Analyst

CRITICAL REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/d6an00140h

Metabolomics in the wild: research opportunities, challenges, and regulatory potential for effects-based environmental monitoring

Georgia M. Sinclair, [a] Sarah L. Green, [a,b] Ryan Lester, [a,c] Katherine J. Jeppe, [d] Steven D. Melvin, [c] Sara M. Long, [e] Oliver A. H. Jones [f] and David J. Beale *[a]

Field-based environmental metabolomics offers a powerful tool for examining how organisms respond to complex mixtures of chemical and environmental stressors. When coupled with traditional ecotoxicology, metabolomics can help reveal mechanistic pathways; however, laboratory-based tests cannot fully replicate the dynamic, multifactorial conditions of natural ecosystems. This review synthesises current advances, challenges, and opportunities in applying metabolomics to ecotoxicology under real-world field conditions. We highlight the growing use of wild-caught organisms, caged exposures, mesocosms, and laboratory studies using field-collected samples to detect sub-lethal metabolic disruptions associated with contaminants such as PFAS, metals, organic pollutants, and wastewater-derived mixtures of compounds. Key themes include the sensitivity of metabolomics to early physiological changes, integration with complementary chemical and ecological data, the challenges in distinguishing natural variability from contaminant effects, the importance of establishing baselines and dose–response relationships, and the need for improved QA/QC and metadata reporting. As the methodological and logistical challenges are overcome, metabolomic profiles from field-exposed organisms are increasingly demonstrating value for environmental risk monitoring and forecasting. Environmental metabolomics has been successfully used for environmental monitoring, supporting regulatory frameworks, and identifying mechanistically grounded biomarkers of exposure and effects. However, to realise its full potential, coordinated efforts among current and future metabolomics practitioners are still needed to advance the current Metabolomics Standards Initiative (MSI) guidance. The MSI should, ideally, be expanded to include common standardised workflows, strengthen bioinformatics infrastructure, expand case studies, and fully embed and integrate metabolomics within routine environmental assessment and decision-making processes, thereby transitioning these *'academic'* approaches into practical regulatory tools.

Received 5th February 2026,
Accepted 1st April 2026

DOI: 10.1039/d6an00140h

rsc.li/analyst

Introduction

Understanding the impacts of climate change and environmental contaminants on wildlife requires approaches that reflect the complexity of natural ecosystems. Field-based ecotoxicology can enhance environmental monitoring by capturing the multifaceted, fluctuating, and competing processes and conditions that organisms experience in the wild—conditions that are difficult, if not impossible, to replicate and control in a laboratory setting.[1,2] In natural environments, organisms are exposed to mixtures of chemical stressors, variable temperatures, changing trophic dynamics, disease and other environmental pressures that interact in ways we do not yet fully understand.[1,3,4] Moreover, many ecologically relevant species cannot be readily maintained under laboratory conditions due to their size, complex life histories, and/or sensitivity to handling. This means that critical biological and exposure data can only be obtained through field-based studies. Studying biological responses under environmentally relevant conditions improves environmental risk assessment. Metabolomics enables linking mechanistic biochemical

[a]*Environment, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Ecoscience Precinct, Dutton Park, Queensland 4102, Australia*
[b]*Department of Ecological, Plant and Animal Sciences, La Trobe University, Albury-Wodonga Campus, Victoria, Australia. E-mail: David.Beale@csiro.au*
[c]*Australian Rivers Institute, Griffith University, Gold Coast Campus, Southport, Queensland 4215, Australia*
[d]*Monash Proteomics & Metabolomics Platform, Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia*
[e]*Aquatic Environmental Stress Research Group (AQUEST), School of Science, RMIT-University, Bundoora, Victoria 3083, Australia*
[f]*Australian Centre for Research on Separation Science (ACROSS), School of Science, RMIT University, Bundoora West Campus, Bundoora, Victoria 3083, Australia*

changes to adverse outcomes and other responses/changes that are not biochemical, while capturing realistic exposure scenarios, thereby enhancing ecological relevance and informing more effective management and conservation strategies.[2]

Environmental metabolomics is maturing as a powerful tool for investigating biochemical responses to environmental stress.[1] Metabolomics provides a comprehensive snapshot of an organism's physiological state and is highly sensitive to early changes induced by contaminants and other stressors.[1] Because metabolites are downstream products of gene expression and protein activity, metabolomics can detect subtle, sub-lethal effects (biomarkers) before traditional endpoints are observed (*i.e.*, reduced growth or impaired reproduction).[5] This sensitivity makes metabolomics particularly valuable for assessing contaminants, including per- and polyfluoroalkyl substances (PFAS), metals, organic pollutants, and other complex mixtures in natural environments. It also highlights its potential role in effects-based monitoring and investigating cumulative effects on wildlife.[1,5–11]

Despite this promise, metabolomics is still not widely used in environmental regulation or risk assessment. Several factors underpin this gap. Natural variability makes it challenging to distinguish contaminant-driven effects from background noise; baseline metabolic ranges are poorly defined for most species; and consistent frameworks linking metabolic changes to adverse outcomes or regulatory thresholds remain limited. In addition, inconsistencies in field sampling, analytical workflows, QA/QC, metadata reporting, and bioinformatic interpretation continue to hinder reproducibility and comparability across studies. As a result, although metabolomics can provide mechanistic insight, its operational use in routine monitoring and regulatory decision-making remains constrained.

Laboratory-based and field-based metabolomics each address different components of the challenge to make metabolomics mainstream. This complementary relationship is illustrated in Fig. 1, which contrasts the controlled mechanistic resolution of laboratory metabolomics with the natural environmental exposure conditions and integrative exposure profiles captured in field-based studies. The figure illustrates how laboratory studies provide controlled, mechanistic insight into metabolic disruptions caused by specific contaminants, while field studies capture ecologically realistic responses shaped by variable environmental conditions and mixed stressors. Together, these complementary perspectives strengthen the interpretation of metabolic biomarkers and enhance the ecological relevance of metabolomics in environmental assessment. When integrated, the two approaches provide a more complete picture of exposure and effect. Field-based metabolomics cannot replace controlled experiments; rather, it complements them by capturing ecologically realistic exposure conditions that help generate and prioritise hypotheses for mechanistic testing. Addressing the practical and interpretive challenges of field-based metabolomics, along with laboratory and mesocosm validation, is essential for advancing metabolomics toward regulatory readiness.

To avoid overinterpretation, field-based metabolomics measurements should, ideally, be embedded within a framework that includes direct environmental chemistry (*e.g.*, water, sediment, soil), organism contaminant burdens (bioaccumulation), and controlled experiments that constrain dose and other confounding variables. In this hybrid continuum, field studies provide exposure realism and relevance, while laboratory and mesocosm studies establish causality, mechanism, and quantitative thresholds. This mirrors the established trajectory of environmental health science, where biological mechanisms are first resolved under controlled laboratory conditions and are subsequently confirmed and contextualised through field observations of exposure, bioaccumulation, and ecological impact. The objective of this review is to synthesise key developments, challenges, and opportunities for the application of metabolomics in field-based ecotoxicology and environmental assessment. Rather than presenting an exhaus-

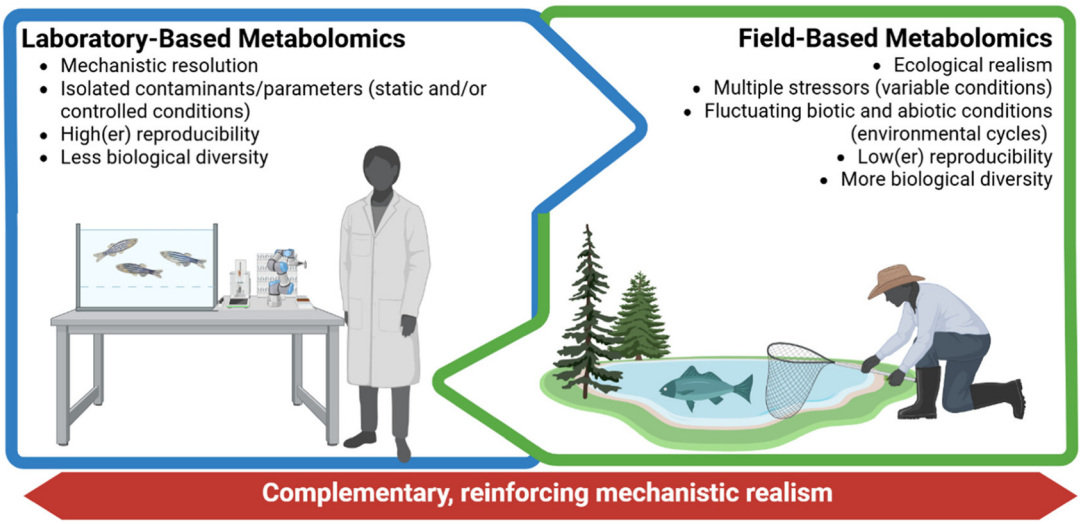

Fig. 1 Comparison of laboratory-based and field-based metabolomics approaches.

tive survey, we focus on major themes emerging from studies using wild-caught organisms, caged deployments, mesocosms, field-derived samples, and wildlife-derived cell lines. We highlight where metabolomics has already demonstrated value in detecting sub-lethal effects, identifying biomarkers, and linking exposure to mechanistic pathways. We also consider where limitations continue to impede adoption. By drawing together methodological insights and identifying areas where standardisation, improved data infrastructure, and clearer interpretative frameworks are needed, we aim to clarify how field-based metabolomics can contribute effectively to environmental monitoring and risk assessment in future.

Metabolomics in ecotoxicology and environmental science

One of the advantages of metabolomics is that it enables the detection of metabolic changes indicative of sub-lethal stress well before more traditional ecotoxicological endpoints, such as reduced reproduction or growth, become apparent.[1] This makes metabolomics useful in ecology and environmental science studies, particularly for examining sub-lethal effects of contaminants of emerging concern, such as micro/nanoplastics[12] and PFAS.[13]

The sensitivity of metabolomics is, however, somewhat of a double-edged '*biochemical*' sword. On the one hand, a specific early warning signal of potential harm to an organism or population would allow action to be taken before permanent damage occurs. On the other hand, how do we know that any observed changes will cause adverse effects, specific and large enough to warrant action? This is analogous to a metabolomics point-of-departure (POD) concept (*i.e.*, a dose, concentration, or exposure level at which a defined biological response is first detected and is then used as the starting point for deriving health- or environment-based guideline values).[14] Changes in metabolism occur constantly, including diurnal fluctuations, circalunar rhythms, seasonal variations in light and temperature, and fluctuations related to feeding and normal homeostatic mechanisms. In a human context, it is possible to detect metabolic changes before and after our morning coffee, or before and after taking medication, food, or alcoholic drinks.[15] In each case, there might be a detectable biological change; it might even be a significant change from the normal functioning, yet remain biologically inconsequential. How do we separate natural changes from damaging alterations in metabolism with potentially irreversible downstream adverse effects, and, if we could, would such biochemical data be sufficient for environmental regulators to act? These challenges, along with suggestions for addressing them, are discussed in Beale *et al.*[16]

The key question for metabolomics is not whether a change is detectable, but whether the exposure to a stressor causes a metabolic shift, and how that shift fits within the broader cause-effect continuum—*i.e.*, contextualised PODs. Addressing this requires asking several follow-up questions, such as:

1. Is the change outside normal ranges for the metabolites involved in the species in question? If so, how many metabolites are outside normal ranges?

2. How long does any observed change persist? Is it temporary or lasting?

3. Do the metabolic shifts lead to downstream biological effects?

4. Do these changes indicate potential harm at the organism, population, or community level, and can they be linked to Adverse Outcome Pathways (AOPs) – linking the biological response following a chemical stressor to adverse outcomes?

The first question is the most challenging because the normal ranges for most metabolites are not well-defined, even in humans. There are some exceptions; the normal range of blood glucose concentrations is well defined, for example, to enable the diagnosis of diabetes. There are also reference ranges for many hormones and enzymes, although these are proteins rather than metabolites. In addition, metabolomics is a static technique that cannot directly capture data on dynamic processes; however, this limitation can be addressed with appropriate experimental designs that incorporate sufficient longitudinal sample collection. In wild populations, such temporal inference is typically derived from repeated cross-sectional sampling rather than true longitudinal tracking of individuals, as repeated sampling of the same organism is often infeasible.

Although the remaining questions are central to interpreting metabolomic perturbations, they are not explored further in this introductory section. Our aim here is to outline the conceptual challenges that arise when attempting to disentangle natural variability from contaminant-driven effects in field-based metabolomics. Detailed consideration of persistence, biological consequences, and links to AOP frameworks requires study-specific evidence—including longitudinal datasets, matched contaminant burdens, mechanistic assays, and ecological context—which are addressed later in the review through dedicated case studies and methodological discussions. The distinction is important because, at this stage, these questions are intended to underscore the inherent complexity of interpreting field-based metabolomics. At the same time, their resolution relies on the acquired data and applied frameworks, which will be addressed in later sections of this review.

To bridge these exposure-driven considerations with analytical interpretation, the following section outlines how different metabolomics acquisition and quantification modalities are selected to address specific biological questions arising from field, caged, and mesocosm study designs.

Targeted *vs.* non-targeted *vs.* widely targeted methods

Metabolomics relies primarily on nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS)-based techniques to measure small-molecule metabolites in biological

and environmental samples. NMR offers highly reproducible, quantitative measurements with minimal sample preparation. In contrast, MS-based approaches (commonly liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS)) provide substantially higher sensitivity and broader metabolite coverage, which is critical for field-based studies where sample mass is limited.[1,17,18] High-resolution MS instruments (*e.g.* QToF and Orbitrap platforms) now underpin most environmental metabolomics workflows, enabling both untargeted discovery and targeted quantification, as well as retrospective data mining and integration with contaminant screening from the same samples.[19,20]

There are two broad approaches to metabolomics, namely non-targeted and targeted. In a non-targeted approach, the analyst attempts to detect, identify and at least semi-quantify as many endogenous metabolites as possible, then looks for any differences between groups, usually *via* some form of multivariate statistical analysis.[1] In a targeted approach, only specific metabolites or biochemical pathways are assessed, allowing precise quantification of known metabolites linked to defined mechanisms.[5,21] There is a third approach, widely targeted metabolomics, which is less commonly used and combines elements of the first two. Here, untargeted metabolomics is used for initial metabolite screening for a particular factor; targeted metabolomics is then used for validation and to explore and identify metabolite classes from step one in more depth.[22] Extremely beneficial to environmental field-based metabolomics are methods that simultaneously perform quantification and discovery (SQUAD) from the same biological extract within a single analytical run.[23] Modern hyphenated or hybrid mass-spectrometry workflows, particularly those using QToF (Quadrupole Time-of-Flight) instruments, now enable acquisition modes that capture both untargeted and targeted metabolite data simultaneously (*e.g.*, Mass Spectrometry with Elevated Energy (MSE), Data-Independent Acquisition (DIA), Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH)). These approaches effectively bridge the gap between traditional methods by broadening metabolite coverage while retaining the sensitivity and quantification accuracy of targeted analysis. It also makes it possible to include non-target contaminant/suspect screening in a metabolomics workflow, helping to characterise the full profile of exposure stressors under study. This framework is illustrated in Fig. 2 (Panels A and B).

Targeted metabolomics is helpful if one knows the specific metabolic pathways that might be affected by a pollutant(s) of interest. Some have argued that targeted metabolomics is not true metabolomics, but rather metabolic profiling, since, by definition, metabolomics (and any omics science) seeks to identify as many things as possible. However, if there are specific metabolites and/or metabolic pathways that are changing, and some that are not, it makes sense to focus on the former and not '*dilute the signal*' (so to speak) with non-relevant data. This focus becomes especially valuable when those metabolites clearly map onto known Molecular Initiating Events (MIEs) or Key Events (KEs) within AOPs,[24] allowing tar-

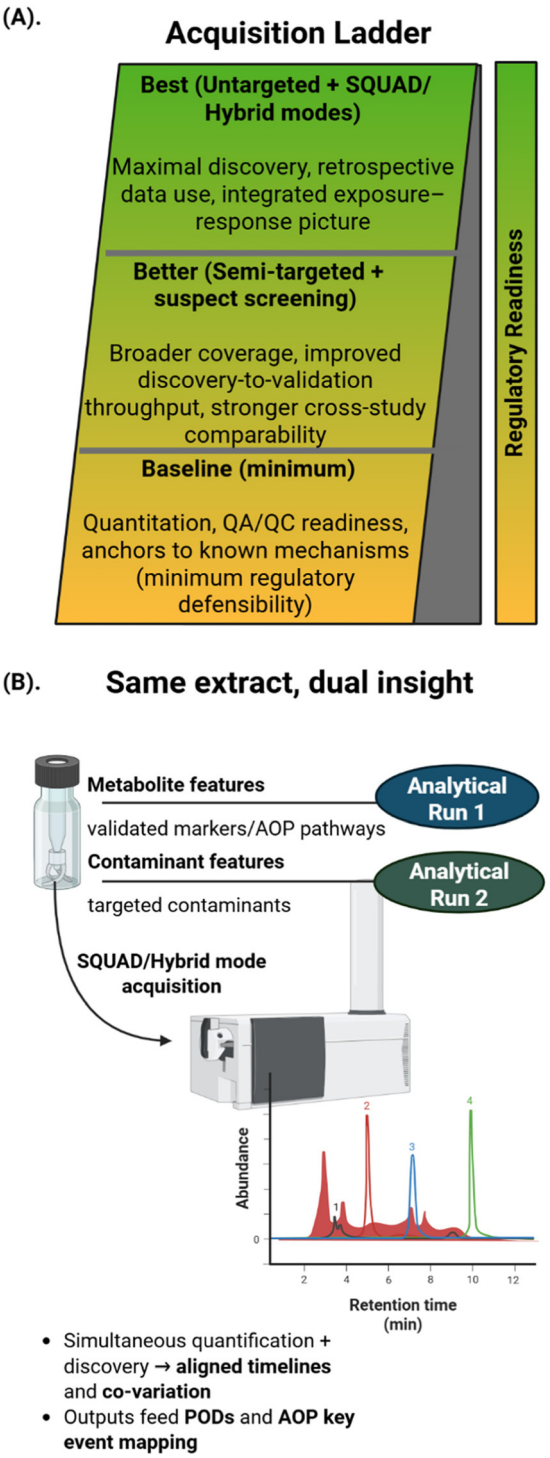

Fig. 2 Acquisition strategy linking metabolomics to contaminant insight. (A) A three-level ladder clarifies good–better–best acquisition choices that are aligned to regulatory needs. As a minimum standard, pair targeted metabolomics with a targeted environmental screen. Semi-targeted/widely-targeted and untargeted approaches (including SQUAD and hybrid HRMS modes) provide broader feature coverage, enable suspect/non-target contaminant discovery from the same extracts, and strengthen the linkage between exposure and biological effect. (B) Proposed workflow showing parallel extraction and analytics producing both metabolite and contaminant information, bridged by SQUAD to support retrospective mining and quantitation.

geted metabolomics to directly quantify AOP-relevant biochemical changes. *Noting that MIEs are the first measurable biological interactions at the molecular level within a biological system and initiate a pathway leading to an adverse effect, KEs represent the necessary biological change along the pathway, and an AOP links a MIE to an adverse outcome through a series of causally connected KEs*. By measuring metabolites already associated with toxicity pathways, which are cellular response pathways that can lead to adverse biological effects when perturbed by chemical stressors, and validated under controlled laboratory conditions, targeted metabolomics applied to a field-based setting strengthens mechanistic interpretation and enhances the ability to link pollutant exposure to potential adverse outcomes. The practical implementation of these analytical strategies in environmental settings introduces additional constraints that influence study design, data quality, and interpretability.

Experimental challenges

Environmental metabolomics presents its own specific challenges in both the laboratory and the field. Concentrations of pollutants are generally present at ng L^{-1} to low μg L^{-1} levels and typically exist as part of complex mixtures in the environment. In contrast, laboratory exposures frequently use more well-characterised μg L^{-1} to mg L^{-1} concentrations. Slight differences in experimental setup can affect the results. Sinclair *et al.*[25] found that using different substrates did not affect survival but did elicit different metabolic responses to the same toxicants in adult and juvenile amphipods (*Austrochiltonia subtenuis*), for example. Insects fed different plastic substrates (*e.g.*, polylactic acid, polypropylene, polyethylene *etc*.) have shown similar results.[26] This raises an important question: when metabolic shifts are observed, does the toxicant itself truly drive them, or are they partly, or even primarily, shaped by the experimental set-up and its interaction(s) with the organism? In other words, does the substrate, or other experimental factors, modulate how the toxicant is experienced and processed, thereby ultimately shaping the organism's metabolomic response? This distinction is crucial because it determines whether observed metabolic changes should be interpreted as biomarkers of contaminant exposure or as artefacts of experimental design. It therefore cannot be ignored.

Correlating metabolite changes with pollutant concentrations is also not straightforward. Most environmental omics-based studies attempt to assess impact after it has occurred, but applying this knowledge outside the lab requires knowing the state of the environment in its '*unaffected*' state. Shah *et al.*[27] addressed this challenge by applying both targeted and untargeted metabolomics to surface sediments (n = 50) from four pristine estuaries along the Western Cape York Peninsula in Far North Queensland, Australia. Their analysis revealed clear taxa–function relationships that predicted microbial community metabolic potential, with pathways such as carbon metabolism and amino acid biosynthesis showing strong positive correlations with community-level metabolic outputs—including 2-oxisocaproate, tryptophan, histidine, citrulline, and succinic acid. These findings establish a valuable baseline microbial metabolic blueprint against which future ecosurveillance efforts can assess environmental change. They also provide a necessary foundation for emerging effects-based and weight-of-evidence approaches in ecological risk assessment, although routine regulatory adoption remains in its early stages.

Subsequent research incorporated targeted screening for both organic contaminants (*e.g.*, pesticides) and inorganic pollutants (*e.g.*, metals) to determine how perturbed systems diverge from this baseline across similar geographical regions.[28–31] This integrative approach enables researchers to link shifts in community metabolism not only to natural environmental variation but also to specific contaminant pressures, strengthening the diagnostic power of ecosurveillance frameworks.[3]

Bioaccumulation is another important factor to consider. This occurs when an organism absorbs pollutants from the environment at a rate greater than it can excrete and/or metabolise them, resulting in a concentration in the body much higher than in the surrounding media.[32–35] Beale *et al.*[36,37] demonstrated a significant bioaccumulation of perfluoroalkyl sulfonic acids (PFSAs), a 100-fold increase in serum relative to surrounding water concentrations, and that these elevated body burdens were associated with a range of adverse biological effects on freshwater turtles (*Emydura macquarii macquarii*). The effects included altered egg composition (particularly in magnesium to calcium ratios, potentially affecting eggshell strength); altered biochemical profiles and hatchling deformities (due to adult females offloading PFAS into eggs). Similar observations have been made in snakes, toads and frogs.[11,38,39] These findings highlight the critical need to overcome such experimental challenges by measuring contaminants directly within the surrounding environment—such as water, sediment, soil, and local flora and fauna—because exposure concentrations, exposure routes (*e.g.*, waterborne uptake, dietary intake, dermal adsorption) and the rates at which organisms take up and eliminate contaminants all vary substantially across species, habitats, and environmental conditions. Without such context, it is difficult to determine whether observed metabolic or developmental effects stem from direct rapid/recent toxicant exposure or long-term, trophic transfer, or cumulative environmental loading.

Field applications

Environmental pollutants may vary in concentration over time. For example, Singh *et al.*[40] showed that the concentrations of PFAS in three estuaries varied by as much as 7 times in response to the tide. Furthermore, Melvin *et al.*[41] found different concentrations of metals corresponding with the '*wet*' and '*dry*' seasons in a metal(oid)-contaminated wetland, and this influenced the metabolic response of mosquitofish (*Gambusia holbrooki*) inhabiting the studied ponds. These studies imply that sampling at different points in the tidal

cycle and season may result in an overestimation or underestimation of the actual concentration of contaminants present in the system being studied.

Matrices, such as soils or sediments, must often be finely crushed under liquid nitrogen and extensively sonicated to ensure that all microbial metabolites are released from the matrix;[42] samples from marine systems may contain high concentrations of salts that need to be removed.[43] Some biofluids may need to be treated with β-glucuronidase enzymes to convert hydroxylated and carboxylated metabolites back to the parent form.[44] Samples from complex matrices, such as plants, may require fractionation or the use of QuEChERS (Quick, Easy, Cheap, Effective, Rugged, and Safe) sample preparation to separate out pigments and other large molecules.[44] This type of work is particularly challenging in remote locations, where logistical constraints can limit sampling frequency, accessibility, and the ability to comprehensively characterise exposure sources. In addition, collected samples must be kept cold to prevent post-collection metabolic changes during transport, making strict cold-chain maintenance essential to preserve sample integrity and ensure reliable metabolomic and contaminant analyses. Several studies have directly compared field-compatible preservation methods (*e.g.* RNAlater™ and chemical quenching) with conventional −80 °C storage, showing that some approaches can yield comparable metabolomic profiles for specific tissues and metabolite classes.[45,46] These findings support their use in field-based metabolomics where immediate ultra-cold storage is impractical, while also highlighting the need for method-specific validation.

Field-based environmental metabolomic approaches

Field-based metabolomics provides a useful framework for understanding how organisms respond to real-world environmental stressors across a gradient of natural systems. This approach encompasses wild-caught organisms, which integrate authentic, long-term exposure histories; caged organisms, which offer controlled exposure windows within natural habitats; and mesocosms, which emulate complex ecosystems under semi-controlled conditions. Complementing these are laboratory exposure experiments using field-collected samples, which allow researchers to isolate specific stressors while still retaining environmental relevance, as well as wildlife-derived cell lines, which extend metabolomics into mechanistic, reductionist systems anchored in the biology of wild species. These complementary exposure platforms create an integrated continuum for interpreting metabolic responses to environmental change, as conceptualised in Fig. 3, and discussed in more detail below. The diagram in Fig. 3 illustrates how diverse exposure platforms—field-based environmental metabolomics approaches (including wild-caught organisms, caged organisms, mesocosm systems, laboratory exposures using field-collected samples, and wildlife-derived cell lines)—capture meta-

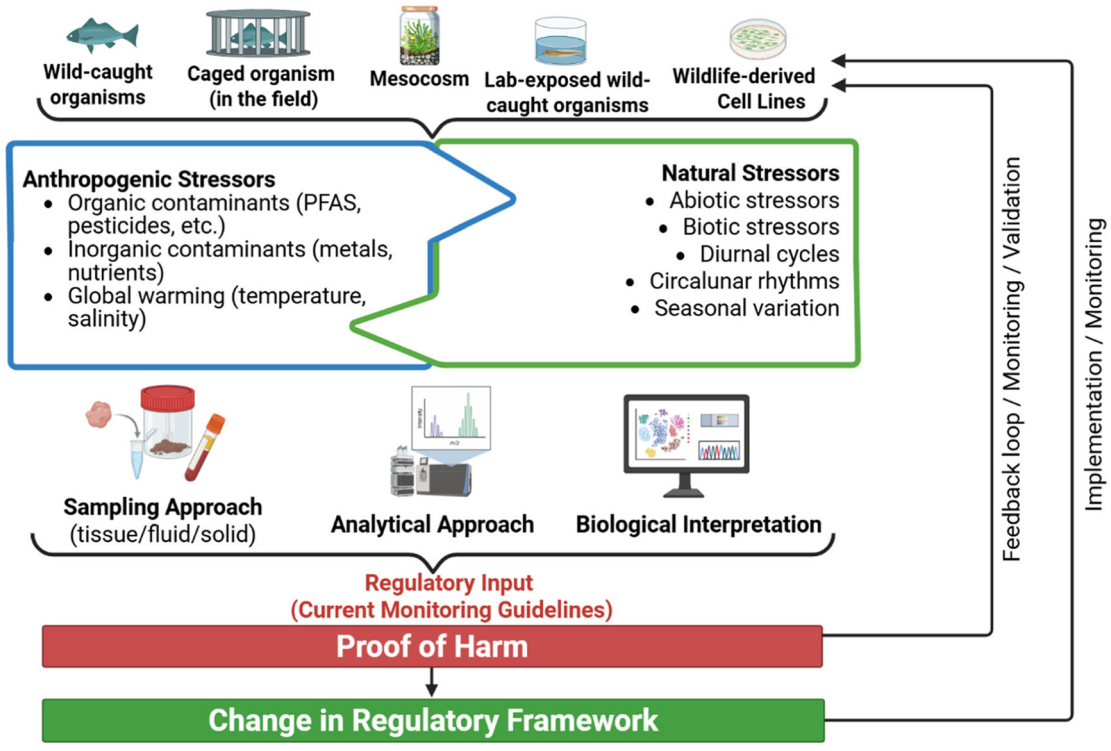

Fig. 3 Conceptual framework for field-based metabolomics approaches and its application to ecological assessment and regulatory decision-making.

bolic responses to complex mixtures of chemical, climate-related, and other anthropogenic stressors. Environmental and biological context (abiotic conditions, biotic interactions, habitat variability) shape sampling strategies, analytical workflows, and interpretation of metabolic endpoints. These components collectively feed into evidence-of-harm assessments and regulatory processes, establishing a feedback loop that strengthens environmental monitoring frameworks and supports the development of new ecologically grounded regulatory guidelines.

Wild caught samples

Wild-caught samples represent studies that have collected organisms/specimens (including tissue, organ and blood samples) from the environment and analysed their metabolomic profiles within the laboratory. Across wildlife studies, metabolomics has emerged as a powerful tool for assessing sub-lethal effects of environmental contaminants. A consistent finding is the disruption of lipid metabolism, particularly involving phospholipids and fatty acids, in response to exposure to PFAS, polychlorinated biphenyls (PCBs), and short-chain chlorinated paraffins (SCCPs), and other persistent pollutants. This was evident in polar bears and ringed seals,[47] belugas,[48] and Arctic charr,[49] suggesting that lipid-related pathways may serve as sensitive indicators of exposure to lipophilic, bioaccumulative contaminants across taxa.

Species-specific metabolic responses have also been observed, with differences in contaminant sensitivity and biotransformation capacity influencing metabolomic outcomes. For example, ringed seals showed stronger metabolic responses to PFAS than polar bears despite lower contaminant concentrations, highlighting the importance of physiological traits in interpreting exposure effects.[47] This cross-species analysis of exposure and metabolome across regions allows for the identification of correlations across broader contamination gradients/profiles and species.

Several studies have proposed oxidative stress and disruptions to energy metabolism as common pathways affected by metals, organohalogens and many other contaminants,[49,50] with some metabolites identified as potential biomarkers of exposure. Gago-Tinoco *et al.*[50] reported that while certain metabolites showed promise as biomarkers of exposure, some of these were tissue-specific, though some were found to be common across all sample types within a given site. Despite discussing metals and metabolites in parallel, the study did not analyse correlations between contaminant concentrations and metabolic profiles, limiting the strength of mechanistic inference. In contrast, Gauthier *et al.*[49] highlighted that PCB concentrations in Arctic charr (*Salvelinus alpinus*) muscle were 29 ng g^{-1}, within ranges known to induce reproductive toxicity in other fish species, and linked the metabolic differences observed between populations in two lakes to these elevated PCB levels. Melvin *et al.*[41] demonstrated, using mosquitofish (*Gambusia holbrooki*) from a metal(loid)-contaminated wetland, that suites of metabolites, owing to their integrative roles within biochemical pathways, are sensitive indicators of metal stress. Nevertheless, their application as environmental biomarkers depend on rigorous validation and quantitative assessment to confirm reliability and ecological relevance in monitoring programmes.

In another example, metabolomics was applied as part of a regulatory investigation into PFAS contamination in freshwater turtles (*Emydura macquarii macquarii*) in Queensland (Australia), providing sensitive biochemical evidence of sub-lethal impacts linked to pollutant exposure.[32,36] Turtles sampled from PFAS-impacted catchments showed extreme PFAS bioaccumulation, with serum PFOS concentrations reaching up to 235-fold higher than surrounding water.[51] Metabolomic analyses revealed consistent alterations in purine metabolism, glycerophosphocholine and lipid metabolism, immune-related pathways, and central carbon metabolism, all of which correlated with elevated PFAS burdens.[52] Integrated microbiome–metabolome studies further showed disruptions to amino acid, butanoate, and nucleotide metabolism, demonstrating combined host–microbiome responses to PFAS exposure.[53,54] Additional sampling identified hormone disruption, increased gout risk in adult turtles (which is deadly for reptiles), altered mineral ratios in eggs, and hatchling deformities, with population modelling predicting long-term decline at highly contaminated sites.[36] Collectively, these findings demonstrate that metabolomics, embedded within a regulatory assessment framework, can provide mechanistic evidence linking PFAS exposure to biological impairment and support improved ecological risk assessments for contaminated wildlife. Other studies have examined reptiles and amphibians in similar regulatory settings,[11,38,39] but unlike the freshwater turtle study, they did not focus on establishing proof of environmental harm.

Despite these insights, several important limitations persist across studies. Many did not establish direct correlations between contaminant concentrations and metabolite responses[50,55] or the correlations were not strong,[11,38,39] limiting mechanistic interpretations.[56,57] For example, Nzabanita *et al.*[55] measured differences in metabolites between Pacific black ducks (*Anas superciliosa*) based on site location, but observed metabolite patterns did not correlate with feather metal concentrations. Importantly, this study did not measure non-metal contaminants and therefore could not assess whether unmeasured chemical exposures contributed to the observed metabolomic variation. This finding raises concerns about the development of biomarkers, as the metabolic impacts of all stressors/variables present in an environment must be understood to infer the drivers of metabolomic responses.

In the study by Nzabanita *et al.*,[55] tissue mismatch introduces additional limitations, with the authors measuring the accumulation of metals in one tissue and the metabolome in another, potentially obscuring relationships due to differential accumulation and metabolic activity. Additionally, small sample sizes, sex bias, and limited geographic coverage reduced statistical power and generalizability.

Analytical constraints can impact data quality. For instance, suboptimal sample storage temperatures reaching higher than

the recommended −80 °C for long-term biobanking[47] and semi-quantitative metabolomics approaches[55] may affect metabolite stability and detection. In some cases, contaminant concentrations were not reported directly but referenced from prior studies,[49] complicating exposure assessment. However, a further methodological challenge lies in the fundamental mismatch between metabolomics workflows, which aim to capture all measurable metabolic changes, and traditional contaminant analyses, which typically rely on targeted methods that measure only a predefined list of chemicals. This discrepancy creates an inherent imbalance in data-acquisition breadth, making it difficult to align wide-scope metabolic responses with narrow contaminant panels. Yet this gap is surmountable: modern high-resolution mass spectrometry within metabolomics pipelines now enables both targeted and untargeted acquisition modes, supporting broad suspect-screening and non-targeted contaminant discovery in the same biological sample. Integrating these complementary datasets, together with improved annotation and mapping of metabolomic pathways, will expand the network of potential biochemical interactions and strengthen our ability to link biological changes directly to environmental stressors, improving correlation and, ultimately, mechanistic inference from wild-caught organisms.

Caged field metabolomics

The use of caged (field-deployed) organisms, particularly bivalves (such as mussels) and insect nymphs, has proven effective for detecting subtle biochemical responses to environmental stressors *in situ*. Field-deployed exposures allow for real-time monitoring of pollution impacts under realistic conditions, offering insights into both contaminant bioavailability and organismal responses. For example, metabolomic profiling of caged mussels (*Dreissena* sp.) deployed along pollution gradients in the Great Lakes (east-central North America) revealed site-specific alterations in energy and amino acid metabolism that were not detectable with traditional biomarkers alone, uncovering early stress under realistic exposure conditions.[58] Despite challenges in standardisation, quantification, and model integration, metabolomic data from caged exposures can enhance ecological management by enabling early detection of stress, guiding conservation efforts, and improving understanding of ecological and evolutionary responses.[1]

Mussels are particularly well-suited for such studies due to their sedentary lifestyle, filter-feeding behaviour, contaminant bioaccumulation capacity, and ease of deployment. Studies using *Mytilus galloprovincialis* in Italy demonstrated that ^{1}H NMR metabolomics, combined with histology and immunohistochemistry, can sensitively detect petrochemical pollutants such as mercury and PAHs.[59–61] Disruptions in osmoregulation, energy metabolism, and neurotransmission were observed across gill, muscle, and digestive gland tissues, with carbonic anhydrase proposed as a biomarker. However, regulatory relevance was limited by single-site exposures, narrow chemical profiling, and a lack of dose–response data.

Similar findings were reported using a clam species (*Ruditapes decussatus*) transplanted to contaminated agricultural and urban runoff sites in Spain. Digestive gland metabolomics using LC-MS revealed elevated levels of amino acids, osmotic protectants, and nucleotides after 7 d of exposure, likely reflecting disturbances in osmoregulation and energy metabolism. Metabolite levels dropped below those of control clams following 22 d, suggesting a two-phase metabolic response: initial compensation followed by metabolic exhaustion.[62] Taurine emerged as a potential biomarker for complex contaminant mixtures, though the authors emphasized that metabolomic responses in field studies should be interpreted by considering environmental parameters such as trophic conditions and potential time-dependent patterns.

Caged zebra mussels (*Dreissena polymorpha*) were used to demonstrate that combining metabolomics with classical biomarkers is beneficial.[63] The study suggested that site-specific metabolic profiles revealed impacts on osmoregulation and anaerobic metabolism, while traditional energy biomarkers were less consistent, lactate was identified as a promising health indicator.

In the Great Lakes, mussel metabolomes correspond closely with local chemical pollution profiles, underscoring the utility of combining chemical and metabolic data[64] to provide a holistic view of stressor impacts, though seasonal metabolic cycles and stressor interactions must be carefully considered.

Recent multi-omics studies using *Sinanodonta woodiana*, commonly known as Chinese pond mussel, revealed pronounced tissue-specific adaptive responses to environmental contamination. These studies identified key metabolites, including alanine, glutamate, and 7-oxocholesterol as consistent biomarkers of exposure across multiple tissues and sampling times, effectively discriminating streams with differing pollution loads.[65,66] The inclusion of antioxidant enzymes and metallothionein responses strengthened causal inference, though biomarker validation remains ongoing.

Studies using other caged macroinvertebrates further highlight the value of field deployable metabolomics frameworks. Notably, work involving Brua and co-authors has advanced the application of caged invertebrates, particularly aquatic insect nymphs and crayfish, in real-world contaminant assessment.[67–69] Izral *et al.*[69,70] demonstrated that caged crayfish exhibit clear, energy-related metabolic shifts in response to food limitation and dissolved oxygen stress, identifying metabolites in tail muscle as sensitive indicators of sublethal physiological disruption. In a complementary *in situ* study in the Athabasca River, transplanted dragonfly nymphs (*Ophiogomphus colubrinus*) housed in baskets were used to assess the combined effects of municipal sewage effluent, polycyclic aromatic hydrocarbons (PAHs) from oil-sands extraction activities, and naturally eroded bitumen-derived contaminants (Fig. 4 illustrates field-deployable cages used to house dragonfly nymphs), for a region characterised by some of the world's largest oil deposits (northeastern Alberta, Canada). Despite exposure to complex mixtures of heavy metals and PAHs, nymph survival remained high and their metabolomes did not

Fig. 4 Example of a field-deployable cage used in an environmental metabolomics study. Each cage is used to house two dragonfly nymphs (*Ophiogomphus colubrinus*). The photographs show the 1 mm mesh basket fully zipped for deployment (left) and open prior to closure (right). Reprinted from Applied Environmental Metabolomics, R. Brua, J. Culp, S. Pomfert, D. Halliwell, Chapter 19 – NMR-based metabolomics of dragonfly nymphs exposed to multiple stressors: An approach for field assessments to diagnose effects, Pages 273–289, Copyright (2022), with permission from Elsevier.

differ significantly among sites after one week. This may reflect both species-specific tolerance to these contaminants and the potential need for longer or more sensitive exposure designs in multi-stressor environments.[68] Together, these studies emphasise the promise of caged macroinvertebrates as field-ready sentinels and demonstrate how metabolomics can deepen ecological interpretation in systems influenced by interacting anthropogenic and natural stressors.

Field-based metabolomics with fish species like fathead minnows (*Pimephales promelas*) and yellow perch (*Perca flavescens*) has also shown promise for identifying contaminant-driven biological responses. In the Great Lakes Basin, liver metabolite shifts were linked to contaminants such as DEET (*N*,*N*-diethyl-*meta*-toluamide) and bisphenol A, though many acted as markers of wastewater effluent rather than direct toxicants.[71] Despite the constraints of single-tissue analysis and the lack of temporal resolution, the integration of metabolomics with chemical analyses offered valuable evidence for prioritising contaminants of emerging concern and emphasised the need for replication, multi-tissue analyses, and consideration of confounding stressors such as temperature.

Ekman *et al.*[72] demonstrated that targeted and untargeted metabolomics could detect both estrogenic and oxidative stress responses in *P. promelas* exposed to wastewater treatment plant (WWTP) effluents. Dose–response relationships were observed for estrone and several pesticides (thiabendazole, malathion, and 1,4-dichlorobenzene), reinforcing the value of untargeted metabolomics in characterising complex mixtures. The authors emphasised that such tools are critical for assessing the diverse effects of anthropogenic contaminants.

Finally, Defo *et al.*[73] highlighted the complexity of interpreting metabolomic data in caged *P. flavescens* exposed to urban WWTP effluent. Differences between caged and lab-reared fish suggested a caging effect, while limited metabolomic separation between effluent-exposed and reference fish raised questions about adaptive responses and biomarker sensitivity. Differences in profiles of liver metabolites emerged by week six, but the authors noted that larger sample sizes were needed to confirm these patterns. These findings underscore that contaminant exposure does not always result in immediate or measurable metabolomic disruption, raising critical questions about the timescales of adaptive responses and the sensitivity of current biomarker approaches in complex field environments.

While field-deployable caged metabolomics has clear value for detecting sublethal stress under realistic field conditions, major challenges remain in disentangling metabolomic signals from natural variability, seasonal cycles, and multi-stressor interactions. These limitations highlight the need for more standardised designs, longer exposures, and multi-tissue, multi-omics integration. Even so, the approach offers substantial opportunity; by coupling metabolomics with chemical analyses and complementary biomarkers of effect and exposure, caged organisms can become a powerful field-ready tool for diagnosing early ecological stress and strengthening environmental assessments within a regulatory framework.

These field observations highlight both the interpretive power of metabolomics and the considerations required when translating biological signals into regulatory or monitoring contexts.

Lab reared organisms exposed to field samples (mesocosms)

Metabolomics has been applied to studies using laboratory-reared or laboratory-maintained organisms exposed to field-derived materials, most commonly through mesocosm experiments. The original focus of this review was on field-deployed mesocosms, where environmental realism is maximised by conducting exposures directly within natural ecosystems. Like the large-scale transcriptomics ExStream System mesocosm experiments[74–77] (Fig. 5).

However, because few such studies currently exist using metabolomics, this section also includes mesocosm-like experiments conducted in greenhouses and aquaria, where environmental factors such as temperature and photoperiod were not fully controlled but still incorporated field-relevant matrices or stressors. It is important to note that field-deployed mesocosm studies remain rare for several reasons, including the substantial logistical effort required to transport materials and maintain experimental systems in remote or uncontrolled environments; the need for strict regulatory and permitting approvals to deploy organisms or experimental infrastructure in natural habitats; and the challenge of preserving metabolite integrity through reliable onsite cold-chain procedures. Moreover, *in situ* mesocosms are vulnerable to unpredictable environmental variability (*e.g.*, storms, heat-

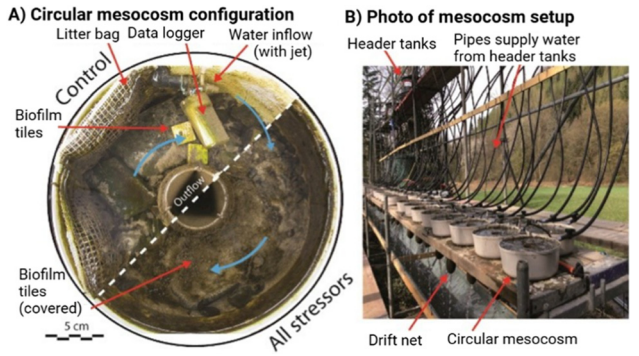

Fig. 5 The ExStream mesocosm system consists of circular flow-through units containing three compartments (streambed substratum, litter bag and drift net) to capture benthic, detrital and drift responses. Panel A illustrates variation between control mesocosms (above the dashed line) contrasted with those receiving stressors (below the line). The full setup includes 64 replicated units (panel B). Reprinted from Science of the Total environment, vol. 610, A. Beermann, V. Elbrecht, S. Karnatz, L. Ma, C. Matthaei, J. Piggot, F. Leese, Multiple-stressor effects on stream macroinvertebrate communities: A mesocosm experiment manipulating salinity, fine sediment and flow velocity, Pages 961–971, Copyright (2018), with permission from Elsevier.

waves, hydrological shifts) and disturbance by wildlife or humans, complicating experimental control, replication, and sample recovery. These combined constraints, along with the high analytical demands of metabolomics, mean that many researchers opt for semi-controlled mesocosm systems as a practical compromise between environmental realism and experimental feasibility.

Across these reported studies, metabolomics was used to characterise organismal responses to a range of environmental stressors. These included rhizosphere microbial communities and plants subjected to nutrient and moisture-limited soil,[78,79] insects exposed to a range of copper concentrations[80] and mussels exposed to oily wastewater discharges before and after biofilm membrane reactor treatment.[81]

The only true field-deployed mesocosm study to date was conducted by Jeppe *et al.*,[80] who spiked reference-site sediments with copper, placed them into tub-based mesocosms, and deployed them back into a wetland to assess impacts on aquatic macroinvertebrates. Laboratory-bred snails (*Potamopyrgus antipodarum* and *Physa acuta*) and the chironomid, *Chironomus tepperi*, were introduced at different time points, enabling assessment of responses across multiple biological levels. Metabolites from *C. tepperi* were quantified using a targeted LC-MS method.[80,82] Importantly, metabolite shifts were detected at 60 mg kg^{-1} copper concentrations, lower than the concentration that was reported to affect larvae dry weight with an EC_{50} of 238 mg kg^{-1} for copper, demonstrating the higher sensitivity of metabolomics for early detection of sediment toxicity.

Other mesocosm studies, though not field-deployed, further highlight the utility of metabolomics for disentangling organism–environment interactions. Baker *et al.*[78] profiled metabolites from rhizosphere soil under nutrient-limited, and moisture-limited conditions over 18 weeks in a greenhouse, revealing distinct shifts in nitrogen-containing metabolites and metabolites associated with moisture stress. Cai *et al.*[79] used a plant-wastewater mesocosm to evaluate how nitrogen speciation (NH_4^+–N *vs.* NO_3^-–N) influences plant metabolic responses in constructed wetlands, showing clear differences in root exudate metabolite profiles that reflected nitrogen uptake, energy metabolism, and growth. Gornati *et al.*[81] examined mussels (*Mytilus galloprovincialis*) exposed to untreated and treated oily wastewater within aquaria-based mesocosms, finding tissue-specific metabolite alterations consistent with differing levels of wastewater treatment efficacy. Gyawali *et al.*[83] performed a similar experiment with mussels exposed to human wastewater and used metabolomics to identify potential biomarkers of exposure related to Norovirus contamination in shellfish, and Nguyen *et al.*[84] to identify acute heat stress responses in Abalone.

Collectively, these studies demonstrate that metabolomics provides sensitive and diagnostically powerful endpoints in mesocosm-based ecotoxicology, even when environmental realism differs across systems. Although relatively few true field-deployed mesocosm studies have incorporated metabolomics, transcriptomic analyses have already been successfully applied in such settings,[77] indicating that the logistical and methodological barriers are not prohibitive. This gap highlights a clear opportunity to expand metabolomics into field-deployable mesocosm experiments that would allow researchers to verify whether biomarker patterns and mechanistic stress responses observed under controlled conditions persist in natural environments. Broadening this work is essential for advancing environmental metabolomics toward routine use in biomonitoring and regulatory assessment frameworks.

Field samples collected and used for experiments in a laboratory

Laboratory experiments using field-collected animals or media offer a middle ground between the ecological realism of natural exposures and the mechanistic control of laboratory toxicology. These organisms/media carry authentic environmental histories, including prior contaminant exposure, natural dietary patterns, and seasonal metabolic states, allowing researchers to probe how real-world conditions shape metabolic responses. When transferred into controlled laboratory settings, field-collected animals can be exposed to defined stressors or environmental media can be used as the induced stressor platform, enabling the establishment of causal links between environmental variables and metabolomic shifts while reducing confounding noise. Furthermore, field-collected organisms are commonly used for laboratory-based exposure studies, particularly when the objective is to investigate toxicity in non-model or locally relevant species for which cultured cell lines may not exist. However, field-collected fauna can be costly to source, transport, and maintain under controlled conditions, and these operational demands, plus ethical requirements, for endangered species or populations

with low abundance can lead researchers to use taxa obtained from commercial suppliers for laboratory exposure experiments.

Acclimation to laboratory conditions prior to experimentation is standard practice with wild caught organisms (and for commercially sourced taxa), to ensure the animals are unstressed and healthy, and to increase both statistical power and the reliability of results by reducing intraspecific variation of the test population. The acclimation process involves transitioning animals to laboratory housing (including water for aquatic species), lighting, feeding and handling. On the one hand, such acclimation may be particularly important for metabolomics research since an abrupt change to environmental conditions or the animals' experience will likely have a physiological toll. On the other hand, little attention has been paid to how the acclimation process, or the use of long-standing laboratory-cultured lineages, might reduce the ecological relevance of metabolomic responses due to the '*de-wilding*' of wild-collected organisms. Depending on the goal of the study (*i.e.*, unravelling mechanistic information *vs.* predicting environmental risk), approaches for acclimating field collected organisms prior to laboratory testing may differ.

Timeframes for acclimating field-collected organisms to laboratory conditions vary widely, spanning anywhere from a few days to several months, and although lab conditions likely differ considerably from the collection site, these are often undisclosed. In a study in the USA, wild oysters (*Crassostrea virginica*) were acclimated for 30 d to constant temperature, salinity, photoperiod, and a commercial food, in filtered UV-sterilised artificial sea water.[85] A similar metabolomics study with pearl oysters (*Pinctada martensii*) collected from Lingshui harbour, China, acclimated animals to laboratory conditions for just 3 d, and used filtered natural sea water.[86] Neither study reported the environmental parameters present at the collection site, limiting interpretation of pre-exposure influences. Wild-caught adult topsmelt (*Atherinops affinis*) collected by beach seine in the USA were acclimated for 3 m to laboratory conditions prior to exposure to crude and dispersed oil, and responses were compared to unacclimated topsmelt embryos obtained from a commercial supplier.[87] The authors attributed disparities in metabolomic responses to life-stage differences but did not consider how organism origin or acclimation history may have contributed to the observed variation.

Depending on the organism, it may be unlikely or unfeasible for metabolomics studies to use organisms that have not been fully acclimated to laboratory conditions. For example, Cladocerans and Chironomids are commonly cultured in ecotoxicology labs and reared for multiple generations to obtain enough healthy individuals for exposures,[88,89] sometimes for as long as a decade.[90] While this may contribute towards enhanced comparability between studies over time, the use of such populations raises uncertainty about how well the metabolism of these organisms reflects that of natural populations.

Failure to disclose the age of cultures from wild-caught populations is also common. For example, Dumas *et al.*[91] described the exposure of a laboratory culture of Mediterranean mussels (*Mytilus galloprovincialis*) to sewage effluent but did not provide any information about when the culture was established, its origin, or its rearing conditions. Similarly, Santos *et al.*[92] investigated metabolomic effects in stickleback (*Gasterosteus aculeatus*) exposed to copper, but while the authors indicate the test population originated from a pristine field site in England, they do not describe the conditions of the collection site or details surrounding the acclimation timeframe and simply state that the fish were maintained until sexually mature. Other studies fail to disclose the source of experimental animals altogether. Tang *et al.*,[93] for instance, reported acclimation conditions for earthworms (*Eisenia fetida*) to environmental chambers, but did not specify if the worms were wild-caught or laboratory-cultured. Beyond the potential for differences in water quality and environmental parameters between field and laboratory to influence the metabolome and its susceptibility to chemical insult, failure to fully characterise the environment where animals are collected could overlook pre-exposure to relevant physical or chemical stressors.

Despite these challenges, the integration of field-collected organisms, tissues and fluids into laboratory-controlled metabolomics experiments remains a promising and rapidly growing area of ecotoxicology. By combining ecological realism with mechanistic clarity, these studies help validate biomarkers, improve causal inference, and strengthen the interpretation of field-based metabolomic datasets. Overall, such studies demonstrate the value of metabolomics in ecotoxicology but highlight the need for standardised reporting of organism origin, acclimation conditions, and environmental context, along with matched tissue analyses and robust statistical frameworks to enhance reproducibility and support biomarker development.

Wildlife-derived cell lines

While metabolomics has yet to see widespread application in wildlife-derived cell lines, it offers a robust approach for investigating the responses of non-model species to environmental stressors, pollutants, and diseases. Moreover, this methodology aligns with the three Rs principles—Replacement, Reduction, and Refinement—by minimizing reliance on whole animal testing. Because many wildlife species possess unique physiological adaptations, metabolomic profiling helps reveal both conserved and species-specific biochemical pathways.[94] This overall workflow is illustrated in Fig. 6, which outlines the process of establishing wildlife cell cultures to assess their metabolic responses to contaminants.

Applications include assessing the cellular effects of pollutants, identifying metabolic signatures linked to ecological stress, and comparing metabolic phenotypes across related species. Studies in primates (*Rhesus macaques*),[95] for example, show that metabolomics can distinguish species-specific metabolic traits, while environmental toxicology research demonstrates its value for detecting biochemical disruptions caused by contaminants. Standardised metabolomics workflows devel-

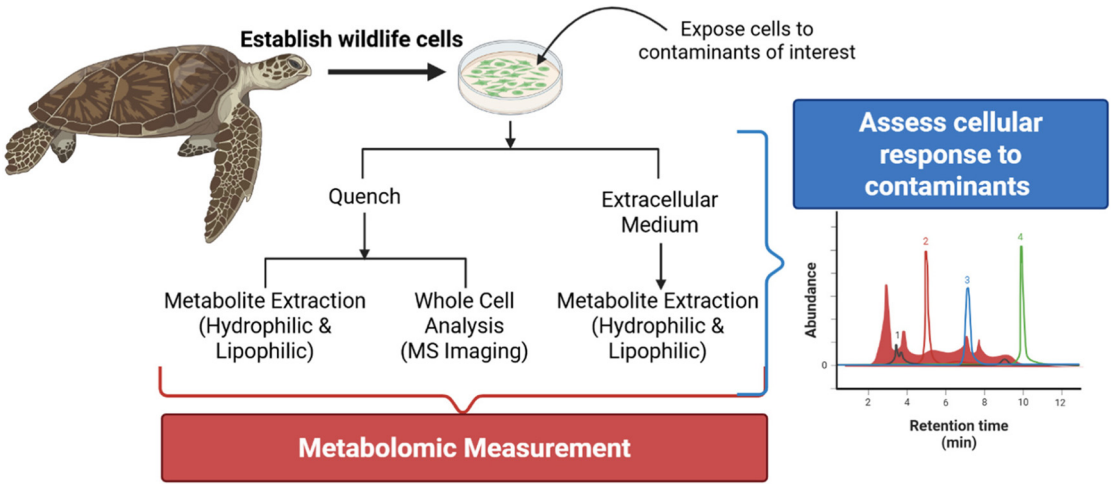

Fig. 6 Workflow for establishing wildlife-derived cell cultures and assessing their metabolic responses to environmental contaminants. Wildlife cells are first established and then exposed to contaminants of interest. Following quenching, intracellular and extracellular metabolites are extracted for hydrophilic and lipophilic profiling or analyzed using whole-cell mass spectrometry imaging. Resulting metabolomic measurements are used to evaluate cellular responses to contaminant exposure.

oped for mammalian cell culture also provide a methodological foundation for applying these techniques to wildlife-derived cells.[96] Research groups are establishing wildlife-derived cell culture biobanks dedicated to toxicity testing,[97,98] such as the Australian ARI-Tox Marine Wildlife Cell Bank[99] (https://aritox.com/mwcb/). While many global wildlife biobanks mainly focus on conservation,[100–104] their wildlife cell lines are available to be used in ecotoxicology, especially mechanistic toxicology and chemical mode-of-action work.

Strengths and limitations of field-based environmental metabolomics

Sensitivity

Field-based studies capture the biological variation that results from organisms' exposure to the dynamic and ever-changing environment. While some environmental variables are easily replicated *in vivo*, the sheer number of such variables makes total simulations challenging and expensive. There are countless attributes of environmental metabolomics studies that exert an effect on the metabolome through, for example, stress,[105,106] energy expenditure, homeostatic effort[107] and cyclic biological rhythms (circadian, circannual, circatidal, *etc.*).[108] Herein lies the major strength of environmental field-based metabolomics.[109] As highlighted throughout the previous section, environmental toxicology studies use metabolomics to measure sublethal effects of chemicals-of-concern through comparison of the same species in polluted test sites and less polluted or unpolluted reference sites. Here, a toxicological effect can be apparent in the metabolome without any overt signs of behavioural change, sickness or lethality. Similarly, biomonitoring is becoming a strong use case for metabolomics owing to its sensitivity.[67]

Known unknowns and unknown unknowns

The utility of metabolomics for measuring biochemical impacts due to often small changes in external factors can become a limitation. In field research, each organism and sampling location has different attributes, and these attributes change day-to-day and hour-to-hour.[110] Consequently, it is near impossible to find sites that are identical in all ways besides the variables of interest, meaning field metabolomics studies are potentially subject to more residual error than a well-controlled laboratory study. To aid in elucidating field effects and understanding residual error, field metabolomics researchers can capture data on the attributes of their sample sites. Data includes physicochemical measurements of water and soil, broad screening for organic pollutants, pharmaceuticals, and metals, and weather records.[111] The Metabolomics Standards Initiative (MSI) asks researchers to report these attributes along with things like the detailed location, altitude, moon phase and sampling time.[112]

Free-ranging organisms are likely to be exposed to complex mixtures of hazardous and non-hazardous chemicals, including natural products, breakdown products, nutrients and pollutants.[113] Such mixtures can cause interactive, additive and non-monotonic effects that complicate analyses and obscure comprehension of field metabolomics findings. Complex mixtures vary between sites, seasons and samples, and are challenging to characterise, leading to further between-site and sample variance and Type II statistical errors (false negatives).

Timing

Circadian rhythms are set and regulated by factors that change throughout the day, such as light quantity and quality, temperature, food consumption, socialisation and oxygen availability.[114–116] Circadian effects on metabolism are profound and easily measured with metabolomics.[117] Circadian

metabolic drift is relatively easily controlled for in laboratory studies, as it is generally quite straightforward to sample all experimental organisms over a short time frame and under the same conditions. In contrast, sampling is rarely so straightforward in the field. In field-based studies, such temporal resolution typically reflects repeated cross-sectional sampling of populations rather than true longitudinal tracking of individuals, as repeated sampling of the same organism is often infeasible and many metabolomic analyses are inherently destructive. Furthermore, field sites are often far apart, forcing researchers to travel between them and sample at different times of day or on different days altogether. This may induce Type I (false positive) errors, where a study measures metabolomic change due to time-of-day rather than site-specific attributes. Similarly, if sampling at any site extends over many hours the within-group variance may increase and lead to Type II (false negative) statistical error. Other biological rhythms (*e.g.* circatidal, circalunar, circasemilunar) may also modulate hormones, behaviour and, through logical extension, the metabolome, although characterisation is limited in the literature.[108]

Weather

The weather at any given time is unlikely to be exactly like the weather at another time, which can reduce repeatability and increase residual error in field metabolomics studies. Weather causes changes to temperature, humidity, evaporation rates, light quantity and quality, UV exposure, and water quality and availability. For example, rainfall can have profound effects on shallow aquatic ecosystems through dilution, alteration of physicochemical traits like oxygenation or pH,[118] or the introduction of pollutants and sediment *via* runoff.[119] A previous study measured significant metabolomic changes spurred by rainfall events in two populations of *Gambusia holbrooki* (Eastern mosquitofish).[41] Changing temperature also modulates metabolism, through enzyme activity rates in ectotherms and poikilotherms, and movement (*e.g.* shivering) and thermogenesis in endotherms.[120]

Addressing these strengths and limitations in practice depends not only on study design, but also on transparent reporting and rigorous analytical quality control.

Reporting in field-based environmental metabolomics

Environmental regulators and regulatory-science frameworks require a far higher degree of analytical performance than is typically seen in academic metabolomics, because policy decisions must be based on evidence that is demonstrably reliable, reproducible, and robust enough to stand up in court.[121] These decisions often draw on multiple lines of evidence, and the consequences of error can be substantial, affecting ecological integrity, human health, and imposing significant financial or legal penalties.[122]

Despite major advances in analytical instrumentation, field-based environmental metabolomics continues to suffer from inconsistent reporting of essential metadata. Compliance with the MSI remains low, with many studies omitting details necessary for assessing analytical rigour and reproducibility.[123] This inconsistency is common amongst metabolomics studies but is particularly consequential for field-based environmental studies, where environmental variability amplifies the need for transparent and comprehensive method reporting.

Data processing and pre-analytical steps

Most field studies employ untargeted metabolomics workflows, leveraging chemometric tools designed for large, complex datasets. Most reviewed studies explicitly reported normalization and scaling steps, commonly total or median intensity normalization, followed by log transformation and auto- or Pareto-scaling.[50,64] However, only a quarter used internal standards, limiting correction for extraction efficiency or ion-suppression artefacts. Blank and QC filtering—critical steps to remove technical noise—were reported by just over half of the papers reviewed, typically *via* exclusion thresholds for blank features[48] or removing variables with poor precision in pooled QC samples.[64] These steps are fundamental to distinguishing true biological signals from analytical artefacts.

Multivariate statistics and model interpretation

Multivariate statistical analysis (MVA) remains the backbone of field-based metabolomics. Principal Component Analysis (PCA) is the dominant exploratory tool, enabling visualization of sample clustering and identification of outliers.[63,71] Partial Least-Squares Discriminant Analysis (PLS-DA) is widely applied for supervised classification, with model validity commonly assessed using permutation tests and predictive Q^2 values, however, it is repeatedly prone to overfitting and misinterpretation when applied without rigorous validation, particularly in high dimensional, low sample size datasets.[124] Variable Importance in Projection (VIP) scores frequently guide the identification of discriminant metabolites. For multi-stressor ecosystems, Partial Least-Squares Regression (PLSR) is increasingly used to correlate metabolite profiles with contaminant concentrations or environmental gradients.[71] While effective, this approach is limited by the fact that neither chemical nor metabolomic datasets comprehensively capture all environmental stressors. Just over half of the review papers complemented multivariate findings with univariate statistics (*e.g.*, ANOVA, ANOVA Simultaneous Component Analysis (ASCA), *t*-tests), and a third incorporated pathway or network analysis. The limited uptake reflects challenges inherent to environmental metabolomics, such as ambiguous metabolite annotation and pathway redundancy. Pathway analysis in metabolomics is problematic because metabolite identification is often ambiguous and pathway databases are incomplete or poorly aligned with real metabolic network complexity. These issues can contribute to misleading or biologically implausible pathway enrichment results.[125]

About a third of the papers considered in this review used machine learning/predictive modelling such as defining predictive accuracy (Q^2) beyond validation or using methods like

concentration-response modelling to determine Benchmark Doses (BMD). These approaches require large independent data sets to build and test models, which is a limitation in many field-based metabolomics studies. Consequently, while AI- and machine-learning-driven predictive modelling represents a promising future direction for the field, its widespread and reliable application in environmental monitoring is likely to remain aspirational in the near term.

Machine learning, predictive modelling and networking

The machine learning methods or predictive models used in the reviewed literature included concentration–response modelling or BMD analysis to quantify metabolite–exposure relationships. These computational approaches are increasingly used to identify sensitive metabolic signals, prioritise biomarkers, and interpret complex multi-stressor exposure patterns. However, their broader adoption remains constrained by the scarcity of sufficiently large, independent training datasets needed to build and validate reproducible predictive models, a challenge made more acute in field-based metabolomics where true replication may be limited. For instance, concentration–response and BMD models typically require many quantified exposure levels that aren't skewed, such as those obtained by sampling repeatedly along a river pollution gradient. In comparison, sampling from discrete ecosystems (*e.g.*, separate lakes or wetlands) often yields only a few distinct exposure conditions, limiting the number of training points available to machine-learning models and reducing predictive power in multi-stressor environments.

The DRomics framework (which uses BMD analysis) streamlines dose-response modelling for omics datasets, like metabolomics.[126] It allows users to import and process omics data, identify responsive metabolites, fit various concentration–response models, and calculate BMD values. Tailored for ecotoxicology and environmental risk assessment, DRomics manages non-replicated datasets collected from field studies and supports comparisons across transcriptomics, proteomics, and metabolomics within a unified model. Tools such as DRomics-Interpreter offer biological annotation and pathway visualization.[127] Lettoof *et al.*[11] investigated dose–response relationships between contaminant concentrations (PFAS and metal(loid)s) and metabolomic alterations, in the livers and tissues of *Ranoidea moorei* (motorbike frogs) collected across contamination gradients in urban wetlands. Wild-caught frogs were used, and their chemical burden linked to whole-organism health metrics derived from metabolomics datasets. Such case studies show DRomics helps identify dose-responsive metabolic pathways, distinguish early *versus* late molecular effects, compare omics sensitivities, and establish quantitative thresholds for AOPs and ecological risk assessments.

As computational power is increasing, untargeted metabolomics analysis has looked toward networking, such as GNPS (Global Natural Products Social Molecular Networking) to address the challenges of identifying unknown metabolites. These networks organise fragmentation spectra into molecular networks that reflect structural similarity, supporting annotation propagation and comparative analyses across datasets.[128] This approach is powerful for exploring chemical space without prior knowledge, but it is constrained by incomplete spectral libraries, leaving a '*dark metabolome*' of unannotated features. The dark metabolome, as well as the spectral libraries, can be confounded by in-source fragments and other artefactual features that form clusters unrelated to true metabolites.[129] Careful feature curation, fragmentation-aware workflows, and orthogonal validation are therefore essential to avoid overinterpretation if these approaches are to be considered.

Deficiencies in quality control (QC) reporting

Quality control reporting remains the weakest area of field-based metabolomics. The analytical QA/QC requirements for metabolomics are broadly similar across laboratory and field studies; however, field-based metabolomics must additionally manage uncertainty associated with uncontrolled exposure conditions, environmental variability, and sampling logistics. These factors necessitate more rigorous documentation of sampling context, biological covariates, preservation methods, and metadata to ensure interpretability and regulatory relevance.

Only a quarter of studies reported using internal standards, half of that reported including pooled QC samples, and only 5% documented system suitability testing, despite MSI and Chemical Analysis Working Group (CAWG) guidance emphasising their importance.[123] Reporting of sample handling—collection conditions, storage regime, freeze–thaw cycles—was similarly low (16%), even though storage conditions significantly affect metabolite stability.[64,123] No reviewed studies explicitly stated that sample preparation or analysis order was randomised, a key requirement for avoiding batch or drift related artefacts.[130] Data deposition in public repositories such as the Metabolomics Workbench,[131] MetaboLights[132] or Dryad Digital Repository[80] was recorded in only 20% of papers, mirroring broader concerns that MSI guidelines are difficult to interpret and inconsistently enforced.[133] The deposition requirement, defined by MSI,[134] aims to make raw data and metadata available for re-analysis, but the submission process remains cumbersome.[130]

Opportunities for improvement

Improving field-based environmental metabolomics requires much greater emphasis on transparent, standardised reporting, particularly regarding quality control procedures and metadata. Mandatory documentation of sample handling, randomisation, internal standards, pooled QC samples, and instrument performance checks is essential for ensuring high data quality, reproducibility and enabling cross-study comparisons. This need is underscored by persistently low compliance with MSI and CAWG guidelines[123] and the high degree of analytical variability observed in field-based studies.[130]

Importantly, these improvements are not only methodological but also critical for regulatory relevance. Regulatory

decisions often carry significant ecological, societal, and financial consequences, including impacts on biodiversity, ecosystem services, human health, and economic penalties associated with environmental noncompliance. For most free-ranging wildlife, particularly large-bodied, long-lived, or disturbance-sensitive species, regulators lack laboratory-derived effects or risk data, meaning that evidence of exposure and biological impacts must necessarily be derived from organisms exposed in the field.

The use of metabolomics data that lack adequate QC controls, transparent reporting, or robust validation therefore poses a risk of misinforming management actions or misattributing environmental harm.

To meet regulatory expectations, field-based metabolomics must adhere to the same data integrity standards historically required in environmental chemistry and toxicology. Strengthening documentation of internal standards and pooled QC samples, improving transparency in storage and sample handling conditions,[135–137] and implementing routine systems suitability testing would markedly improve analytical defensibility. Likewise, wider adoption of streamlined metadata templates and checklists—such as those proposed for lipidomics[130] could reduce reporting burdens and improve compliance.

Increased editorial oversight is needed to ensure compliance with standards for internal protocols and quality controls. Thorough documentation supports reliable assessment of metabolite data, while guiding researchers toward best practices in interpreting pathways and biological processes connecting metabolite changes to significant outcomes. While online data repositories are valuable, the community must address the '*inconvenience*' of metadata recording.[133] Streamlined tables or checklists, such as those from McDonald *et al.*[138] for lipidomics, can simplify metadata reporting, improve compliance, and encourage the reuse of high-quality environmental datasets. By enhancing methodological consistency, data robustness, and biological interpretability, environmental metabolomics will be better positioned to contribute to regulatory decision making, risk assessment, and the development of mechanistically grounded biomarkers in multi stressor ecological contexts.

Despite these analytical advances, several methodological and data-integration gaps continue to limit broader application and regulatory uptake of metabolomics data.

Knowledge gaps and future directions

Despite significant progress in the development and application of field-based metabolomics, several key knowledge gaps remain that limit broader implementation, standardisation, and integration into environmental monitoring frameworks. Addressing these gaps will be essential for advancing the field toward operational readiness and regulatory uptake. The Metabolomics Society recently endorsed the creation of an Environmental Metabolomics Task Group (EcoMet Task Group; https://metabolomicssociety.org/board-committees/scientific-task-groups/) to tackle and create such tools, which will help harmonise practice and reduce barriers for non-specialist users.

Decision-support tools and workflow guidance

There is a need for clear, accessible decision-support frameworks to guide practitioners through the complexities of field-based metabolomics. A visual flow chart or decision-support tool could assist users in selecting appropriate sampling strategies, analytical platforms, data-processing approaches, and interpretation pathways based on study design and environmental context.

Centralised bioinformatics infrastructure

Currently, bioinformatics capacity in environmental metabolomics is fragmented across institutions and platforms. Developing a roadmap toward centralised data infrastructure —including shared repositories, standardised metadata frameworks, and harmonised processing pipelines—would enhance interoperability and long-term data value. Centralisation would also support reproducibility, facilitate cross-study comparisons, and enable large-scale meta-analyses.

Field metabolomics often produces large, multidimensional datasets that require robust computational solutions. Efficient workflows for handling large datasets in field contexts—including mobile computing platforms, automated QC, and scalable cloud solutions—are still emerging. Research into lightweight, high-throughput data processing pipelines would accelerate field adoption.

Further improvement in bioinformatics tools, particularly those designed to operate efficiently in field contexts, is needed. Challenges include handling low-resolution or partially annotated datasets, limited internet connectivity, and the need for near-real-time feedback. Improvements to platforms such as other machine-learning-based annotation pipelines could enhance automated identification, reduce uncertainty, and streamline interpretation.

Standardisation of methods and protocols

The research domain still lacks widely accepted, standardised protocols for field metabolomics, spanning sample collection, preservation, extraction, instrumental analysis, QA/QC, and reporting. Methodological heterogeneity complicates comparisons across studies and limits regulatory confidence. The last iteration of the MSI was published nearly two decades ago,[112] and is therefore in urgent need of an update to address challenges associated with global uptake and to ensure alignment with current regulatory requirements. Establishing consensus-driven protocols, including minimum reporting standards, is a critical next step for advancing field readiness.

Addressing protocols unique to field conditions

Field environments present unique challenges—including variable temperatures, limited power, contamination risks, and logistical constraints—that differ from traditional laboratory workflows. Protocols specifically designed and validated for field-adapted metabolomics remain underdeveloped. There is a need for systematic evaluation of field-based sampling

kits, portable extraction methods, and on-site stabilisation technologies.

Field environments inherently introduce variability (*e.g.*, temperature shifts, fluctuating exposure mixtures, light variability, *etc.*) that cannot (and should not) be fully controlled. The goal of field-adapted protocols is therefore not to recreate the laboratory-like control outdoors but to ensure that variability is transparently documented, analytically defensible, and interpretable alongside matched environmental chemistry and validation studies. Standardised field workflows (sampling kits, preservation and quenching approaches, portable extraction options, and harmonised metadata capture) should be framed as tools to strengthen reproducibility and inference, while retaining the ecological realism that motivates field metabolomics in the first place.

Metabolomic research benefits from other overlapping omics field sampling approaches, such as transcriptomics-based protocols for rapid tissue preservation that can be readily adopted by metabolomics to improve comparability and integration across datasets. Tools and technological platforms that can capture onsite field metadata *via* smart phones would also be advantageous.

Policy, regulatory, and monitoring applications

Most metabolomics studies remain cross-sectional. There is a substantial knowledge gap in repeated cross-sectional studies that track metabolic responses over time. Temporal datasets are essential for understanding baseline variability, resilience, recovery trajectories, and early-warning indicators of stress. It is important to note that, in field-based environmental metabolomics, most temporal datasets necessarily reflect repeated cross-sectional or population-level designs rather than true longitudinal sampling of individuals, given logistical constraints and the destructive nature of many analytical workflows. Expanded efforts in monitoring changes over time would significantly strengthen ecological interpretation. Integrating metabolomics-based approaches in current and ongoing monitoring programs would facilitate the collection of such data. Future advances may expand the feasibility of individual-level longitudinal metabolomics in select taxa through non-lethal or minimally invasive sampling; however, repeated cross-sectional designs will remain the dominant and most broadly applicable approach for temporal field studies. To achieve broader environmental management impact, metabolomic biomarkers of exposure must be translated into actionable policy and regulatory frameworks. This includes establishing threshold values, determining acceptable ranges of biological variability, and defining metabolomic endpoints compatible with national monitoring programs. Engagement with regulators and end-users is critical for co-developing fit-for-purpose applications.

There remains a limited understanding of dose–response relationships linking metabolomic changes to measured contaminant concentrations. Building stronger mechanistic evidence through controlled exposure studies, combined with sentinel field sampling, will support causal inference and improve relevance to risk assessment.

Importantly, field-based metabolomics should not be positioned as a stand-alone replacement for controlled testing. Regulatory decision-making will be strengthened when metabolomics is integrated with direct environmental sampling (*i.e.*, water, sediment, soil chemistry), exposure reconstruction, and laboratory or mesocosm experiments that validate candidate pathways and establish quantitative dose–response relationships. In practice, field metabolomics is most powerful for identifying ecologically relevant exposure scenarios and early biological signals, while controlled experiments are required to confirm mechanisms, eliminate confounding drivers, and derive defensible thresholds that can be mapped to AOPs and monitoring endpoints.

Consistent with historical regulatory practice, field-derived biological signals gain evidentiary weight only when interpreted alongside the consideration of mechanisms and dose–response relationships established through controlled experimentation, reinforcing the need for paired laboratory, mesocosm, and field-based approaches.

Need for additional case studies

While several demonstrations of field-based metabolomics exist, there is an ongoing need for more diverse, well-documented case studies. These should span multiple ecosystems, taxa, contaminants, and environmental conditions to illustrate robustness, build confidence, and demonstrate broad applicability. Expanded case studies will also help validate standardised methods and bioinformatic workflows.

Conclusions

Field-based metabolomics is rapidly emerging as a transformative approach within effects-based environmental monitoring, providing a window into biochemical responses of organisms directly within their natural ecosystems. This review has explored how advances in technologies, refined analytical methods, and integrated ecological study designs are enabling researchers to capture molecular signatures of stress and exposure. These capabilities position field metabolomics as a critical tool for identifying early-warning indicators, supporting more nuanced ecological assessments, and enhancing our ability to manage environmental change.

At the same time, the discipline faces a suite of challenges that must be addressed before its full regulatory potential can be realised. Variability in field conditions, limited standardisation across sampling and analytical workflows, and the complexity of large metabolomics datasets all present barriers to reproducibility and uptake. Furthermore, the ecological interpretation of metabolomic signals—particularly in relation to contaminant exposure and harmful effects, baseline variability, and dose–response relationships—requires improved bioinformatics capacity and closer integration with established monitoring frameworks. These challenges define clear research opportunities and invite coordinated efforts to develop infrastructure, tools, and protocols that support consistent application in real-world settings.

Addressing these gaps offers significant potential for strengthening regulatory and policy decision-making. Centralised data repositories, validated field-ready protocols, and interoperable bioinformatics pipelines will enable metabolomic endpoints to be assessed alongside traditional biomarkers and ecological indicators. As the evidence base grows, metabolomics could play a vital role in effect-based monitoring programs by providing sensitive measures of sub-lethal stress, supporting cumulative impact assessment, and informing adaptive management strategies. Expanding case studies across diverse species, environments, and contaminant profiles will be key to demonstrating robustness, building confidence, and aligning research outputs with regulatory needs.

Taken together, the opportunities, challenges, and regulatory potential outlined in this review highlight a pivotal moment for the field. Realising the promise of field-based metabolomics will require sustained investment, cross-disciplinary collaboration, and active engagement between researchers, regulators, industry, and monitoring agencies. We call for coordinated national and international efforts to standardise methods, strengthen data infrastructure, and embed metabolomic indicators into environmental monitoring frameworks. By acting now, the research community can accelerate the transition of field-based metabolomics from experimental innovation to an operational, policy-ready tool for safeguarding ecosystem health.

Author contributions

Georgia M. Sinclair: conceptualization, data curation, visualization, writing – original draft, writing – review & editing. Sarah L. Green: conceptualization, data curation, visualization, writing – original draft, writing – review & editing. Ryan Lester: conceptualization, data curation, visualization, writing – original draft, writing – review & editing. Katherine J. Jeppe: conceptualization, data curation, visualization, writing – original draft, writing – review & editing. Steve Melvin: conceptualization, data curation, visualization, writing – original draft, writing – review & editing. Sara M. Long: conceptualization, data curation, visualization, writing – original draft, writing – review & editing. Oliver A.H. Jones: conceptualization, data curation, visualization, writing – original draft, writing – review & editing. David J. Beale: conceptualization, data curation, visualization, writing – original draft, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

Supplementary information (SI): the study-level data extracted and synthesized for this review are provided as a SI Excel workbook. The workbook includes (i) a guidance tab describing the workbook structure, (ii) a thematic summary tab containing summary statistics and category-level syntheses, and (iii) a detail summary tab containing the complete study-by-study extraction table of all variables used in the review. See DOI: https://doi.org/10.1039/d6an00140h.

Acknowledgements

The authors thank the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for supporting this work. We are especially grateful to the two internal CSIRO reviewers who provided thoughtful and constructive feedback on earlier versions of this manuscript. Their insights significantly strengthened the clarity, accuracy, and overall quality of the review.

References

1 J. G. Bundy, M. P. Davey and M. R. Viant, *Metabolomics*, 2009, **5**, 3–21.
2 M. E. Franco, *Ecotoxicology*, 2025, **34**, 1–10.
3 D. J. Beale, O. A. H. Jones, U. Bose, J. A. Broadbent, T. K. Walsh, J. van de Kamp and A. Bissett, *Emerg. Top. Life Sci.*, 2022, **6**, 185–199.
4 V. Soni, N. Bartelo, R. Venkataraman and B. S. Rawat, in *Metabolomics: Recent Advances and Future Applications*, ed. V. Soni and T. E. Hartman, Springer International Publishing, Cham, 2023, pp. 307–362, DOI: 10.1007/978-3-031-39094-4_10.
5 M. S. H. Akash, A. Yaqoob, K. Rehman, M. Imran, M. A. Assiri, F. Al-Rashed, F. Al-Mulla, R. Ahmad and S. Sindhu, *Front. Mol. Biosci.*, 2023, **10**, 1218497.
6 A. O. De Silva, J. M. Armitage, T. A. Bruton, C. Dassuncao, W. Heiger-Bernays, X. C. Hu, A. Kärrman, B. Kelly, C. Ng, A. Robuck, M. Sun, T. F. Webster and E. M. Sunderland, *Environ. Toxicol. Chem.*, 2021, **40**, 631–657.
7 D. K. MacMillan, B. A. Wetmore, S. Dasgupta and W. S. Baldwin, *Toxics*, 2025, **13**, 19.
8 D. J. Beale, G. M. Sinclair, R. Shah, A. M. Paten, A. Kumar, S. M. Long, S. Vardy and O. A. H. Jones, *Sci. Total Environ.*, 2022, **845**, 157255.
9 S. L. Green, D. J. Beale, K. Finlayson, M. Shackleton, E. Silvester, J. Stauber, J. van de Merwe and A. Holland, *Mar. Pollut. Bull.*, 2026, **222**, 118790.
10 K. Ilyas, H. Iqbal, M. S. H. Akash, K. Rehman and A. Hussain, *Environ. Sci. Pollut. Res.*, 2024, **31**, 37963–37987.
11 D. C. Lettoof, R. Wang, B. Clarke, N. Gray, S. Sala, N. Bourne, W. R. Richmond, H. E. Nice, P. Novak, B. von Takach, B. L. Webber, A. H. Kaksonen, T. Walsh and D. J. Beale, *Environ. Pollut.*, 2025, **385**, 127097.
12 L. Zhang, P. García-Pérez, B. Muñoz-Palazon, A. Gonzalez-Martinez, L. Lucini and A. Rodriguez-Sanchez, *Sci. Total Environ.*, 2024, **932**, 172915.

13 G. M. Sinclair, O. A. H. Jones, N. Singh and S. M. Long, *Environ. Toxicol. Pharmacol.*, 2024, **109**, 104494.
14 J. M. Malinowska, T. Palosaari, J. Sund, D. Carpi, R. J. M. Weber, G. R. Lloyd, M. Whelan and M. R. Viant, *Arch. Toxicol.*, 2023, **97**, 721–735.
15 M. Castro-Guarda and R. D. Evans, *Surgery*, 2025, **43**, 6–15.
16 *Applied Environmental Metabolomics Community Insights and Guidance from the Field*, ed. D. J. Beale, K. E. Hillyer, A. C. Warden and O. A. H. Jones, Academic Press, Cambridge, Massachusetts, United States, 2022.
17 W. B. Dunn, A. Erban, R. J. M. Weber and D. J. Creek, *Metabolomics*, 2011, **7**, 4–16.
18 A. H. Emwas, *Metabolites*, 2019, **9**, 123.
19 E. L. Schymanski, H. P. Singer and J. Slobodnik, *Environ. Sci. Technol.*, 2014, **48**, 2097–2098.
20 M. R. Viant, I. J. Kurland, M. R. Jones and W. B. Dunn, *Curr. Opin. Chem. Biol.*, 2017, **36**, 64–69.
21 M. Chen, *Toxics*, 2024, **12**, 647.
22 Y. Li, Y. Tian, X. Zhou, X. Guo, H. Ya, S. Li, X. Yu, C. Yuan and K. Gao, *PLoS One*, 2024, **19**, e0298194.
23 B. Amer, R. R. Deshpande and S. S. Bird, *Metabolites*, 2023, **13**, 648.
24 M. T. McBride, in *A Systems Biology Approach to Advancing Adverse Outcome Pathways for Risk Assessment*, ed. N. Garcia-Reyero and C. A. Murphy, Springer International Publishing, Cham, 2018, pp. 177–198, DOI: 10.1007/978-3-319-66084-4_9.
25 G. M. Sinclair, M. D. Giannantonio, O. A. H. Jones and S. M. Long, *Environ. Monit. Assess.*, 2023, **195**, 344.
26 D. J. Beale, R. M. Shah, A. Marcora, A. Hulthen, A. V. Karpe, K. Pham, G. Wijffels and C. Paull, *Sci. Total Environ.*, 2022, **831**, 154840.
27 R. M. Shah, K. E. Hillyer, S. Stephenson, J. Crosswell, A. V. Karpe, E. A. Palombo, O. A. H. Jones, D. Gorman, L. Bodrossy, J. van de Kamp, A. Bissett, A. S. Whiteley, A. D. L. Steven and D. J. Beale, *Sci. Total Environ.*, 2021, **781**, 146526.
28 K. E. Hillyer, E. Raes, A. Bissett and D. J. Beale, *Environ. Pollut.*, 2023, **318**, 120857.
29 D. J. Beale, J. Crosswell, A. V. Karpe, W. Ahmed, M. Williams, P. D. Morrison, S. Metcalfe, C. Staley, M. J. Sadowsky, E. A. Palombo and A. D. L. Steven, *Sci. Total Environ.*, 2017, **609**, 842–853.
30 D. J. Beale, J. Crosswell, A. V. Karpe, S. S. Metcalfe, P. D. Morrison, C. Staley, W. Ahmed, M. J. Sadowsky, E. A. Palombo and A. D. L. Steven, *Sci. Total Environ.*, 2018, **631–632**, 1328–1341.
31 K. E. Hillyer, U. Bose, J. A. Broadbent, A. Bissett and D. J. Beale, *Environ. Pollut.*, 2023, **319**, 120993.
32 S. Vardy, B. Baddiley, C. Braun, C. Limpus, D. J. Limpus, M. Du Plessis, S. Nilsson, V. Gonzalez-Astudillo and D. Beale, *J. Hazard. Mater.*, 2024, **469**, 133885.
33 M. d. C. Gómez-Regalado, J. Martín, J. L. Santos, I. Aparicio, E. Alonso and A. Zafra-Gómez, *Sci. Total Environ.*, 2023, **861**, 160638.
34 L. Wang, F. F. Sead, M. Khoshmanesh, A. M. Sanati, K. Su and B. Ramavandi, *Mar. Pollut. Bull.*, 2026, **222**, 118882.
35 Q. Yu, L. Cai, H. Liu, Z. Hu, H. Wu, X. Chen, C. Ren and J. Zhang, *J. Water Process Eng.*, 2025, **75**, 108065.
36 D. J. Beale, D. Limpus, G. Sinclair, U. Bose, N. Bourne, S. Stockwell, D. C. Lettoof, R. Shah, T. V. Nguyen, V. Gonzalez-Astudillo, C. Braun, A. Myburgh, B. Baddiley, T. Shimada, C. Limpus and S. Vardy, *Sci. Total Environ.*, 2024, **956**, 176313.
37 D. J. Beale, T. V. Nguyen, U. Bose, R. Shah, J. L. D. Nelis, S. Stockwell, J. A. Broadbent, S. Nilsson, R. Rane, L. Court, D. C. Lettoof, G. Pandey, T. K. Walsh, S. Shaw, J. Llinas, D. Limpus, C. Limpus, C. Braun, B. Baddiley and S. Vardy, *Sci. Total Environ.*, 2024, **926**, 171743.
38 D. C. Lettoof, T. V. Nguyen, W. R. Richmond, H. E. Nice, M. M. Gagnon and D. J. Beale, *Sci. Total Environ.*, 2023, **897**, 165260.
39 D. C. Lettoof, M. Suzuki, S. Nilsson, T. V. Nguyen, N. Bourne, C. L. Pegg, S. Stockwell, U. Bose, J. Devine, T. Contor, B. L. Webber, A. H. Kaksonen, T. Walsh, S. Vardy and D. J. Beale, *Environ. Pollut.*, 2025, **372**, 125968.
40 N. Singh, J. Paz-Ferreiro, M. P. J. Askeland, T. L. Coggan and O. A. H. Jones, *Sci. Total Environ.*, 2025, **974**, 179196.
41 S. D. Melvin, C. M. Lanctôt, N. J. C. Doriean, A. R. Carroll and W. W. Bennett, *Environ. Pollut.*, 2018, **243**, 1096–1105.
42 O. A. H. Jones, in *Applied Environmental Metabolomics*, ed. D. J. Beale, K. E. Hillyer, A. C. Warden and O. A. H. Jones, Academic Press, 2022, pp. 45–59, DOI: 10.1016/B978-0-12-816460-0.00017-4.
43 E. M. Sogin, E. Puskás, N. Dubilier and M. Liebeke, *mSystems*, 2019, **4**, e00638-19.
44 A. Di Trana, G. Sprega, G. Kobidze, O. Taoussi, A. F. Lo Faro, G. Bambagiotti, E. Montanari, M. S. Fede, J. Carlier, A. Tini, F. P. Busardò, A. Di Giorgi and S. Pichini, *Molecules*, 2024, **29**, 3440.
45 T. V. Nguyen, A. Alfaro, E. Frost, D. Chen, D. J. Beale and C. Mundy, *Metabolomics*, 2021, **18**, 7.
46 Evaluation of Tissues Preserved in RNAlater for Metabolomics Analysis, https://www.metabolon.com/resources/evaluation-of-tissues-preserved-in-rnalater-for-metabolomics-analysis/.
47 A. Remili, A. D. Morris, D. C. G. Muir, M. Houde, T. M. Brown, S. H. Ferguson, D. A. D. Blair and R. J. Letcher, *Environ. Res.*, 2025, **269**, 120862.
48 A. E. Simond, M. Houde, V. Lesage, R. Michaud and J. Verreault, *Sci. Total Environ.*, 2020, **717**, 137204.
49 P. T. Gauthier, A. Evenset, G. N. Christensen, E. H. Jorgensen and M. M. Vijayan, *Environ. Sci. Technol.*, 2018, **52**, 868–876.
50 A. Gago-Tinoco, R. Gonzalez-Dominguez, T. Garcia-Barrera, J. Blasco-Moreno, M. J. Bebianno and J. L. Gomez-Ariza, *Environ. Sci. Pollut. Res. Int.*, 2014, **21**, 13315–13323.
51 D. J. Beale, K. Hillyer, S. Nilsson, D. Limpus, U. Bose, J. A. Broadbent and S. Vardy, *Sci. Total Environ.*, 2022, **806**, 151264.

52 D. J. Beale, S. Nilsson, U. Bose, N. Bourne, S. Stockwell, J. A. Broadbent, V. Gonzalez-Astudillo, C. Braun, B. Baddiley, D. Limpus, T. Walsh and S. Vardy, *Sci. Total Environ.*, 2022, **817**, 153019.
53 D. J. Beale, T. V. Nguyen, R. M. Shah, A. Bissett, A. Nahar, M. Smith, V. Gonzalez-Astudillo, C. Braun, B. Baddiley and S. Vardy, *Metabolites*, 2022, **12**, 747.
54 D. J. Beale, A. Bissett, S. Nilsson, U. Bose, J. L. D. Nelis, A. Nahar, M. Smith, V. Gonzalez-Astudillo, C. Braun, B. Baddiley and S. Vardy, *Sci. Total Environ.*, 2022, **838**, 156324.
55 D. Nzabanita, H. Shen, S. Grist, J. O. Hampton, J. Hufschmid and D. Nugegoda, *Environ. Toxicol. Chem.*, 2025, **44**, 92–102.
56 E. Saccenti, in *Metabolomics: Recent Advances and Future Applications*, ed. V. Soni and T. E. Hartman, Springer International Publishing, Cham, 2023, pp. 59–92, DOI: 10.1007/978-3-031-39094-4_3.
57 S. Wei, Y. Wei, Y. Gong, Y. Chen, J. Cui, L. Li, H. Yan, Y. Yu, X. Lin, G. Li and L. Yi, *Metabolomics*, 2022, **18**, 35.
58 E. Legrand, A. L. Bayless, D. W. Bearden, F. Casu, M. Edwards, A. Jacob, W. E. Johnson and T. B. Schock, *Environ. Sci. Technol.*, 2023, **57**, 19169–19179.
59 T. Cappello, M. Maisano, A. Giannetto, V. Parrino, A. Mauceri and S. Fasulo, *Comp. Biochem. Physiol., Part C: Toxicol. Pharmacol.*, 2015, **169**, 7–15.
60 T. Cappello, M. Maisano, A. Mauceri and S. Fasulo, *Ecotoxicol. Environ. Saf.*, 2017, **142**, 417–422.
61 R. Caricato, M. E. Giordano, T. Schettino, M. Maisano, A. Mauceri, A. Giannetto, T. Cappello, V. Parrino, S. Ancora, I. Caliani, N. Bianchi, C. Leonzio, G. Mancini, S. Cappello, S. Fasulo and M. G. Lionetto, *Sci. Total Environ.*, 2019, **690**, 140–150.
62 J. A. Campillo, A. Sevilla, M. Albentosa, C. Bernal, A. B. Lozano, M. Cánovas and V. M. León, *Sci. Total Environ.*, 2015, **524**, 136–147.
63 Y. M. I. Hani, S. M. Prud'Homme, J. M. Nuzillard, I. Bonnard, C. Robert, K. Nott, S. Ronkart, O. Dedourge-Geffard and A. Geffard, *Environ. Pollut.*, 2021, **270**, 116048.
64 E. Legrand, A. L. Bayless, D. W. Bearden, F. Casu, M. Edwards, A. Jacob, W. E. Johnson and T. B. Schock, *Environ. Sci. Technol.*, 2023, **57**, 19169–19179.
65 A. P. Nugroho, M. Rivaldi, D. S. Yudha, T. Ukasha and S. Hadisusanto, *Environ. Sci. Pollut. Res. Int.*, 2025, **32**, 20375–20394.
66 M. Rivaldi, M. Nomura and A. P. Nugroho, *Environ. Toxicol. Pharmacol.*, 2025, **117**, 104741.
67 S. M. Pomfret, R. B. Brua, N. M. Izral and A. G. Yates, *Environ. Rev.*, 2020, **28**, 89–98.
68 R. B. Brua, J. M. Culp, S. M. Pomfret and D. B. Halliwell, in *Applied Environmental Metabolomics*, ed. D. J. Beale, K. E. Hillyer, A. C. Warden and O. A. H. Jones, Academic Press, 2022, pp. 273–289, DOI: 10.1016/B978-0-12-816460-0.00002-2.
69 N. M. Izral, R. B. Brua, J. M. Culp and A. G. Yates, *Environ. Sci. Pollut. Res. Int.*, 2018, **25**, 36184–36193.
70 N. M. Izral, R. B. Brua, J. M. Culp and A. G. Yates, *Sci. Total Environ.*, 2021, **760**, 143322.
71 D. R. E. John, M. Davis, Q. Teng, G. T. Ankley, J. P. Berninger, J. E. Cavallin, K. M. Jensen, M. D. Kahl, A. L. Schroeder, D. L. Villeneuve, Z. G. Jorgenson, K. E. Lee and T. W. Collette, *Environ. Toxicol. Chem.*, 2016, **35**, 2493–2502.
72 D. R. Ekman, K. Keteles, J. Beihoffer, J. E. Cavallin, K. Dahlin, J. M. Davis, A. Jastrow, J. M. Lazorchak, M. A. Mills, M. Murphy, D. Nguyen, A. M. Vajda, D. L. Villeneuve, D. L. Winkelman and T. W. Collette, *Environ. Pollut.*, 2018, **239**, 706–713.
73 M. A. Defo, L. Mercier, C. Beauvais, R. B. Brua, G. Tétreault, A. Fontaine, P. Couture, J. Verreault and M. Houde, *Ecotoxicol. Environ. Saf.*, 2021, **222**, 112483.
74 I. Madge Pimentel, D. Baikova, D. Buchner, A. Burfeid Castellanos, G. M. David, A. Deep, A. Doliwa, U. Hadžiomerović, N. A. S. Mayombo, S. Prati, M. A. Spyra, A.-M. Vermiert, D. Beisser, M. Dunthorn, J. J. Piggott, B. Sures, S. D. Tiegs, F. Leese and A. J. Beermann, *Sci. Total Environ.*, 2024, **926**, 171849.
75 I. Madge Pimentel, P. M. Rehsen, A. J. Beermann, F. Leese, J. J. Piggott and S. Schmuck, *Limnol. Oceanogr.: Methods*, 2024, **22**, 135–148.
76 A. J. Beermann, V. Elbrecht, S. Karnatz, L. Ma, C. D. Matthaei, J. J. Piggott and F. Leese, *Sci. Total Environ.*, 2018, **610–611**, 961–971.
77 N. A. S. Mayombo, A. M. Burfeid-Castellanos, A.-M. Vermiert, I. M. Pimentel, P. M. Rehsen, M. Dani, C. Jasinski, M. A. Spyra, M. Kloster, D. Vidaković, D. Buchner and B. Beszteri, *Sci. Total Environ.*, 2024, **943**, 173670.
78 N. R. Baker, K. Zhalnina, M. Yuan, D. Herman, J. A. Ceja-Navarro, J. Sasse, J. S. Jordan, B. P. Bowen, L. Wu, C. Fossum, A. Chew, Y. Fu, M. Saha, J. Zhou, J. Pett-Ridge, T. R. Northen and M. K. Firestone, *Proc. Natl. Acad. Sci. U. S. A.*, 2024, **121**, e2303439121.
79 Z. X. Cai, Q. S. Li, H. Bai, C. Y. Zhu, G. H. Tang, H. Z. Zhou, J. W. Huang, X. S. Song and J. F. Wang, *Environ. Res.*, 2022, **213**, 113716.
80 K. J. Jeppe, J. Yang, S. M. Long, M. E. Carew, X. Zhang, V. Pettigrove, A. A. Hoffmann and J. Piggott, *J. Appl. Ecol.*, 2016, **54**, 1331–1342.
81 R. Gornati, M. Maisano, C. Pirrone, T. Cappello, F. Rossi, M. Borgese, A. Giannetto, S. Cappello, G. Mancini, G. Bernardini and S. Fasulo, *Mar. Biotechnol.*, 2019, **21**, 773–790.
82 B. A. Boughton, D. L. Callahan, C. Silva, J. Bowne, A. Nahid, T. Rupasinghe, D. L. Tull, M. J. McConville, A. Bacic and U. Roessner, *Anal. Chem.*, 2011, **83**, 7523–7530.
83 P. Gyawali, A. V. Karpe, K. E. Hillyer, T. V. Nguyen, J. Hewitt and D. J. Beale, *Sci. Total Environ.*, 2021, **771**, 145363.
84 T. V. Nguyen, A. C. Alfaro, T. N. Ngo, A. Kumar and D. Beale, *Fish. Res.*, 2025, **287**, 107426.

85 D. W. Brew, M. C. Black, M. Santos, J. Rodgers and W. M. Henderson, *Environ. Toxicol. Chem.*, 2020, **39**, 419–436.
86 H. Chen, X. Diao, H. Wang and H. Zhou, *Ecotoxicol. Environ. Saf.*, 2018, **156**, 330–336.
87 A. R. Van Scoy, B. S. Anderson, B. M. Philips, J. Voorhees, M. McCann, H. D. Haro, M. J. Martin, J. McCall, C. R. Todd, D. Crane, M. L. Sowby and R. S. Tjeerdema, *Ecotoxicol. Environ. Saf.*, 2012, **78**, 99–109.
88 B. Zhang, H. Zhang, C. Du, Q. X. Ng, C. Hu, Y. He and C. N. Ong, *Water Res.*, 2017, **114**, 135–143.
89 P. Wang, Q. X. Ng, H. Zhang, B. Zhang, C. N. Ong and Y. He, *Ecotoxicol. Environ. Saf.*, 2018, **163**, 266–273.
90 S. W. Lee, N. Chatterjee, J. E. Im, D. Yoon, S. Kim and J. Choi, *Ecotoxicol. Environ. Saf.*, 2018, **163**, 111–116.
91 T. Dumas, B. Bonnefille, E. Gomez, J. Boccard, N. A. Castro, H. Fenet and F. Courant, *Sci. Total Environ.*, 2020, **712**, 136551.
92 E. M. Santos, J. S. Ball, T. D. Williams, H. Wu, F. Ortega, R. van Aerle, I. Katsiadaki, F. Falciani, M. R. Viant, J. K. Chipman and C. R. Tyler, *Environ. Sci. Technol.*, 2010, **44**, 820–826.
93 R. Tang, C. Ding, F. Dang, Y. Ma, J. Wang, T. Zhang and X. Wang, *Environ. Pollut.*, 2018, **239**, 428–437.
94 T. V. Nguyen, P. N. Trang and A. Kumar, *Environ. Int.*, 2024, **186**, 108620.
95 D. Stefanoni, H. K. H. Shin, J. H. Baek, D. P. Champagne, T. Nemkov, T. Thomas, R. O. Francis, J. C. Zimring, T. Yoshida, J. A. Reisz, S. L. Spitalnik, P. W. Buehler and A. D'Alessandro, *Haematologica*, 2020, **105**, 2174–2186.
96 S. Hayton, G. L. Maker, I. Mullaney and R. D. Trengove, *Cell. Mol. Life Sci.*, 2017, **74**, 4421–4441.
97 J. C. Kenney, D. White-Kiely, J. P. van de Merwe, C. J. Limpus and K. A. Finlayson, *Mar. Pollut. Bull.*, 2025, **212**, 117589.
98 K. A. Finlayson, F. D. L. Leusch, C. J. Limpus and J. P. van de Merwe, *Ecotoxicol. Environ. Saf.*, 2019, **173**, 63–70.
99 ARI-TOX Marine Wildlife Cell Bank, https://aritox.com/mwcb/, (accessed 29/01, 2026).
100 San Diego Zoo Wildlife Alliance – Frozen Zoo, https://sandiegozoowildlifealliance.org/frozen-zoo, (accessed 29/01, 2026).
101 Species360 Global Cryobankhttps://species360.org/, (accessed 29/01, 2026).
102 Smithsonian Pan-Smithsonian Cryo-Initiative, https://nationalzoo.si.edu/center-for-species-survival/cryo-initiative, (accessed 2026, 29/01).
103 CRYOZOO – Animal Cell Line Biobank, https://www.embl.org/about/info/planetary-biology/projects/cryozoo/, (accessed 29/01, 2026).
104 San Diego Zoo Wildlife Alliance Biodiversity Bank, https://sandiegozoowildlifealliance.org/biodiversity-banking, (accessed 29/01, 2026).
105 G. Noctor, C. Lelarge-Trouverie and A. Mhamdi, *Phytochemistry*, 2015, **112**, 33–53.
106 H. Sies, *Curr. Opin. Toxicol.*, 2018, **7**, 122–126.
107 R. L. Smith, M. R. Soeters, R. C. I. Wüst and R. H. Houtkooper, *Endocr. Rev.*, 2018, **39**, 489–517.
108 N. S. Häfker and K. Tessmar-Raible, *Curr. Opin. Neurobiol.*, 2020, **60**, 55–66.
109 S. M. Rappaport, D. K. Barupal, D. Wishart, P. Vineis and A. Scalbert, *Environ. Health Perspect.*, 2014, **122**, 769–774.
110 J. M. Prokkola and M. Nikinmaa, *J. Exp. Biol.*, 2018, **221**, jeb179267.
111 D. J. Beale and O. A. H. Jones, in *Applied Environmental Metabolomics*, ed. D. J. Beale, K. E. Hillyer, A. C. Warden and O. A. H. Jones, Academic Press, 2022, ch. 25, pp. 379–397.
112 N. Morrison, D. Bearden, J. G. Bundy, T. Collette, F. Currie, M. P. Davey, N. S. Haigh, D. Hancock, O. A. H. Jones, S. Rochfort, S.-A. Sansone, D. Štys, Q. Teng, D. Field and M. R. Viant, *Metabolomics*, 2007, **3**, 203–210.
113 B. I. Escher, H. M. Stapleton and E. L. Schymanski, *Science*, 2020, **367**, 388–392.
114 S. Panda, *Science*, 2016, **354**, 1008–1015.
115 Y. Adamovich, B. Ladeuix, M. Golik, M. P. Koeners and G. Asher, *Cell Metab.*, 2017, **25**, 93–101.
116 R. Ch, O. Chevallier and C. T. Elliott, *Trends Anal. Chem.*, 2020, **130**, 115986.
117 K. A. Dyar, D. Lutter, A. Artati, N. J. Ceglia, Y. Liu, D. Armenta, M. Jastroch, S. Schneider, S. de Mateo, M. Cervantes, S. Abbondante, P. Tognini, R. Orozco-Solis, K. Kinouchi, C. Wang, R. Swerdloff, S. Nadeef, S. Masri, P. Magistretti, V. Orlando, E. Borrelli, N. H. Uhlenhaut, P. Baldi, J. Adamski, M. H. Tschöp, K. Eckel-Mahan and P. Sassone-Corsi, *Cell*, 2018, **174**, 1571–1585.
118 C. Abhishek, G. Nikola and U. Gayh, *IgMin Res.*, 2024, **2**, 121–126.
119 T. Zhang, Y. Xiao, D. Liang, H. Tang, S. Yuan and B. Luan, *Front. Earth Sci.*, 2020, **8**, 305.
120 A. D. Flouris and C. Piantoni, *Temperature*, 2015, **2**, 73–85.
121 T. C. M. Brock, K. C. Elliott, A. Gladbach, C. Moermond, J. Romeis, T.-B. Seiler, K. Solomon and G. Peter Dohmen, *Integr. Environ. Assess. Manag.*, 2021, **17**, 1229–1242.
122 R. J. Wenning, in *Human and Ecological Risk Assessment*, 2024, pp. 833–864. DOI: 10.1002/9781119742975.ch22.
123 L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden and M. R. Viant, *Metabolomics*, 2007, **3**, 211–221.
124 Y. Xu and R. Goodacre, *TrAC, Trends Anal. Chem.*, 2025, **183**, 118064.
125 K. S. Lee, X. Su and T. Huan, *Nat. Metab.*, 2025, **7**, 858–861.
126 M. L. Delignette-Muller, A. Siberchicot, F. Larras and E. Billoir, *Peer Community J.*, 2023, **3**, e90.
127 F. Larras, E. Billoir, V. Baillard, A. Siberchicot, S. Scholz, T. Wubet, M. Tarkka, M. Schmitt-Jansen and M.-L. Delignette-Muller, *Environ. Sci. Technol.*, 2018, **52**, 14461–14468.

128 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, P. C. A. Boya, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nat. Biotechnol.*, 2016, **34**, 828–837.

129 M. Giera, A. Aisporna, W. Uritboonthai and G. Siuzdak, *Nat. Metab.*, 2024, **6**, 1647–1648.

130 S. Alseekh, A. Aharoni, Y. Brotman, K. Contrepois, J. D'Auria, J. Ewald, J. C. Ewald, P. D. Fraser, P. Giavalisco, R. D. Hall, M. Heinemann, H. Link, J. Luo, S. Neumann, J. Nielsen, L. Perez de Souza, K. Saito, U. Sauer, F. C. Schroeder, S. Schuster, G. Siuzdak, A. Skirycz, L. W. Sumner, M. P. Snyder, H. Tang, T. Tohge, Y. Wang, W. Wen, S. Wu, G. Xu, N. Zamboni and A. R. Fernie, *Nat. Methods*, 2021, **18**, 747–756.

131 M. Sud, E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, K. S. Nair, S. Sumner and S. Subramaniam, *Nucleic Acids Res.*, 2016, **44**, D463–D470.

132 O. Yurekten, T. Payne, N. Tejera, F. X. Amaladoss, C. Martin, M. Williams and C. O'Donovan, *Nucleic Acids Res.*, 2023, **52**, D640–D646.

133 R. A. Spicer, R. Salek and C. Steinbeck, *Sci. Data*, 2017, **4**, 170138.

134 K. Haug, R. M. Salek, P. Conesa, J. Hastings, P. de Matos, M. Rijnbeek, T. Mahendraker, M. Williams, S. Neumann, P. Rocca-Serra, E. Maguire, A. Gonzalez-Beltran, S. A. Sansone, J. L. Griffin and C. Steinbeck, *Nucleic Acids Res.*, 2013, **41**, D781–D786.

135 J. Braisted, T. Henderson, J. W. Newman, S. C. Moore, J. Sampson, K. McClain, S. Ross, D. J. Baer, E. A. Mathé and K. A. Zanetti, *Am. J. Epidemiol.*, 2024, kwae347.

136 C. Walton-Doyle, E. Sinclair, H. Begum, K. Hollywood, D. Trivedi and P. Barran, *Sci. Rep.*, 2024, **14**, 21707.

137 S. Wandro, L. Carmody, T. Gallagher, J. J. LiPuma and K. Whiteson, *mSystems*, 2017, **2**, e00100–e00117.

138 J. G. McDonald, C. S. Ejsing, D. Kopczynski, M. Holcapek, J. Aoki, M. Arita, M. Arita, E. S. Baker, J. Bertrand-Michel, J. A. Bowden, B. Brugger, S. R. Ellis, M. Fedorova, W. J. Griffiths, X. Han, J. Hartler, N. Hoffmann, J. P. Koelmel, H. C. Kofeler, T. W. Mitchell, V. B. O'Donnell, D. Saigusa, D. Schwudke, A. Shevchenko, C. Z. Ulmer, M. R. Wenk, M. Witting, D. Wolrab, Y. Xia, R. Ahrends, G. Liebisch and K. Ekroos, *Nat. Metab.*, 2022, **4**, 1086–1088.