Interrogating metabolic plasticity in marine organisms: A framework for best practices using metabolomic and lipidomic approaches

Yaamini R. Venkataraman[1]* and Ariana S. Huffmyer[2]
[1]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA
[2]School of Aquatic and Fisheries Sciences, University of Washington, Seattle, WA, USA

Both authors contributed equally to this manuscript.
* = corresponding author, yaamini.venkataraman@gmail.com

Abstract

Understanding the mechanisms that underlie resilience in marine invertebrates is critical as climate change and human impacts transform coastal ecosystems. Metabolic plasticity, or an organism's capacity to modulate energy production, allocation, and use, plays a central role in mediating resilience under environmental stress. While research on marine invertebrate stress responses has grown, integrative studies that examine metabolic plasticity by connecting molecular, physiological, and organismal scales remain limited. In this Perspective, we advocate for the rigorous and thoughtful use of metabolomic and lipidomic approaches to understand resilience in marine systems through the lens of metabolic plasticity. We provide recommendations for experimental design, summarize current methodologies, and provide an overview of commonly used data analysis approaches. Advances in other molecular approaches such as genomics, epigenomics, and transcriptomics can be harnessed to further explore stress responses through multi-omic integrative analyses. As quantitative integrative analysis remains limited in marine fields, we call for a stronger integration of molecular, metabolomic, physiological, and organismal data sets to link mechanisms to phenotypes. We explore the use of these approaches in studies of marine invertebrates and highlight promising areas of multi-omic research that deserve exploration. By embracing metabolic complexity and scaling from molecules to phenotypes, we suggest that the marine invertebrate research community will be better equipped to understand, anticipate, and mitigate the impacts of environmental change on marine ecosystems.

Keywords: phenotypic plasticity, metabolomics, lipidomics, metabolic flexibility, metabolic plasticity

Introduction

Energy metabolism is a central component of organismal responses to environmental stressors. Under a particular environmental condition, the Oxygen- and Capacity-Limitation of Thermal Tolerance (OCLTT) hypothesis suggests that the optimal temperature for an organism is one that maximizes its aerobic scope. Within the upper and lower bounds of tolerance, or "pejus" temperatures, organisms must respond to environmental stressors using metabolic compensation (Pörtner 2010; Pörtner et al. 2017). Outside this range, organisms shift to anaerobic processes, and stress responses are characterized by energy conservation and essential function maintenance at the expense of growth or reproduction (Pörtner 2010; Pörtner et al. 2017). Metabolic compensation can occur prior to the onset of physiological or organismal manifestations of stress. In order to understand transitions between active and passive tolerance (sensu (Pörtner et al. 2017)), whole-organism physiology metrics do not provide a complete picture of organism stress responses .

Molecular examinations of metabolic responses using metabolomics and lipidomics (see definitions in **Box 1**) are increasingly being used to investigate plasticity in response to environmental stressors in marine invertebrates (**Figure 1A)**. Metabolomic and lipidomic fields emerged in the 20th century allowing for increased high-throughput profiling through advances in mass spectrometry and nuclear magnetic resonance (NMR) techniques (Viant 2008; Lindon and Wilson 2016; Beale et al. 2018). However, use of these approaches in non-model systems was not more prevalent until the 2000s due to challenges and limitations in protocol development and compound identification in non-model systems (Viant 2008; Williams et al. 2011; Schock et al. 2014; Carriot et al. 2021). In recent years, improved instrument sensitivity, expanded databases and libraries, reduced costs, and improved computational approaches have made metabolomics and lipidomics more accessible (Putri et al. 2013; Beale et al. 2018; Munjal et al. 2022).

Recent work has investigated energetic constraints that lead to stressor susceptibility, and highlighted pathways of resilience and resistance to sublethal stressors. For example, American lobster (*Homarus americanus*) exposure to ocean acidification resulted in broad metabolic reprogramming that was not associated with changes to resting metabolism, suggesting that metabolic homeostasis was maintained by plasticity in energy usage (Noisette et al. 2021). Poorer survival and lower concentrations of stress-associated lipid classes highlights the susceptibility of staghorn coral *Acropora cervicornis* to low pH and irradiance conditions in deeper waters; however, increased diversity of various lipid classes also suggests deepwater corals employ heterotrophy more than shallow reef counterparts to meet energetic demands in stressful conditions (Rodriguez-Casariego et al. 2023). Additionally, examination of metabolic responses to stress can reveal sublethal impacts that are not detectable at the whole-organism level. For example, a study in coral larvae detected metabolic reprogramming under elevated temperature without a decrease in survival (Huffmyer et al. 2024). In the blue mussel, *Mytilus edulis*, 1H-NMR metabolomics revealed differential energetic responses to OA stress in males as compared to females (Ellis et al. 2014). Further, manila clams (*Ruditapes philippinarum)*, exhibit variable responses to environmental toxins indicated by changes in major metabolite abundance (Liu et al. 2011) and metabolite abundance is indicative of thermal stress effects in soft corals (Farag et al. 2018). Sublethal changes in lipidomic responses are also seen in blue mussels (*M. edulis)* in which the composition of lipids shifts with diet in newly settled spat (Laudicella et al. 2020).

Due to their direct connection to energy metabolism, metabolomic and lipidomic approaches can be used to quantify metabolic plasticity. Metabolic plasticity can be achieved by processing the same compounds in different pathways to achieve similar results for cellular metabolism (Fendt et al. 2020), or through wholesale shifts in metabolic pathways, particularly in changing conditions or hostile environments (Jia et al. 2019). While the concept of metabolic plasticity has its origin in cancer biology, we encourage its use in organismal biology to guide

proper use of metabolomic and lipidomic approaches. Placing metabolic plasticity at the center of investigations provides critical insight into mechanisms of organismal response to stress.

This perspective highlights the applications of these technologies to the interrogation of metabolic plasticity. We propose a framework to navigate experimental and analytical decisions centering these concepts. We also demonstrate the power of combining these methods with the study of other molecular mechanisms, such as gene expression. Finally, we highlight the importance of understanding these molecular mechanisms in the context of whole-organism physiological metrics. As metabolomics and lipidomics technologies become widely-used in organismal biology, establishing consensus around these practices will allow for rigorous, reproducible, and biologically meaningful analyses to examine plasticity in important ecosystems.

Experimental design choices influence the capacity to characterize metabolic plasticity

Sampling considerations

As is the case with any molecular tool, experimental design choices will influence data interpretation and robustness. Choices should be made in the context of hypotheses and budget for each individual experiment. We encourage readers to refer to published literature when determining the appropriate number of technical and biological replicates (Blaise et al. 2016; Jacyna et al. 2019; Lee et al. 2022). Mass spectrometry labs can also provide guidance on sample sizes, but most facilities recommend six biological replicates per treatment for adequate analytical power. Samples used should be chosen with the scientific question in mind, especially if sampling involves tissue isolation (**Figure 2**). For example, gill tissue may be appropriate for studying aerobic respiratory responses to stress while the gonad tissue is appropriate for understanding stress effects on reproductive tissues (Downey-Wall et al. 2020; Venkataraman

et al. 2024). One study in the softshell clam (*Mya arenaria*) examined metabolomic responses in gill, mantle, and adductor muscle tissues following temperature exposure and found significant variation between tissue types (Beaudreau et al. 2024). In abalone (*Haliotis iris*) (Nguyen et al. 2021) examined tissue-specific metabolomes in haemolymph and muscle tissue and found that muscle tissues were less affected by thermal stress than the haemolymph. In corals, single polyp approaches allow for examination of spatial biochemical structuring in complex holobiont systems (Roach et al. 2021) with separate host and symbiont analyses showing distinct metabolic responses between the partners (e.g., (Gamba et al. 2022)). For organisms in which tissue specific sampling is conducted or not possible, researchers should interpret metabolome responses at the organism scale.

Samples should be processed and preserved in a way that minimizes enzyme activity during metabolite extraction (Liu and Locasale 2017). In order to quench metabolic activity and prevent degradation of compounds, samples should be isolated (e.g., seawater removed) and immediately snap-frozen in liquid nitrogen, stored at -80°C, and transported using dry ice or liquid nitrogen. Samples should not be stored in reagents such as RNALater to avoid any alterations to metabolic state during preservation. Avoid freeze thaw cycles and perform any necessary processing or extraction steps on ice or dry ice as required for specific protocols. Researchers should refer to Liu and Locasale (2017) for a review of compound extraction protocols.

Time series, or time course, metabolomic experiments can provide key insights into metabolic regulation, plasticity, and stress responses by examining the dynamic nature of the metabolome and gaining a more complete view of relationships between key metabolites or pathways of interest (Sriyudthsak et al. 2016). These experiments must be designed with consideration of the appropriate time scale to test hypotheses. For example, metabolism of glucose through central carbon metabolism occurs on the order of seconds to minutes while changes in lipid store composition may take hours to days, with each of these processes

requiring a different time series design to capture on appropriate time scales. Previous work has conducted metabolomics time series studies on the order of seconds to minutes to characterize compound synthesis (Sekar et al. 2018) and metabolic responses to starvation in microbial systems (Link et al. 2015). Longer time scales such as weekly to monthly sampling is better suited to capture seasonal changes, as done to characterize cold acclimation in plants (Angelcheva et al. 2014; Rathore et al. 2021). Therefore, researchers should design time series experiments with prior knowledge of the rate of metabolic pathways of interest, the turnover rate of target metabolites, and/or the time scale of physiological responses of interest. We direct readers to previous reviews and studies that discuss data analysis and considerations specific to time series and dynamic metabolomic studies (Smilde et al. 2010; Nägele et al. 2016; Sriyudthsak et al. 2016).

Analytical considerations for data acquisition

Analytical platform choice is an important methodological consideration and should be selected based on the target compounds of interest, their chemical composition, desired output data format, and the robustness of the databases used for compound identification (**Figure 2**). Prior to identifying a platform for data acquisition, researchers should determine what kind of data is best suited for addressing their hypotheses. Metabolomic and lipidomic data can be acquired in three different formats: targeted, semi-targeted, or untargeted. Targeted experiments provide specific concentrations of molecules (absolute quantitation), allowing researchers to investigate hypotheses associated with specific pathways or compounds of interest (e.g., glycolysis) (Bennett et al. 2008; Cajka and Fiehn 2016; Park et al. 2016; Liu and Locasale 2017; Lee and Yokomizo 2018; Georgoulis et al. 2022). Targeted lipidomics revealed how changes in membrane remodeling were associated with physiological tipping points in response to low pH in the Pacific oyster (*Crassostrea gigas*) (Lutier et al. 2022). However, targeted experiments require intimate knowledge of metabolic pathways in the organism of interest. Researchers,

especially those working in non-model marine systems, should consider why they require exact concentrations of molecules to answer their research questions, and if the nomenclature for molecules of interest is conserved between their organisms and those used to generate compound databases. In contrast, untargeted experiments tend to provide relative feature abundance differences between experimental conditions or populations (semiquantitation) (Doroghazi et al. 2014; Cajka and Fiehn 2016; Liu and Locasale 2017; Lee and Yokomizo 2018). The majority of studies identified in **Appendix A** used untargeted data acquisition approaches (**Figure 1B**). This approach may be useful in non-model systems, where several molecules are likely uncharacterized by existing databases and can allow for novel compound identification. For example, a study in reef-building corals used untargeted metabolomics and compound identification to identify lipid classes (e.g., betaine lipids) that distinguished between thermally resilient and sensitive colonies (Roach et al. 2021). Also in corals, untargeted metabolomics identified dipeptides that were important in heat stress responses (Williams, Chiles, et al. 2021). While untargeted data acquisition can enable novel compounds identification, feature annotation is time consuming and requires comprehensive reference databases and organismal knowledge (Liu and Locasale 2017). Identification of unknown compounds may be easier for lipidomics due to conserved nomenclature conventions based on compound structure. When collecting relative abundance data, we encourage readers to confirm measurements with facilities and interpret data in terms of relative changes rather than absolute quantification. Semi-targeted data acquisition may be a good alternative to targeted or untargeted assays (Breitling et al. 2006; Gika et al. 2016; Liu and Locasale 2017; Reisz et al. 2019). Diversity of waxy ester and triglyceride compounds detected with semi-targeted lipidomics in the coral *Acropora cervicornis* highlight how outplanting in deep environments promotes heterotrophy (Rodriguez-Casariego et al. 2023). These experiments identify and absolutely quantify a large number of molecules, enabling comparisons of specific molecule abundances and pathway-level analyses. Unlike targeted analyses, it does not require the

researchers to specify in their hypotheses which molecule(s) may differ between treatments or populations.

Metabolomic and lipidomic analyses most commonly are conducted to estimate metabolite absolute or relative concentrations at a particular point in time, known as “steady-state” measurements (**Figure 1B**). However, a key challenge in employing these methods is the limitation in interpreting “metabolic flux”, or the rate at which metabolites pass through a metabolic pathway (Jang et al. 2018). Although characterizing shifts in metabolite concentration with steady-state metabolomics can inform researchers of relative differences in concentration of metabolites, concentration alone does not directly relate to metabolic flux. For example, increased pool size of a metabolite may be the result of either increased production or decreased downstream metabolism, resulting in accumulation (Jang et al. 2018), as seen in a study of coral larvae (Huffmyer et al. 2024). Therefore, we caution researchers from making strong conclusions regarding metabolic flux from steady-state analyses. Stable isotope tracing offers an approach to quantify metabolic flux of pathways of interest by tracking the incorporation of labeled atoms from stable isotope tracers into metabolites (Jang et al. 2018). Tracing metabolic flux of pathways of interest using probes targeted for specific hypotheses (e.g., ^{13}C labeled carbon or ^{15}N labeled nitrogen), provides detailed insights into pathway activity and regulation that cannot be obtained with steady-state metabolomics alone (Jang et al. 2018). We direct the reader to previous literature that describes stable isotope tracing methods in detail (Creek et al. 2012; Fan et al. 2012; Chokkathukalam et al. 2014; Jang et al. 2018; Balcells et al. 2019). For example, stable isotope tracing approaches have been employed across non-human taxa and systems including *Drosophila* (Wang et al. 2022), soils (Wilhelm et al. 2022), and plants (Freund and Hegeman 2017), but are far less utilized than steady-state methods in marine invertebrates (**Figure 1B**). One area of research that features stable isotope metabolomic tracing is the investigation of symbiotic nutritional exchange and nutrient metabolism in reef-building corals. For example, (Chiles et al. 2022) utilized nitrogen (^{15}N)

tracing experiments to investigate nutrient shifts during heat stress and (Huffmyer et al. 2024) conducted carbon (^{13}C) tracing to examine changes in symbiotic relationships under high temperatures. Also in corals, (Hillyer et al. 2018) documented widespread changes in carbon metabolism during coral bleaching using carbon (^{13}C) metabolomic tracing. In the blue crab (*Callinectes sapidus*) (Holt and Kinsey 2002; Kinsey and Lee 2003) and red abalone (Haliotis rufescens) (Tjeerdema et al. 1993), ^{31}P NMR studies track flux through central energy metabolism reactions and provide insights on shifts in energetic state under environmental stress. Application of stable isotope tracing metabolomic studies can provide rich information on metabolic flux and can provide a more complete understanding of metabolic plasticity in marine invertebrates. However, isotopic tracing studies are more expensive and researchers should carefully consider advantages and limitations of selected methods for addressing hypotheses of interest.

Metabolomic analyses are commonly performed using nuclear magnetic resonance (NMR) or mass spectrometry (MS), or a combination of the two (Ren et al. 2015). Most MS analyses are conducted gas chromatography mass spectrometry (GC-MS) or liquid chromatography mass spectrometry (LC-MS), the latter of which is commonly analyzed using high performance liquid chromatography (HPLC) or ultra high performance liquid chromatography (UHPLC). The specification of these platforms and technical descriptions of the analytical pipelines has been described elsewhere (Naz et al. 2014; Beale et al. 2018). LC-MS is generally preferred for analyzing a wider range of polar, non-volatile compounds due to the capacity for characterization of a wider range of molecule types (e.g., fatty acids and lipids), while GC-MS is best suited for volatile molecules (Ren et al. 2015). In metabolomics applications, GC-MS platforms are considered to provide a robust and reproducible approach with more highly developed and universal databases for identification (Beale et al. 2018), while lipidomic analyses are conducted more commonly using LC-MS platforms (Cajka and Fiehn 2014). NMR (commonly, 1H-NMR), on the other hand, is used to quantify metabolites by placing

a sample in a magnetic field and using the inherent magnetic properties to identify the metabolites (Markley et al. 2017; Bingol 2018; Emwas et al. 2019). NMR offers highly reproducible measurements and high resolution to quantify metabolite concentrations and can detect a broader range of metabolites, but is limited in sensitivity and can be higher in cost (Emwas 2015; Ren et al. 2015; Markley et al. 2017; Bingol 2018). While NMR and MS both present limitations and challenges, recent efforts have emphasized the advantages of using both methods for complete characterization of the metabolome (Nagana Gowda and Raftery 2015). We recommend that researchers determine whether targeted, semi-targeted, untargeted analyses are required, then select the platform or combination of platforms best suited for the size and nature of compounds of interest. Researchers should consider the robustness of reference databases and standards to allow for reliable and accurate compound identification from spectral data. Previous work has discussed considerations for compound identification datasets and potential challenges in dataset nomenclature (Kind et al. 2009; Neumann and Böcker 2010; Blaženović et al. 2018; Sindelar and Patti 2020; Misra 2021; de Jonge et al. 2022). If researchers are resourcing analyses to an external facility, we encourage readers to discuss analytical choices and database resources with the facility team.

Data analysis to effectively address questions on marine invertebrate plasticity

Using appropriate analytical methods to address hypotheses and challenges with commonly used approaches

The choice of analytical method to address questions and hypotheses is a critical decision analyzing metabolomic or lipidomic data (**Figure 3**). Each approach has unique advantages and challenges and we argue that researchers should use multiple approaches to analyze data. Here, we provide an overview of analytical approaches to answer commonly asked questions and examine pitfalls and challenges. We focus on the general concepts of

statistical approaches using data that has been combined from negative and positive ion modes, and previously undergone necessary peak alignment and quantification, spectral deconvolution, and corrections (Ren et al. 2015) and point readers to previous discussions of these aspects of data analysis (Issaq et al. 2009; Li et al. 2014; Smith et al. 2014; Zhao et al. 2019). Choice of statistical method to analyze metabolomic and lipidomic data must be appropriate to address the scientific question and statistical hypotheses. Prior to conducting statistical analyses, it is critical to conduct biologically appropriate normalization, assess quality controls (i.e., pooled biological quality control samples), and control for batch and confounding effects. We also point readers to previous reviews that discuss analytical approaches in more detail (Worley and Powers 2013; Checa et al. 2015; Ren et al. 2015; Zhao et al. 2019).

How does the composition of the metabolome/lipidome or the concentration of a metabolite/lipid of interest change across groups or treatments?

Some lines of questioning may require testing the concentrations of particular metabolites. If single metabolite tests are necessary, analysis of variance (ANOVAs) or linear models (general and generalized linear models) are useful for testing specific hypotheses and are robust and widely used. However, metabolomic and lipidomic data frequently violate test assumptions of normality and heteroskedasticity and exhibit collinearity, which further requires data transformation or the use of non-parametric tests (e.g., Kruskal-Wallis or Mann-Whitney U tests) (Vinaixa et al. 2012). If multiple ANOVA tests are used to evaluate differences in metabolites or lipids, multiple comparison p-value adjustments and corrections for false discovery rate are necessary (Broadhurst and Kell 2007; Vinaixa et al. 2012). When identification of a single metabolite is desired, laboratory assays may be more appropriate than whole metabolome characterization (e.g., succinate quantification in (Zittier et al. 2018) and glycogen quantification in (Chen et al. 2022)).

Many studies evaluate the composition of the metabolome or lipidome as a multivariate response and examine variation in these responses between groups, treatments, or across time. The most commonly used unsupervised multivariate statistical approach is performing a permutational analysis of variance (PERMANOVA), which is non-parametric and well-suited for highly dimensional and non-normal data (Anderson 2017). However, PERMANOVA tests are sensitive to differences not only in centroid location between groups of interest, but also to differences in dispersion, or spread (Anderson 2017). Therefore, PERMANOVAs should be paired with permutational analyses of dispersion (PERMDISP) to evaluate whether multivariate differences between groups are a product of centroid location (i.e., significant PERMANOVA but non-significant PERMDISP) or centroid location and/or dispersion (i.e., significant PERMANOVA and significant PERMDISP) (Anderson 2017). For example, Beauclercq et al. (2023) utilized PERMANOVA and PERMDISP analyses to examine the influence of saxitoxin on metabolites and lipids in *M. edulis* immune cells. They found significant differences in fatty acid profiles when mussels were fed the toxin-producing *Alexandrium catenella* versus the non-toxic *Tetraselmis suecia* algae. Since PERMANOVA tests are sensitive to unbalanced sample sizes and the distance metric (e.g., Euclidian, Bray-Curtis) used, researchers should select distance metrics appropriate for their data sets. PERMANOVAs are also not intended to identify drivers of differences between groups, which should be addressed using supervised and other methods outlined below.

How do metabolomic or lipidomic features correlate with quantitative responses or time?

Examining the relationships between metabolomic and lipidomic features with quantitative responses, phenotypes, or time can be accomplished through several correlation-based and network approaches. We point the reader to work discussing the nature of dynamic metabolomic datasets and analyses in more detail (Smilde et al. 2010). Here, we discuss several approaches utilized in biological studies.

First, weighted gene co-expression network analyses (WGCNA) are commonly used in gene expression studies to identify modules, or groups, of genes that share expression patterns (i.e., co-expression) (Langfelder and Horvath 2008). This approach can be applied not only to gene expression data, but also to metabolomic and lipidomic data, which is less frequently utilized (Pei et al. 2017). For example, WGCNA has been applied to characterize metabolomic responses in tomato plants (DiLeo et al. 2011), dinoflagellate algae (Sui et al. 2014), and pathogenic fungi (Sun et al. 2024) but has yet to be applied in the study of marine invertebrate metabolomic or lipidomic analyses. WGCNA analyses are useful to identify groups of metabolites or lipids that share abundance patterns and are useful for constructing networks of highly dimensional datasets. Following identification of feature modules, module expression values can then be correlated against quantitative traits of interest including time, physiological responses, or survival (Langfelder and Horvath 2008; Pei et al. 2017). For example, a study in corals identified metabolite modules using WGCNA and correlated these modules to gene ontology terms obtained from transcriptomics analyses, although they found no significant correlations (Drury et al. 2022). Exploration of the utility of WGCNA approaches in the study of marine organism responses is warranted. However, WGCNA/WCNA approaches are dependent on user-defined parameters (e.g., soft thresholding power) and may be confounded by multicollinearity, autocorrelation, and/or missing values. Correlation is not causation, and researchers should state conclusions from correlation-based approaches appropriately.

ANOVA-simultaneous components analyses (ASCA) can also be useful to examine multivariate responses across time or multiple levels of factors of interest (e.g., modeling the effects of time and experimental levels) (Jansen et al. 2005; Smilde et al. 2005; Bertinetto et al. 2020). The strength of ASCA analyses is the ability to decompose multivariate data according to factors or variables of interest and visualize the effects and is particularly well suited for characterizing changes in -omic responses across time (Jansen et al. 2005). For example, this approach has been used to identify metabolites that contributed to differences by treatment,

lifestage, and their interaction in cuttlefish exposed to ocean acidification conditions (Minet et al. 2025). Similar to unsupervised Principal Components Analyses (PCA), ASCA approaches are sensitive to unbalanced sample design, small sample sizes, and assume linear relationships between multivariate responses and experimental factors (Bertinetto et al. 2020). Additional analyses are required to identify the differential compounds contributing to the effects of experimental variables.

Which compounds drive differences between treatments or groups?

After data exploration through unsupervised analyses and examining experimental effects such as time and treatment variables, the next step is often identification of individual metabolites or lipids (i.e., features) that drive significant differences. Partial least squares discriminant analysis (PLS-DA) can identify metabolites or lipids that distinguish between groups of interest (Kalivodová et al. 2015; Saccenti and Timmerman 2016). PLS-DA analyses are a supervised approach that use group labels (e.g., treatments) to reduce dimensionality by identifying latent variables that explain relationships between predictors and multivariate responses (Saccenti and Timmerman 2016). Feature importance can then be examined as Variable Importance in Projection, or VIP, values that quantify the discriminatory power of individual metabolites or lipids in the PLS-DA model (Galindo-Prieto et al. 2014). It is important to note that a feature's high VIP value indicates statistical importance, but not necessarily biologically importance. Conclusions and interpretations of biological importance of a particular metabolite or lipid must be made by conducting functional or pathway analyses and when contextualized with phenotypic or physiological responses (see *Enabling biological interpretation of metabolic plasticity through enrichment analyses*). PLS-DA analyses are useful for highly dimensional datasets and for biomarker selection, but can be sensitive to sample size imbalance and overfitting when sample size is low (Gromski et al. 2015). Supervised analyses, such as PLS-DA, should be paired with unsupervised analyses such as PERMANOVA tests and

PCA visualizations, or other alternative approaches (Gromski et al. 2015), to validate the influence of experimental variables of interest.

Significance Analysis of Microarray (or Metabolites; SAM) models provide an additional method to identify differential features between treatment groups of interest (Nadon and Shoemaker 2002; Xia and Wishart 2011). SAM models use feature-specific modified t-tests combined with permutation analyses and account for false discovery rate to determine whether a specific feature is differentially present (Nadon and Shoemaker 2002). Use of SAM in conjunction with other multivariate methods demonstrated that heat-hardening upregulates metabolic pathways to promote homeostasis in elevated temperatures in *Mytilus galloprovincialis* mussels (Georgoulis et al. 2022). In contrast, SAM methods identified metabolites that differed by symbiont profiles, but not heat stress, in the coral *Pocillopora acuta* (Haydon et al. 2023). This approach can be useful with highly dimensional datasets and when features may not be independent of each other or when sample sizes are small (Nadon and Shoemaker 2002). However, if sample size is low and differences in feature concentration or relative abundance are minimal, SAM models can fail to detect significant differences (Nadon and Shoemaker 2002).

Machine learning (ML) approaches are increasing in use as “big data” becomes more readily available for biological studies (Greener et al. 2022). Machine learning approaches, broadly, are computer systems that learn or adapt using statistical models to mimic human behavior and recognize patterns and are useful when datasets are too complex, too large, or require automation beyond the capacity of human analysis (Greener et al. 2022). A description of ML approaches and a guide for use in biological sciences has been previously published (Greener et al. 2022) and we point readers to this resource if ML analyses are of interest. Traditional machine learning methods are those that perform tasks including regression, classification, dimensionality reduction, or clustering using limited layers and are therefore appropriate for small data sets and provide the advantage of increased interpretability (Reel et

al. 2021; Greener et al. 2022). Indeed, traditional ML approaches include analyses we have previously discussed, including PCA and PLS-DA. Deep learning models, however, are a subset of ML that utilize neural networks and include many layers to learn hierarchical representations of data (Reel et al. 2021; Greener et al. 2022). Deep learning is appropriate for large and complex datasets that are beyond the scope of human ability to recognize patterns or model high complexity (Reel et al. 2021; Greener et al. 2022). Deep learning approaches require large amounts of data — the more complex the problem, the more data is required — and are "black box" approaches that result in reduced interpretability (Reel et al. 2021). In the sections below we discuss the use of deep learning ML for multi-omic integration. However, just because ML approaches can be used in biological sciences does not mean that they are appropriate for every study. ML approaches require careful attention to choice of models (e.g., supervised or unsupervised), objectives (e.g., clustering, regression, or classification) and proper design of test and training datasets and procedures for training, validating, and testing models (Greener et al. 2022).

Best practices in quantitative analyses: Contextualizing results and applying complementary approaches

Given the large number of data analysis tools available to researchers and the diverse sets of hypotheses tested using lipidomic and metabolomic data, we strongly encourage the field to use multiple complementary analysis approaches to validate findings. Applying multiple statistical approaches to the same metabolomic or lipidomic dataset enhances the reliability, depth, and interpretability of findings. Different statistical methods capture distinct aspects of data structure — univariate analyses identify individual metabolites or lipids that differ significantly between conditions, while multivariate techniques (ex. PCA, PLS-DA) (Kalivodová et al. 2015; Saccenti and Timmerman 2016) and ML approaches reveal patterns and interactions across multiple variables (Reel et al. 2021; Greener et al. 2022). Correlation

network analyses can further uncover biochemical pathway relationships (Langfelder and Horvath 2008; Pei et al. 2017), while time series approaches track dynamic shifts over experimental conditions of interest (Jansen et al. 2005; Smilde et al. 2005; Bertinetto et al. 2020). By integrating multiple statistical strategies, researchers can validate key findings, minimize biases inherent to any single method, and examine broad scale responses as well as mechanistic underpinnings of those responses. For example, combination of PCA, PPLS-DA, and pairwise tests showed that the bryozoan *Bugula neritina* metabolome was largely unchanged after heat stress, demonstrating this species' resilience to high temperature (Gauff et al. 2025). Sensitivity of *Mya arenaria* and *Mya truncata* clams to marine heat waves was examined using PERMANOVA, linear mixed effects models, PCA, and PLS-DA analysis. Using these multiple methods, researchers found that differences in heat tolerance between species was associated by differential use of metabolic pathways (Beaudreau et al. 2024).

Metabolomic and lipidomic data provide valuable insights into the biochemical state of an organism, but these data are effectively interpreted only when contextualized with phenotypic or physiological data. A total of 41 studies of the 68 represented in **Figure 1** paired molecular data with whole-organism physiology or phenotypic data (**Appendix A**). Metabolites and lipids are dynamic molecules that reflect an organism's immediate response to environmental conditions, stressors, or developmental changes (Wenk 2005; Roessner and Bowne 2009; Rey et al. 2022). Without integrating physiological responses and phenotypes such as metabolic rate, growth, reproduction, or survival, it is difficult to determine whether observed molecular shifts correspond to metabolic plasticity and result in either adaptive or maladaptive responses. Clams (*Sinonovacula constricta*) had an increase in Arrhenius breakpoint temperature after heat hardening, and increased glycerophospholipid abundance suggests homeoviscous adaptation at higher temperatures (Zhang and Dong 2021). Marine copepods (*Apocyclops royi*) reared in hyposaline conditions for multiple generations demonstrated reproductive resilience, but metabolomics analysis shows an increase in anaerobic stress is a "cost" to this resilience

(Winding Hansen et al. 2022). The addition of physiological or phenotypic data enables researchers to move beyond descriptive metabolomic or lipidomic profiles and instead contextualize omic data with organismal function, helping to uncover the mechanistic basis of metabolic changes.

Enabling biological interpretation of metabolic plasticity through enrichment analyses

Researchers often interpret complex datasets by linking metabolites and lipids to known biological functions, metabolic pathways, and responses. One approach is to manually map compounds to known pathways of interest to interpret larger scale metabolic shifts. This approach is best used when metabolite and lipid pathways are well characterized and conserved across organisms and when researchers have a hypothesis regarding a specific pathway. Of the 68 studies included in **Figure 1**, 47 manually mapped compounds to known pathways. Many studies used databases like the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000) and Human Metabolome Database (HMDB) (Wishart et al. 2007) to obtain pathway information. In reef-building corals, (Chiles et al. 2022) examined shifts in nitrogen metabolism by examining 15N enrichment of target metabolites and, similarly, (Huffmyer et al. 2024) examined shifts in central carbon metabolism via shifts in enrichment of glycolytic metabolites. Other studies used other tools like MetaMapp (Barupal et al. 2012). Wanamaker et al. (2019) used MetaMapp to assign chemical and biochemical relationships to metabolites of interest, then visualized all known relationships using Cytoscape (Shannon et al. 2003). Using this approach, they found that amino acid metabolism was significantly impacted in Dungeness crab (*Cancer magister*) juveniles exposed to low pH and dissolved oxygen conditions (Wanamaker et al. 2019). This targeted examination of pathways of interests provides a method for testing hypotheses regarding a particular pathway or function.

Enrichment analysis facilitates biological interpretation of molecular datasets by linking changes in individual metabolomic and lipidomic responses with large-scale shifts in broad biological processes. Originally developed for microarrays and RNA-Seq data (Khatri et al. 2012; Zhao and Rhee 2023), enrichment analyses identify biological pathways that are significantly represented in a dataset based on annotation information available for specific features. There are three different kinds of enrichment analyses: ranking-based enrichment, overrepresentation-based enrichment, and network topology-based enrichment (Wright et al. 2015; Ihnatova et al. 2018; Nguyen et al. 2019; Geistlinger et al. 2021; Zhao and Rhee 2023). We focus our discussion on utilization of the latter two methods as overrepresentation and network topology are more commonly employed with metabolomic and lipidomic datasets (Zhao and Rhee 2023).

Overrepresentation-based enrichment methods determine if specific pathways or functions are observed in a target dataset more than expected by chance in comparison to a background dataset (Das et al. 2020; Maleki et al. 2020). Network-topology based enrichment incorporates additional factors that impact pathway activity, such as feature position in a pathway or feature-feature interactions (Bayerlová et al. 2015; Ihnatova et al. 2018; Yang et al. 2019). With both methods, the dataset used to conduct enrichment analysis will impact the interpretation of affected pathways, so it is critical to make an informed analytical choice to identify significantly different metabolites or lipids (see discussion of analyses above) (Chicco and Agapito 2022; Zhao and Rhee 2023). For example, a particular module of metabolites or lipids that correlate with physiological responses (i.e., WCNA analysis) or features that significantly differentiate between treatment groups (i.e., PLSDA and VIP analyses) may be used as input data for enrichment tests. In contrast, using all detected features without prior selection based on biological hypotheses can produce pathway enrichment results that are biologically misleading (Chicco and Agapito 2022). Further, the dataset used as the background context for enrichment analyses can impact the interpretation of the results. A common practice

is to use all detected compounds above a noise threshold (Zhao and Rhee 2023). However, different background sets can be used depending on the question. For example, using all detected compounds as a background may be useful for an untargeted analysis, but may not be an appropriate background for a targeted assay focusing on a specific pathway.

The most commonly used enrichment platform is Metabolomics Pathway Analysis (Xia and Wishart 2010) through the web-based GUI MetaboAnalyst (Xia et al. 2009; Pang et al. 2024) (**Figure 4**). MetaboAnalyst enables overrepresentation analysis (Enrichment Analysis) or a combined overrepresentation and network topology analysis (Pathway Analysis). Users can upload a target list into the Enrichment Analysis module, and choose between a user-defined background or MetaboAnalyst-provided background reference metabolomes based on human data. Significance values for Enrichment Analysis are obtained using Fisher's tests (Xia et al. 2009; Xia and Wishart 2010; Pang et al. 2024). No studies from **Figure 1** used the Enrichment Analysis module from MetaboAnalyst as the sole method of linking compound differences with biological function. The Pathway Analysis module from MetaboAnalyst is more commonly used, with ten studies in **Figure 1** employing this method to assist in biological interpretation of results. In this module, similar options to upload targets and define background datasets are present. Users must select a KEGG pathway library from a species most relevant for their system (last updated by MetaboAnalyst December 2024), and choose between hypergeometric or Fisher's tests for significance metrics. It is important to note that results can vary depending on the pathway reference database selected (**Figure 4**). Therefore, thoughtful selection of the reference database and a clear statement on which database was used is required. Pathway analysis using the *Drosophila* KEGG pathway library identified tricarboxylic acid cycle and amino acid metabolism as significantly impacted in northern shrimp (*Pandalus borealis*) exposed to ocean acidification and warming. (Guscelli et al. 2023). These complementary MetaboAnalyst methods can be employed together: Noisette et al. (2021) used both over-representation analysis and pathway topology analysis to identify metabolites underpinning

differences in ocean acidification tolerance in *H. americanus*, and similarly Nguyen et al. (2021) used both methods to investigate heat tolerance in the abalone *Haliotis iris*. MetaboAnalyst can be used in conjunction with other enrichment methods. Hillyer et al. (2017) used the Pathway Activity Profiling (PAPi) R package (Aggio et al. 2010) to obtain pathway information for metabolite data from heat-stressed *Acropora aspera* corals, then used that information as input with MetaboAnalyst to get statistical information. The use of MetaboAnalyst is limited by compound annotation databases and prior network topology knowledge (Zhao and Rhee 2023).

While specific platforms have been developed primarily for metabolomics analyses, they can be used for lipid enrichment as well. For example, users can upload a list of target lipids into the MetaboAnalyst Enrichment Analysis module for overrepresentation analysis. Lipid-specific enrichment tools are becoming more common. The Lipid Ontology (LION) enrichment analysis web application (LION/web) is comparable to MetaboAnalyst for lipidomics datasets (Molenaar et al. 2019). LION/web facilitates overrepresentation- and ranking-based enrichment for lipid ontology terms associated with a target list of lipids (Molenaar et al. 2019). The overrepresentation-based Target List analysis allows users to define a background list, and uses Fisher's tests to assess statistical significance. While none of the studies in **Figure 1** conducted enrichment analysis for lipidomics data, LION/web has been employed to understand the lipidome of other organisms and tissue types such rat hepatic cells (Molenaar et al. 2023) and cetacean blubber (Bories et al. 2021).

Compound nomenclature is a source of variability and inconsistency for metabolomic and lipidomic enrichment analyses, especially in non-model organisms (**Figure 4**). Several enrichment platforms, including those listed above, use data from humans and other model-systems for their reference databases, and web-based enrichment platforms can also have outdated pathway annotations (Wadi et al. 2016). This can lead to two issues when using enrichment platforms for non-model marine systems: artificially inflated pathway size and nomenclature mismatches (Zhao and Rhee 2023). Reference databases can use different

ontologies (eg. KEGG) to define pathways, which can change the number of compounds in a specific pathway. When more molecules are assigned to a pathway, then more differentially abundant targets are needed to identify that pathway as significantly enriched, changing the biological interpretation of metabolic plasticity for a specific dataset (Karp et al. 2021). In addition to potential pathway misclassification, compound nomenclature variation can lead to data not being used in enrichment. For example, if a named compound does not have a matching name in the MetaboAnalyst database and sequencer data, and cannot be identified through alternative methods (eg. PubChem, KEGG), then that compound is discarded from analysis. For example, compounds that have a general nomenclature in a data set (e.g., “glucose”) may not map to expected pathways in MetaboAnalyst due to variation in specific nomenclature of isomers (e.g., “alpha-d-glucose”), which may alter interpretation of pathway enrichment (**Figure 4**).

Regardless of which kind of enrichment analysis is performed, researchers should interrogate how appropriate that approach is for their study system and hypotheses. The study hypotheses should guide which features are used for enrichment analysis. Researchers should consider the appropriateness of the background set used for enrichment, and when possible, define study-specific background sets for enrichment. In addition to using enrichment platforms with appropriate and recently updated reference databases, multiple enrichment tools should be used for analyses to corroborate results and increase confidence in enrichment findings (Chicco and Agapito 2022). Analysis output should be groundtruthed by examining changes in features involved in enriched pathways, and by considering how molecular changes correspond with physiology data. We encourage referencing previously published papers on enrichment analysis, including those that compare methods for metabolomic datasets, to assist in decision-making (Ma et al. 2019; Chicco and Agapito 2022; Mubeen et al. 2022; Wijesooriya et al. 2022; Zhao and Rhee 2023).

Linking responses across different levels of biological organization through multi-omic integration

Increasing generation and availability of large molecular datasets presents a challenge in effectively integrating these data to understand organismal responses with improved mechanistic interpretations. Analyzing data using molecular datasets at different levels of biological organization enhances the robustness and depth of scientific conclusions, allows researchers to bridge molecular mechanisms with functional outcomes, and reveals interactions that may be overlooked in single-omics studies. Each approach provides unique insights: transcriptomics identifies gene expression patterns; metabolomics provides insights on shifts in metabolic pathways; lipidomics captures membrane dynamics and energy storage; and epigenomics reveals regulatory modifications.

Integration of two molecular approaches is a commonly used approach (**Figure 1C**; **Appendix A**). Of the 68 studies in **Figure 1**, 19 used gene expression and metabolomics to understand the molecular underpinnings of metabolic responses. For example, concurrent analysis of transcripts and metabolites can elucidate the relevant level of biological organization impacted by environmental stress in *C. gigas*, which exhibited an altered amino acid, carbohydrate, and fatty acid metabolite profiles in response to ocean acidification (Liu et al. 2020). These changes were not only associated with downregulation of corresponding genes, but also reductions in calcification gene expression (Liu et al. 2020). Similarly, thermal stress elicited changes to gene expression and metabolites associated with redox pathways in the rice coral *Montipora capitata* (Williams, Chiles, et al. 2021) and analysis of genes and metabolites in early developmental stages in *M. capitata* reveal developmental shifts in metabolism (Huffmyer et al. *in press*). In the Pacific white shrimp (*Penaeus vannamei*), correlation and network analyses of metabolomic and transcriptomic data revealed that regulation of amino acid and

lipid metabolism increased energy availability under cold stress (Zhu et al. 2024). Use of metabolomic and lipidomic responses is also common, with five studies employing these approaches together (**Figure 1C**; **Appendix A**). Combined metabolomic and lipidomic analysis (e.g., (Reddy et al. 2023)) can reveal changes in active metabolic pathways and energy storage. Examination of lipidomic responses, in particular, complements metabolomic data and provides a detailed view of lipid storage, lipid metabolism, signaling, and cellular membrane state (see reviews in (Imbs et al. 2021; Rey et al. 2022)). For example, Costa et al. (2024) utilized lipidomic and metabolomic approaches to assess the impact of red tides on core metabolic pathways in reef building corals. In the coral *Pocillopora damicornis*, lipids and metabolites provided predictive biomarkers of organism performance (Sogin et al. 2016), and combined lipid and metabolomic analyses revealed shifts in lipid metabolism during reproductive maturation in the mud crab *Scylla paramamosain* (Fu et al. 2022).

While it is clear that integrative multi-omic approaches improve our mechanistic understanding of organism responses, there are significant barriers in conducting this work. For example, examining organismal response to environmental stress would ideally include measurements of molecular mechanisms (e.g., epigenetics and gene expression), metabolic responses (e.g., metabolomics and lipidomics), and physiological and phenotypic measurements. This is often not feasible due to limitations in biological material available for sampling, limited time, limited personnel, and high cost of molecular approaches. Here, we discuss the state of multi-omic integration in the study of marine invertebrates and offer recommendations to move the study of metabolic plasticity towards integrative approaches.

Methodologies for integrative analysis

The intended outcome of integrative approaches will depend on the specific questions and hypotheses. Approaches to integrate multiple -omic data sets generally fall into two categories: 1) individual analysis of each data set followed by qualitative integrative

interpretation; and 2) quantitative integration of data sets in joint statistical analyses (see review in (Santiago-Rodriguez and Hollister 2021)). Here, we will highlight examples of each approach and provide recommendations on using these integrative approaches in marine invertebrate systems. We propose that multi-omic examinations of organism responses should include both individual -omic examination and quantitative integration of multi-omic data when appropriate and relevant to biological hypotheses.

Individual analysis of single-omic layers is a necessary step to identify strong signals and patterns at each level of biological organization and ensures proper outlier identification and correction of batch effects prior to more complex multi-omic approaches (Santiago-Rodriguez and Hollister 2021). The patterns detected in single-omic analyses can help form biologically informed hypotheses that may then be pursued through multi-omic integration. One approach to considering data across multiple layers is to conduct a qualitative comparison then narrate a biological story using conclusions from single-omic analyses. The approach is more common with marine invertebrate studies (18 of studies in **Figure 1** used qualitative integrative interpretation). For example, transcriptomics, metabolomics, and physiological data were used to study metabolic shifts across reef-building coral development, in which the authors discuss potential interactions and relationships between molecular data (Huffmyer et al. *in press*). Also in corals, Putnam et al. (2016) utilized metabolomics and DNA methylation to examine plasticity in response to ocean acidification and qualitatively discussed relationships between the two -omics. Some studies use integrative figures that show molecular data layers mapped onto shared pathways (Wanamaker et al. 2019; Ren et al. 2020; Sun et al. 2021; Zhu et al. 2024), which can assist in making sense of highly dimensional data. It is important to shape these investigations using biology-driven questions and fully report the limitations of qualitative comparisons when providing evidence for mechanistic explanations. Further, it is critical to consider the interplay and interactions of multiple partners in holobiont systems, such as corals (Williams 2024).

Quantitative multi-omic integration offers a powerful way to move beyond side-by-side single-omic layer comparisons and address questions and hypotheses on how organisms respond across multiple biological scales. In practice, this includes using statistical and computational methods to combine omic layers (e.g., transcriptomics, lipidomics, and metabolomics) and uncover relationships and patterns that may not be apparent when analyzing each dataset separately (Santiago-Rodriguez and Hollister 2021; Greener et al. 2022). There is underexplored potential to utilize quantitative integration through statistical analyses, which are more common in biomedical contexts (e.g., see review in (Reel et al. 2021)). Only nine of studies in **Figure 1** used a quantitative or statistical approach to integrate molecular datasets and are largely correlation based in approach.

Several statistical approaches used for individual molecular datasets can be applied towards integrative analysis. Coexpression and correlation-based approaches, such as WGCNA and network analyses, are well suited for identifying groups of genes, lipids, or other features that change together across samples and highlight shared biological functions or coordinated pathways (Sun et al. 2022; Geng et al. 2024; Zhou et al. 2024; Zhu et al. 2024). Other methods, including DIABLO and PLS-DA analyses (e.g., (Sun et al. 2022; Jing et al. 2023; Zhou et al. 2024)), focus on selecting the most important features that differentiate between treatments or groups of interest, which can provide a tool for multi-omic biomarker discovery or building predictive models (Zhang et al. 2011; Young and Alfaro 2018; Sweet et al. 2021). For example, Sun et al. (2021) utilized pairwise correlations to examine metabolic responses to salinity in the clam *R. philippinarum* and Geng et al. (2024) conducted correlation network analyses to identify genes and metabolites correlated with biomarkers in the blue mussel (*M. galloprovincialis*).

Machine learning (ML) approaches are becoming increasingly common in multi-omic studies and provide an approach to identify hidden or complex patterns, networks, and relationships in large data sets, automate analyses, and make forecasts and predictions (Reel et al. 2021; Greener et al. 2022). While some ML methods remain difficult to interpret (i.e.,

“black box” methods), ML offers unique advantages when working with highly dimensional data or when uncovering relationships or patterns that are complex (Reel et al. 2021; Greener et al. 2022). A variety of integration strategies exist depending on how and when the data are brought together and we refer readers to recent reviews that explain these approaches in detail (Reel et al. 2021; Greener et al. 2022; Manochkumar et al. 2023). Concatenation-based methods combine all raw features into a single matrix while model-based approaches analyze each omic independently before integrating model outputs (Reel et al. 2021). Transformation-based strategies, in contrast, reduce each data set into lower dimensions or conduct omic-specific transformations prior to building a joint model (Reel et al. 2021). Integration methods can be unsupervised (e.g., cluster analyses, factor analyses, Bayesian approaches), which is used to discover patterns and structure in a data sets, or supervised (e.g., Bayesian networks, support vector machines, hierarchical classifiers, ensemble-based methods) in order to make predictions and classifications (Reel et al. 2021; Greener et al. 2022). For example, a metabolomics and transcriptomics study in corals used “MAGI”, which provides a method for integration of metabolite and gene information (Erbilgin et al. 2019), to study metabolite-gene interactions in *M. capitata* under thermal stress (Williams, Pathmanathan, et al. 2021). Using gene-metabolite interaction networks, the authors identified redox pathways involved in quenching reactive oxygen species during stress (Williams, Pathmanathan, et al. 2021). Ultimately, the choice of method should be guided by the biological question, the size and quality of data, and the interpretability of the output (Greener et al. 2022; Manochkumar et al. 2023). Machine learning approaches for multi-omic integration have promising applications for the study of marine invertebrate plasticity, but are not widely employed. We direct readers to previous reviews that discuss challenges in multi-omic integration in marine systems for a more detailed discussion (Manochkumar et al. 2023).

Best practices for integrative analysis

As with any analytical approach, best practices for multi-omic integration must be grounded in clear biological questions and hypotheses. Supervised methods are best suited for predictive tasks, such as classifying phenotypes or forecasting physiological outcomes; unsupervised methods are appropriate for discovering structure, patterns, or molecular subtypes within the data; and network-based or regression models may be useful when the goal is to infer mechanisms or relationships among layers of data (Reel et al. 2021; Greener et al. 2022; Manochkumar et al. 2023). Regardless of the approach, thoughtful preprocessing (e.g., normalization, transformation, filtering) is essential as molecular data often differ in scale, distribution, and feature count and high-dimensional data can easily overfit small datasets if not properly constrained with cross-validation or regularization techniques (Reel et al. 2021; Manochkumar et al. 2023).

We recommend building a foundation of single-omic analyses before layering in complexity in order to develop biologically relevant hypotheses and drive the responsible use of more complex integration approaches. Multi-omic integration should be used to explore biologically driven hypotheses in greater depth and uncover patterns that aren't detectable in analysis of single-omic levels. However, relying solely on single-omic approaches or one integration approach may lead to over simplification of the results or one integration approach may lead to overly simplistic conclusions or missing deeper relationships. Correlative strategies, while informative, must be interpreted appropriately and the limitations of correlation approaches need to be acknowledged. Another challenge is that many integration platforms are designed for human or model system datasets and are not always compatible with non-model organisms or complex experimental designs (e.g., MetaboAnalyst). However, we can learn from biomedical research, where multi-omic integration has driven advances in health and medical research (Acharjee et al. 2016; Beaulieu-Jones et al. 2019; Triantafyllidis and Tsanas 2019;

Ghassemi et al. 2020; Rubinger et al. 2023; Jain and Jain 2024). These approaches are beginning to be applied to ecological and evolutionary biology (see reviews in (Olden et al. 2008; Christin et al. 2019; Lürig et al. 2021; Greener et al. 2022; Pichler and Hartig 2023)), and we argue that they are particularly needed in the study of non-model systems, such as marine invertebrates where stress responses involve complex coordination across biological levels and are not always apparent through a single-omic layer lens. With careful application, these tools can help identify regulatory drivers of resilience, build predictive models of organismal health, and illuminate new layers of biological complexity in systems where mechanistic understanding has traditionally been limited.

Conclusion

Metabolomics and lipidomics are powerful tools for examining metabolic plasticity in non-model marine invertebrates. We encourage researchers to design clear, testable hypotheses and use them to guide molecular investigations. Given the complexity of these data types, appropriate statistical analysis may include complementary univariate and multivariate approaches to identify compounds of interest, and pairing manual and programmatic pathway mapping and enrichment methods to understand the biological significance of results. Pairing molecular data with physiology and/or phenotype information may elucidate sublethal impacts of stress and provide a holistic understanding of organismal resilience. When appropriate, we also encourage researchers to pair metabolomic or lipidomic data with metrics at different levels of biological organization such as transcriptomics or proteomics.

We also identify areas of growth for the application of these methodologies to organismal biology. First and foremost, thorough reporting in manuscripts is necessary to provide context and improve reproducibility. Of the 68 studies identified in **Figure 1**, several were not explicit about whether or not targeted, semi-targeted, or untargeted data acquisition methods were used, or presented pathway or enrichment results without specifying methods. The lack of this

necessary information makes it difficult for newer researchers to understand best practices for the field. Raw data should ideally be housed in publicly accessible data repositories (Santiago-Rodriguez and Hollister 2021), similar to the NCBI Short Read Archive or Gene Expression Omnibus. Existing databases include Metabolomics Workbench (https://www.metabolomicsworkbench.org/) and MassIVE (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp). Effort should be made to improve annotation databases for non-model systems to use for compound identification. More accurate databases can facilitate improved enrichment analysis or quantitative integration with other datasets.

Data availability

All appendices, figures, and metadata are available at https://github.com/yaaminiv/ICB-perspective-piece.

Author contributions

YRV and ASH jointly conceived, wrote the initial draft, and revised the manuscript. YRV created the figures with input from ASH. All authors reviewed and approved the final manuscript.

Conflict of Interest Statement

The authors have no conflicts of interest to report.

Acknowledgements

This work was supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2209018 to YRV. National Science Foundation Rules of Life-Epigenetics (award no. EF-1921465) provided salary support for ASH. We would like to thank members of

the Tepolt Lab at Woods Hole Oceanographic Institution for their feedback on a draft of this manuscript.

Boxes

BOX 1 | Glossary

Metabolomics: The study of chemical processes involving the small molecules that are the direct and indirect products of metabolic pathways. Metabolites are often classified into primary metabolites involved in growth, development, and reproduction, and secondary metabolites that are more important for ecological function.

Lipidomics: The study of complete fatty acid and lipid profiles in a sample. Lipid molecules are crucial for short- and long-term energy storage.

Plasticity: Ability of one genotype to produce more than one phenotype.

Metabolic plasticity: An organism's capacity to modulate energy production, allocation, and use.

Steady state: Absolute or relative concentrations of a compound at a particular point in time.

Metabolic flux: Rate at which compounds pass through a metabolic pathway.

Liquid chromatography mass spectrometry (LC-MS): Combines physical separation capability of liquid chromatography with mass analysis capability of mass spectrometry. Can characterize a wide range of molecule types, and is considered more robust for lipidomic applications.

Gas chromatography mass spectrometry (GC-MS): Combines features of gas chromatography with mass spectrometry. Best suited for volatile molecules, and is considered more robust for metabolomic applications.

Nuclear magnetic resonance (NMR): Quantify compounds by placing a sample in a magnetic field and using the inherent magnetic properties to identify the compounds. 1H-NMR is most commonly used with metabolomics data.

Untargeted: Provides relative feature abundance differences between experimental conditions for all detected compounds.

Targeted: Provides absolute quantitation, or specific concentrations of molecules, based on an input list of known compounds.

Semi-targeted: Quantifies known compounds and detects unknown or unidentified compounds.

Permutational analysis of variance (PERMANOVA): Unsupervised multivariate approach used to test if centroids of groups are significantly different from each other. Often paired with permutational analyses of dispersion (PERMDISP).

Weighted gene co-expression network analysis (WGCNA): Correlation-based approach originally developed to identify groups of genes that shared expression patterns. Can be applied to identify metabolites or lipids with shared abundance patterns.

ANOVA-simultaneous components analysis (ASCA): Decomposes multivariate data according to variables of interest. Useful for examining multivariate responses across time or multiple additional factors.

Principal components analysis (PCA): Unsupervised molecular approach generally used for exploratory analysis. Reduces the number of dimensions in large datasets to principal components that retain most of the original information.

Partial least squares discriminant analysis (PLS-DA): Supervised approach that uses group labels to reduce dimensionality by identifying variables that explain relationships between predictors and responses. Orthogonal PLS-DA (OPLS-DA) is a variation commonly used with metabolomic and lipidomic data.

Variable Importance in Projection (VIP): Quantifies the discriminatory power of a compound from a PLS-DA model.

Significance Analysis of Microarray (SAM): Feature-specific modified t-tests combined with permutation analysis used to determine in a specific compound has differential abundance between experimental conditions.

Machine learning (ML): Computer systems that learn or adapt using statistical models to mimic human behavior and recognize patterns. Useful when datasets are complex, large, or require automation.

Overrepresentation-based enrichment: Determines if specific pathways or functions are observed in a target dataset more than expected by chance in comparison to a background dataset.

Network topology-based enrichment: Incorporates additional factors that impact pathway activity, such as feature position in a pathway or feature-feature interactions, into an enrichment analysis.

Figures

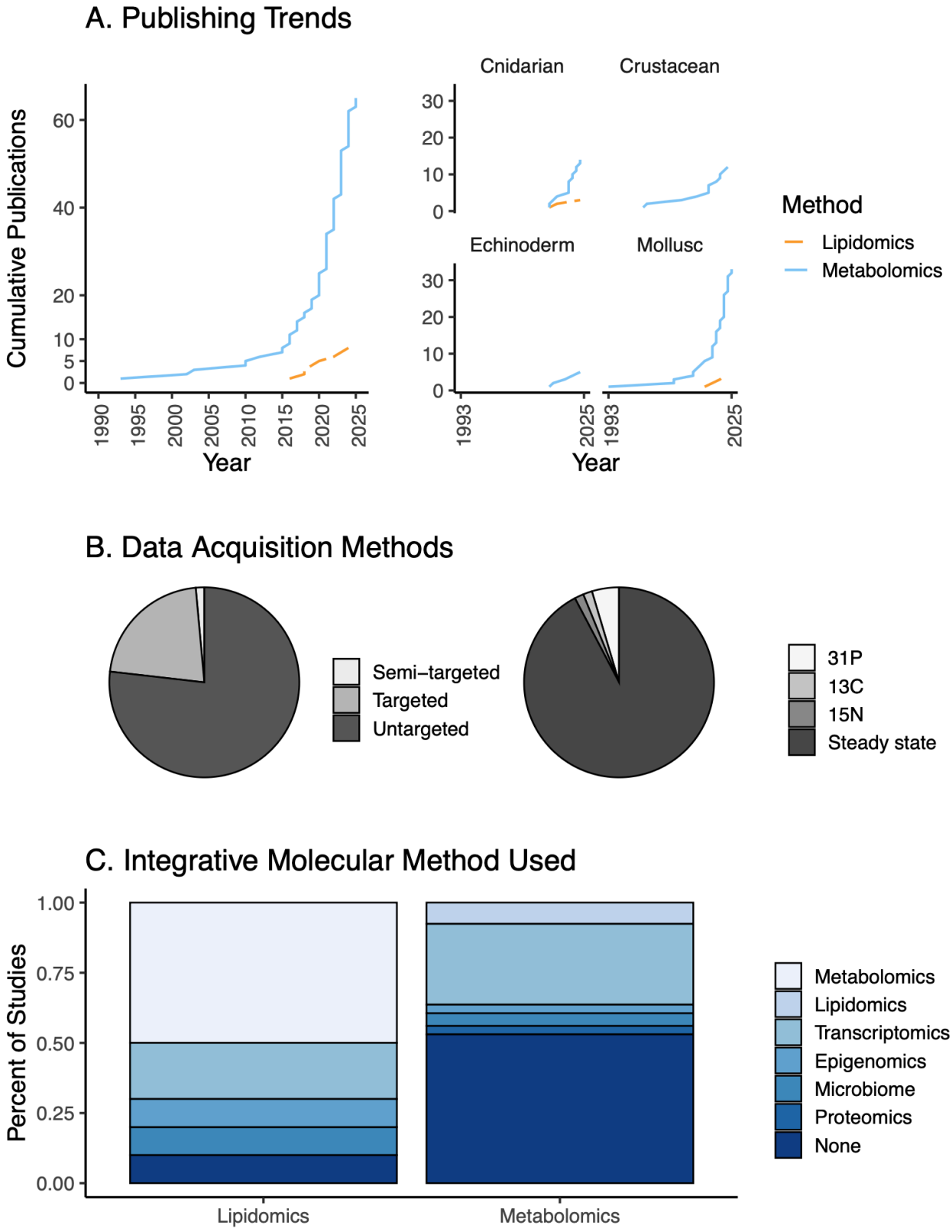

A. Publishing Trends
Cumulative Publications
Year
Cnidarian
Crustacean
Echinoderm
Mollusc
Method
Lipidomics
Metabolomics
B. Data Acquisition Methods
Semi−targeted
Targeted
Untargeted
31P
13C
15N
Steady state
C. Integrative Molecular Method Used
Percent of Studies
Lipidomics
Metabolomics
Metabolomics
Lipidomics
Transcriptomics
Epigenomics
Microbiome
Proteomics
None

Figure 1. Metadata for 68 published studies examining metabolic plasticity. Studies were papers examining metabolome and/or lipidome responses to environmental stress in marine invertebrates. See **Appendix A** for search terms and results. **A**) Cumulative metabolomic (blue line) and lipidomics (orange dashed line) papers published between 1993 and 2025. Published research papers were identified through Web of Science and ProQuest searches and supplemented with manual searches through Google Scholar. In this timeframe, 65 used metabolomics and eight studies used lipidomics. Publishing trends are also shown by phylum for cnidarian (17), crustacean (13), echinoderm (5), and molluscs (37). Two or fewer papers were published for annelids, brachiopods, and bryozoans each over this time frame, and therefore are not visualized separately. **B**) Data acquisition methods of papers in **Appendix A**. Metabolomics and lipidomics data were collected primarily using untargeted experiments (77.9%), followed by targeted studies (22.1%), then semi-targeted studies (1.5%). One study used both targeted and untargeted methods. The majority of studies (92.6%) examined steady state responses, while 7.4% used 31P, 13C, or 15N labeling methods to understand metabolic flux. **C**). Other molecular methods used to integratively study metabolic plasticity with either lipidomics or metabolomics. A total of 30 of 65 metabolomics studies integrated an additional molecular method, while 7 of 8 lipidomics studies used an additional molecular method. Only two studies used more than two molecular methods (Rodriguez-Casariego: (2023): lipidomics, epigenomics, transcriptomics, microbiome; Wei et al. (2015): metabolomics, transcriptomics, proteomics).

Experimental design considerations

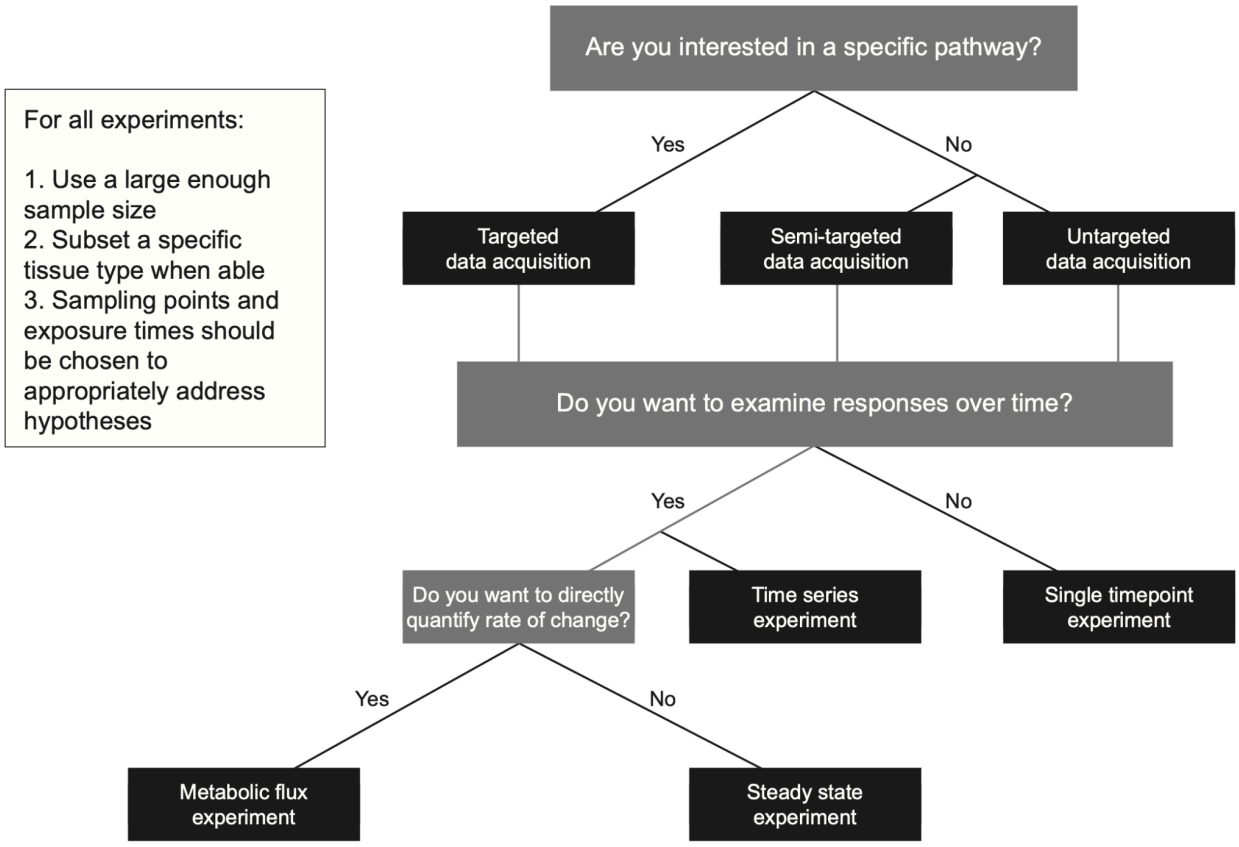


Figure 2. Decision tree for metabolomic and lipidomic experimental design. Researchers should consider if they are interested in a specific pathway, if they want to examine changes in the metabolome or lipidome over time, or quantify rate of change of compounds. For all experiments, researchers should use appropriate sample sizes, tissue types, and sampling points to address hypotheses.

Analytical considerations

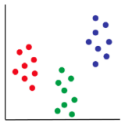
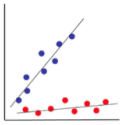
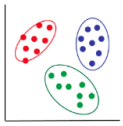
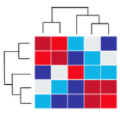
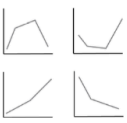
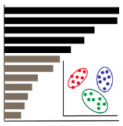
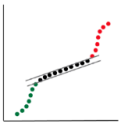
	Single metabolite tests		Unsupervised analyses			Supervised analyses		
	ANOVA	Linear models	PERMANOVA	WGCNA	ASCA	PLS-DA + VIP	SAM	ML
								
How does concentration or composition change between treatments?	✓	✓				✓	✓	✓
How do compounds correlate with quantitative responses or time?			✓	✓	✓			✓
What compounds drive differences between treatments?						✓	✓	✓

Figure 3. Analytical considerations for metabolomic and lipidomic experiments. This table indicates suitable analytical options (*ie.*, single metabolite tests, unsupervised analysis, or supervised analyses) for different experimental objectives (*ie.*, examining changes to concentration or composition between treatments; examining correlations between compounds and either quantitative responses or time; or identifying compounds that drive differences between treatments), and provides an example of what the output visualization may look like. Researchers should determine which methods are most appropriate for their data and hypotheses.

Enrichment with MetaboAnalyst

A. Pros and cons of MetaboAnalyst for enrichment

Pros	Cons
• Tools available to analyze data from raw spectra through enrichment and visualization • Interactive platform with no coding required • Provides the foundation for other analytical packages	• Extended capabilities for untargeted data, but better suited for targeted data • Issues with shared nomenclature or inability to identify molecules • Limited available reference databases

B. Different KEGG databases impact output

Human Pathway	Match Status	FDR	Impact
Starch and sucrose metabolism	6/17	0.003	0.62
Nitrogen metabolism	3/6	0.053	0
Galactose metabolism	4/15	0.059	0.05
Valine, leucine, and isoleucine biosynthesis	3/8	0.059	0

C. elegans Pathway	Match Status	FDR	Impact
Starch and sucrose metabolism	6/16	0.005	0.54
Phenylalanine metabolism	3/6	0.07	0.5
Amino sugar and nucleotide sugar metabolism	6/31	0.07	0.31
One carbon pool by folate	5/23	0.07	0.11

C. Nomenclature inconsistencies impact enrichment results

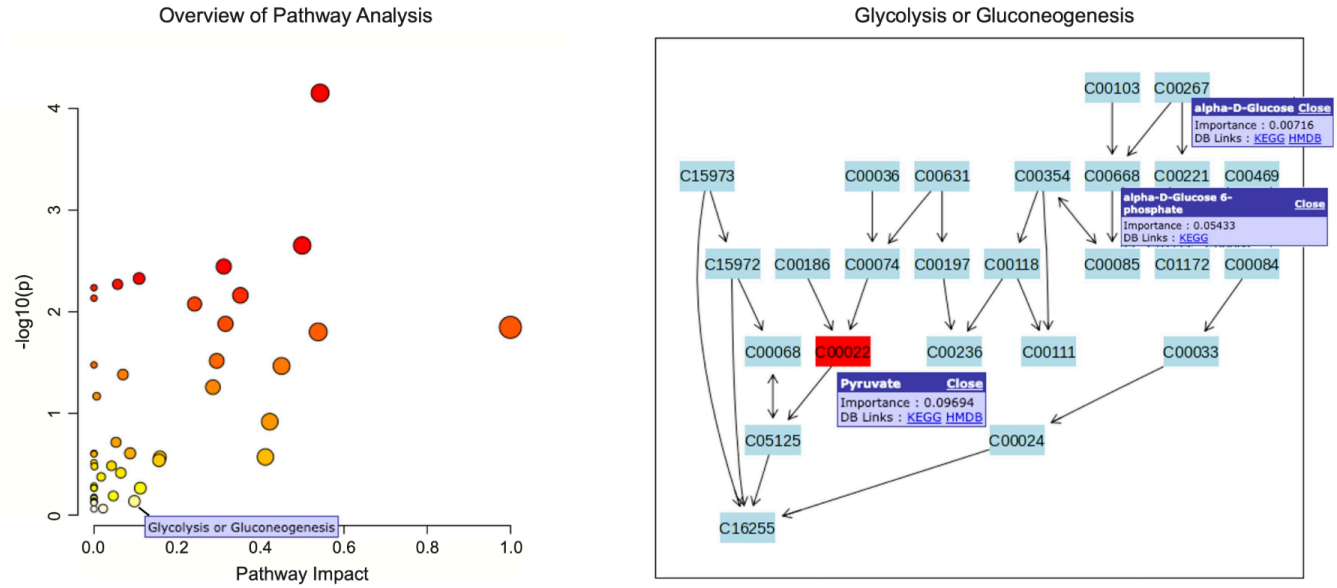


Figure 4. Enrichment with MetaboAnalyst. **A**) Pros and cons of using MetaboAnalyst. Panels **B**) and **C**) show MetaboAnalyst (v6.0) Pathway Analysis output for an example differential metabolite dataset generated in a previous study of reef-building coral early life stages (Huffmyer et al. *in press*) to highlight potential challenges in using MetaboAnalyst with non-model marine invertebrate species. This dataset is available in **Appendix B**. **B**) Differences in pathway analysis output based on reference KEGG database. Pathway analysis was conducted using either humans (*Homo sapiens*) or *Caenorhabditis elegans* as KEGG references. The top four pathway results are shown. While the top pathway did not change based on the database, there are differences in the remaining pathways identified, the number of metabolites in the dataset that match the database (Match Status), FDR, and pathway impact values. **C**) KEGG pathway analysis results using *C. elegans* as a reference. Arrow and label box indicate the enrichment of glycolysis or gluconeogenesis pathways. FDR *P*-value is indicated by color. These pathways were not significantly enriched (*P*-value = 1.0) and were considered low impact (impact = 0.10). Details of the glycolysis and gluconeogenesis KEGG

pathway are shown, with red boxes indicating metabolites in the test set that matched to the KEGG pathway. While the dataset included core metabolites in the glycolysis and gluconeogenesis pathways such as “glucose”, “glucose-6-phosphate”, and “pyruvate,” only pyruvate was recognized as a hit by MetaboAnalyst. This is due to differences in the nomenclature of glucose and glucose-6-phosphate required by Metaboanalyst to match to pathways (“alpha-D-Glucose” and “alpha-D-Glucose 6-phosphate”, respectively, as indicated in text label boxes). These results demonstrate that nomenclature and specificity of nomenclature can limit pathway analysis results in databases that rely on particular nomenclature. Results were not different when running against the *C. elegans* (nematode), *Strongylocentrotus purpuratus* (urchin)*, Mus musculus* (mouse), or human KEGG databases. Note that in the study (Huffmyer et al. *in press*) acknowledged this limitation and additionally examined glycolytic metabolic pathways through individual metabolite abundance.

Supplementary Information

Appendix A. Web of Science and ProQuest search terms and results for marine invertebrate studies examining metabolic plasticity in response to environmental stressors. Searches were conducted using a University of Washington login for Web of Science and ProQuest on March 24, 2025 and April 4, 2025, respectively. An additional 27 papers were added manually from Google Scholar searches.

Appendix B. Test case metabolite dataset from (Huffmyer et al. *in press*) for illustrative purposes using Metaboanalyst (v6.0) platform.

References

Acharjee A, Ament Z, West JA, Stanley E, Griffin JL. 2016. Integration of metabolomics, lipidomics and clinical data using a machine learning method. BMC Bioinformatics 17:440.

Aggio RBM, Ruggiero K, Villas-Bôas SG. 2010. Pathway Activity Profiling (PAPi): from the metabolite profile to the metabolic pathway activity. Bioinformatics 26:2969–76.

Anderson MJ. 2017. Permutational multivariate analysis of variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online.

Angelcheva L, Mishra Y, Antti H, Kjellsen TD, Funk C, Strimbeck RG, Schröder WP. 2014. Metabolomic analysis of extreme freezing tolerance in Siberian spruce (*Picea obovata*). New Phytol 204:545–55.

Balcells C, Foguet C, Tarragó-Celada J, de Atauri P, Marin S, Cascante M. 2019. Tracing metabolic fluxes using mass spectrometry: Stable isotope-resolved metabolomics in health and disease. Trends Analyt Chem 120:115371.

Barupal DK, Haldiya PK, Wohlgemuth G, Kind T, Kothari SL, Pinkerton KE, Fiehn O. 2012. MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. BMC Bioinformatics 13:99.

Bayerlová M, Jung K, Kramer F, Klemm F, Bleckmann A, Beißbarth T. 2015. Comparative study on gene set and pathway topology-based enrichment methods. BMC Bioinformatics 16:334.

Beale DJ, Pinu FR, Kouremenos KA, Poojary MM, Narayana VK, Boughton BA, Kanojia K, Dayalan S, Jones OAH, Dias DA. 2018. Review of recent developments in GC-MS approaches to metabolomics-based research. Metabolomics 14:152.

Beauclercq S, Grenier O, Arnold AA, Warschawski DE, Wikfors GH, Genard B, Tremblay R, Marcotte I. 2023. Metabolomics and lipidomics reveal the effects of the toxic dinoflagellate *Alexandrium catenella* on immune cells of the blue mussel, *Mytilus edulis*. Harmful Algae 129:102529.

Beaudreau N, Page TM, Drolet D, McKindsey CW, Howland KL, Calosi P. 2024. Using a metabolomics approach to investigate the sensitivity of a potential Arctic-invader and its Arctic sister-species to marine heatwaves and traditional harvesting disturbances. Sci Total Environ 170167.

Beaulieu-Jones B, Finlayson SG, Chivers C, Chen I, McDermott M, Kandola J, Dalca AV, Beam A, Fiterau M, Naumann T. 2019. Trends and focus of machine learning applications for health research. JAMA Netw Open 2:e1914051.

Bennett BD, Yuan J, Kimball EH, Rabinowitz JD. 2008. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat Protoc 3:1299–1311.

Bertinetto C, Engel J, Jansen J. 2020. ANOVA simultaneous component analysis: A tutorial

review. Anal Chim Acta X 6:100061.

Bingol K. 2018. Recent advances in targeted and untargeted metabolomics by NMR and MS/NMR methods. High Throughput 7:9.

Blaise BJ, Correia G, Tin A, Young JH, Vergnaud A-C, Lewis M, Pearce JTM, Elliott P, Nicholson JK, Holmes E, Ebbels TMD. 2016. Power analysis and sample size determination in metabolic phenotyping. Anal Chem 88:5179–88.

Blaženović I, Kind T, Ji J, Fiehn O. 2018. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites 8.

Bories P, Rikardsen AH, Leonards P, Fisk AT, Tartu S, Vogel EF, Bytingsvik J, Blévin P. 2021. A deep dive into fat: Investigating blubber lipidomic fingerprint of killer whales and humpback whales in northern Norway. Ecol Evol 11:6716–29.

Breitling R, Ritchie S, Goodenowe D, Stewart ML, Barrett MP. 2006. Ab initio prediction of metabolic networks using Fourier transform mass spectrometry data. Metabolomics 2:155–64.

Broadhurst DI, Kell DB. 2007. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2:171–96.

Cajka T, Fiehn O. 2014. Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. Trends Analyt Chem 61:192–206.

Cajka T, Fiehn O. 2016. Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. Anal Chem 88:524–45.

Carriot N, Paix B, Greff S, Viguier B, Briand J-F, Culioli G. 2021. Integration of LC/MS-based molecular networking and classical phytochemical approach allows in-depth annotation of the metabolome of non-model organisms - The case study of the brown seaweed *Taonia atomaria*. Talanta 225:121925.

Checa A, Bedia C, Jaumot J. 2015. Lipidomic data analysis: tutorial, practical guidelines and applications. Anal Chim Acta 885:1–16.

Chen L, Yu F, Shi H, Wang Q, Xue Y, Xue C, Wang Y, Li Z. 2022. Effect of salinity stress on respiratory metabolism, glycolysis, lipolysis, and apoptosis in Pacific oyster (*Crassostrea gigas*) during depuration stage. J Sci Food Agric 102:2003–11.

Chicco D, Agapito G. 2022. Nine quick tips for pathway enrichment analysis. PLoS Comput Biol 18:e1010348.

Chiles EN, Huffmyer AS, Drury C. 2022. Stable isotope tracing reveals compartmentalized nitrogen assimilation in scleractinian corals. Frontiers in Marine.

Chokkathukalam A, Kim D-H, Barrett MP, Breitling R, Creek DJ. 2014. Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. Bioanalysis 6:511–24.

Christin S, Hervet É, Lecomte N. 2019. Applications for deep learning in ecology. Methods Ecol Evol 10:1632–44.

Costa KA, Maldonado A, Donaldson J, Guingab-Cagmat J, Garrett T, Bowden JA. 2024. Examining the sublethal impacts of environmentally relevant concentrations of Karenia brevis on three South Florida reef-building coral species using a multi-omics workflow. Coral Reefs 43:1703–16.

Creek DJ, Chokkathukalam A, Jankevics A, Burgess KEV, Breitling R, Barrett MP. 2012. Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. Anal Chem 84:8442–47.

Das S, McClain CJ, Rai SN. 2020. Fifteen years of gene set analysis for high-throughput genomic data: A review of statistical approaches and future challenges. Entropy (Basel) 22:427.

de Jonge NF, Mildau K, Meijer D, Louwen JJR, Bueschl C, Huber F, van der Hooft JJJ. 2022. Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools. Metabolomics 18:103.

DiLeo MV, Strahan GD, den Bakker M, Hoekenga OA. 2011. Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. PLoS One 6:e26683.

Doroghazi JR, Albright JC, Goering AW, Ju K-S, Haines RR, Tchalukov KA, Labeda DP, Kelleher NL, Metcalf WW. 2014. A roadmap for natural product discovery based on large-scale genomics and metabolomics. Nat Chem Biol 10:963–68.

Downey-Wall AM, Cameron LP, Ford BM, McNally EM, Venkataraman YR, Roberts SB, Ries JB, Lotterhos KE. 2020. Ocean Acidification Induces Subtle Shifts in Gene Expression and DNA Methylation in Mantle Tissue of the Eastern Oyster (*Crassostrea virginica*). Frontiers in Marine Science 7:828.

Drury C, Bean NK, Harris CI, Hancock JR, Huckeba J, H CM, Roach TNF, Quinn RA, Gates RD. 2022. Intrapopulation adaptive variance supports thermal tolerance in a reef-building coral. Commun Biol 5:486.

Ellis RP, Spicer JI, Byrne JJ, Sommer U, Viant MR, White DA, Widdicombe S. 2014. (1)H NMR metabolomics reveals contrasting response by male and female mussels exposed to reduced seawater pH, increased temperature, and a pathogen. Environ Sci Technol 48:7044–52.

Emwas A-HM. 2015. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. Methods Mol Biol 1277:161–93.

Emwas A-H, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, Raftery D, Alahmari F, Jaremko L, Jaremko M, Wishart DS. 2019. NMR spectroscopy for metabolomics research. Metabolites 9:123.

Erbilgin O, Rübel O, Louie KB, Trinh M, Raad M de, Wildish T, Udwary D, Hoover C, Deutsch S, Northen TR, Bowen BP. 2019. MAGI: A Method for Metabolite Annotation and Gene

Integration. ACS Chem Biol 14:704–14.

Fan TW-M, Lorkiewicz PK, Sellers K, Moseley HNB, Higashi RM, Lane AN. 2012. Stable isotope-resolved metabolomics and applications for drug development. Pharmacol Ther 133:366–91.

Farag MA, Meyer A, Ali SE, Salem MA, Giavalisco P, Westphal H, Wessjohann LA. 2018. Comparative metabolomics approach detects stress-specific responses during coral bleaching in soft corals. J Proteome Res 17:2060–71.

Fendt S-M, Frezza C, Erez A. 2020. Targeting metabolic plasticity and flexibility dynamics for cancer therapy. Cancer Discov 10:1797–1807.

Freund DM, Hegeman AD. 2017. Recent advances in stable isotope-enabled mass spectrometry-based plant metabolomics. Curr Opin Biotechnol 43:41–48.

Fu Y, Zhang F, Ma C, Wang W, Liu Z, Chen W, Zhao M, Ma L. 2022. Comparative metabolomics and lipidomics of four juvenoids application to *Scylla paramamosain* hepatopancreas: Implications of lipid metabolism during ovarian maturation. Front Endocrinol (Lausanne) 13:886351.

Galindo-Prieto B, Eriksson L, Trygg J. 2014. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). J Chemom 28:623–32.

Gamba A, Petras D, Little M, White B, Dorrestein PC, Rohwer F, Foster RA, Hartmann AC. 2022. Applying tissue separation and untargeted metabolomics to understanding lipid saturation kinetics of host mitochondria and symbiotic algae in corals under high temperature stress. Front Mar Sci 9.

Gauff RPM, Greff S, Bohner O, Loisel S, Lejeusne C, Davoult D. 2025. Fouling community shows high resistance and metabolic resilience towards experimental high intensity heatwave. Mar Environ Res 203:106813.

Geistlinger L, Csaba G, Santarelli M, Ramos M, Schiffer L, Turaga N, Law C, Davis S, Carey V, Morgan M, Zimmer R, Waldron L. 2021. Toward a gold standard for benchmarking gene set enrichment analysis. Brief Bioinform 22:545–56.

Geng Q, Zou L, Guo M, Peng J, Li F, Bi Y, Jiang S, Qin H, Tan Z. 2024. Insights into the combined toxicity and mechanisms of BDE-47 and PFOA in marine blue mussel: An integrated study at the physiochemical and molecular levels. Aquat Toxicol 273:106999.

Georgoulis I, Bock C, Lannig G, Pörtner H-O, Feidantsis K, Giantsis IA, Sokolova IM, Michaelidis B. 2022. Metabolic remodeling caused by heat hardening in the Mediterranean mussel *Mytilus galloprovincialis*. J Exp Biol 225.

Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. 2020. A review of challenges and opportunities in machine learning for health. AMIA Summits Transl Sci Proc 2020:191–200.

Gika HG, Zisi C, Theodoridis G, Wilson ID. 2016. Protocol for quality control in metabolic profiling of biological fluids by U(H)PLC-MS. J Chromatogr B Analyt Technol Biomed Life

Sci 1008:15–25.

Greener JG, Kandathil SM, Moffat L, Jones DT. 2022. A guide to machine learning for biologists. Nat Rev Mol Cell Biol 23:40–55.

Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, Goodacre R. 2015. A tutorial review: Metabolomics and partial least squares-discriminant analysis--a marriage of convenience or a shotgun wedding. Anal Chim Acta 879:10–23.

Guscelli E, Chabot D, Vermandele F, Madeira D, Calosi P. 2023. All roads lead to Rome: inter-origin variation in metabolomics reprogramming of the northern shrimp exposed to global changes leads to a comparable physiological status. Front Mar Sci 10.

Haydon TD, Matthews JL, Seymour JR, Raina J-B, Seymour JE, Chartrand K, Camp EF, Suggett DJ. 2023. Metabolomic signatures of corals thriving across extreme reef habitats reveal strategies of heat stress tolerance. Proc Biol Sci 290:20221877.

Hillyer KE, Dias DA, Lutz A, Wilkinson SP, Roessner U, Davy SK. 2017. Metabolite profiling of symbiont and host during thermal stress and bleaching in the coral *Acropora aspera*. Coral Reefs 36:105–18.

Hillyer KE, Dias D, Lutz A, Roessner U, Davy SK. 2018. 13C metabolomics reveals widespread change in carbon fate during coral bleaching. Metabolomics 14:12.

Holt SM, Kinsey ST. 2002. Osmotic effects on arginine kinase function in living muscle of the blue crab *Callinectes sapidus*. J Exp Biol 205:1775–85.

Huffmyer AS, Ashey J, Strand E, Chiles EN, Su X, Putnam HM. 2024. Coral larvae increase nitrogen assimilation to stabilize algal symbiosis and combat bleaching under increased temperature. PLoS Biol 22:e3002875.

Huffmyer AS, Wong KS, Becker DM, Strand E, Mass T, Putnam HM. *In press*. Shifts and critical periods in coral metabolism reveal energetic vulnerability during development. Curr Biol

Ihnatova I, Popovici V, Budinska E. 2018. A critical comparison of topology-based pathway analysis methods. PLoS One 13:e0191154.

Imbs AB, Ermolenko EV, Grigorchuk VP, Sikorskaya TV, Velansky PV. 2021. Current progress in lipidomics of marine invertebrates. Mar Drugs 19:660.

Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. 2009. Analytical and statistical approaches to metabolomics research. J Sep Sci 32:2183–99.

Jacyna J, Kordalewska M, Markuszewski MJ. 2019. Design of Experiments in metabolomics-related studies: An overview. J Pharm Biomed Anal 164:598–606.

Jain S, Jain V. 2024. Machine learning and deep learning methods in brain tumor classification: A decade: Systematic literature review. Intell Data Anal 1–32.

Jang C, Chen L, Rabinowitz JD. 2018. Metabolomics and Isotope Tracing. Cell 173:822–37.
Jansen JJ, Hoefsloot HCJ, van der Greef J, Timmerman ME, Westerhuis JA, Smilde AK. 2005.

ASCA: analysis of multivariate data obtained from an experimental design. J Chemom 19:469–81.

Jia D, Lu M, Jung KH, Park JH, Yu L, Onuchic JN, Kaipparettu BA, Levine H. 2019. Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. Proc Natl Acad Sci U S A 116:3909–18.

Jing H, Zhou L, Gao Y, Liu Z, Wu B, Sun X, Tu K. 2023. Transcriptomics and metabolomics reveal the molecular and metabolic adaptation to heat stress in Manila clam *Ruditapes philippinarum*. Front Mar Sci 10.

Kalivodová A, Hron K, Filzmoser P, Najdekr L, Janečková H, Adam T. 2015. PLS-DA for compositional data with application to metabolomics: PLS-DA for Compositional Data. J Chemom 29:21–28.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30.

Karp PD, Midford PE, Caspi R, Khodursky A. 2021. Pathway size matters: the influence of pathway granularity on over-representation (enrichment analysis) statistics. BMC Genomics 22:191.

Khatri P, Sirota M, Butte AJ. 2012. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 8:e1002375.

Kind T, Scholz M, Fiehn O. 2009. How large is the metabolome? A critical analysis of data exchange practices in chemistry. PLoS One 4:e5440.

Kinsey ST, Lee BC. 2003. The effects of rapid salinity change on in vivo arginine kinase flux in the juvenile blue crab, *Callinectes sapidus*. Comp Biochem Physiol B Biochem Mol Biol 135:521–31.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559.

Laudicella VA, Beveridge C, Carboni S, Franco SC, Doherty MK, Long N, Mitchell E, Stanley MS, Whitfield PD, Hughes AD. 2020. Lipidomics analysis of juveniles' blue mussels (*Mytilus edulis* L. 1758), a key economic and ecological species. PLoS One 15:e0223031.

Lee AY, Troisi J, Symes SJK. 2022. Experimental design in metabolomics. In: Metabolomics Perspectives Elsevier. p. 27–61.

Lee H-C, Yokomizo T. 2018. Applications of mass spectrometry-based targeted and non-targeted lipidomics. Biochem Biophys Res Commun 504:576–81.

Li M, Yang L, Bai Y, Liu H. 2014. Analytical methods in lipidomics and their applications. Anal Chem 86:161–75.

Lindon JC, Wilson ID. 2016. The development of metabolic phenotyping—A historical perspective. In: Metabolic Phenotyping in Personalized and Public Healthcare Elsevier. p. 17–48.

Link H, Fuhrer T, Gerosa L, Zamboni N, Sauer U. 2015. Real-time metabolome profiling of the metabolic switch between starvation and growth. Nat Methods 12:1091–97.

Liu X, Locasale JW. 2017. Metabolomics: A primer. Trends Biochem Sci 42:274–84.

Liu X, Zhang L, You L, Yu J, Zhao J, Li L, Wang Q, Li F, Li C, Liu D, Wu H. 2011. Differential toxicological effects induced by mercury in gills from three pedigrees of Manila clam *Ruditapes philippinarum* by NMR-based metabolomics. Ecotoxicology 20:177–86.

Liu Z, Zhang Y, Zhou Z, Zong Y, Zheng Y, Liu C, Kong N, Gao Q, Wang L, Song L. 2020. Metabolomic and transcriptomic profiling reveals the alteration of energy metabolism in oyster larvae during initial shell formation and under experimental ocean acidification. Sci Rep 10:6111.

Lürig MD, Donoughe S, Svensson EI, Porto A, Tsuboi M. 2021. Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. Front Ecol Evol 9.

Lutier M, Di Poi C, Gazeau F, Appolis A, Le Luyer J, Pernet F. 2022. Revisiting tolerance to ocean acidification: Insights from a new framework combining physiological and molecular tipping points of Pacific oyster. Glob Chang Biol 28:3333–48.

Ma J, Shojaie A, Michailidis G. 2019. A comparative study of topology-based pathway enrichment analysis methods. BMC Bioinformatics 20:546.

Maleki F, Ovens K, Hogan DJ, Kusalik AJ. 2020. Gene set analysis: Challenges, opportunities, and future research. Front Genet 11:654.

Manochkumar J, Cherukuri AK, Kumar RS, Almansour AI, Ramamoorthy S, Efferth T. 2023. A critical review of machine-learning for “multi-omics” marine metabolite datasets. Comput Biol Med 165:107425.

Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, Wishart DS. 2017. The future of NMR-based metabolomics. Curr Opin Biotechnol 43:34–40.

Minet A, Melvin S, Metian M, Taylor A, Oberhänsli F, Lefrançois C, Swarzenski P, Bustamante P, Lacoue-Labarthe T. 2025. Metabolomic and phenotypic effects of ocean acidification on cuttlefish differ across early life stages. Mar Environ Res 205:107013.

Misra BB. 2021. New software tools, databases, and resources in metabolomics: updates from 2020. Metabolomics 17:49.

Molenaar MR, Haaker MW, Vaandrager AB, Houweling M, Helms JB. 2023. Lipidomic profiling of rat hepatic stellate cells during activation reveals a two-stage process accompanied by increased levels of lysosomal lipids. J Biol Chem 299:103042.

Molenaar MR, Jeucken A, Wassenaar TA, van de Lest CHA, Brouwers JF, Helms JB. 2019. LION/web: a web-based ontology enrichment tool for lipidomic data analysis. Gigascience 8.

Mubeen S, Tom Kodamullil A, Hofmann-Apitius M, Domingo-Fernández D. 2022. On the

influence of several factors on pathway enrichment analysis. Brief Bioinform 23:bbac143.
Munjal Y, Tonk R, Sharma R. 2022. Analytical Techniques Used in Metabolomics: A Review. Syst Rev Pharm 13:550–56.

Nadon R, Shoemaker J. 2002. Statistical issues with microarrays: processing and analysis. Trends Genet 18:265–71.

Nagana Gowda GA, Raftery D. 2015. Can NMR solve some significant challenges in metabolomics? J Magn Reson 260:144–60.

Nägele T, Fürtauer L, Nagler M, Weiszmann J, Weckwerth W. 2016. A strategy for functional interpretation of metabolomic time series data in context of metabolic network information. Front Mol Biosci 3:6.

Naz S, Moreira dos Santos DC, García A, Barbas C. 2014. Analytical protocols based on LC-MS, GC-MS and CE-MS for nontargeted metabolomics of biological tissues. Bioanalysis 6:1657–77.

Neumann S, Böcker S. 2010. Computational mass spectrometry for metabolomics: identification of metabolites and small molecules. Anal Bioanal Chem 398:2779–88.

Nguyen T-M, Shafi A, Nguyen T, Draghici S. 2019. Identifying significantly impacted pathways: a comprehensive review and assessment. Genome Biol 20:203.

Nguyen TV, Alfaro A, Frost E, Chen D, Beale DJ, Mundy C. 2021. Investigating the biochemical effects of heat stress and sample quenching approach on the metabolic profiling of abalone (*Haliotis iris*). Metabolomics 18:7.

Noisette F, Calosi P, Madeira D, Chemel M, Menu-Courey K, Piedalue S, Gurney-Smith H, Daoud D, Azetsu-Scott K. 2021. Tolerant Larvae and Sensitive Juveniles: Integrating Metabolomics and Whole-Organism Responses to Define Life-Stage Specific Sensitivity to Ocean Acidification in the American Lobster. Metabolites 11.

Olden JD, Lawler JJ, Poff NL. 2008. Machine learning methods without tears: a primer for ecologists. Q Rev Biol 83:171–93.

Pang Z, Lu Y, Zhou G, Hui F, Xu L, Viau C, Spigelman AF, MacDonald PE, Wishart DS, Li S, Xia J. 2024. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. Nucleic Acids Res 52:W398–406.

Park JO, Rubin SA, Xu Y-F, Amador-Noguez D, Fan J, Shlomi T, Rabinowitz JD. 2016. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat Chem Biol 12:482–89.

Pei G, Chen L, Zhang W. 2017. WGCNA application to proteomic and metabolomic data analysis. Methods Enzymol 585:135–58.

Pichler M, Hartig F. 2023. Machine learning and deep learning—A review for ecologists. Methods Ecol Evol 14:994–1016.

Pörtner H-O. 2010. Oxygen- and capacity-limitation of thermal tolerance: a matrix for integrating

climate-related stressor effects in marine ecosystems. J Exp Biol 213:881–93.

Pörtner H-O, Bock C, Mark FC. 2017. Oxygen- and capacity-limited thermal tolerance: bridging ecology and physiology. J Exp Biol 220:2685–96.

Putnam HM, Davidson JM, Gates RD. 2016. Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. Evol Appl 9:1165–78.

Putri SP, Yamamoto S, Tsugawa H, Fukusaki E. 2013. Current metabolomics: technological advances. J Biosci Bioeng 116:9–16.

Rathore N, Thakur D, Kumar D, Chawla A, Kumar S. 2021. Time-series eco-metabolomics reveals extensive reshuffling in metabolome during transition from cold acclimation to de-acclimation in an alpine shrub. Physiol Plant 173:1824–40.

Reddy MM, Goossens C, Zhou Y, Chaib S, Raviglione D, Nicolè F, Hume BCC, Forcioli D, Agostini S, Boissin E, Boss E, Bowler C, de Vargas C, Douville E, Flores M, Furla P, Galand PE, Gilson E, Lombard F, Pesant S, Reynaud S, Sullivan MB, Sunagawa S, Troublé R, Thurber RV, Wincker P, Zoccola D, Voolstra CR, Allemand D, Planes S, Thomas OP, Banaigs B. 2023. Multi-omics determination of metabolome diversity in natural coral populations in the Pacific Ocean. Commun Earth Environ 4:1–10.

Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. 2021. Using machine learning approaches for multi-omics data analysis: A review. Biotechnol Adv 49:107739.

Reisz JA, Zheng C, D'Alessandro A, Nemkov T. 2019. Untargeted and semi-targeted lipid analysis of biological samples using mass spectrometry-based metabolomics. Methods Mol Biol 1978:121–35.

Ren S, Hinzman AA, Kang EL, Szczesniak RD, Lu LJ. 2015. Computational and statistical analysis of metabolomics data. Metabolomics 11:1492–1513.

Ren X, Yu Z, Xu Y, Zhang Y, Mu C, Liu P, Li J. 2020. Integrated transcriptomic and metabolomic responses in the hepatopancreas of kuruma shrimp (*Marsupenaeus japonicus*) under cold stress. Ecotoxicol Environ Saf 206:111360.

Rey F, Melo T, Lopes D, Couto D, Marques F, Domingues MR. 2022. Applications of lipidomics in marine organisms: progress, challenges and future perspectives. Mol Omics 18:357–86.

Roach TNF, Dilworth J, H CM, Jones AD, Quinn RA, Drury C. 2021. Metabolomic signatures of coral bleaching history. Nat Ecol Evol 5:495–503.

Rodriguez-Casariego JA, Mercado-Molina A, Lemos LS, Quinete NS, Bellantuono A, Rodriguez-Lanetty M, Sabat A, Eirin-Lopez JM. 2023. Multi-omic characterization of mechanisms contributing to rapid phenotypic plasticity in the coral *Acropora cervicornis* under divergent environments. Coral Reefs.

Roessner U, Bowne J. 2009. What is metabolomics all about? Biotechniques 46:363–65.

Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. 2023. Machine learning and artificial intelligence in research and healthcare. Injury 54 Suppl 3:S69–73.

Saccenti E, Timmerman ME. 2016. Approaches to sample size determination for multivariate data: Applications to PCA and PLS-DA of omics data. J Proteome Res 15:2379–93.

Santiago-Rodriguez TM, Hollister EB. 2021. Multi 'omic data integration: A review of concepts, considerations, and approaches. Semin Perinatol 45:151456.

Schock T, Keller J, Rice M, Balazs G, Bearden D. 2014. Metabotyping of a protected non-model organism, Green Sea turtle (*Chelonia mydas*), using ^{1}H NMR spectroscopy and optimized plasma methods for metabolomics. Curr Metabolomics 1:279–90.

Sekar K, Rusconi R, Sauls JT, Fuhrer T, Noor E, Nguyen J, Fernandez VI, Buffing MF, Berney M, Jun S, Stocker R, Sauer U. 2018. Synthesis and degradation of FtsZ quantitatively predict the first cell division in starved bacteria. Mol Syst Biol 14:e8623.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504.

Sindelar M, Patti GJ. 2020. Chemical discovery in the era of metabolomics. J Am Chem Soc 142:9097–9105.

Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers R-JAN, van der Greef J, Timmerman ME. 2005. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. Bioinformatics 21:3043–48.

Smilde AK, Westerhuis JA, Hoefsloot HCJ, Bijlsma S, Rubingh CM, Vis DJ, Jellema RH, Pijl H, Roelfsema F, van der Greef J. 2010. Dynamic metabolomic data analysis: a tutorial review. Metabolomics 6:3–17.

Smith R, Mathis AD, Ventura D, Prince JT. 2014. Proteomics, lipidomics, metabolomics: a mass spectrometry tutorial from a computer scientist's point of view. BMC Bioinformatics 15.

Sogin EM, Putnam HM, Anderson PE, Gates RD. 2016. Metabolomic signatures of increases in temperature and ocean acidification from the reef-building coral, *Pocillopora damicornis*. Metabolomics 12.

Sriyudthsak K, Shiraishi F, Hirai MY. 2016. Mathematical modeling and dynamic simulation of metabolic reaction systems using metabolome time series data. Front Mol Biosci 3:15.

Sui X, Niu X, Shi M, Pei G, Li J, Chen L, Wang J, Zhang W. 2014. Metabolomic analysis reveals mechanism of antioxidant butylated hydroxyanisole on lipid accumulation in Crypthecodinium cohnii. J Agric Food Chem 62:12477–84.

Sun J, Lu L, Liu J, Cui Y, Liu H, Zhang Y, Zheng Z, Yang W. 2024. Metabolomics and WGCNA analyses reveal the underlying mechanisms of resistance to *Botrytis cinerea* in hazelnut. Genes (Basel) 16:2.

Sun X, Tu K, Li L, Wu B, Wu L, Liu Z, Zhou L, Tian J, Yang A. 2021. Integrated transcriptome and metabolome analysis reveals molecular responses of the clams to acute hypoxia. Mar Environ Res 168:105317.

Sun X, Wu B, Tu K, Zhou L, Yang A, Liu Z. 2022. Transcriptome and metabolome analyses provide insights into the salinity adaptation of clam *Ruditapes philippinarum*. Aquac Rep 27:101368.

Sweet M, Bulling M, Varshavi D, Lloyd GR, Jankevics A, Najdekr L, Weber RJM, Viant MR, Craggs J. 2021. Species-specific variations in the metabolomic profiles of *Acropora hyacinthus* and *Acropora millepora* mask acute temperature stress effects in adult coral colonies. Front Mar Sci 8.

Tjeerdema RS, Kauten RJ, Crosby DG. 1993. Interactive effects of pentachlorophenol and temperature in the abalone (*Haliotis rufescens*) as measured by in vivo 31P-NMR spectroscopy. Aquat Toxicol 26:117–32.

Triantafyllidis AK, Tsanas A. 2019. Applications of machine learning in real-life digital health interventions: Review of the literature. J Med Internet Res 21:e12286.

Venkataraman YR, Huffmyer AS, White SJ, Downey-Wall A, Ashey J, Becker DM, Bengtsson Z, Putnam HM, Strand E, Rodríguez-Casariego JA, Wanamaker SA, Lotterhos KE, Roberts SB. 2024. DNA methylation correlates with transcriptional noise in response to elevated pCO2 in the eastern oyster (*Crassostrea virginica*). Environ Epigenet 10:dvae018.

Viant MR. 2008. Recent developments in environmental metabolomics. Mol Biosyst 4:980–86.

Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. 2012. A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. Metabolites 2:775–95.

Wadi L, Meyer M, Weiser J, Stein LD, Reimand J. 2016. Impact of outdated gene annotations on pathway enrichment analysis. Nat Methods 13:705–6.

Wanamaker SA, McElhany P, Maher M, Perez D, Busch DS, Nichols KM. 2019. Uncovering mechanisms of global ocean change effects on the Dungeness crab (*Cancer magister*) through metabolomics analysis. Sci Rep 9:10717.

Wang R, Yin Y, Li J, Wang H, Lv W, Gao Y, Wang T, Zhong Y, Zhou Z, Cai Y, Su X, Liu N, Zhu Z-J. 2022. Global stable-isotope tracing metabolomics reveals system-wide metabolic alternations in aging *Drosophila*. Nat Commun 13:3518.

Wei L, Wang Q, Ning X, Mu C, Wang C, Cao R, Wu H, Cong M, Li F, Ji C, Zhao J. 2015. Combined metabolome and proteome analysis of the mantle tissue from Pacific oyster *Crassostrea gigas* exposed to elevated pCO2. Comp Biochem Physiol Part D Genomics Proteomics 13:16–23.

Wenk MR. 2005. The emerging field of lipidomics. Nat Rev Drug Discov 4:594–610.

Wijesooriya K, Jadaan SA, Perera KL, Kaur T, Ziemann M. 2022. Urgent need for consistent standards in functional enrichment analysis. PLoS Comput Biol 18:e1009935.

Wilhelm RC, Barnett SE, Swenson TL, Youngblut ND, Koechli CN, Bowen BP, Northen TR, Buckley DH. 2022. Tracing carbon metabolism with stable isotope metabolomics reveals the legacy of diverse carbon sources in soil. bioRxiv.

Williams A. 2024. Multiomics data integration, limitations, and prospects to reveal the metabolic activity of the coral holobiont. FEMS Microbiol Ecol 100:fiae058.

Williams A, Chiles EN, Conetta D, Pathmanathan JS, Cleves PA, Putnam HM, Su X, Bhattacharya D. 2021. Metabolomic shifts associated with heat stress in coral holobionts. Sci Adv 7:eabd4210.

Williams A, Pathmanathan JS, Stephens TG, Su X, Chiles EN, Conetta D, Putnam HM, Bhattacharya D. 2021. Multi-omic characterization of the thermal stress phenome in the stony coral *Montipora capitata*. PeerJ 9:e12335.

Williams TD, Turan N, Diab AM, Wu H, Mackenzie C, Bartie KL, Hrydziuszko O, Lyons BP, Stentiford GD, Herbert JM, Abraham JK, Katsiadaki I, Leaver MJ, Taggart JB, George SG, Viant MR, Chipman KJ, Falciani F. 2011. Towards a system level understanding of non-model organisms sampled from the environment: a network biology approach. PLoS Comput Biol 7:e1002126.

Winding Hansen B, Ciappini G, Malmendal A, Rayner TA. 2022. Can we adjust a marine cyclopoid copepod to freshwater?—First step towards a “universal” live feed product for fish and shrimp larvae. Aquac Res 53:178–90.

Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly M-A, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L. 2007. HMDB: The Human Metabolome Database. Nucleic Acids Res 35:D521–26.

Worley B, Powers R. 2013. Multivariate analysis in metabolomics. Curr Metabolomics 1:92–107.

Wright RM, Aglyamova GV, Meyer E, Matz MV. 2015. Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. BMC Genomics 16:371.

Xia J, Psychogios N, Young N, Wishart DS. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res 37:W652–60.

Xia J, Wishart DS. 2010. MetPA: a web-based metabolomics tool for pathway analysis and visualization. Bioinformatics 26:2342–44.

Xia J, Wishart DS. 2011. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nat Protoc 6:743–60.

Yang Q, Wang S, Dai E, Zhou S, Liu D, Liu H, Meng Q, Jiang B, Jiang W. 2019. Pathway enrichment analysis approach based on topological structure and updated annotation of pathway. Brief Bioinform 20:168–77.

Young T, Alfaro AC. 2018. Metabolomic strategies for aquaculture research: a primer. Rev Aquac 10:26–56.

Zhang L, Liu X, You L, Zhou D, Wu H, Li L, Zhao J, Feng J, Yu J. 2011. Metabolic responses in

gills of Manila clam Ruditapes philippinarum exposed to copper using NMR-based metabolomics. Mar Environ Res 72:33–39.

Zhang W, Dong Y. 2021. Membrane lipid metabolism, heat shock response and energy costs mediate the interaction between acclimatization and heat-hardening response in the razor clam *Sinonovacula constricta*. J Exp Biol 224.

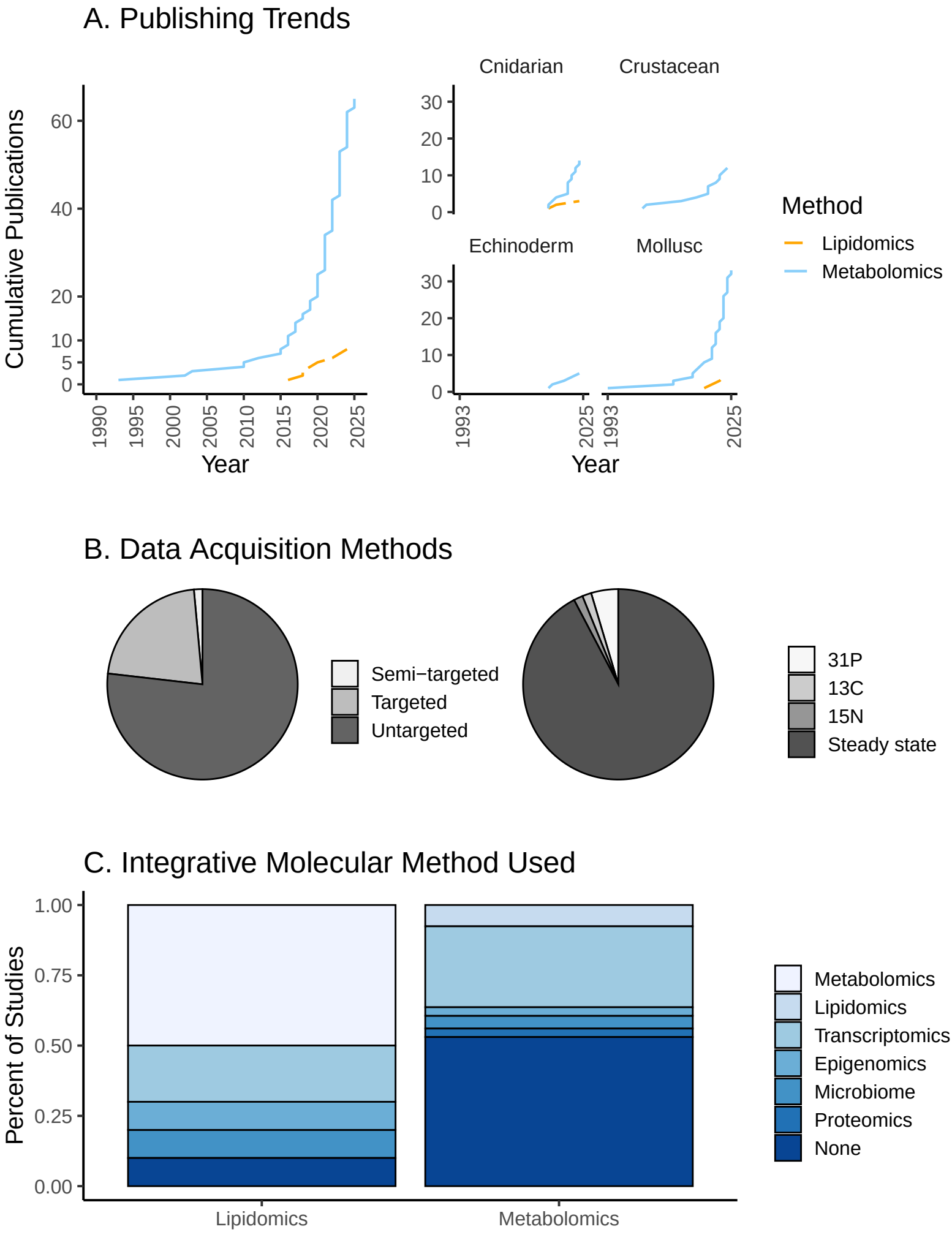
Zhao K, Rhee SY. 2023. Interpreting omics data with pathway enrichment analysis. Trends Genet 39:308–19.

Zhao X, Niu L, Clerici C, Russo R, Byrd M, Setchell KDR. 2019. Data analysis of MS-based clinical lipidomics studies with crossover design: A tutorial mini-review of statistical methods. Clin Mass Spectrom 13:5–17.

Zhou Y, Li Q, Zhang Q, Yuan M, Zhu X, Li Y, Li Q, Downs CA, Huang D, Chou L-M, Zhao H. 2024. Environmental concentrations of herbicide prometryn render stress-tolerant corals susceptible to ocean warming. Environ Sci Technol 58:4545–57.

Zhu J, Shi W, Zhao R, Gu C, Li H, Wang L, Wan X. 2024. Effects of cold stress on the hemolymph of the pacific white shrimp *Penaeus vannamei*. G 9:36.

Zittier ZMC, Bock C, Sukhotin AA, Häfker NS, Pörtner HO. 2018. Impact of ocean acidification on thermal tolerance and acid–base regulation of *Mytilus edulis* from the White Sea. Polar Biol 41:2261–73.


A. Publishing Trends
Cumulative Publications
60
40
20
10
5
0
1990
1995
2000
2005
2010
2015
2020
2025
Year
Cnidarian
Crustacean
Echinoderm
Mollusc
30
20
10
0
1993
2025
Year
Method
Lipidomics
Metabolomics
B. Data Acquisition Methods
Semi−targeted
Targeted
Untargeted
31P
13C
15N
Steady state
C. Integrative Molecular Method Used
Percent of Studies
1.00
0.75
0.50
0.25
0.00
Lipidomics
Metabolomics
Metabolomics
Lipidomics
Transcriptomics
Epigenomics
Microbiome
Proteomics
None

BOX 1 | Glossary

Metabolomics: The study of chemical processes involving the small molecules that are the direct and indirect products of metabolic pathways. Metabolites are often classified into primary metabolites involved in growth, development, and reproduction, and secondary metabolites that are more important for ecological function.

Lipidomics: The study of complete fatty acid and lipid profiles in a sample. Lipid molecules are crucial for short- and long-term energy storage.

Plasticity: Ability of one genotype to produce more than one phenotype.

Metabolic plasticity: An organism's capacity to modulate energy production, allocation, and use.

Steady state: Absolute or relative concentrations of a compound at a particular point in time.

Metabolic flux: Rate at which compounds pass through a metabolic pathway.

Liquid chromatography mass spectrometry (LC-MS): Combines physical separation capability of liquid chromatography with mass analysis capability of mass spectrometry. Can characterize a wide range of molecule types, and is considered more robust for lipidomic applications.

Gas chromatography mass spectrometry (GC-MS): Combines features of gas chromatography with mass spectrometry. Best suited for volatile molecules, and is considered more robust for metabolomic applications.

Nuclear magnetic resonance (NMR): Quantify compounds by placing a sample in a magnetic field and using the inherent magnetic properties to identify the compounds. 1H-NMR is most commonly used with metabolomics data.

Untargeted: Provides relative feature abundance differences between experimental conditions for all detected compounds.

Targeted: Provides absolute quantitation, or specific concentrations of molecules, based on an input list of known compounds.

Semi-targeted: Quantifies known compounds and detects unknown or unidentified compounds.

Permutational analysis of variance (PERMANOVA): Unsupervised multivariate approach used to test if centroids of groups are significantly different from each other. Often paired with permutational analyses of dispersion (PERMDISP).

Weighted gene co-expression network analysis (WGCNA): Correlation-based approach originally developed to identify groups of genes that shared expression patterns. Can be applied to identify metabolites or lipids with shared abundance patterns.

ANOVA-simultaneous components analysis (ASCA): Decomposes multivariate data according to variables of interest. Useful for examining multivariate responses across time or multiple additional factors.

Principal components analysis (PCA): Unsupervised molecular approach generally used for exploratory analysis. Reduces the number of dimensions in large datasets to principal components that retain most of the original information.

Partial least squares discriminant analysis (PLS-DA): Supervised approach that uses group labels to reduce dimensionality by identifying variables that explain relationships between predictors and responses. Orthogonal PLS-DA (OPLS-DA) is a variation commonly used with metabolomic and lipidomic data.

Variable Importance in Projection (VIP): Quantifies the discriminatory power of a compound from a PLS-DA model.

Significance Analysis of Microarray (SAM): Feature-specific modified t-tests combined with permutation analysis used to determine in a specific compound has differential abundance between experimental conditions.

Machine learning (ML): Computer systems that learn or adapt using statistical models to mimic human behavior and recognize patterns. Useful when datasets are complex, large, or require automation.

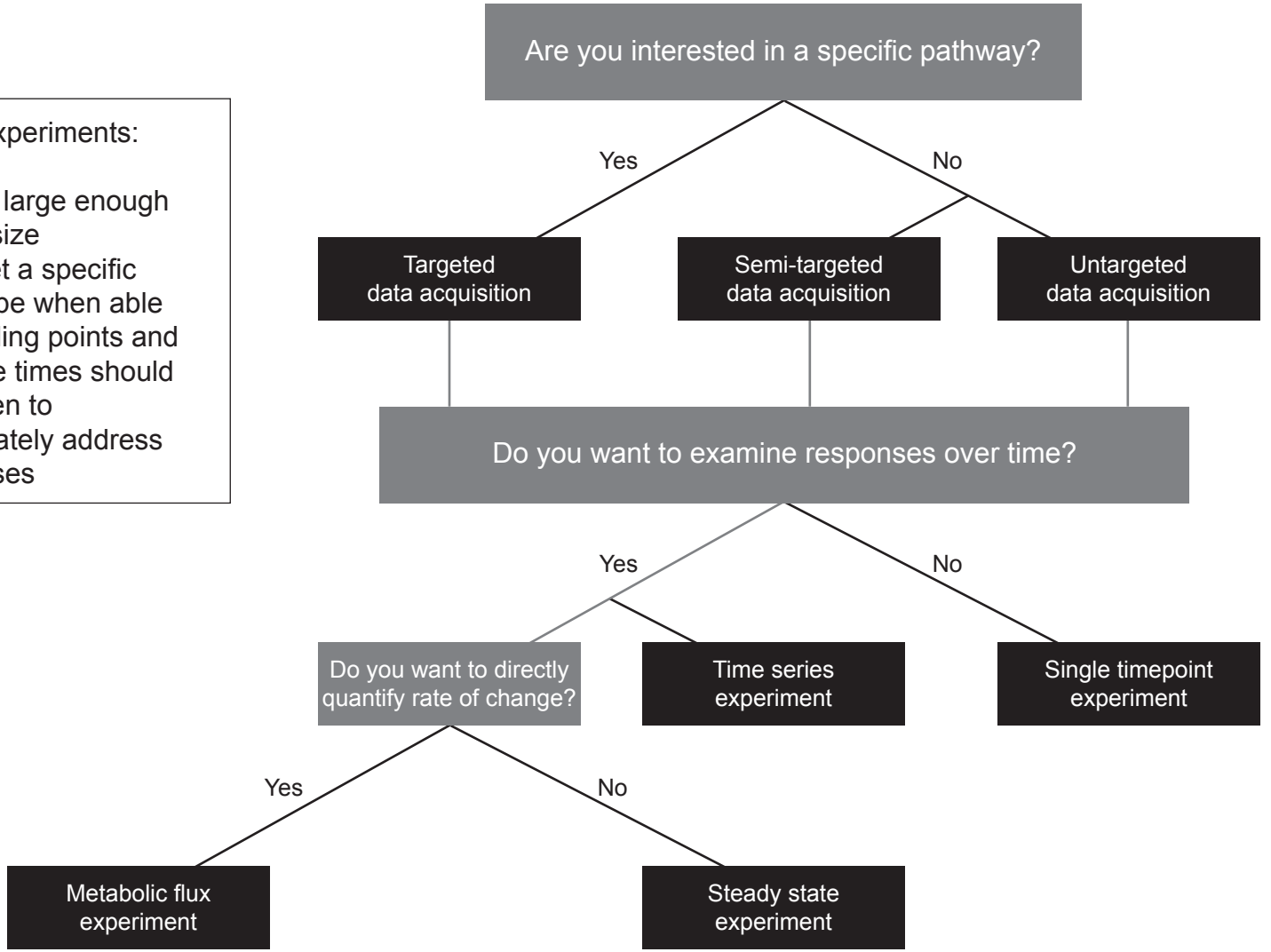
Overrepresentation-based enrichment: Determines if specific pathways or functions are observed in a target dataset more than expected by chance in comparison to a background dataset.

Network topology-based enrichment: Incorporates additional factors that impact pathway activity, such as feature position in a pathway or feature-feature interactions, into an enrichment analysis.

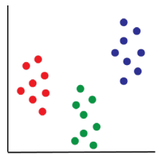
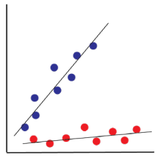
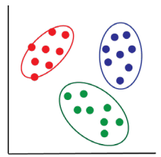
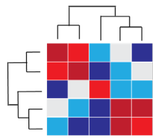
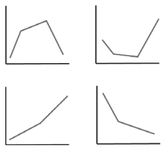
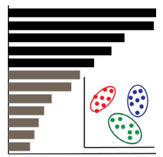
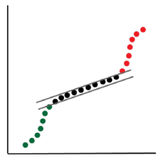
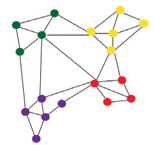
Experimental design considerations

For all experiments:

1. Use a large enough sample size
2. Subset a specific tissue type when able
3. Sampling points and exposure times should be chosen to appropriately address hypotheses



Analytical considerations

	Single metabolite tests		Unsupervised analyses			Supervised analyses		
	ANOVA	Linear models	PERMANOVA	WGCNA	ASCA	PLS-DA + VIP	SAM	ML
								
How does concentration or composition change between treatments?	✓	✓				✓	✓	✓
How do compounds correlate with quantitative responses or time?			✓	✓	✓			✓
What compounds drive differences between treatments?						✓	✓	✓

Enrichment with MetaboAnalyst

A. Pros and cons of MetaboAnalyst for enrichment

Pros	Cons
• Tools available to analyze data from raw spectra through enrichment and visualization • Interactive platform with no coding required • Provides the foundation for other analytical packages	• Extended capabilities for untargeted data, but better suited for targeted data • Issues with shared nomenclature or inability to identify molecules • Limited available reference databases

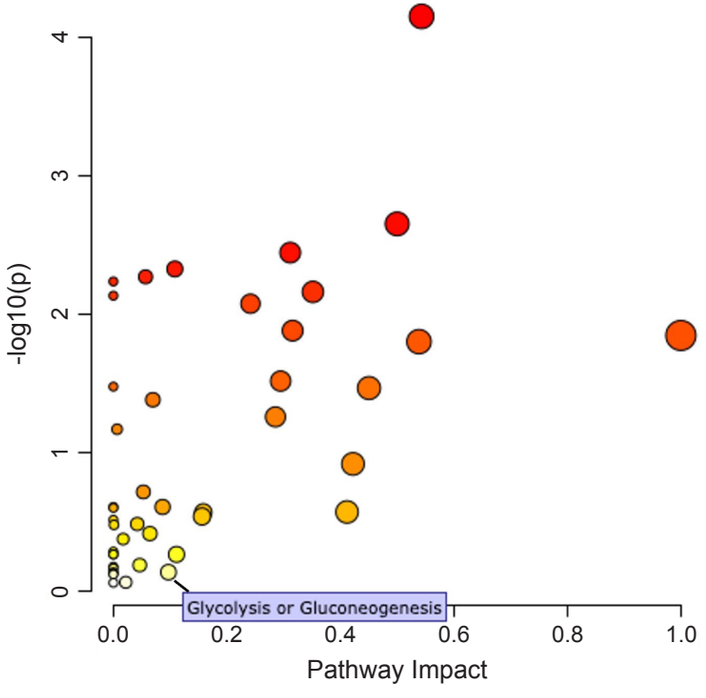
B. Different KEGG databases impact output

Human Pathway	Match Status	FDR	Impact
Starch and sucrose metabolism	6/17	0.003	0.62
Nitrogen metabolism	3/6	0.053	0
Galactose metabolism	4/15	0.059	0.05
Valine, leucine, and isoleucine biosynthesis	3/8	0.059	0

C. elegans Pathway	Match Status	FDR	Impact
Starch and sucrose metabolism	6/16	0.005	0.54
Phenylalanine metabolism	3/6	0.07	0.5
Amino sugar and nucleotide sugar metabolism	6/31	0.07	0.31
One carbon pool by folate	5/23	0.07	0.11

C. Nomenclature inconsistencies impact enrichment results

Overview of Pathway Analysis



Glycolysis or Gluconeogenesis