Trends in Microbiology

Capturing the microbial volatilome: An oft overlooked “ome”

--Manuscript Draft--

Manuscript Number:	TIMI-D-21-00210R2
Article Type:	Opinion
Keywords:	Volatile Organic Compounds; Volatilome; Metabolome; Microbe; Realtime measurements; Omics
Corresponding Author:	Laura K Meredith, Ph.D. University of Arizona Tucson, Arizona UNITED STATES
First Author:	Laura K Meredith, Ph.D.
Order of Authors:	Laura K Meredith, Ph.D.
	Malak M Tfaily, Ph.D.
Abstract:	Among the diverse metabolites produced by microbial communities, some are volatile. Volatile organic compounds (VOCs) are vigorously cycled by microbes as metabolic substrates and products and as signaling molecules. Yet, current microbial metabolomic studies predominantly focus on nonvolatile metabolites and overlook VOCs, which therefore represent a missing component of the metabolome. Advances in VOC detection now allow simultaneous observation of the numerous VOCs constituting the ‘volatilome’ of microbial systems. We present a roadmap for integrating and advancing VOC and other ‘omics approaches and highlight the potential for realtime VOC measurements to help overcome limitations in discrete ‘omics sampling. Including volatile metabolites in metabolomics, both conceptually and in practice, will build a more comprehensive understanding of microbial processes across ecological communities.

Highlights

H1: Volatile organic compounds (VOCs) should be considered metabolites because microbes use and produce compounds in different states including as gases.

H2: Current approaches in metabolite profiling (metabolomics) do not capture the vast majority of volatile metabolites, leaving our metabolic understanding of microbial systems incomplete.

H3: Recent advances in VOC detection allow fast, high-resolution characterization of entire collections of volatile metabolites in a microbial system (volatilome) and new VOC sampling systems unlock new means for eavesdropping on microbial activity in realtime.

H4: Complementary strengths in the analytics and interpretation of VOCs with other 'omics approaches could be combined to help improve underlying pathways and databases.

H5: Integrating VOC with other multi-omics approaches would add a new temporal dimension and a more unified view of microbial systems.

Title: Capturing the microbial volatilome: An oft overlooked "ome"

Authorship

Laura K. Meredith[1,2], Malak M. Tfaily[2,3]

[1]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA; [2]BIO5 Institute, University of Arizona, Tucson, AZ, USA; [3]Department of Environmental Science, University of Arizona, Tucson, AZ, USA.

*Correspondence: laurameredith@arizona.edu https://twitter.com/DrLauraMeredith (L.K. Meredith) and tfaily@arizona.edu https://twitter.com/Tfaily_Lab (M.M. Tfaily).

Keywords Volatile Organic Compound, Volatilome, Metabolome, Microbe, Realtime measurements, Omics

Abstract

Among the diverse metabolites produced by microbial communities, some are volatile. Volatile organic compounds (VOCs) are vigorously cycled by microbes as metabolic substrates and products and as signaling molecules. Yet, current microbial metabolomic studies predominantly focus on nonvolatile metabolites and overlook VOCs, which therefore represent a missing component of the metabolome. Advances in VOC detection now allow simultaneous observation of the numerous VOCs constituting the 'volatilome' of microbial systems. We present a roadmap for integrating and advancing VOC and other 'omics approaches and highlight the potential for realtime VOC measurements to help overcome limitations in discrete 'omics sampling. Including volatile metabolites in metabolomics, both conceptually and in practice, will build a more comprehensive understanding of microbial processes across ecological communities.

Main text

The rich and untapped world of microbial volatiles

Microbes produce and consume organic and inorganic **volatile compounds** (see **Glossary**) that drive microbial metabolism [1,2], signalling [3,4], and ecosystem-to-global interactions [5–7]. Growing recognition of the importance of **volatile inorganic compounds** for microbial survival (e.g. molecular hydrogen; H_2) [8–10] challenges the view that volatile compounds are simply byproducts of metabolism. Rather, volatile compounds are themselves important metabolites with untapped potential for investigation. Current microbial **metabolomics** studies predominantly focus on the pool of nonvolatile organic metabolites, and volatile organic metabolites in particular remain an unexplored missing component [11] of microbial metabolism. This in turn represents a disconnect in our understanding and conceptualization of microbial systems, interactions, and function. Recent studies stress the importance of bridging these divides to understand **volatile organic compounds** (VOCs) as overall mediators of interactions in microbial systems [3,12,13], yet most focus has been placed on standalone **volatilomics** approaches that profile VOCs in microbial systems to fingerprint or decode signalling rather than as a frontier to metabolism.

Here, we argue that (1) VOCs are an integral yet underrepresented and largely unexplored part of the comprehensive metabolome. Fundamentally, volatile metabolites are a part of the broader metabolome, but most metabolomics studies overlook the volatile component [11], leading to an incomplete picture of intra- and inter- species-to-ecosystem interactions. We show that the situation is hopeful because (2) once embraced as metabolites, VOC measurements offer exciting opportunities to monitor microbial metabolism at temporal scales inaccessible to current

metabolomics techniques. Further, recent developments of new approaches for measuring many volatile components at once provide a valuable opportunity to not only further the understanding of the function of microbiomes but also to track microbial populations in real time through the use of VOCs as tracers. Finally, (3) we outline new opportunities for integration of **'omics** and VOC measurements, and provide a summary of the number of substantial, but solvable, roadblocks through a roadmap that emphasizes lever points for progress.

Volatile metabolites are an integral part of the comprehensive metabolome

Volatile compounds (organic and inorganic) should be treated as metabolites. Microbial metabolism harnesses substrates across different states of matter, including volatile compounds. For example, some sulfur (S)-oxidizing bacteria 'breathe' inorganic hydrogen sulfide (H_2S) gas to aqueous sulfate (SO_4^{2-}) then onto solid elemental sulfur (S_0) [14], while other microbes produce VOCs like dimethyl sulfide (DMS; C_2H_6S) from assimilated S pools [15]. Volatile metabolites are often relegated to belonging to specialized functions with low relevance to overall metabolism, or attention is placed on only a few volatile metabolic by-products (e.g. biogeochemists focus on greenhouse gases that contribute directly to climate warming). Contrary to this narrow view, compelling evidence now suggests widespread importance of volatile inorganic compounds in microbial energy, carbon, and nutrient metabolism (e.g. H_2 supports energy needs [3–6] and carbon fixation [7]) in modern [18] and ancient [16,17] environments. Likewise, VOCs are widespread as substrates and products of microbial **primary** and **secondary metabolism**, such as microbial use of plant-emitted isoprene (C_5H_8) as a carbon source (primary metabolism) [19] and production of sesquiterpenes by fungi infecting human respiratory tracts (secondary metabolism) [20]. Beyond metabolism, VOCs serve as messenger compounds that drive intra- and inter-kingdom

70 communication that can traverse distances that separate organisms by diffusing through the gas
71 phase [3]. In contrast to metabolites in the aqueous and solid phase, gas-phase metabolites
72 efficiently diffuse across microbial cell membranes along concentration gradients, allowing VOCs
73 to move into the cell for consumption or out as products. Here, we focus in particular on the
74 important and understudied role of VOCs as part of the comprehensive metabolism.
75

76 *VOC metabolism is overlooked in microbial metabolism.* Volatile metabolism remains
77 conceptually separate from standard studies on microbial metabolism, leaving the relevance of
78 VOCs to metabolism underappreciated. This divide continues with current approaches in
79 metabolomics (**Box 1**) that do not routinely incorporate techniques to constrain VOCs in the study
80 of metabolic pathways in microbial cultures and populations. For example, we recently found in a
81 soil system using a multi-approach metabolomics dataset, that volatile organic metabolites were
82 disproportionately missing from metabolic pathways due to biases against VOCs during sample
83 collection, processing, and analysis [11]. In general, techniques used in routine metabolomic
84 analysis tend to target few volatile metabolites or have low sensitivity, resulting in poor coverage
85 of the volatile subset of the metabolome. Therefore, most understanding derived across microbial
86 systems has been missing large swaths of the volatile portion of the metabolome, with certain but
87 unclear ramifications on our comprehensive understanding of microbial metabolism and its
88 responses.
89

90 It is time to fill out our picture of microbial metabolism. Research in microbiology needs a
91 complete picture of a comprehensive microbial metabolism that is inclusive of volatile and non-

volatile metabolites (**Fig. 1, Key Figure**). There is an opportunity to complement traditional metabolite profiling methods with emerging VOC-measuring methods to more comprehensively describe entire metabolomes (**Fig. 1a**) and individual metabolic pathways (**Fig. 1b**). For example, the first step in microbial isoprene degradation [19] links a volatile substrate and product detectable by **PTR-TOF-MS** (a specific type of mass spectrometer used for monitoring of VOCs) [22] with downstream aqueous metabolites that can be detected by other metabolomic methods (**Box 1**). Given the prevalence of metabolites with a tendency towards volatility (**Fig. 1c**), a significant fraction of information has been overlooked. Illuminating this current void will improve understanding of microbial activity and interactions, connections to other biomolecular indicators of microbial activity (e.g. gene, transcript, and protein sequences), and closure of carbon and nutrient balances at microbe-to-ecosystem scales.

Tracking microbial activity by eavesdropping on the volatilome

In striking parallel to other metabolomic techniques, the volatile subset of microbial metabolism can now be increasingly well captured with modern VOC detection methods. This opens the door to a new complementary 'omics' tool that can also provide unique opportunities for characterizing microbial systems: **volatilomics**. Volatilomics detects volatile metabolites with **targeted** and **untargeted methods**, as in traditional metabolomics (**Box 1**). Yet, an important contrasting feature is that VOCs can be measured in **realtime** allowing noninvasive eavesdropping on microbial communities (**Box 2**). Volatile compounds partition to the gas phase and diffusively mix to provide a continuous source of representative metabolites that can be 'sniffed' using VOC detectors. This contrasts sharply with other 'omics techniques that use destructive sampling at **discrete** time points followed by sample preparation for analysis (**Box 2**). Multiple cultures, treatments, or locations

115 can be sampled at relatively high throughput with discrete VOC sampling techniques, while VOC
116 detectors can be integrated with sample transfer systems to track microbial volatile metabolism in
117 near realtime from multiple locations (**Box 2**) using multiplexing valves (e.g. [23]). For example,
118 recent work in plants has used volatilomics to track metabolic processes [24] including responses
119 to pathogen [25] and heat [26] stress. Further, in microbial systems these methods have been used
120 to phenotype microbial and fungal traits [27,28] and profile soil microbiomes [5,29–31] by
121 measuring VOC production potential through the release of VOCs into VOC-free air. Finally,
122 VOC composition of human breath is being used to detect pathogens such as invasive fungal
123 infections [20]. Realtime gas sampling systems could also be designed to quantify net VOC
124 exchange to also capture microbial VOC degradation (uptake), which is an important part of
125 microbial metabolism.
126
127 Metabolomics is at the forefront of discoveries in soil systems, capturing the responses of soil
128 primary and secondary metabolism to ecosystem disturbances, and helping guide recent adoption
129 in other systems such as secondary metabolite 'exposome' research in human systems [32]. In soil,
130 volatilomics provides a much-needed window into understanding microbial activity and
131 communication. Soil hosts diverse microbial assemblages that vigorously cycle volatile
132 compounds in ways that have been linked to patterns in microbial community composition and
133 diversity [6,33,34]. The porous soil matrix linked by gas- and solution-filled pores tends to
134 disconnect microbes from each other and resources, especially under dry conditions where water
135 films break and soil aggregation increases [35]. Under these conditions, however, VOCs can
136 efficiently diffuse through soil, representing a mode of communication and interaction between
137 distant biological actors (e.g. roots and microbes interacting in the extended rhizosphere [12,13])

that can be tapped into using VOC analyzers. Our recent advances in soil gas probe development **(Box 2, Fig. Ib**; [36,37]) now allow volatile metabolites to be monitored belowground in realtime and *in situ* with cm-scale resolution. This method extends temporal coverage of offline probing [38,39] and fills a gap in soil VOC measurements that primarily quantify VOCs at the soil-atmosphere interface [6] or through destructive soil extractions [34]. Real-time gas probe sampling can track the responses of microbial metabolism to environmental changes, including soil wetting and redox shifts [36,37], and to observe hotspots, hot moments, and interactions that would otherwise be missed.

Integrating volatilomics with multi-omics to enhance understanding in microbial systems

Integrating volatilomics and other metabolomics data would generate a more comprehensive picture of microbial metabolism. Recent studies from soil systems illustrate complementary paths for converging on the comprehensive metabolome. Honeker et al. (2021) predicted compound volatility along metabolic pathways and showed that soil metabolomics data sets were disproportionately missing volatile metabolites [11], illustrating a gap that can be filled with a volatilomics approach. Conversely, Kim et al. (2021) measured VOC emission potential (PTR-TOF-MS) from simulated snowmelt soil mesocosms and demonstrated the potential to map volatile compound time series onto metabolic pathways [29]. These studies take the first steps in a conceptual shift by the scientific community to consider volatilomics an integral part of the metabolome in all microbial systems. With this new mindset, opportunities open to unify advances across volatilomics and other more established metabolomics approaches (**Box 2**).

Volatilomics has been missing from multi-omics integration efforts. Microbial research has made significant advances in integrating microbial metabolomics with other 'omics' approaches. For example, working in peatland soils, Wilson and Tfaily et al. (2021) used multi-omics (i.e. metagenomics, metaproteomics, metabolomics, but not volatilomics) to show that CO_2 and CH_4 production in response to climate warming were fueled by the availability of reactive carbon substrates released by surface vegetation to soil microbiomes [40]. The approach powerfully showed how microbes drive biogeochemical cycling in soil systems and helped identify new potential organic matter decomposition pathways that link metabolites and enzymes in a framework that can be filled in with the measurable data (e.g. **Fig. 1b**). Using this approach, they found that the genomic and proteomic approaches indicated that a number of volatile metabolites should have been present in the system, but were undetected in the metabolome, emphasizing the potential and need for VOCs to build a comprehensive understanding.

Integrating VOCs with multi-omics faces a number of challenges. The availability, quality, and connectivity of pathway information in most commonly used 'omics databases are missing critical information. For example, whether a metabolite is volatile and detectable by different tools is not directly connected to these databases. While the majority of microbial VOC cycling pathways are unknown, even those that are described are inconsistently represented in databases. For example, isoprene degradation (**Fig. 1b**) is represented in some (MetaCyc [41]) but not all (Kyoto Encyclopedia of Genes and Genomes; KEGG [42] and mVOC 2.0 fingerprint [43]) databases. Furthermore, genes that codify for key enzymes in VOC pathways are not clearly annotated—the isoprene monooxygenase (MetaCyc) is buried within a broader enzyme group (i.e., alkene monooxygenase EC:1.14.13.69) and protein family (i.e. Rieske [2Fe-2S] domain PF00355) that

makes it difficult to search for isoprene degradation in genome databases. While the number of identified isoprene degrading organisms is increasing [44], there remains uncertainty as to whether the traits for VOC cycling are conserved; and isoprene degradation is not represented in the phylogeny-based prediction tools (e.g., PICRUST [45]) that rely on the aforementioned databases. Fortunately, these challenges are not unique to volatile metabolites, and unifying features between volatilomics and other 'omics techniques that can be leveraged to advance both approaches (**Box 2**).

Despite the challenges, we suggest a navigable roadmap to break through current roadblocks to applying volatilomics to more comprehensively understand microbial systems (**Fig. 2**). To increase knowledge, we argue that VOC-'omics interactions (**Box 2**) could be more strongly leveraged to resolve unknowns in volatile metabolites and pathways on one hand, and iteratively improve data analytics and interpretation tools on the other. We believe that addressing these roadblocks will help unlock the unique strengths of volatilomics on a common path of multi-omics science.

Concluding Remarks and Future Perspectives

Volatile compounds are critical to microbial metabolism, but have been overlooked as part of the comprehensive metabolome, despite unique opportunities for volatilomics measurements to inform our understanding of microbiomes. We argue that the first step in this emerging frontier is transformation of our conceptual view of metabolomics into one that embraces volatile compounds—organic and inorganic— and in particular VOCs from experimental design through interpretation.

206 There is much work to be done to enable volatilomics approaches. Fortunately, rapid advances in
207 metabolomics and multi-omics approaches can help lead the way, and future developments should
208 focus on frameworks and pipelines to link VOC data with other 'omics' information and pathways,
209 for example by incorporating volatility predictions [11] into automated metabolomic pipelines.
210 Hypothesis-driven experiments should be designed to identify the pathways responsible for
211 production and consumption of specific VOCs in pure cultures and various ecosystems to fill
212 persistent gaps in databases. Volatilomics approaches used to study plant metabolism [24,25] can
213 be translated to microbial systems, for example by evaluating metabolic responses to bioactive
214 small molecules [25] and stable isotope tracers [23]. In particular, pulse labeling with position-
215 specific ^{13}C-pyruvate moieties and realtime monitoring of ^{13}C-CO_2 and ^{13}C-VOCs has been used
216 to elegantly probe branches of central plant metabolism [46] and could be especially helpful for
217 tracking complex metabolic interactions in microbiomes involving cross-feeding through the
218 metabolic exchange of compounds.
219
220 We should make use of enhanced temporal coverage of volatilomics, especially in systems such
221 as soils where other 'omics approaches are destructive and discrete, and in situ soil gas sensing is
222 increasingly feasible [36,37], and in human health where developing non-invasive sampling
223 approaches is paramount [20]. High frequency time series can be mined to infer metabolic
224 connections from VOC data alone, as in the innovative information theory approach used by Kim
225 et al. (2021) to infer time-dependent metabolic connections between volatile compounds by
226 considering the mutual information shared between compounds over different timelags [29].
227 Moreover, once uncovered, metabolic connections and volatile biomarkers of specific organisms
228 could be used to generate predictive frameworks for microbial processes. While challenging,

studies should make full use of the mass-resolving power of modern VOC analyzers and account for fragments and other interferences. Continued advancements in VOC analysis software, technology, and methods will be valuable in this regard, particularly efforts to combine approaches to disambiguate the identity of observed volatile masses (e.g., by GC-PTR-TOF-MS). Given the high cost and specialized training needed to operate most current VOC analyzers, there is a need to develop inexpensive, user-friendly, and scalable VOC sensing platforms that target one or more VOC analytes or functional classes. Finally, as in metabolomics [47], we will need to address the influence of abiotic transformations and environmental conditions on volatile compounds, as these factors will also influence the cycling of VOCs within and across ecosystems.

The future for volatilomics is bright, and volatile. Many questions remain unanswered (see **Outstanding Questions Box**), but already studies are implementing a new lens into microbial activity in the environment. Embracing and enabling volatilomics will open new opportunities for VOCs as harbingers of shifting organismal interactions and microbiome health and function in systems relevant to agroecosystems, health care, climate, and beyond.

Text Boxes

Box 1. Existing and emerging methods in metabolomics and volatilomics

Metabolomics: Metabolomics refers to the large-scale study of metabolites in different ecosystems that are produced and transformed during metabolism [48] by either targeting a specific subset of metabolites in a sample (targeted metabolomics) [49] and/or examining the complete metabolome with untargeted approaches (untargeted metabolomics). Untargeted approaches often drive

252 hypothesis generation, followed by targeted profiling for more confident quantitation of
253 metabolites [50]. Multiple analytical platforms may be used to cover the full spectrum of
254 metabolites including **mass spectrometry** (MS), nuclear magnetic resonance (NMR)
255 spectroscopy, and Fourier transform infrared spectroscopy (FT-IR) [48]. The development of high-
256 resolution accurate mass MS workflows with high throughput and quantitative capabilities has
257 expanded our knowledge of metabolite presence and composition [51]. Electrospray ionization
258 (ESI) is frequently used for chromatography interfaced- and direct injection- (DI) MS for many
259 environmental fields. ESI uses soft ionization that transfers solubilized ions from the liquid to the
260 gaseous phase with nominal compound fragmentation to analyze the intact molecular ion.
261 Meanwhile, versatile mass analyzers are coupled to ESI to aid metabolite identification by
262 acquiring highly resolved and accurate MS/MS spectra achieved mainly through collision-induced
263 dissociation [52]. Commonly used MS include time-of-flight (TOF), orbitraps, quadrupoles and
264 ion traps, each with specific characteristics [53].
265
266 ***Volatilomics:*** VOCs are poorly measured by most metabolomics methods (see Table 1 in [11]).
267 High sensitivity VOC analyzers are increasingly applied to monitor biological systems. DI proton
268 transfer reaction mass spectrometers (PTR-MS) ionize chemical compounds with a proton affinity
269 higher than water and record tens to hundreds of mass-to-charge ratios (m/z) in realtime [54,55].
270 PTR-MS quadrupole and ion trap mass analyzers required preselecting masses, but PTR-TOF-MS
271 [56] acquires entire m/z spectra within microseconds with sufficient mass resolving power to
272 separate compounds with the same nominal m/z integer mass (e.g. isoprene and furan), but not the
273 same chemical formula (e.g. monoterpene isomers α-pinene and myrcene). **Gas chromatography**
274 (GC) MS methods resolve isomers, and even molecular enantiomers [57], in reference to standards

and retention time databases, when they exist. GC-MS can be used to quantify VOCs collected at discrete time points using controlled gas sample flow through adsorptive media in cartridges [58] or by passive adsorption for semi-quantitative insights [59]. Coupling GC and PTR-TOF-MS in an experimental design [5], and even the same instrument (GC-PTR-TOF-MS), integrates benefits of both high temporal resolution and structural identification to track volatile metabolites.

Box 2. VOC-'omics interactions: unifying and contrasting features

VOC measurements can be added to experiments as a compatible and distinct resource. Already, microbial studies use gases to track activity (e.g. CO_2 for respiration) and subsample for 'omics analyses (**Fig. Ia**).

A contrasting feature of volatilomics is that it can be measured at temporal scales inaccessible to discrete 'snapshot' 'omics techniques (**Fig. Ib**). Realtime detection of volatile metabolites can help keep a finger on the pulse of metabolic pathways and bridge the gaps in other approaches.

Unifying features in the analysis and interpretation of VOC and other 'omics data provide leverage points to accelerate their combined progress (**Fig. Ic**). A key challenge in untargeted 'omics is the need to annotate and identify observed biological molecules. **Metabolite annotation** and **structural identification** is a lengthy process involving multiple tools and databases that has been greatly enabled with a decade of rapid expansion in new software. Similarly, VOC analysis tools can now automate analysis of entire mass spectra including peak detection, integration, and calibration to efficiently process data collected over hours to years [60]. However, volatilomics approaches lack standardized pipelines for metabolite annotation, representing an opportunity to adapt metabolomics pipelines to include VOCs.

297 A unifying challenge to the interpretation of volatilomics and other 'omics approaches are gaps in
298 databases and metabolic models used to link genes, proteins, and metabolites. Databases lag
299 behind the most current understanding of biological pathways and molecular processes, and many
300 pathways are still incomplete [61]. While many metabolites map to overall metabolic modules
301 (e.g. in KEGG) a high percentage do not. Further, the majority of metabolites are yet to be
302 discovered, structurally elucidated, and functionally characterized [62]. Even with the recent
303 development in technology and computational approaches, metabolome and volatilome coverage
304 is considerably lower than for other 'omics, in part because chemical extraction methods fail to
305 capture all metabolites in a sample, including volatiles [11]. VOCs add new information that will
306 help fill database gaps and improve tools. VOC and multi-omics data can be integrated using
307 existing approaches to: 1) elucidate the underlying regulatory networks and identify genes that
308 influence the metabolome through direct and indirect involvement; and 2) confirm whether
309 pathways are up- or down-regulated by superimposing on the predicted pathway. Finally, VOC
310 data and other realtime information (e.g. environmental sensors) can be used to drive predictive
311 models of microbial activity to select sampling time points or inform management approaches.
312
313 **Glossary**
314 - **Discrete sampling:** Sampling that occurs at one particular point in time used to generate a
315 discontinuous 'snapshot' of processes and biological systems.
316 - **Gas chromatography:** Method for separating molecules travelling in a column based on
317 differential interactions with the matrix before detection.
318 - **Mass spectrometry:** Analytical tool for measuring the mass of molecules.

- **Metabolite annotation:** Assignment of tentative metabolite candidates based on matching their masses and fragmentation pattern with database or library entries.
- **Metabolite structural identification:** Use of chromatographic and spectral data of an organic compound to identify the structure of a compound.
- **Metabolomics:** The large-scale profile study of low molecular weight organic compounds that are produced and transformed during the metabolism.
- **'Omics:** Approaches that characterize of pools of biological molecules that reflect the collective set of genes (metagenomics), transcripts (metatranscriptomics), proteins (proteomics), and metabolites (metabolomics), including volatile metabolites (volatilomics), that relate to structure, function, and dynamics of microbial systems.
- **Primary metabolites:** Metabolites essential for the survival of an organism as they are essential for normal growth, development, and reproduction.
- **PTR-TOF-MS:** Proton Transfer Reaction Time-of-Flight Mass Spectrometry is a soft ionization technique that generates intact VOC ions for accurate mass analysis based on their travel time to the detector used for realtime quantification of VOCs in air.
- **Realtime sampling:** Sampling that generates continuous undelayed data on processes and biological systems by directly transferring samples to online analyzers.
- **Secondary metabolites:** Metabolites part of a specialized metabolism that are not essential to survival; the majority of which remain unknown.
- **Targeted methods:** Identification and quantification of a selective set of biological molecules in microbial systems.
- **Untargeted methods:** Identification and quantification of the full set of biological molecules in microbial systems.

- **Volatile compounds:** Low molecular weight, high vapor pressure compounds that partition to the gas phase, some of which are directly involved in metabolic pathways as metabolites.
- **Volatile inorganic compounds:** Volatile compounds that are inorganic (lack of a C-H bond) such as H_2, CO_2, and H_2S.
- **Volatile organic compounds:** Volatile compounds that are organic and are often biologically produced and consumed as metabolites, which are often referred to as BVOCs and mVOCS in more system-specific applications, respectively.
- **Relative Volatility Index:** A relative, unitless measure of the tendency of a molecule to be volatile estimated using empirical models based on the contributions of molecular functional groups [11].
- **Volatilomics:** An 'omics approach that characterizes the full set of volatile compounds in an ecosystem.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. 2034192 and 2045332 to LM and by the Department of Energy's Office of Biological and Environmental Research Grant, DE-SC0021349 to MT. We thank Viviana Freire Zapata for contributions to the figures. We appreciate fruitful discussions with Jordan Krechmer.

References

1 Tyc, O. *et al.* (2017) The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends Microbiol.* 25, 280–292

2 McGenity, T.J. *et al.* (2018) Microbial cycling of isoprene, the most abundantly produced biological volatile organic compound on Earth. *ISME J.* 12, 931–941

3 Weisskopf, L. *et al.* (2021) Microbial volatile organic compounds in intra-kingdom and inter-kingdom interactions. *Nat. Rev. Microbiol.19, 391–404*

4 Schmidt, R. *et al.* (2015) Microbial Small Talk: Volatiles in Fungal-Bacterial Interactions. *Front. Microbiol.* 6, 1495

5 Bourtsoukidis, E. *et al.* (2018) Strong sesquiterpene emissions from Amazonian soils. *Nat. Commun.* 9, 2226

6 Trowbridge, A.M. *et al.* Soil Biogenic Volatile Organic Compound Flux in a Mixed Hardwood Forest: Net Uptake at Warmer Temperatures and the Importance of Mycorrhizal Associations. , *Journal of Geophysical Research: Biogeosciences*, 125. (2020)

7 Rinnan, R. and Albers, C.N. (2020) Soil Uptake of Volatile Organic Compounds: Ubiquitous and Underestimated? *J. Geophys. Res. Biogeosci.* 125, 1034

8 Greening, C. *et al.* (2021) Hydrogen is a major lifeline for aerobic bacteria. *Trends Microbiol. xx, 1–8*

9 Bay, S.K. *et al.* (2021) Chemosynthetic and photosynthetic bacteria contribute differentially to primary production across a steep desert aridity gradient. *ISME J.15, 3339–3356*

10 Piché-Choquette, S. and Constant, P. (2019) Molecular Hydrogen, a Neglected Key Driver of Soil Biogeochemical Processes. *Appl. Environ. Microbiol.* 85,

11 Honeker, L.K. *et al.* (2021) The Volatilome: A Vital Piece of the Complete Soil Metabolome. *Front. Environ. Sci.* 9, 113

12 Raza, W. *et al.* (2021) Extended Plant Metarhizobiome: Understanding Volatile Organic Compound Signaling in Plant-Microbe Metapopulation Networks. *mSystems*

13 de la Porte, A. *et al.* (2020) A Gaseous Milieu: Extending the Boundaries of the Rhizosphere. *Trends Microbiol.* 28, 536–542

14 Shively, J.M. (2006) *Inclusions in Prokaryotes*, Springer Berlin Heidelberg.

15 Carrión, O. *et al.* (2017) Methanethiol-dependent dimethylsulfide production in soil environments. *ISME J. 11, 2379–2390*

16 Vignais, P.M. and Billoud, B. (2007) Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* 107, 4206–4272

17 Garcia, A.K. *et al.* (2020) Reconstructing the evolutionary history of nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. *Geobiology* 18, 394–411

18 Bay, S.K. *et al.* (2021) Trace gas oxidizers are widespread and active members of soil microbial communities. *Nat Microbiol* 6, 246–256

19 Crombie, A.T. *et al.* (2019) Genetics and Ecology of Isoprene Degradation. In *Aerobic Utilization of Hydrocarbons, Oils, and Lipids* (Rojo, F., ed), pp. 557–571, Springer International Publishing

20 Acharige, M.J.T. *et al.* (2018) Breath-based diagnosis of fungal infections. *J. Breath Res.* 12, 027108

21 Smith, K.S. and Ferry, J.G. (2000) Prokaryotic carbonic anhydrases. *FEMS Microbiol. Rev.* 24, 335–366

22 Pagonis, D. *et al.* (2019) A Library of Proton-Transfer Reactions of H3O+ Ions Used for Trace Gas Detection. *J. Am. Soc. Mass Spectrom. 30, 7, 1330–1335*

23 Weitz, K.K. *et al.* (2021) Real-Time Mass Spectrometry Measurements of Respiration Rates in Biological Systems. *J. Am. Soc. Mass Spectrom.* 32, 648–652

24 Majchrzak, T. *et al.* (2020) Real-Time Volatilomics: A Novel Approach for Analyzing

411 Biological Samples. *Trends Plant Sci.* 25, 302–312
412 25 Jud, W. *et al.* (2018) Volatilomics: a non-invasive technique for screening plant phenotypic
413 traits. *Plant Methods* 14, 109
414 26 Yáñez-Serrano, A.M. *et al.* (2019) Heat stress increases the use of cytosolic pyruvate for
415 isoprene biosynthesis. *J. Exp. Bot.* 70, 5827–5838
416 27 Misztal, P.K. *et al.* (2018) Emission Factors of Microbial Volatile Organic Compounds
417 from Environmental Bacteria and Fungi. *Environ. Sci. Technol.* 52, 8272–8282
418 28 Guo, Y. *et al.* Sniffing fungi – phenotyping of volatile chemical diversity in Trichoderma
419 species. , *New Phytologist*, 227. (2020) , 244–259
420 29 Kim, J. *et al.* 12-Mar-(2021) , Measurement of Volatile Compounds for Real-time Analysis
421 of Soil Microbial Metabolic Response to Simulated Snowmelt. *Front. Microbiol.*12, 1673
422 30 Rossabi, S. *et al.* (2018) Volatile Organic Compound Emissions From Soil Following
423 Wetting Events. *J. Geophys. Res. Biogeosci.* 123, 1988–2001
424 31 Mäki, M. *et al.* (2019) Interannual and Seasonal Dynamics of Volatile Organic Compound
425 Fluxes From the Boreal Forest Floor. *Front. Plant Sci.* 10, 191
426 32 Walker, D.I. *et al.* (2019) The metabolome: A key measure for exposome research in
427 epidemiology. *Curr Epidemiol Rep* 6, 93–103
428 33 Abis, L. *et al.* (2020) Reduced microbial diversity induces larger volatile organic compound
429 emissions from soils. *Sci. Rep.* 10, 6104
430 34 Schenkel, D. *et al.* (2019) Linking soil's volatilome to microbes and plant roots highlights
431 the importance of microbes as emitters of belowground volatile signals. *Environ.*
432 *Microbiol.21(9), 3133–3327*
433 35 Schimel, J.P. (2018) Life in Dry Soils: Effects of Drought on Soil Microbial Communities

and Processes. *Annu. Rev. Ecol. Evol. Syst.* 49, 409–432

36 Gil-Loaiza, J. *et al.* (2021) , Versatile soil gas concentration and isotope monitoring: optimization and integration of novel soil gas probes with online trace gas detection. Accepted in *Biogeosci.*

37 Roscioli, J.R. *et al.* (2021) Soil gas probes for monitoring trace gas messengers of microbial activity. *Sci. Rep.* 11, 8327

38 Wester- Larsen, L. *et al.* (2020) Biogenic Volatile Organic Compounds in Arctic Soil: A Field Study of Concentrations and Variability With Vegetation Cover. *J. Geophys. Res. Biogeosci.* 125, 36

39 Mäki, M. *et al.* (2019) Boreal forest soil is a significant and diverse source of volatile organic compounds. *Plant Soil* 441, 89–110

40 Wilson, R.M. *et al.* (2021) Soil metabolome response to whole-ecosystem warming at the Spruce and Peatland Responses under Changing Environments experiment. *Proc. Natl. Acad. Sci. U. S. A.* 118,

41 Caspi, R. *et al.* (2018) The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* 46, D633–D639

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

43 Lemfack, M.C. *et al.* (2018) mVOC 2.0: a database of microbial volatiles. *Nucleic Acids Res.* 46, D1261–D1265

44 Larke-Mejía, N.L. *et al.* (2019) Novel Isoprene-Degrading Proteobacteria From Soil and Leaves Identified by Cultivation and Metagenomics Analysis of Stable Isotope Probing Experiments. *Front. Microbiol.* 10, 2700

45 Langille, M.G.I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. , *Nature Biotechnology*, 31. (2013) , 814–821

46 Fasbender, L. *et al.* (2018) Real-time carbon allocation into biogenic volatile organic compounds (BVOCs) and respiratory carbon dioxide (CO_2) traced by PTR-TOF-MS, $^{13}CO_2$ laser spectroscopy and ^{13}C-pyruvate labelling. *PLoS One* 13, e0204398

47 AminiTabrizi, R. *et al.* (2022) Elevated temperatures drive abiotic and biotic degradation of organic matter in a peat bog under oxic conditions. *Sci. Total Environ.* 804, 150045

48 Patti, G.J. *et al.* (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13, 263–269

49 Dudley, E. *et al.* Targeted metabolomics and mass spectrometry. , *Advances in Protein Chemistry and Structural Biology*. (2010) , 45–83

50 Fudyma, J.D. *et al.* Untargeted metabolomic profiling of Sphagnum fallax reveals novel antimicrobial metabolites. , *Plant Direct*, 3. (2019)

51 Schrimpe-Rutledge, A.C. *et al.* (2016) Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.* 27, 1897–1905

52 Banerjee, S. and Mazumdar, S. (2012) Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* 2012, 282574

53 Zubarev, R.A. and Makarov, A. Orbitrap Mass Spectrometry. , *Analytical Chemistry*, 85. (2013) , 5288–5296

54 Hansel, A. *et al.* Proton transfer reaction mass spectrometry: on-line trace gas analysis at the ppb level. , *International Journal of Mass Spectrometry and Ion Processes*, 149–150. (1995) , 609–619

480 55 Lindinger, W. *et al.* On-line monitoring of volatile organic compounds at pptv levels by
481 means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food
482 control and environmental research. , *International Journal of Mass Spectrometry and Ion*
483 *Processes*, 173. (1998) , 191–241
484 56 Yuan, B. *et al.* A high-resolution time-of-flight chemical ionization mass spectrometer
485 utilizing hydronium ions (H_3O ToF-CIMS) for measurements of volatile organic
486 compounds in the atmosphere. , *Atmospheric Measurement Techniques Discussions*. (2016)
487 , 1–43
488 57 Williams, J. *et al.* 13-Aug-(2021) , Chiral monoterpenes reveal forest emission mechanisms
489 and drought responses. , *Research Square*, Research Square
490 58 Teuber, M. *et al.* (2008) VOC emissions of Grey poplar leaves as affected by salt stress and
491 different N sources. *Plant Biol.* 10, 86–96
492 59 Kallenbach, M. *et al.* A robust, simple, high-throughput technique for time-resolved plant
493 volatile analysis in field experiments. , *The Plant Journal*, 78. (2014) , 1060–1072
494 60 Holzinger, R. (2015) PTRwid: A new widget tool for processing PTR-TOF-MS data.
495 *Atmos. Meas. Tech.* 8, 3903–3922
496 61 Pinu, F.R. *et al.* Systems Biology and Multi-Omics Integration: Viewpoints from the
497 Metabolomics Research Community. , *Metabolites*, 9. (2019) , 76
498 62 Tsugawa, H. *et al.* Metabolomics and complementary techniques to investigate the plant
499 phytochemical cosmos. , *Natural Product Reports*. (2021)

500 **Figure Legends**

501

Figure 1. Volatile metabolites: an important, but overlooked component of the metabolome. (a) Many organic compounds are present in nature (large circle; volatile in orange, non-volatile in yellow), some of which are directly involved in pathways as metabolites and therefore intersect along metabolic pathways with associated genes (DNA; metagenomics), transcripts (RNA; metatranscriptomics), and proteins (PRO; proteomics) that are now routinely characterized using 'omics approaches. These metabolites comprise a comprehensive metabolome (inside black line) that not only includes predominantly non-volatile metabolites characterized by traditional metabolomic techniques but also the volatilome, which is a missing component in the current metabolomics concept. An example metabolic pathway, (b) isoprene degradation (adapted from [41]) connects metabolites of varying volatility, predicted here using the **relative volatility index** (RVI; colored dots), a unitless volatility scale described in [11]. Volatile metabolites (e.g. isoprene and degradation product (3R)-3,4-epoxy-3-methylbut-1-ene) are missed by common metabolomics methods that capture large metabolites (* indicates mass > 200 Da), but may be measured by VOC analyzers (**Box 1**) that detect their protonated ions (grey text) to fill in currently incomplete pathways. (c) Volatile compounds are prevalent in metabolic pathways as shown by the distribution of estimated metabolite volatility (RVI) across multiple metabolic pathways (8 metabolic maps; 556 metabolites [11]). Depending on the location (atmosphere in blue; soil in brown), different proportions of compounds may partition to the gas phase where they may be measured as VOCs. While thresholds for volatility in the soil system are not well understood, 86 (15%) and 31 (6%) metabolites in these pathways exceeded estimated semi-volatile (dashed) and volatile (solid) thresholds in soil, respectively (brown), and 262 (47%) metabolites could be volatile in the atmosphere (blue).

Figure 2. Roadmap to break through current roadblocks facing volatilomics. Volatilomics is a new approach for addressing research questions and guiding application-based research. Experiments can be observational (surveys, gradients, environmental responses) or experimental (substrate addition, isotope probing) and use a suite of VOC-resolving methods and data collection strategies, including realtime VOC approaches that pave the way for unprecedented resolution of the volatilome. Critical steps in analytics and data processing involve identifying observed volatile compounds through annotation, and interpreting their metabolic roles by referencing databases and pathways. Here, we encounter roadblocks to progress where tools and databases are lacking, reducing our current ability to describe the complete metabolome. Yet, there is an opportunity to leverage VOC-'omics interactions and integrate multiple data types to improve tools and resolve unknowns while simultaneously building new knowledge that addresses motivating questions. In parallel, analytics can be developed to build predictive models around realtime VOC data streams to guide sampling for other 'omics approaches and provide more extensive data coverage of microbial systems. Guided by this roadmap and its iterative approach, we envision a path to building knowledge and bolstering tools in volatilomics and integrative multi-omics.

Figure I (Box 2). Unifying and contrasting features in volatilomics and multi-omics to leverage. On the journey from experimental design to data integration, there are features of VOC-based volatilomics approaches that unify or contrast with other metabolomics and sequence-based omics approaches, which serve as potential acceleration points for progress. (a) Already, volatile compound data is collected from cultures, mesocosms, chambers, and cores alongside discrete -omics sampling, representing existing experimental frameworks that incorporate multiple data streams. (b) Contrasting features of volatile analysis can bolster multi-omics approaches. Namely,

548 the ability to measure realtime VOC profiles, including via innovative in situ soil gas probing
549 [36,37], generates mass spectra at high temporal resolution, providing signals of specific masses
550 concurrent with the pace of the experiment. VOC data can thus inform timing for 'omics methods
551 that rely on sampling specific 'snapshots'. (c) Once generated, unifying requirements for analytics
552 and interpretation align VOC and 'omics approaches. In particular volatile and nonvolatile
553 metabolites analysis use similar annotation and interpretation approaches, and these pipelines and
554 tools could be synergized and help bolster a common toolbox of databases and pathways.
555 Moreover, existing and emerging approaches for integrating multi-omics data including
556 correlation networks and metabolic pathways can also integrate VOC data. Finally, realtime
557 signals from VOC analyzers and other environmental sensors may be used to develop predictive
558 models for management and ecological forecasting.

Outstanding Questions Box

Q1: What is the evolutionary history of microbial VOC metabolism and how susceptible are different VOC metabolisms to impending environmental change?

Q2: What are the dynamics of microbial VOC cycling in different ecosystems? Are some VOCs production and consumption processes constitutive and ubiquitous, while others are only observed infrequently in response to dynamic changes in environmental conditions (e.g. moisture availability) or biological stimuli (e.g. plant interactions)?

Q3: Does incorporating VOCs change our interpretation of microbial systems or provide a missing component to close elemental budgets?

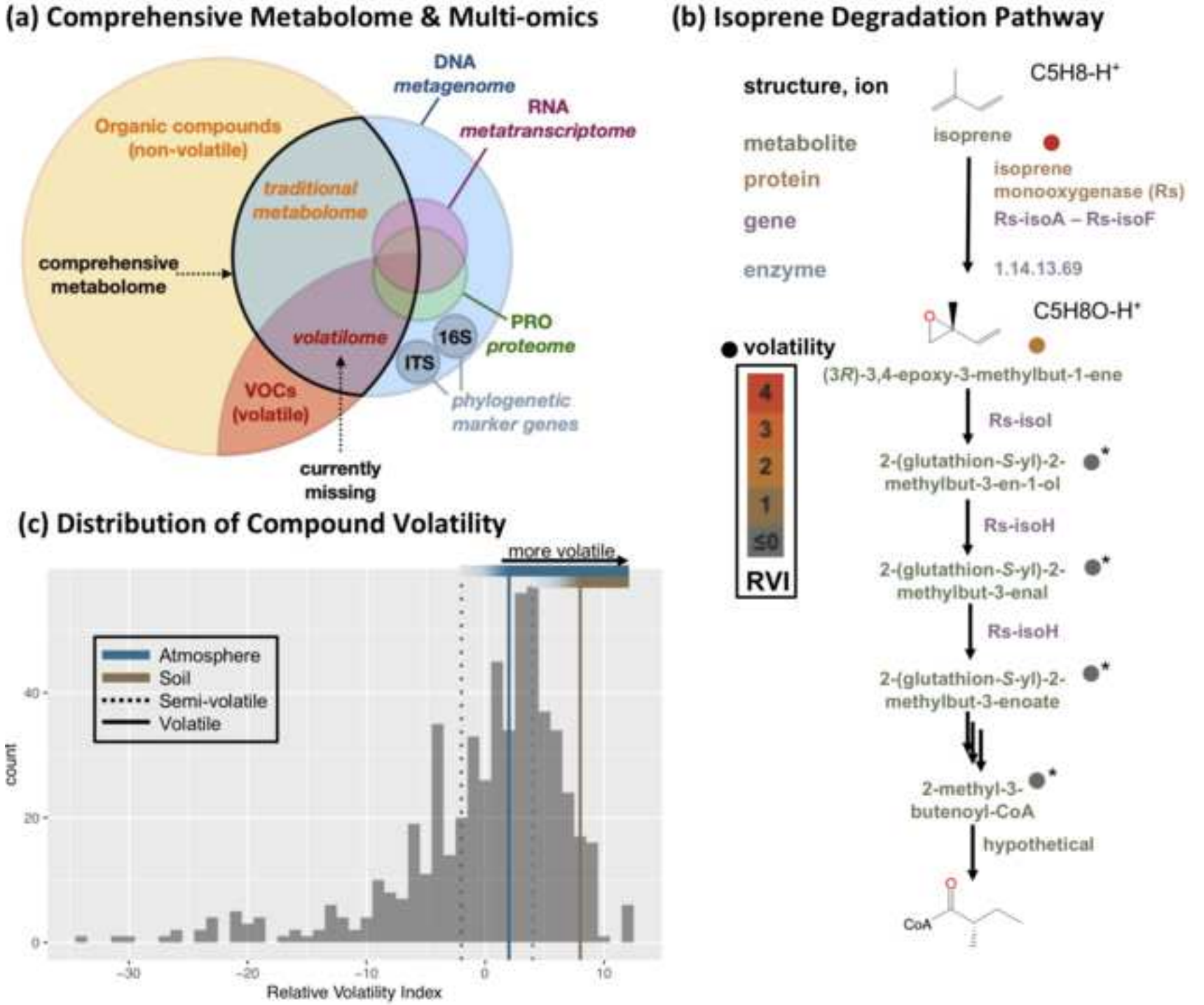
Q4: Can VOCs be used as conserved traits to infer microbial processes that are directly involved in metabolism and/or as markers of interactions such as signaling and communication?

Q5: How can we enhance the spatial resolution of VOC detection to smaller spatial scales consistent with the processes occurring in aggregates, pores, and microbes, while maintaining high temporal fidelity?

Q6: Is the nature of the volatilomics response to drastic environmental disturbance consistent with the major transformations observed in the pools of other biological molecules? For example, are rapid shifts in the metabolome, transcriptome, and proteome in soils rewetted after dry periods (e.g. Birch effect) paralleled in the volatilome?

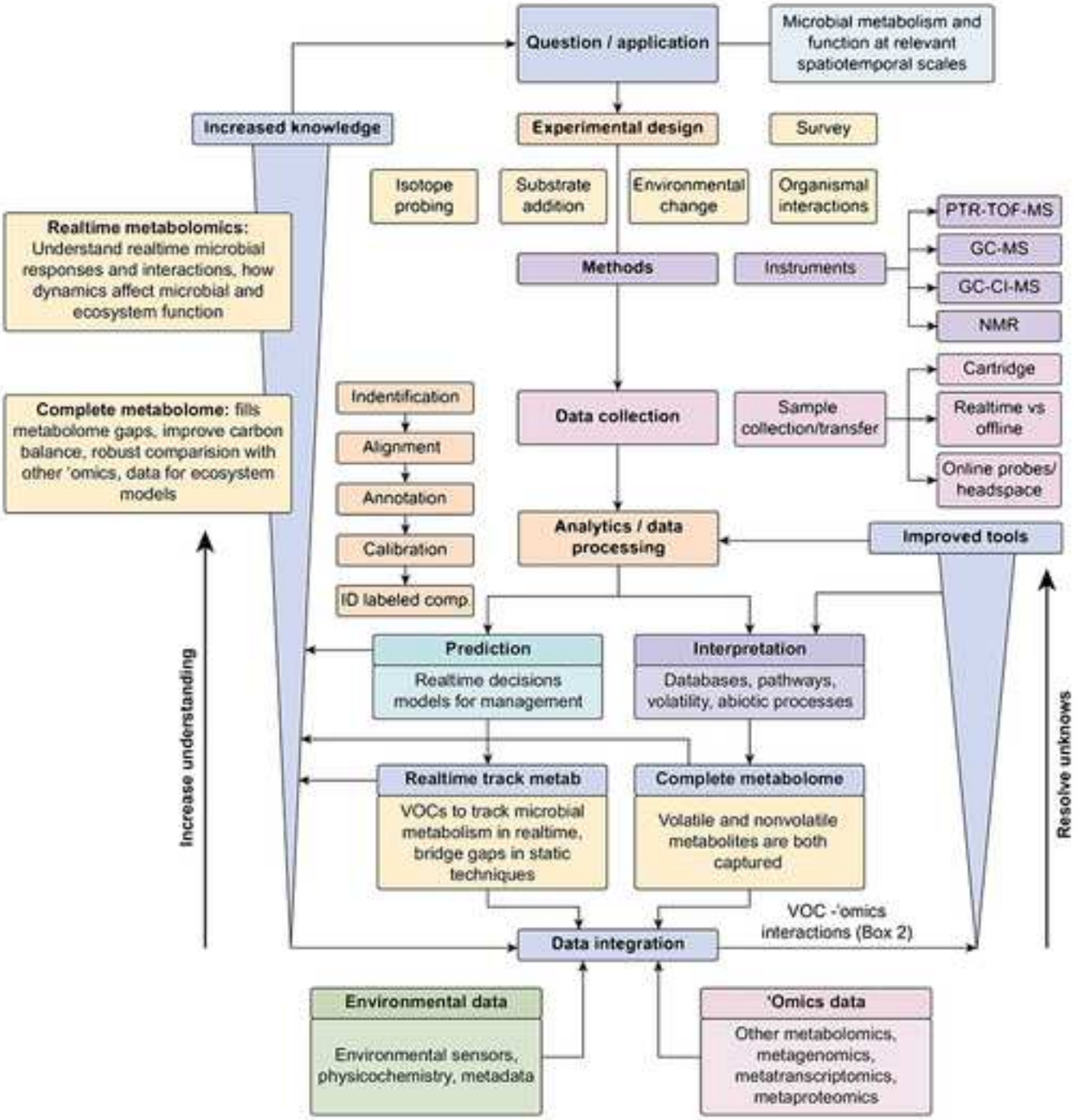
Q7: How can realtime VOC detection yield new insights into the accuracy and limitations of 'snapshot' 'omics sampling approaches, and how efficiently will it help inform when to conduct snapshot sampling?

Key Figure



Figure

Click here to access/download;Figure;Webshop_Volatilome_Figure 2.jpg



Figure

Click here to access/download;Figure;Webshop_Volatilome_Figure 3_ Figure I Box 2.jpg