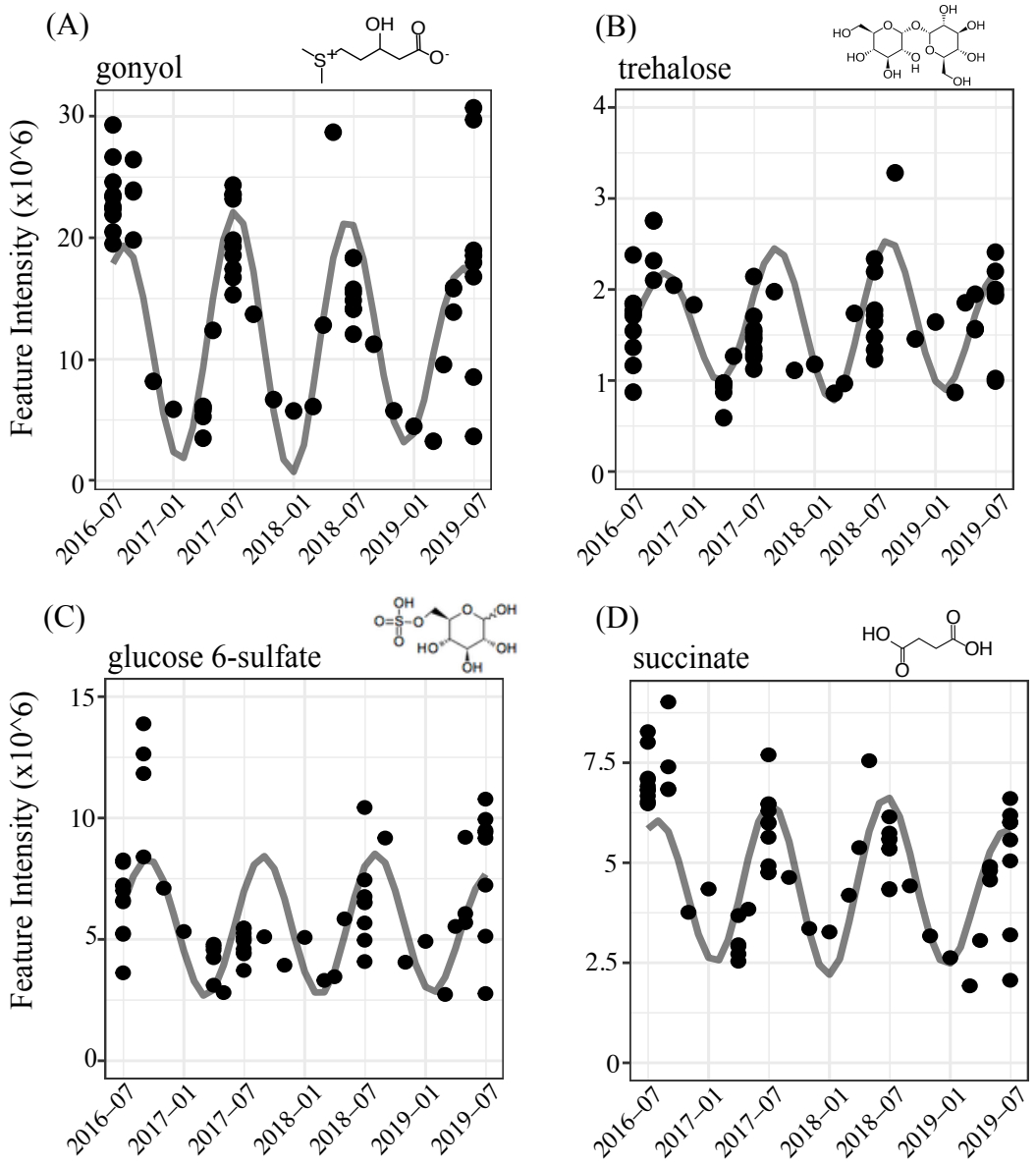
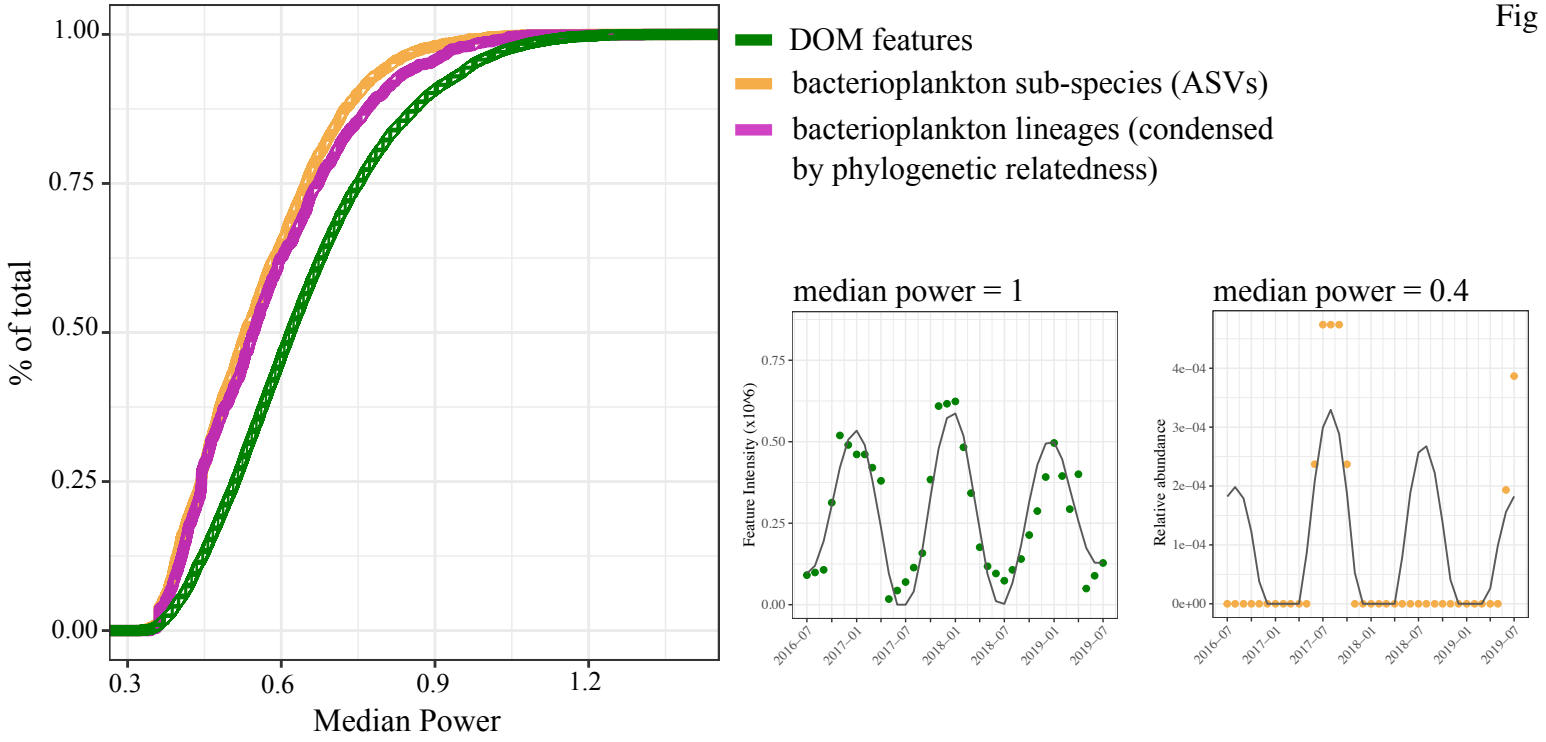

(A)
gonyol
Feature Intensity (x10^6)
(B)
trehalose
(C)
glucose 6-sulfate
Feature Intensity (x10^6)
(D)
succinate
2016–07
2017–01
2017–07
2018–01
2018–07
2019–01
2019–07

Fig 3


% of total
1.00
0.75
0.50
0.25
0.00
0.3
0.6
0.9
1.2
Median Power
DOM features
bacterioplankton sub-species (ASVs)
bacterioplankton lineages (condensed by phylogenetic relatedness)
median power = 1
Feature Intensity (x10^6)
median power = 0.4
Relative abundance
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07

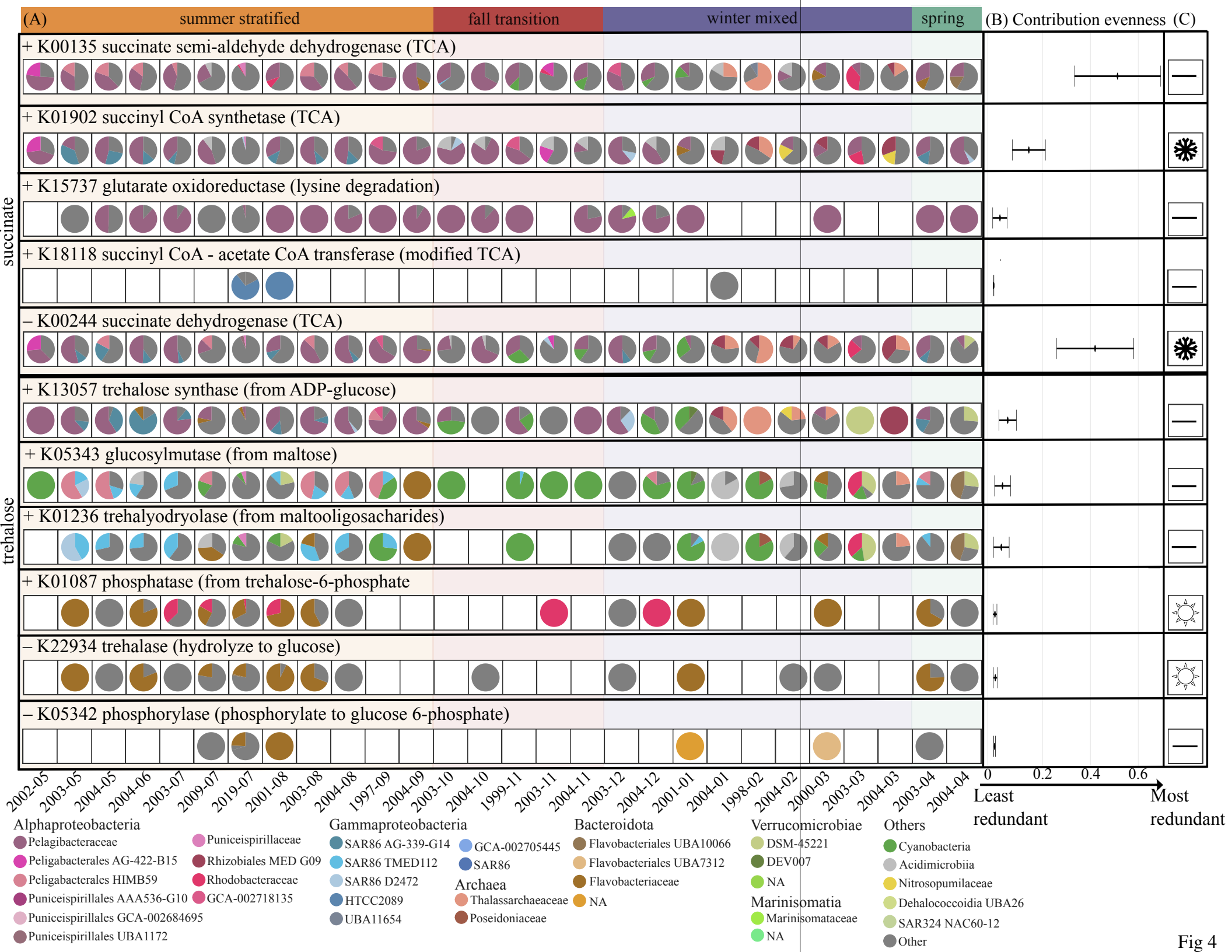

(A)
summer stratified
fall transition
winter mixed
spring
(B) Contribution evenness
(C)
succinate
+ K00135 succinate semi-aldehyde dehydrogenase (TCA)
+ K01902 succinyl CoA synthetase (TCA)
+ K15737 glutarate oxidoreductase (lysine degradation)
+ K18118 succinyl CoA - acetate CoA transferase (modified TCA)
– K00244 succinate dehydrogenase (TCA)
trehalose
+ K13057 trehalose synthase (from ADP-glucose)
+ K05343 glucosylmutase (from maltose)
+ K01236 trehalyodryolase (from maltooligosacharides)
+ K01087 phosphatase (from trehalose-6-phosphate
– K22934 trehalase (hydrolyze to glucose)
– K05342 phosphorylase (phosphorylate to glucose 6-phosphate)
2002-05
2003-05
2004-05
2004-06
2003-07
2009-07
2019-07
2001-08
2003-08
2004-08
1997-09
2004-09
2003-10
2004-10
1999-11
2003-11
2004-11
2003-12
2004-12
2001-01
2004-01
1998-02
2004-02
2000-03
2003-03
2004-03
2003-04
2004-04
0
0.2
0.4
0.6
Least redundant
Most redundant
Alphaproteobacteria
Pelagibacteraceae
Peligabacterales AG-422-B15
Peligabacterales HIMB59
Puniceispirillales AAA536-G10
Puniceispirillales GCA-002684695
Puniceispirillales UBA1172
Puniceispirillaceae
Rhizobiales MED G09
Rhodobacteraceae
GCA-002718135
Gammaproteobacteria
SAR86 AG-339-G14
SAR86 TMED112
SAR86 D2472
HTCC2089
UBA11654
GCA-002705445
SAR86
Archaea
Thalassarchaeaceae
Poseidoniaceae
Bacteroidota
Flavobacteriales UBA10066
Flavobacteriales UBA7312
Flavobacteriaceae
NA
Verrucomicrobiae
DSM-45221
DEV007
NA
Marinisomatia
Marinisomataceae
NA
Others
Cyanobacteria
Acidimicrobiia
Nitrosopumilaceae
Dehalococcoidia UBA26
SAR324 NAC60-12
Other

Fig 4

Fig S1

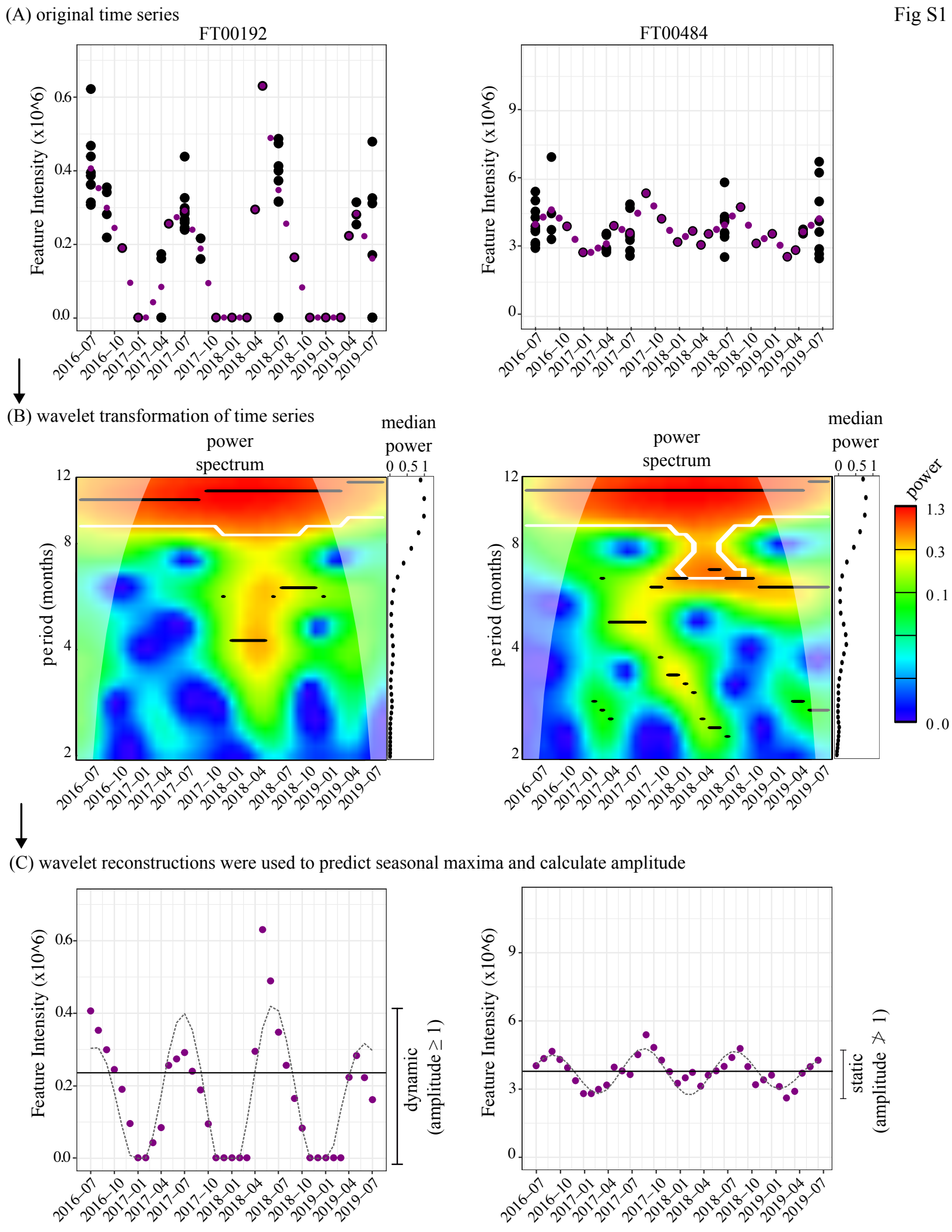

(A) original time series
FT00192
FT00484
Feature Intensity (x10^6)
(B) wavelet transformation of time series
power spectrum
median power
period (months)
power
(C) wavelet reconstructions were used to predict seasonal maxima and calculate amplitude
dynamic (amplitude ≥ 1)
static (amplitude ≯ 1)

Fig S2

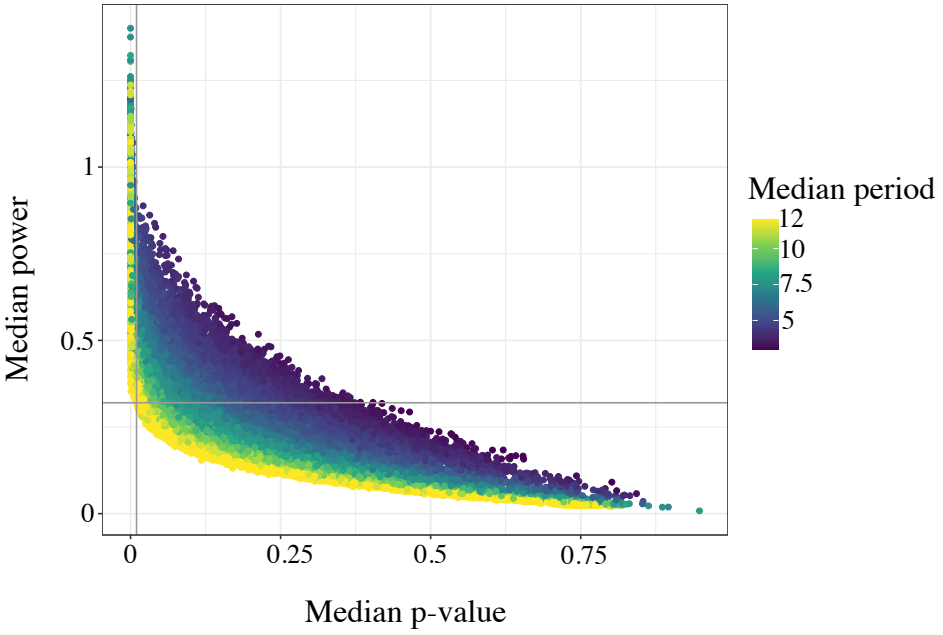

Median power
1
0.5
0
0
0.25
0.5
0.75
Median p-value
Median period
12
10
7.5
5

Fig S3

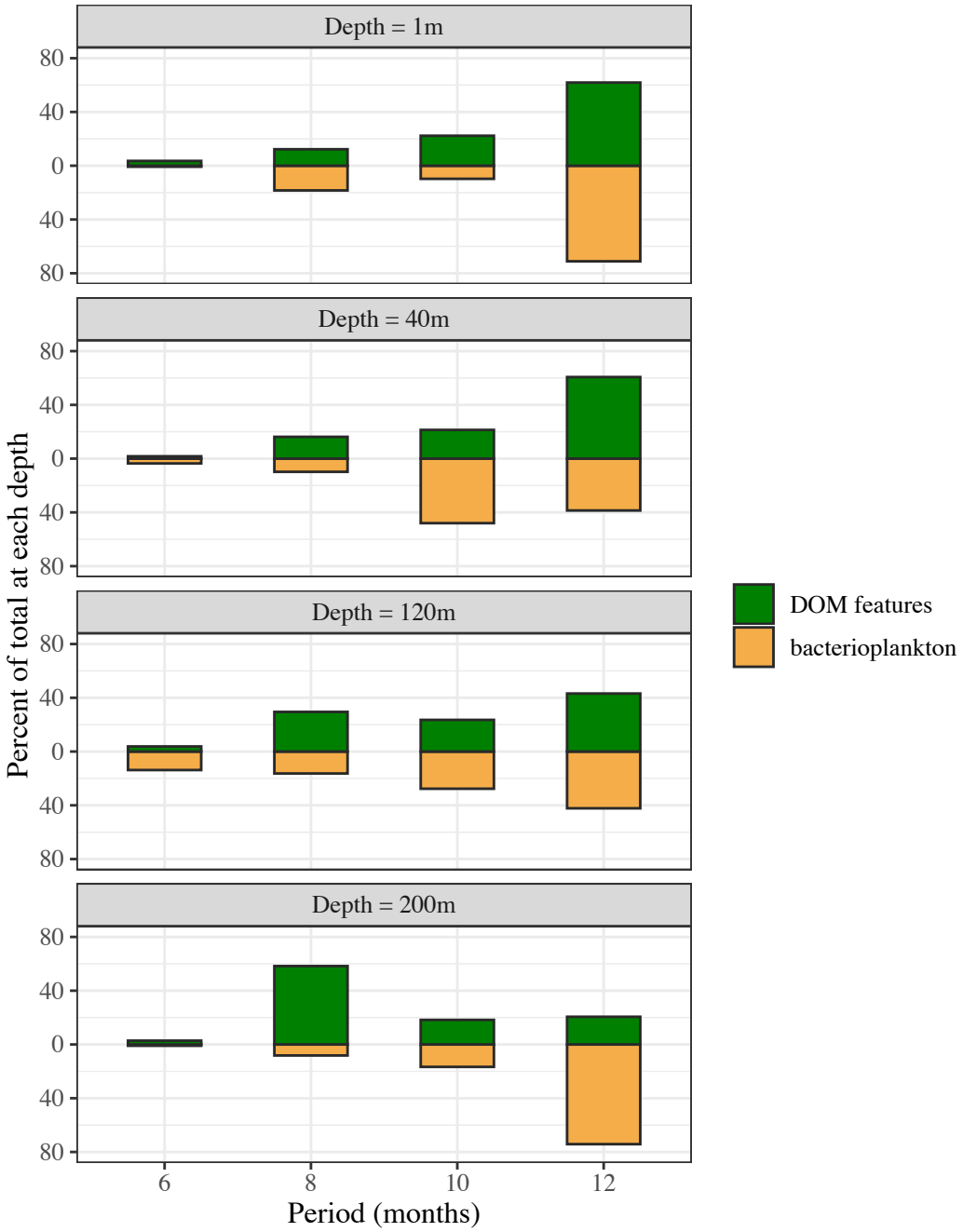

Depth = 1m
Depth = 40m
Depth = 120m
Depth = 200m
80
40
0
40
80
Percent of total at each depth
6
8
10
12
Period (months)
DOM features
bacterioplankton

Fig S4

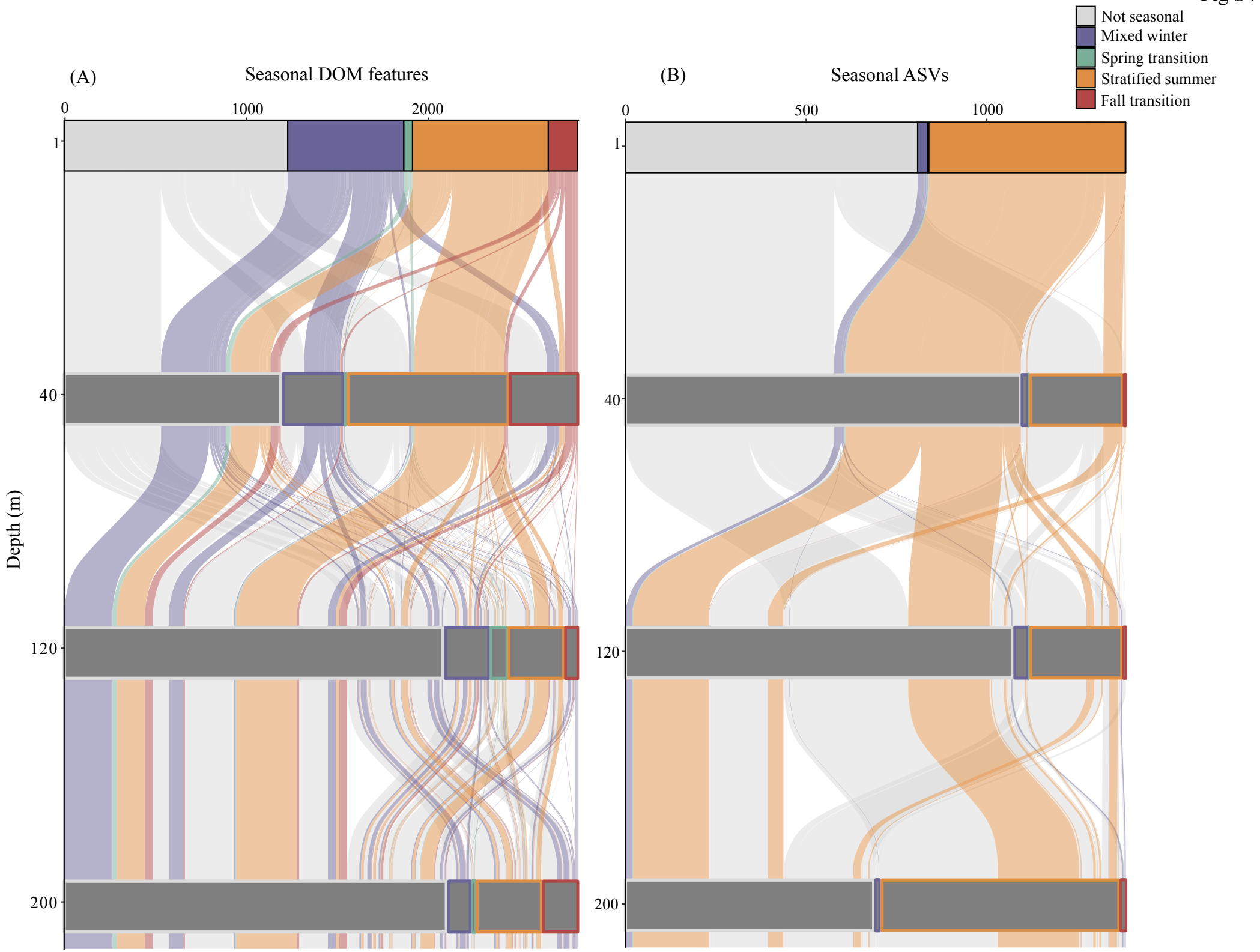
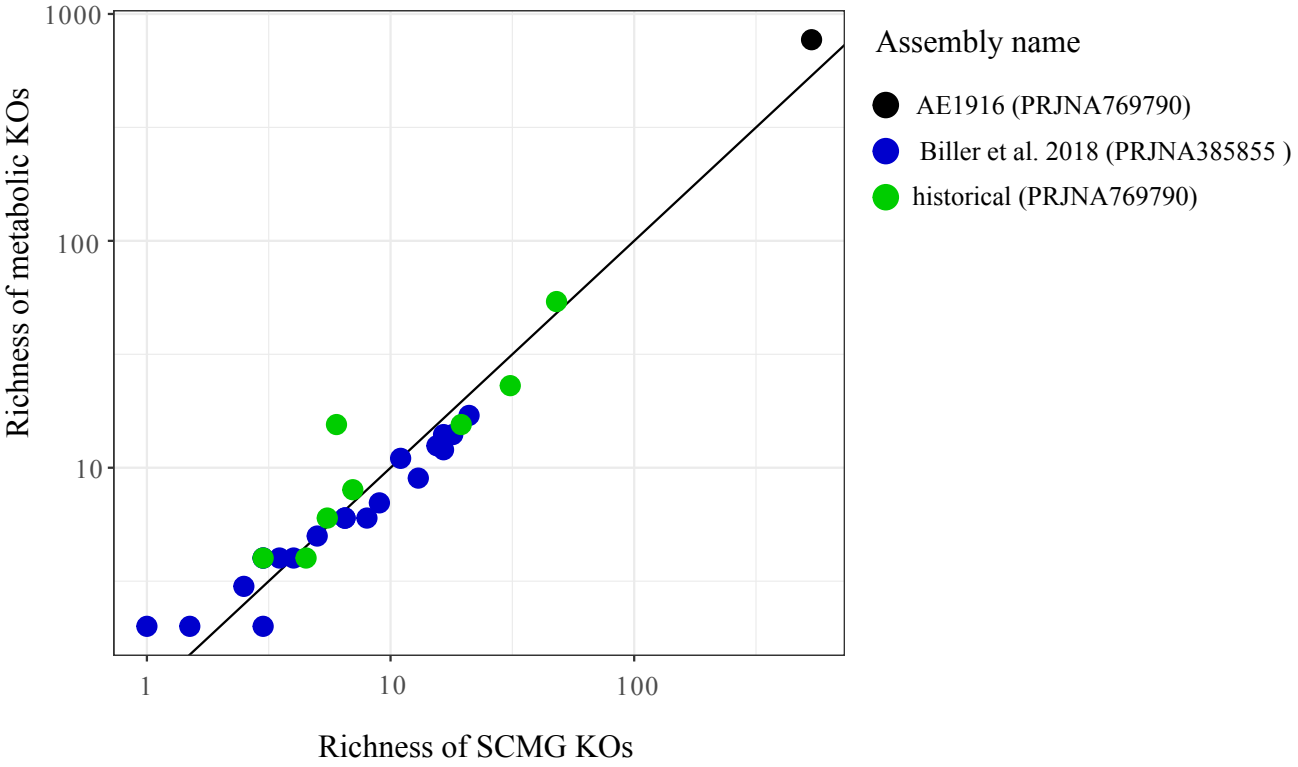
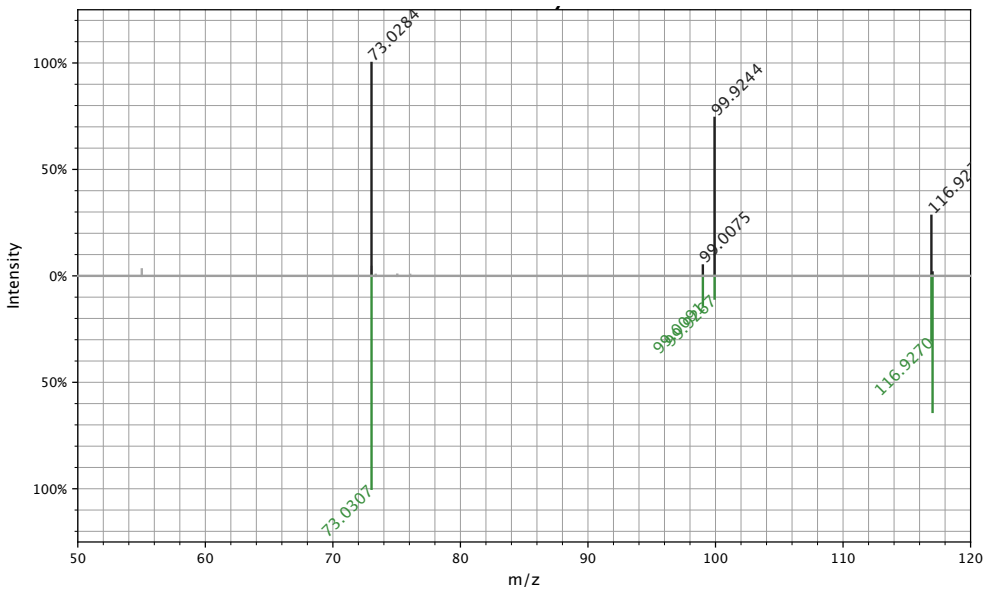

Not seasonal
Mixed winter
Spring transition
Stratified summer
Fall transition
(A)
Seasonal DOM features
0
1000
2000
(B)
Seasonal ASVs
0
500
1000
Depth (m)
1
40
120
200

Fig S5


Richness of metabolic KOs
1000
100
10
1
10
100
Richness of SCMG KOs
Assembly name
AE1916 (PRJNA769790)
Biller et al. 2018 (PRJNA385855)
historical (PRJNA769790)

succinate

(a) MS2 fragments comparison between sample (black) and reference standard (green)



(b) EIC of samples (blue) and standard mix (black)

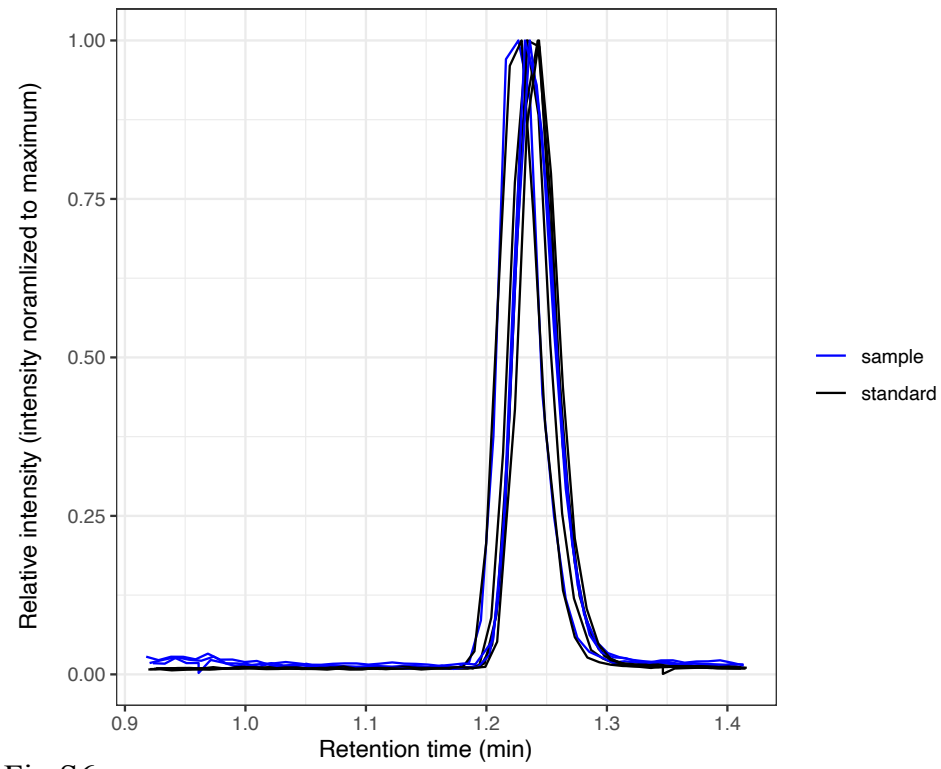
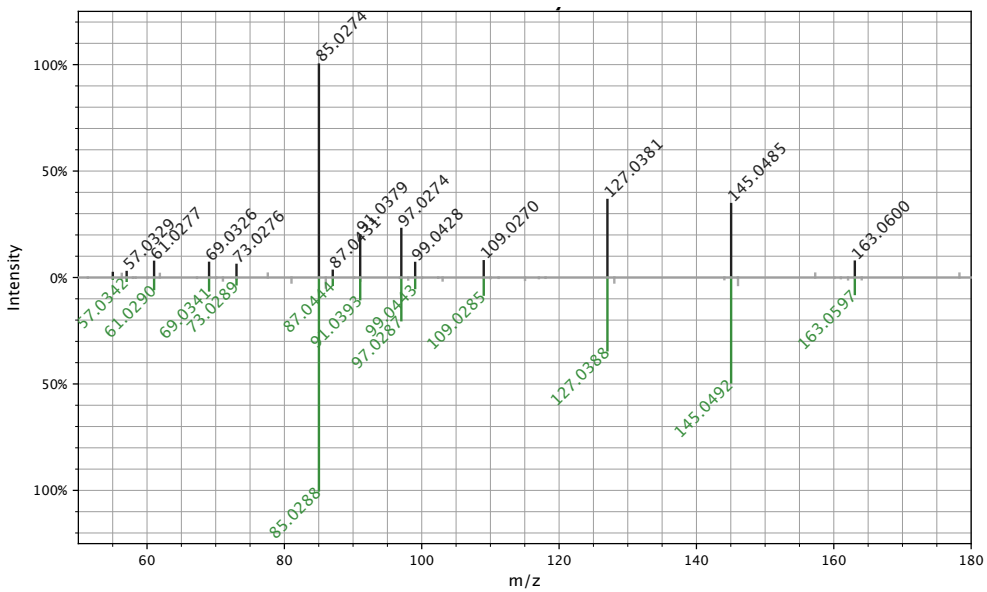


Fig S6

Trehalose

(a) MS2 fragments comparison between sample (black) and reference standard (green)



(b) EIC of samples (blue) and standard mix (black)

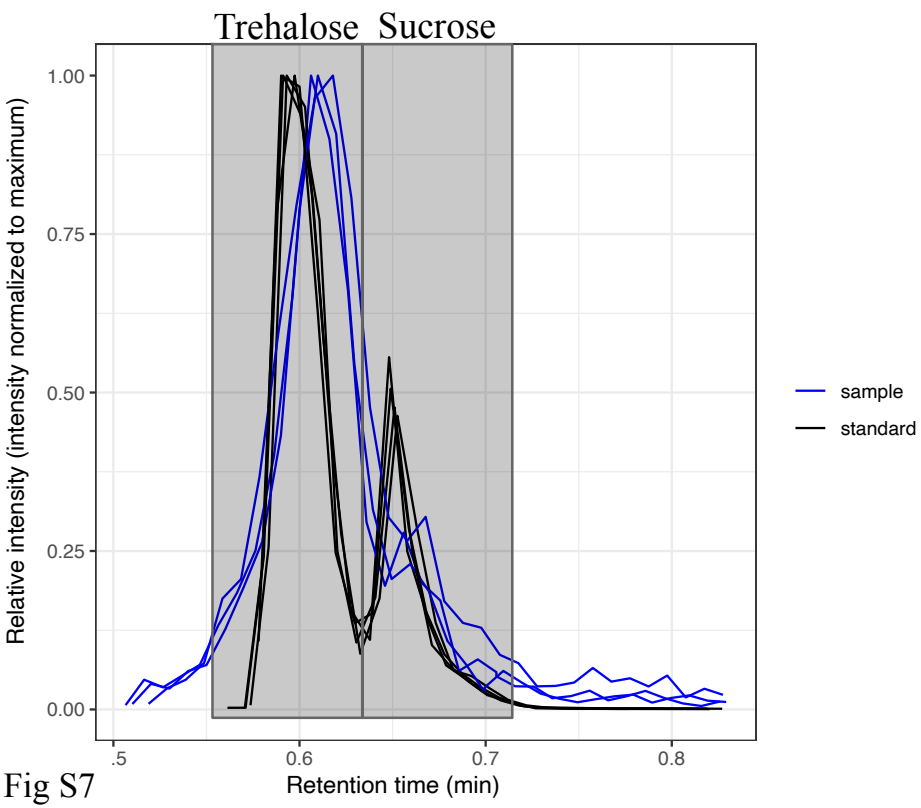
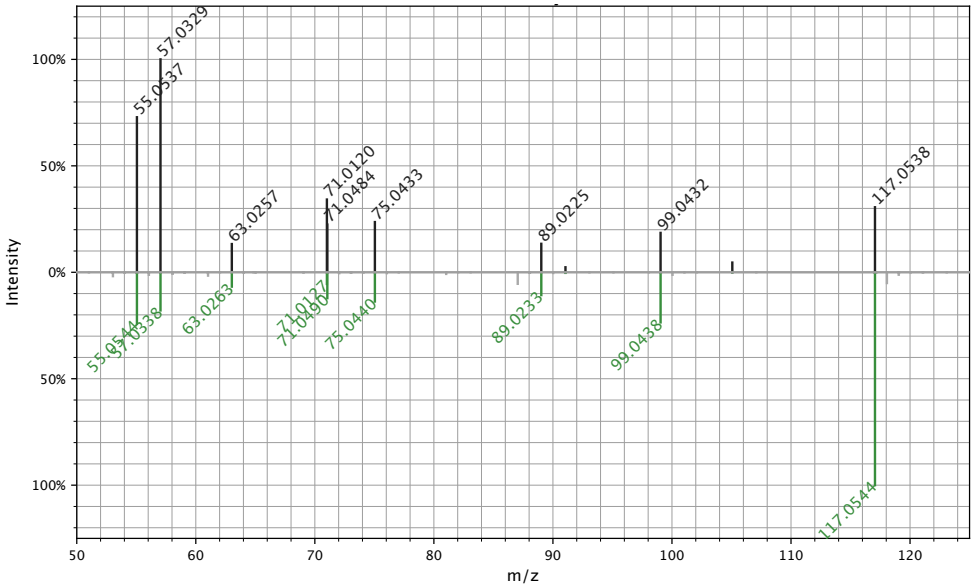


Fig S7

Gonyol

(a) MS2 fragments comparison between sample (black) and reference standard (green)



(b) isotopic fine structure

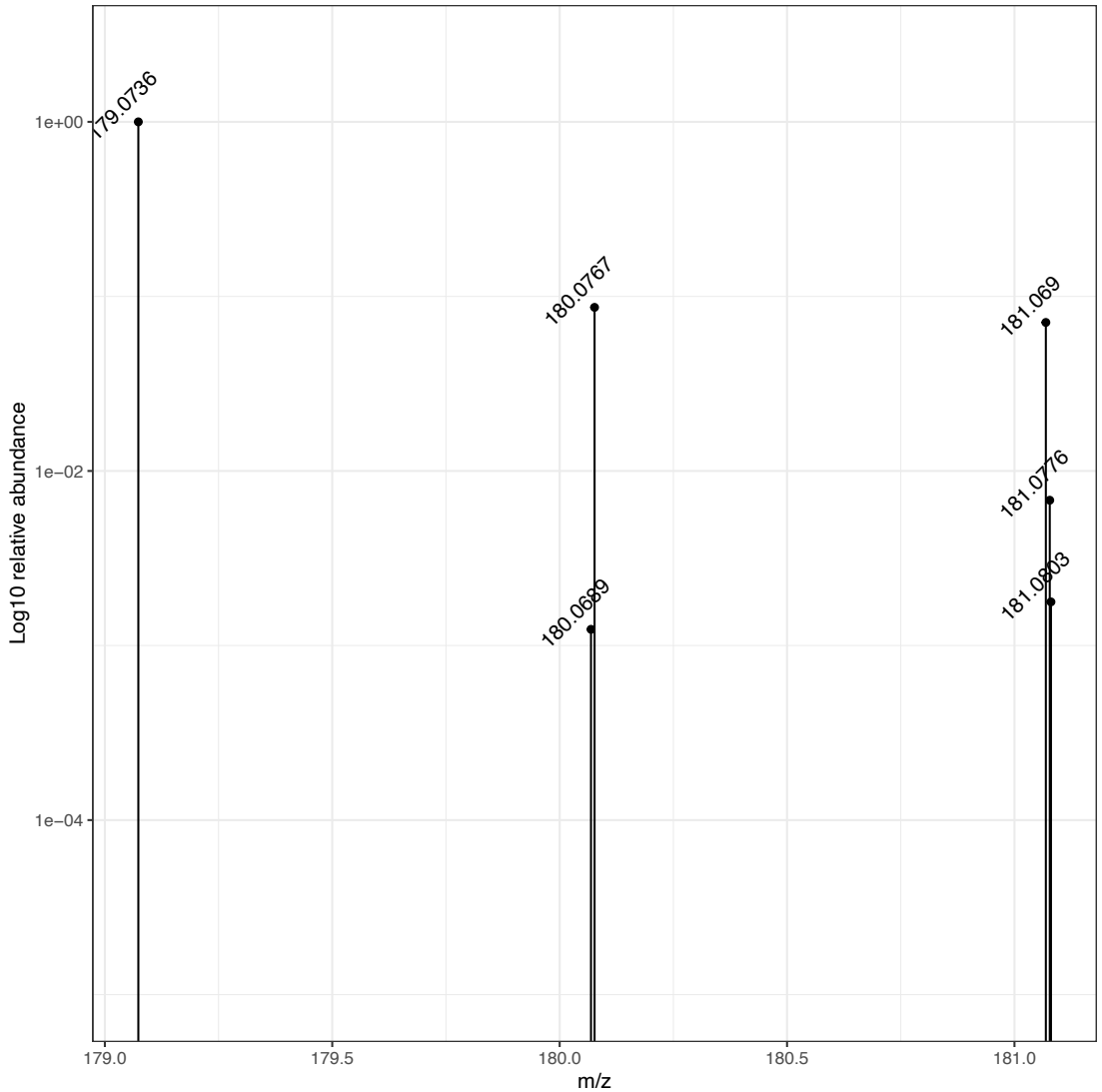


Fig S8

glucose 6-sulfate

(a) MS2 fragments comparison between sample (black) and reference standard (green)

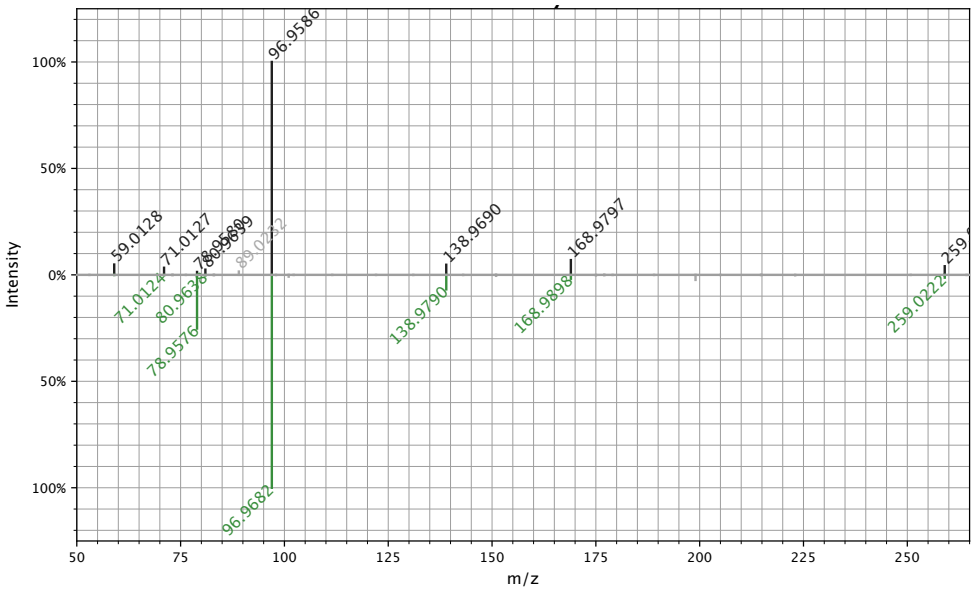


Fig S9

Table S1

Compound	Formula	Id level	Nominal feature id	Ion mode	m/z	Adduct	Ret time (min)
gonyol	$C_7H_{14}O_3S$	2	FT01262	Pos	179.0736	$[M+H]^+$	0.57
trehalose	$C_{12}H_{22}O_{11}$	1	FT07227	Pos	360.1504	$[M+NH_4]^+$	0.61
succinate	$C_4H_6O_4$	1	FT00128	Neg	117.0192	$[M-H]^-$	1.24
glucose 6-sulfate (or	$C_6H_{12}O_9S$	2	FT03840	Neg	259.0128	$[M-H]^-$	0.61

Table S2

compound_name	KO	enzyme_name	direction	group
Succinate	K00135	succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]	production	succinate-semialdehyde dehydrogenase
Succinate	K00139	succinate-semialdehyde dehydrogenase [EC:1.2.1.24]	production	succinate-semialdehyde dehydrogenase
Succinate	K17761	succinate-semialdehyde dehydrogenase, mitochondrial [EC:1.2.1.24]	production	succinate-semialdehyde dehydrogenase
Succinate	K08324	succinate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.24]	production	succinate-semialdehyde dehydrogenase
Succinate	K01902	succinyl-CoA synthetase alpha subunit [EC:6.2.1.5]	production	succinyl CoA synthetase
Succinate	K01899	succinyl-CoA synthetase alpha subunit [EC:6.2.1.4 6.2.1.5]	production	succinyl CoA synthetase
Succinate	K15737	glutarate dioxygenase [EC:1.14.11.64]	production	glutarate oxidoreductase
Succinate	K18118	succinyl-CoA:acetate CoA-transferase [EC:2.8.3.18]	production	succinyl CoA - acetate CoA transferase
Succinate	K00244	succinate dehydrogenase flavoprotein subunit [EC:1.3.5.1]	consumption	succinate dehydrogenase
Succinate	K00234	succinate dehydrogenase (ubiquinone) flavoprotein subunit [EC:1.3.5.1]	consumption	succinate dehydrogenase
Succinate	K00239	succinate dehydrogenase flavoprotein subunit [EC:1.3.5.1]	consumption	succinate dehydrogenase
Trehalose	K13057	trehalose synthase	production	glycosyltransferase
Trehalose	K05343	maltose alpha-D-glucosyltransferase / alpha-amylase	production	glucosyltransferase
Trehalose	K01236	maltooligosyltrehalose trehalohydrolase	production	trehalohydrolase
Trehalose	K01087	trehalose 6-phosphate phosphatase	production	phosphatase
Trehalose	K22934	alpha,alpha-trehalase	consumption	trehalase
Trehalose	K01194	alpha,alpha-trehalase	consumption	trehalase
Trehalose	K05342	alpha,alpha-trehalose phosphorylase	consumption	trehalose phosphorylase

Table S3

	Assembly name	Date (year-month-day)	Depth (m)	sample
1	hist	1997-09-01	1	108_0
2	hist	1998-02-01	1	113_0
3	hist	1999-11-01	1	134_0
4	hist	2000-01-01	1	136_0
5	hist	2000-03-01	1	138_0
6	hist	2001-08-01	1	155_0
7	hist	2002-05-01	1	164_0
8	Biller	2003-02-21	1	SRR5720233
9	Biller	2003-03-22	1	SRR5720238
10	Biller	2003-04-22	1	SRR5720327
11	Biller	2003-05-20	1	SRR5720283
12	Biller	2003-07-15	1	SRR5720235
13	Biller	2003-08-12	1	SRR5720286
14	Biller	2003-10-07	1	SRR5720332
15	Biller	2003-11-04	1	SRR5720276
16	Biller	2003-12-02	1	SRR5720262
17	hist	2003-03-01	1	174A_0
18	Biller	2004-01-27	1	SRR5720338
19	Biller	2004-02-24	1	SRR5720322
20	Biller	2004-03-23	1	SRR5720337
21	Biller	2004-04-21	1	SRR5720256
22	Biller	2004-05-18	1	SRR5720257
23	Biller	2004-06-15	1	SRR5720260
24	Biller	2004-08-17	1	SRR5720321
25	Biller	2004-09-14	1	SRR5720251
26	Biller	2004-10-13	1	SRR5720307
27	Biller	2004-11-12	1	SRR5720278
28	Biller	2004-12-08	1	SRR5720342
29	Biller	2009-07-14	1	SRR6507279
30	AE1916	2019-07-09	1	5_1_S27

Table S4

compound	label	ng/ml	ion mode	ret time (min)	exact mass
alpha-keto-3-methylbutyric acid	$^{13}C5$	150	negative	2.5	121.06408
alpha-ketoglutaric acid	D6	2000	negative	0.91	152.05916
betaine	D11	21	positive	0.57	128.14805
citric acid	D4	250	negative	0.97	196.05211
L-cysteine	D3	1000	positive	0.5	124.03853
guanosine	D2	50	both	2.31	285.10425
DL-malic acid	D3	100	negative	0.77	137.04033
L-phenylalanine	D8	50	both	2.66	173.12922
L-proline	$^{13}C5,^{15}N$	25	positive	0.62	121.07711
L-lysine	D4	750	positive	0.45	150.13061
xanthine	$^{15}N2$	50	both	1.05	154.02747
4-aminobenzoic acid	D4	2000	positive	3.1	141.07281
sn-glycerol-3-phosphate	$^{13}C3$	500	both	0.6	175.02377
pyridoxine (vitamin B6)	$^{13}C4$	25	both	0.95	173.08732
pantothenate	$^{13}C3,^{15}N$	25	both	3.12	223.1178
L-leucine	D3	50	both	1.32	134.11343
adenosine 5`-monophosphate	$^{15}N5$	100	both	0.89	352.04828
L-methionine	D3	50	both	0.93	152.06983
biotin	D2	50	both	4.44	246.10075
succinic acid	D6	500	negative	1.22	124.06426
cholic acid	D4	50	negative	7.37	412.31271
indole-3-acetic acid	D7	1000	positive	5.55	182.10724