Trends in
Plant Science

CellPress
OPEN ACCESS

Review

High-throughput plant phenotyping: a role for metabolomics?

Robert D. Hall [1,2,3,*], John C. D'Auria [4], Antonio C. Silva Ferreira [5,6,7], Yves Gibon [8,9], Dariusz Kruszka [10], Puneet Mishra [11] and Rick van de Zedde [12]

High-throughput (HTP) plant phenotyping approaches are developing rapidly and are already helping to bridge the genotype–phenotype gap. However, technologies should be developed beyond current physico-spectral evaluations to extend our analytical capacities to the subcellular level. Metabolites define and determine many key physiological and agronomic features in plants and an ability to integrate a metabolomics approach within current HTP phenotyping platforms has huge potential for added value. While key challenges remain on several fronts, novel technological innovations are upcoming yet underexploited in a phenotyping context. In this review, we present an overview of the state of the art and how current limitations might be overcome to enable full integration of metabolomics approaches into a generic phenotyping pipeline in the near future.

Highlights

Advanced facilities for automated phenotyping have already been established but there is a growing demand to integrate additional sensing modalities to allow deep phenotyping. Ongoing developments in automated (noninvasive) metabolomics show great promise.

Approaches to online volatile metabolomics are progressing well although nonvolatile analyses continue to pose significant challenges in terms of practicality and cost.

Integration of multiple sensors such as optical molecular spectroscopy, imaging, and mass spectroscopy in automated phenotyping facilities shall provide enhanced complementary insights into the dynamics of plant phenotype.

In the domains of metabolomics, chemometrics, transcriptomics, and imaging, recent data fusion techniques such as statistical multiblock data analysis, network and pathway modelling, and deep learning are beginning to realize systems-type approaches.

The growing desire for deeper plant phenotyping

Considerable recent progress has been made in the development of large-scale, **HTP (see** Glossary) plant phenotyping facilities – now also for large crop plants [1,2] (Figure 1, Key figure). The main characteristics of these facilities are advanced automation, combined sensing technologies, and high-end computing facilities to handle and process the large-scale data generated [3–6]. However, most phenotypic characteristics which can be rapidly and repeatedly monitored, relate either to physical properties (plant height, leaf size, shape) or spectral properties such as chlorophyll fluorescence, **near-infrared (NIR)** reflectance, and color [4,7–9]. Such analyses are relatively easy to perform in a nondestructive, automated fashion and permit following the same plants repeatedly in time. However, there is now an increasing demand to expand upon these noninvasive technologies to deepen analyses to the substructural level.

Plant metabolites can be the cause and/or the consequence of plant phenotype. Therefore, there would be great value in being able to combine extensive physical and spectral data with additional chemical data across time. This is especially relevant when the aim is to decipher key features such as the biochemical dynamics of plant response to (a)biotic stress, product (leaf, fruit) quality, or to follow organ development such as fruit ripening [10,11]. Integrating a generic **metabolomics** platform would be ideal, but this brings with it a set of technical, temporal, scalability, and data management challenges. Progress on each of these fronts is being made, and in this review, we combine state-of-the-art descriptions of the relevant components as well as some emerging innovations, with predictions and foresights into the future integration of metabolomics in **HTP phenotyping (HTPP)** pipelines.

Current state of the art: metabolomics

Over the past 20 years, significant developments have occurred in the different fields of HTP omics technologies. Specifically for metabolomics, remarkable progress has been made, and

[1]BU Bioscience, Wageningen University & Research, 6700 AA, Wageningen, The Netherlands
[2]Laboratory of Plant Physiology, Wageningen University, 6700 AA, Wageningen, The Netherlands
[3]Netherlands Metabolomics Centre, Einsteinweg 55, Leiden, The Netherlands
[4]Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK Gatersleben), Gatersleben, Corrensstraße 3, 06466 Seeland, Germany
[5]Universidade Católica Portuguesa, CBQF—Centro de Biotecnologia e Química Fina—Laboratório Associado, Escola Superior de Biotecnologia, Rua Arquiteto Lobão Vital, Apartado 2511, 4202-401 Porto, Portugal

Key figure

An example of a state-of-the-art automated and modulated plant phenotyping facility installed at The Netherlands Plant Eco-phenotyping Centre (NPEC)

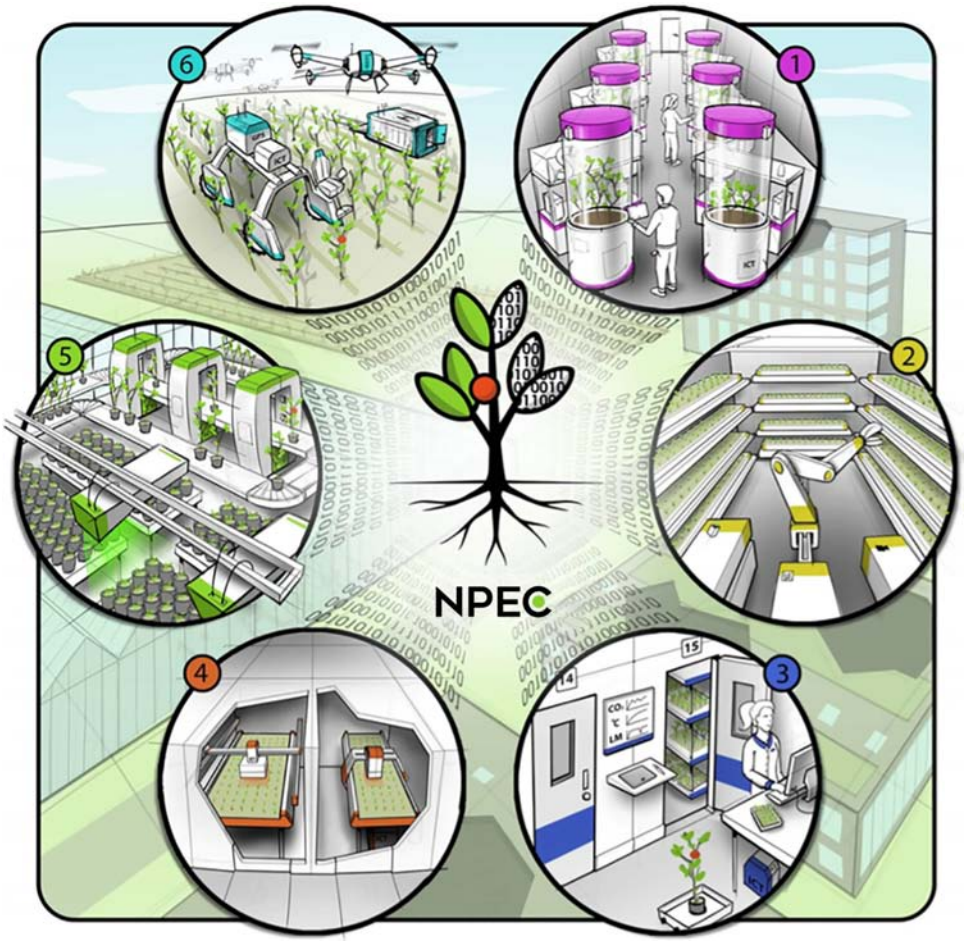

Figure 1. Together, six modules currently form an unbroken chain of phenotypic analyses from the molecular level through to micro-phenotyping and physiological phenotyping. This chain aims to translate mechanistic outputs from controlled conditions to open field situations, as well as feeding back results from the field into more controlled conditions where fundamental mechanisms can be unraveled.

[6]Faculty of AgriSciences, University of Stellenbosch, Matieland 7602, South Africa
[7]Cork Supply Portugal, S.A., Rua Nova do Fial, 4535, Portugal
[8]UMR 1332 Biologie du Fruit et Pathologie, INRAE, Univ. Bordeaux, INRAE Nouvelle Aquitaine – Bordeaux, Avenue Edouard Bourlaux, Villenave d'Ornon, France
[9]Bordeaux Metabolome, MetaboHUB, INRAE, Univ. Bordeaux, Avenue Edouard Bourlaux, Villenave d'Ornon, France
PMB-Metabolome, INRAE, Centre INRAE de Nouvelle, Aquitaine-Bordeaux, Villenave d'Ornon, France
[10]Institute of Plant Genetics, Polish Academy of Sciences, 60-479 Poznan, Poland
[11]Food and Biobased Research, Wageningen University & Research, 6708 WE, Wageningen, The Netherlands
[12]Plant Sciences Group, Wageningen University & Research, 6700 AA, Wageningen, The Netherlands

*Correspondence:
robert.hall@wur.nl (R.D. Hall).

many new tools have been implemented that offer clearer mechanistic insights by enabling the correlation of (bio)chemical changes with phenotype [12,13]. Unlike genes, transcripts, and proteins, metabolites are less easy to define and are chemically highly diverse [14]. Nevertheless, the focus of metabolomics studies is increasingly shifting from the categorization of chemical structures – feature annotation – to finding biochemical narratives.

Metabolomics is an integrative platform of competences from different disciplines, ranging from analytical chemistry, statistics, signal processing to (bio)chemical expertise. There are two main

categories: targeted (hypothesis-driven) and untargeted (data-driven) studies. Untargeted studies focus on the profiling of the total complement of molecules ('fingerprint') in a sample – and is often a valuable starting point – while targeted approaches focus on the quantification of groups of molecules once preselection is possible and appropriate. However, it is often more productive to study an entire system following an integrated approach (untargeted), rather than to apply a reductionist approach by analyzing its parts (targeted). In both cases, the main challenges are chemical complexity and heterogeneity, extraction efficiency, analytical dynamic range, and measurement throughput [13].

Current HTP analytical techniques result in complex metabolomics data sets which are challenging to visualize and interpret. Analytical success is dependent not only on the performance of specific hardware but also on the subsequent treatment of data using advanced modelling approaches [15]. Univariate and multivariate analysis techniques, chosen appropriately on the basis of experimental design and biological question, are routinely used to extract relevant information from the data and convert this into chemical knowledge. However, such approaches lack the systematic environment of metabolites and their interdependencies [16]. Irrespective of the techniques used for data acquisition, unequivocal metabolite identification remains a crucial technical challenge as it is only after biomarker identification that biological significance can be fully determined.

In metabolomics, five levels of molecular identification have now been established by the Metabolomics Society [17] and much effort is currently being extended to advance our skills in computational metabolomics and develop novel workflows for peak annotation [18,19]. All available data mining options need to be exploited for the most robust annotation strategy. Putative metabolite identification can be made based on mass-to-charge ratio (*m/z*) of the mass spectral ion and options for complementary approaches for further annotation: mass spectral database searching, *in silico* fragmentation tools, and orthogonal coupled techniques including retention time matching and **ion mobility spectrometry (IMS)** are advancing [17,19]. Particularly, in a HTPP context, new ultrafast separation approaches giving shorter retention times coupled to high-resolution mass spectrometry (MS) are potentially valuable. Several IMS variants have so far been poorly exploited for crops [20]. However, for certain applications, IMS could potentially replace, or be used in parallel to, **liquid chromatography (LC)-MS** methods since IMS-MS has advantages relating to increased resolution and analysis speed and has greater potential to separate and annotate structurally similar (isomeric) metabolites not possible with standard LC-MS [21]. There are also indications the technology may be more suited for portable devices [20]. The most recent variant, structures for lossless IM (SLIM), using a highly advanced form of IMS may become orders of magnitude faster and could make it highly attractive in a HTPP context [22].

The chemical space of small molecules is currently spread across multiple databases, and the number of compounds reported exceeds 120 million. Current estimations of the number of metabolites in living organisms is approaching 1 million [13], yet the majority of detected metabolites remains unknown. Standardization practices are also still lacking although increasing efforts are being made to rectify this [13,23]. The first significant steps have already been taken but there is still some way to go before we reach an automated approach towards fully annotated metabolite profiling [19,24].

Current state of the art: automated plant phenotyping

Just a few decades ago, plant phenotyping tasks were considered as particularly taxing, requiring laborious human interventions – from germinating seeds and monitoring plant growth to measuring final yield [25]. Through recent advancements in sensing technologies, automation and machine learning, and artificial intelligence approaches, the task of plant phenotyping has become fully

Glossary

Canonical correlation analysis (CCA): a statistical method for inferring information from cross-covariance matrices.
Direct injection (DI): a rapid method for MS-based analyses where no separation method (gas chromatography or liquid chromatography – see later) is used and so analysis times can be greatly reduced but with significant loss of resolution.
Flux balance analysis (FBA): a mathematical approach for analyzing the flow of metabolites through a metabolic network.
Gas chromatography–mass spectrometry (GC-MS): a dedicated separation and detection technique for the analysis of volatile compounds, often present in complex mixtures as typified by flower perfumes, fruit odour, and degradation or fermentation processes.
High throughput (HTP): an approach referring to the execution of many repeat measurements in a relatively short time.
HTP phenotyping (HTPP): an approach where an integrated set of equipment is used in parallel and/or in series to evaluate rapidly specific physical or spectral parameters of a substantial number of whole plants *in situ*. Measurements usually should only take a few seconds to 1–2 min and are repeated at appropriate time intervals (hours – days – weeks) according to the goals of the experiment.
Interdependent component analysis (ICA): a statistical and computational technique for revealing hidden factors underlying sets of random variables.
Ion mobility spectrometry (IMS): an analytical technique used to separate and identify ionized molecules in a gas phase based on their mobility in a carrier buffer gas.
Liquid chromatography–mass spectrometry (LC-MS): a dedicated separation and detection technique for nonvolatile, usually semi-polar metabolites such as plant secondary metabolites (phenolics, alkaloids, etc.) or for example, lipids or fatty acids (lipidomics).
Metabolomics: the technology designed to provide unbiased analyses of the biochemical composition of complex biological extracts.
Near-infrared spectroscopy: in the wavelength range from 780 nm to 2500 nm.

automated, requiring minimal human intervention [3,4,6,26]. Automated phenotyping is now a widely desired approach for whole-plant analysis [27] and can be performed both in the uncontrolled environment of the open field using, for example, aerial imaging or automatic robotic vehicles, as well under greenhouse-controlled conditions using fully automated HTP setups [28] (Figure 1). The primary aim of automated phenotyping is to track the physicochemical changes throughout growth as plants interact with their environment. Although open field phenotyping is widely executed for the assessment of plant performance under true crop conditions, often beforehand, significant time will have been spent using greenhouse-based systems and controlled environments. By exploiting greenhouse-based conclusions, thousands of genotypes can be narrowed down by preselection, yielding only a small number to be carried forward to field trials. Greenhouse and open field trials are thus components of a complementary process. Recently, all around the world, there has been massive development of new HTP greenhouse phenotyping facilities at research institute level such as the European centers: PKH facility[i] in Gatersleben, Germany; Phenovision[ii] in Ghent, Belgium [2]; NEPC[iii] in Wageningen, The Netherlands [7]; and PhenoArch[iv] at INRAE, France [29]. Beyond Europe, there are now also key local or national centers, for example, in the USA (Danforth[v]), Australia, (APPF[vi]), and China[vii]. Industrial facilities, such as at Bayer CropScience, Germany [30,31], depict not only scientific but also commercial interest in automated HTPP.

In the current state of the art, recent applications of HTPP suggest that automation can be performed for a wide variety of traits including drought tolerance [2,32], salt tolerance [33], and biotic stress [34,35] as well as for assessing the efficiency of plant protection agents [30]. Furthermore, facilities are not only used for phenotyping small model plants such as *Arabidopsis thaliana* [36,37] but also can be used for large plants such as *Zea mays* [2,32] and even, tree species [38]. Currently, the main sensors integrated into HTP setups are vision-based sensors for tasks such as 3D shape estimation which allows morphological monitoring of features such as height, width, leaf area, and leaf development, along with general plant growth parameters. There is also a developing trend of integrating spectral sensors such as fluorescence imaging and visible and NIR spectral imaging to monitor the dynamics of certain physicochemical processes [7,39,40]. Here also the technology choice is determined both by the biological question and current technological limitations. There is recently also a growing interest in the integration of, for example, volatile sensors since these can play a key role in understanding metabolites released by plants particularly during (a)biotic stress. However, currently, there is no automated HTP setup that offers the possibility to measure volatiles at frequencies matching those of imaging sensors. There are some reports that have correlated visible and NIR spectroscopy data with secondary metabolites [41]. However, applications are still lacking from the perspective of HTPP experiments.

Developments in rapid, online, microscale volatile metabolomics

Volatile organic compounds (VOCs) play a multitude of roles during the plant life cycle. They can be continuously present, having a protective function through anti-insect or antimicrobial bioactivity [42,43]. However, their release can also be exogenously induced or developmentally regulated. Plants under attack often respond by releasing specialized volatiles to reduce herbivory (e.g., α, β-pinene and limonene in conifers; myrcene, carene, and ocimene, in *Citrus spp.*), attract tritrophic predators [e.g., (S)-(+)-linalool in *Nicotiana spp.]* or inhibit microbial growth such as multiple monoterpene derivatives in thyme (*Thymus vulgaris*) [43–45]. The attacker itself may also release detectable volatiles thus betraying its presence. Volatiles including many monoterpene derivatives are powerful harbingers of the arrival of developmental stages such as flower maturation and fruit ripening as well as the onset of senescence [46]. Consequently, particular volatiles can be exploited as markers for a specific plant development or health status.

Orthogonal partial least squares (OPLS): a statistical modelling tool variant of PLS which uses orthogonal signal correction.
Partial least squares (PLS): a statistical modelling tool used to find the fundamental relations between two matrices.
Partial least squares-discriminant analysis (PLS-DA): a statistical modelling tool variant of PLS.
Principal component analysis (PCA): a statistical modelling tool giving insights into separations between experimental groups.
Proton transfer reaction mass spectrometry (PTR-MS): a technique for sensitive, online measurements of trace amounts of volatile organic compounds (VOCs) in air.
Solid-phase microextraction (SPME): it uses an adsorbent fiber (various materials are available) to trap volatiles mainly in a static system after which the fiber is directly desorbed within the GC.
Time-of-flight mass spectrometry (TOF-MS): used for the measurement of accurate atomic masses.
Two-way orthogonal partial least squares (O2PLS): a statistical modelling tool which is a symmetric method, modeling both predictive and systematic variation.
Ultra-high-performance liquid chromatography (UHPLC): used to reduce runtimes during chromatographic separation and increase molecular resolution.
Volatile organic compounds (VOCs): metabolites which are usually in gaseous form at room temperature and which can be naturally released (or induced) into the atmosphere. Many can be bioactive in that they react with aroma receptors (on the human tongue, insect antennae, etc.) and can lead to behavioral effects related to, for example, repellence (toxicity), attractance, taste, off-flavor, etc.

One advantage of targeting VOCs for *in vivo* phenotyping is their natural release from the plant, making nondestructive analytical approaches relatively straightforward. Minimum handling gives minimum damage, yielding true *in vivo* profiles. Sampling could be performed using small pumps to suck air away from the organ of choice, driven by positioning robots (Figure 2) to a precise location for dynamic headspace trapping using for example, Tenax tubes. A tube-switching station could be used to enable sampling to be done on a plant-by-plant basis after which the adsorbent tubes could be analyzed using a laboratory **gas chromatography (GC)- MS**. Although laborious, this would overcome a key logistical issue. While sampling times could be kept short and held in line with the timing typically needed for the other phenotypic analyses, analysis times for a standard GC approach are usually much longer (15–45 min). Uncoupling these two steps provides a workable solution. **Solid-phase microextraction (SPME)** fibers and stir-bar (Twister®) systems [43] could also be considered for VOC trapping. The former is available with a range of adsorbents (and hence chemical specificities) but has very limited trapping volume while the latter has a much larger adsorbent volume and hence would be more appropriate for trace compound trapping. However, expense (SPME fibers) and practical limitations (Twisters®) may limit automated application strategies. Even with these options, for true HTP approaches, GC-MS analyses should ideally happen *in situ* and online.

Real-time VOC analysis is possible using **direct injection (DI)** approaches such as **proton transfer reaction (PTR)-MS** [43] or by using an electronic nose [47]. Both are fast as they require no separation time and have a broad chemical detection range. The former also has

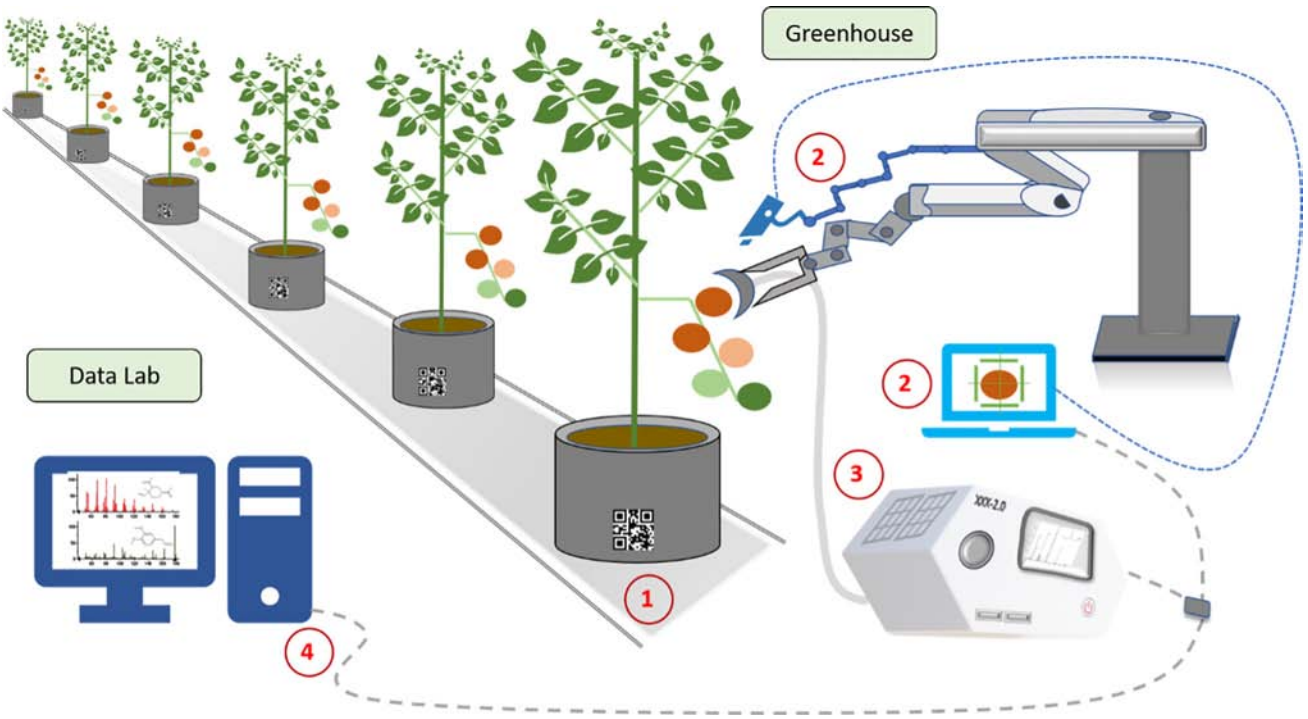

Figure 2. Conceptualization of an online VOC analysis pipeline within a standard automated plant phenotyping facility. All the technologies are principally available but have yet to be installed and integrated in such a setup to test for feasibility and applicability. (1) Standard conveyor-belt system for circulating plants through a greenhouse as is typical of most plant phenotyping facilities; (2) Programmable, image-driven robot arm used to (re)locate specific organs on each plant for localized collection of naturally released volatiles; (3) Mobile GC-MS device for rapid VOC analysis; (4) Laboratory-based data analysis and interpretation. Abbreviations: GC-MS, gas chromatography-mass spectrometry; VOC, volatile organic compounds.

excellent broad sensitivity in the low ppt range [48]; for a full review see [49]. Mozaffar *et al.* successfully used PTR-MS to follow the dynamics of young to senescing *Z. mays* leaf VOC emissions [50,51] which were found to comprise 31% methanol, 30% acetic acid, and 11% aldehyde. Using PTR-MS cycle times of 45 sec, a fivefold methanol emission difference across the day was observed for young maize leaves (3–26 μg/gDW/h) and for developing and senescing leaves, emission rates differed ≤400 fold (1.4–570 μg/gDW/sec) depending on the metabolite and developmental stage [50,51].

PTR-MS can rapidly detect and quantify mass features but cannot assign structural identity unless they are equipped with a high-resolution **time-of-flight (TOF)-MS**. Consequently, it is usually used to screen for preidentified compounds or following the dynamics of unannotated mass features [43,49]. The same limitation applies to E-nose instruments [47], as both rely on the separate use of GC-MS for full annotation. E-nose instruments are portable, but current PTR-MS methods are cumbersome, making them inappropriate for a HTPP facility. Furthermore, the glass chambers resembling large bell jars, regularly used to enclose the plants fully to confine the VOCs, are impractical and have the disadvantage of creating an artificial environment, as pointed out for the VOC-SCREEN system [48]. However, as the imaging cabins in HTPP facilities are often sealed for light exclusion, this may offer concomitant trapping opportunities. VOC-SCREEN is an excellent example of what has become possible with online volatile phenotyping, but our desire is to go further. Alternative methods, including fastGC [52], can enable separation (and hence identification) and shows great potential. Near-to-real-time separation of six isomeric monoterpenes, α- and β-pinene, limonene, 3-carene, camphene, and myrcene, was achieved with high sensitivity [1.2 parts per billion (ppb)] in a separation taking just 80 sec [52].

The goal is to have rapid, online, *in situ*, high sensitivity, and high-resolution VOC analyses with full annotation capacity using a small-scale, mobile instrument (Figure 2). Recent innovative developments in portable GC-MS instruments could already make this possible for plant-based applications. Various instrument makers have products on the market including Teledyne FLIR (G510), Bruker (E^2M), and PerkinElmer (T-9). Philips also have a micro-GC prototype under development (personal communication). Advances including low vacuum operation, capillary GC, and miniaturized iontrap MS have enabled the design of small instruments weighing less than 15 kg. Detection in the low **parts per million**–ppb range is possible. Envisaged primary users are however, 'first responders', the military and crime officers, so unfortunately, no reports of these instruments being tested for plant applications are yet available. Success will depend not only on instrumentation but also on sensitivity as natural volatile emission rates can vary extensively [43,50] entailing that some online applications will be easier to implement than others.

The challenges of nonvolatile metabolomics

The nonvolatile metabolome represents rich information reflecting past (e.g., slow turnover metabolites accumulated in response to past stress), present (e.g., high turnover metabolic intermediates), and future (e.g., precursors of biomass under construction) events. Accordingly, a growing number of top-down studies have shown that this metabolome can be correlated with performance in panels of genetic diversity [53,54], and when metabolic traits are obtained from plants grown under controlled conditions, it becomes possible to predict yield [55] or stress resistance in the field [56]. Information with high predictive value, capable of competing with genomic selection, can be obtained nondestructively with spectral methods such as NIR [57]. However, such noninvasive methods are limited to certain polymers and abundant molecules [58]. It would be attractive to integrate a destructive analysis using limited, targeted tissue sampling, to give maximum coverage of the nonvolatile metabolome. Until recently, cost issues limited metabolomics applications in large-scale phenotyping. However, high-resolution MS

[TOF-MS, Orbitrap, and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS)], enabling one to distinguish different structures with the same nominal mass, using ultrafast chromatography now makes it possible to combine HTP with high resolution [59]. Systems biology approaches and working with large sample numbers (Figure 3) have become possible and increased observation numbers will enable the development of new prebreeding strategies based on predictive models [60].

Sampling is critical – several factors need to be addressed, including (i) the stage of development of the plant or organ, (ii) the position on the organ if it is not sampled entirely, (iii) the time of day because the metabolome fluctuates [61] and therefore (iv) the duration of the sampling step, and (v) the necessity to quench biochemical reactions. Manual sampling is constrained and often a source of error [62] and automation is rarely envisioned. However, agricultural robots are under development, particularly for harvesting fruit [63]. Similar robots, equipped with 3D cameras and capable of taking leaf samples, could soon pave the way for destructive observations [64]. For sample preparation, lyophilizing is a slow step but is possible on a large scale while still preserving sample quality [65] and facilitates the grinding and weighing steps. Grinding and weighing robots have been under recent development [66] but are rarely used routinely. Such sampling approaches are appropriate to deliver samples for both subsequent metabolomics and transcriptomics applications.

The most widely used extraction method, giving a wide metabolome coverage and acceptable polarity compromise, uses methanol-water [65]. Less toxic ethanol could replace methanol making HTP extraction easier to implement [67]. A robot station equipped with a 96-channel head can extract hundreds of samples per day – exceeding the current throughput of a mass spectrometer – so the challenge is to increase analysis throughput. For liquid chromatography, the development of ultrafast **ultra-high-performance liquid chromatography (UHPLC)** can reduce runtime by a factor of 10, to a few minutes, but requires mass spectrometers with rapid data acquisition rates [59]. If the goal is to screen for known markers (as opposed to profiling), direct injection methods (DI-MS) potentially offer huge time savings thus maximizing sample throughput. However, DI methods may also entail technical limitations related to ion cosuppression,isomeric or isobaric structural similarities, and annotation confidence. However, Sarvin *et al.* [68] describe a method for human application, where the trade-off between analysis time and sensitivity has allowed

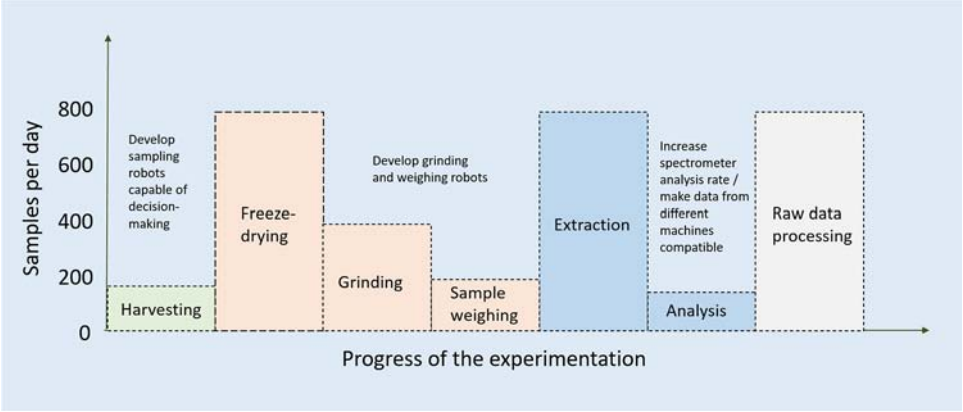

Figure 3. Current sample harvesting and LC-MS-based analysis capacities with indications of where additional developments are required to meet the needs for integration into an *in vivo* HTP (molecular) phenotyping facility. Abbreviations: HTP, high throughput; LC-MS, liquid chromatography-mass spectrometry.

reproducible detection of 19,000 *m/z* features in a total run time of 0.45 min. LC miniaturization strategies also offer potential for increased (cost) efficiency [69]. However, the challenges for miniaturisation of sampling/extraction of plant materials, compared with plasma or urine, cannot be underestimated.

Annotation, the recurring pet peeve of nonvolatile metabolomics, has made tremendous progress in recent years, with algorithms using artificial intelligence capable of naming thousands of metabolites in hundreds of samples per day [70,71]. However, perhaps the biggest challenge is to achieve interoperability of metabolomic data [72,73]. Measurements are not easily repeatable from one analytical pipeline to another or even from day to day on the same pipeline, concerning both metabolome coverage and metabolite (semi) quantification [74]. As stated previously, it is evident that most steps of such 'full-range phenotyping' can be automated, from the cultivation of plants to the processing of metabolomic data. However, there is no turnkey solution that combines all these steps into one pipeline. The time has therefore come to move on to the technological demonstrator stage by building or adapting platforms for the full range of procedures, which will make it possible to identify and solve the bottlenecks that would arise in such a pipeline in a step-wise manner.

Metabolomics data analysis and integration challenges

Metabolomics is much more than a new provider of proxies for performance predictions and has the potential to facilitate deep insights into the genotype–phenotype interconnection. While there is a plethora of data being collected over a multitude of omics-based approaches in plants, the higher-level integration of these data still requires a set of strategies and approaches spanning various disciplines. To better grasp this challenging task of combining multiple data sets in a biologically relevant context, it is critical to describe fundamentally, three levels of data integration: conceptual, statistical, and model-based methods [75]. Conceptual integrations are where omics data sets are analyzed separately, and the results are compared and matched to reach biologically relevant conclusions. Statistical integration is commonly used for transcriptomic and metabolomic data analysis. These methods include correlation-based, concatenation-based, and multivariate-based methods (Figure 4A–C). This also leads to more advanced network-based (Figure 4D)

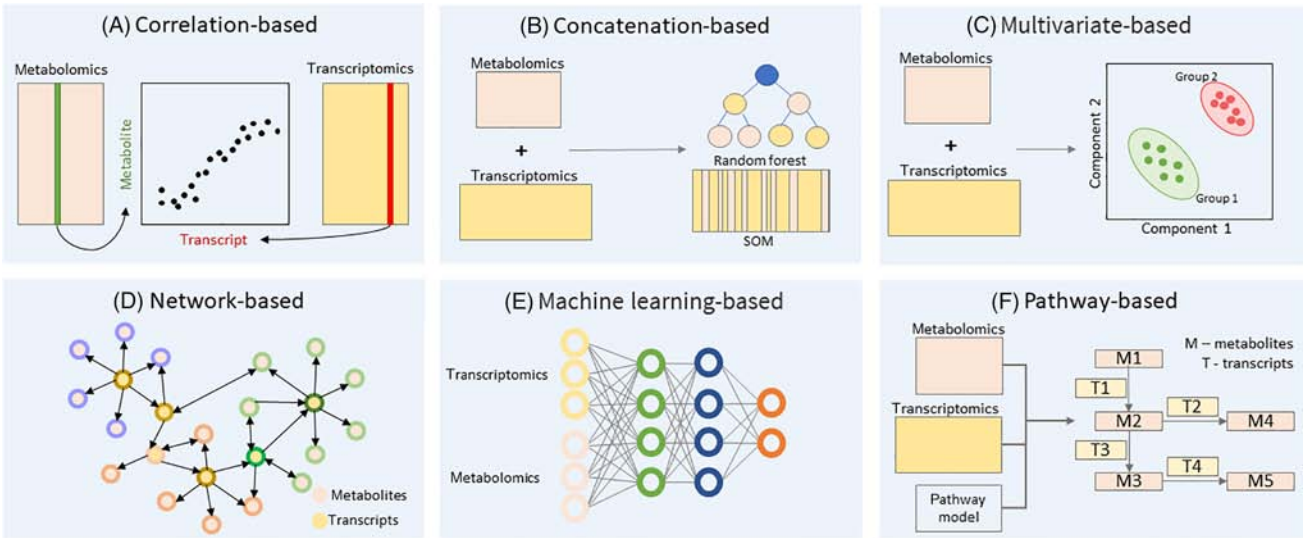

Figure 4. Alternative methods of transcriptomic and metabolomic data integration being (A) correlation based; (B) concatenation based; (C) multivariate based; (D) network based; (E) machine learning based and (F) pathway based.

and machine learning approaches (Figure 4E). The model-based method includes some integrative tools for pathway analysis and reaction flux modelling (Figure 4F and Box 1). Regardless of further downstream analyses, the first step of integration of transcriptomics, proteomics, and metabolomics data is data processing. This step implements quality checks (e.g., removing outliers, absent values, unification of gene and metabolite symbols or names) and statistical normalization and transformation procedures [76].

How does one correlate the up/downregulated transcript profile of a plant tissue with changing metabolite levels? Here, correlation-based methods, that can address these questions, include the Pearson's or Spearman's correlation test and the more advanced correlation measuring methods: the Goodman and Kruskal's gamma test, partial correlation, linear models, or **canonical correlation analysis (CCA)** [75]. All these methods focus on the relationships between paired transcripts – metabolite levels to find pairs either positively (both variables increased) or negatively (one increased and one decreased) correlated. Perez de Souza *et al.* [77] used Pearson's correlation for integration of metabolites with transcripts to identify genes in the common bean participating in specialized metabolism. These types of calculation are supported by the R package IntLIM, which utilizes a linear model-based approach (Tools 1–3, Table 1). A simple method of data integration is the concatenation-based approach, which applies random forest, self-organizing maps, and k-means clustering to combine two data matrices (see Tools 4–6, Table 1). Glaubitz *et al.* [78] employed self-organizing maps to identify clusters of genes and metabolites connected to high temperature tolerance in rice.

Omics approaches generate large quantities of variables (metabolite features, proteins, and genes). Many experimental questions do not focus initially on one variable, but aim to understand how multiple constituents are related. Multivariate latent space-based integration resolves these problems by using statistical methods such as **principal component analysis (PCA)**, consensus PCA, **independent component analysis (ICA)**, **partial least square (PLS)**, **orthogonal two-way PLS (O2PLS)**, multiblock PLS, **PLS-discriminant analysis (DA)**, or orthogonal PLS-DA (OPLS-DA) [79]. However, O2PLS is often used for integration, because it employs a joint variation model to describe the connections between for example, genes and metabolites [80,81]. This approach was applied for integrating membrane lipid data with transcripts from *Arabidopsis* [82] and has also been used to connect soybean (*Glycine max*) primary metabolism

Box 1. Towards an *in silico* plant model

In silico plant modelling relies upon a computational model that links gene networks, metabolic pathways, and cellular, tissue and organ compartments and enables simulations that range from metabolite flux to plant development. The creation and development of *in silico* models is the subject of large international projects [92]. To achieve the goal of a virtual plant, models are constructed using two components: analytical data (transcripts, proteins, metabolites) alongside a computational component. The model may include parametric metabolic networks for the plant, which can be used to study the reaction flows based on **flux balance analysis (FBA)** [93]. The curated and accurate models for FBA can be deposited as SBML files in repositories such as PlantSEED. This repository currently contains models for 39 organisms [94] and can be used by anyone wishing to build their own models. The FBA methods use the COBRA (constraint-based reconstruction and analysis) approach implemented in the MATLAB or Python software environment [95]. In contrast to tools requiring one to master programming languages, MetExplore [96] and Escher-FBA [97] are user-friendly online platforms for calculating fluxes, their visualizations and sharing with other users. FBA could be used to discover the alternative pathways or identify mutations of genes involved in metabolomic pathways [93]. The number of plant organisms with reconstructed metabolism models or pathways is limited [94]. Moreover, a vast number of metabolites, particularly specialized metabolites and lipids, do not currently have annotated predicted biosynthetic pathways. In order to address this issue, network-based integration can assist in connecting the metabolites with enzymes or genes in pre-existing networks, enabling their further analysis and visualization [98]. Networks can be constructed using constraint-, correlation-, and co-expression-based models [99,100]. Network analysis was used to discover a gene cluster associated with anthocyanin biosynthesis and potato pigmentation [101], phenolic biosynthesis [55,102], or genes/metabolites involved in tomato resistance to pathogens [103]. Network-based integration is part of several online and R-based tools including Tools 16, 18–25 (Table 1) or implemented in the CytoScape environment [99].

Table 1. Main software tools designed and currently in use for the integration of omics data sets

Tool #	Software	Omics[a]	Functionalities	Comments	Repositories	Refs
1	IntLIM	T, M	Linear models	Identification of gene-metabolite pairs in relation to phenotype	R/GitHub	[104]
2	integrOmics	T, P, M, R	Multivariant analysis	Selection of variables using CCA, PLS, and machine learning tools	R/CRAN	[105]
3	mixOmics ggmixOmics	T, P, M, R	Multivariant-based framework (PCA, CCA, PLS-DA, etc.)	Dimensions reduction, extraction of variable subgroups connected with traits and visualizations	R/CRAN	[106]
4	MarVis	T, M	Self-organizing maps and pathway enrichment analysis	Identification of profiles and markers connected with experimental conditions, output visualization	Matlab/app	[107]
5	SOMbrero	T, P, M	Tools for self-organizing maps	Identification of profiles connected to traits	R/CRAN	[108]
6	MetaGeneAlyse	T, M	Concatenation and multivariant-based tools	Statistical approach for normalization, clustering, and PCA/ICA	online	[109]
7	omicade4	T, R	Multivariant approach (cPCA, MICA)	Integration of data from multiple platforms, independent of annotation feature selections, visualization of output	R/BioC	[110]
8	GO2PLS	T,R	Implementation of multivariant analysis	Creating the O2PLS analysis and their validation and visualization	R/CRAN	[80]
9	OmicsPLS	T, M			R/CRAN	[81]
10	xMWAS	T, P, M	Multivariant and network-based framework	Application for paired and unpaired study	R/GitHub	[111]
11	metaboGSE	T, M	Connection of network-based approaches and gene set enrichment analysis	Creation of subnetworks in the context of experimental condition	R/CRAN	[112]
12	multiGSEA	T, P, M	Enrichment analysis	Supporting many pathway databases, combined multiomics enrichment test	R/GitHub	[113]
13	FELLA	M	Network-based enrichment analysis of metabolites lists	Supporting KEGG database	R/BioC	[114]
14	GAIT-GM	T,M	Galaxy tool for mapping genes and metabolites data into KEGG pathways	Text-mining algorithm improved annotation; metabolite abundance as a function of gene expression	Python/ Online	[115]
15	RaMP	T, M	Enrichment analysis for genes and metabolites	Clustering of over-represented pathways by pathway similarity	R/GitHub	[85]
16	MetaboAnalyst	T, M	Pathway-based analysis for metabolite and gene list using enrichment analysis (hypergeometric or Fisher's exact tests)	Support limited to a number of plant species (*Arabidopsis thaliana* and *Oryza sativa*)	Online/R	[116]
17	MapMan4	T,P,M	Pathway-based software supported by enrichment analysis	Annotation of transcripts and proteins	Online	[117]
18	PaintOmics 3	T, P, M, R	Pathway-based visualization and analysis of multiomics data including regulatory and region-based omics	Matching the metabolite annotations with KEGG names, mapping to KEGG pathways, creating the multiomics pathway interaction network and visualization	Online	[118]
19	MetExplore	T, P, M	Network-based analysis, pathway mapping, flux balance modelling and analysis	Easy way for network creation, visualization, curation, and metabolite mapping	Online	[96]
20	OmicsAnalyst	T, P, M	Univariant, partial, and multivariant analysis and visualization	Data processing (normalization, filtering), differential analysis, creation of correlation network, and reduction of network dimensionality, data visualization	Online	[119]
21	OmicsNet	T, P, M, R	Network and pathway-based approach	Building, visualization, and exploration of biological networks in 3D space	Online	[120]
22	MiBiOmics	T, P, M	Correlation-based tool for creating, dimensionality reduction, and exploration of networks	Provide the tools for data processing (filtration, normalization, and transformation)	Online	[121]

Table 1. (continued)

Tool #	Software	Omics[a]	Functionalities	Comments	Repositories	Refs
23	NetMet	T, M	Graph-based predicting metabolic capacities for microbial species	A useful tool for experiments with microbial systems	Online	[122]
24	MetaBridge	T, M	Network-based pathway mapping	Identification of connections between metabolites and enzymes, visualization of data and results	R/online	[123,124]
25	NetPathMiner	T,M	Framework for network and pathway exploration and management	Summarized the metabolic pathway using machine learning approach, visualization of data	R/BioC	[125]
26	Escher-FBA	T,P,M	Flux analysis and visualization	Possible implementation of other models, simulation of different environmental scenario, and knockouts of genes	Online	[97]
27	OmicsTIDE	T, P, M	Clustering and visualization	The comparing between trends in the omics experiments	Online	[126]
28	COBRA	T,P,M	'Constraint-based reconstruction and analysis' environmental for FBA	Creating, administrating, calculating, and visualization of flux models and analysis	Python/ Mathlab	[95]
29	DTW4Omics	T, M	Approach for time series alignment	Improving correlation analysis by the matches between time points	R/CRAN	[75]
30	MultiDataSet	T, P, M, R	Multiomics data management framework	Assists in the administration of omics data sets in R	R/BioC	[127]
31	mixKernel	T,P,M	Machine learning approach for data integration	Pattern discovery and phenotype predictions	R/CRAN	[90]

[a]Omics type: T, transcriptomics; P, proteomics; M, metabolomics; R, regulatory omics (e.g., TF, μRNA, epigenetics).

and transcripts after exposure to pathogen attack [83]. These methods are implemented in the free software tools 6–10 (Table 1).

In contrast to statistically based data integration, pathway-based methods provide information about the metabolites and enzymes localized in those pathways. This type of integration is valuable because it is similar to gene set enrichment analysis and returns annotated biochemical maps. Integrated metabolomics and transcriptomics analysis of infected soybean showed upregulation of the phenylpropanoid pathway [84]. Pathway repositories such as KEGG, MetaCyc, BioCyc, Reaxys, or WikiPathways are commonly used in pathway enrichment analysis [85]. This methodology is now used by multiple software packages (Tools 11–18, Table 1). These tools use the Fisher's or Wilcoxon's test for calculating p values for each pathway and for generating conclusions about the up- or downregulated or over-represented pathways. Validation is necessary using knock-out or overexpression mutants followed by a more detailed and targeted analysis.

In the current state-of-the-art HTP facilities, one of the widely integrated sensors for real-time physicochemical analysis of plants is optical molecular spectroscopy such as NIR. Optical molecular spectroscopy sensing provides complementary information to that of metabolomics approaches, particularly the physicochemical properties. In future developments where several sensing modalities are also integrated, it will be of crucial importance to analyse jointly the spectra and metabolite data to understand the complementary and distinct information generated by the different sensing modalities. Furthermore, several of the metabolites may carry a direct or indirect correlation to the optical molecular spectroscopy signal of the plant [41]. Hence, for efficient exploration of spectroscopy data with (non)volatiles, advanced data integration techniques such as multiblock data analysis has recently been advised [86].

The application of machine learning methods is the future of multiomics data integration and systems biology. Two papers describe many methods for use when applying machine learning

approaches to plant-based studies [87,88]. The application of machine learning and network analysis has already been able to predict accurately biochemical pathways in tomato using metabolite data [89]. The free R package mixKernel is one example that implements machine learning methods for use in data integration [90]. For example, Knoch *et al.* [91] compared the canola phenotype predictability for different data sets (genes, transcripts, metabolites) and their combinations. The metabolomics data showed low prediction accuracy. However, combining metabolomics and transcriptomics data improved prediction.

Concluding remarks and future perspectives

Following on from HTP molecular genotyping which has revolutionised plant breeding and plant physiology research, performing whole plant physicochemical phenotyping is the next key target. Techniques for noninvasive imaging have already been successful but we now need to develop our capacities further and initiate the implementation of molecular deep phenotyping for integration into the pipeline for a true systems analysis. Recent analytical developments show great promise while many have yet to be rigorously tested in the context of (crop) plant analysis. Key challenges do remain (see Outstanding questions) but developments are progressing at a rapid rate.

Acknowledgments

The authors acknowledge financial support from the European Commission through partnership in the STARGATE project (H2020/No. 952339): https://stargate-hub.eu/.

Declaration of interests

All authors declare to have no conflicts of interest.

Author contributions

R.D.H. designed the concept and together with all other authors wrote and edited the manuscript. All authors support the final version.

Resources

[i]www.ipk-gatersleben.de/en/research/molecular-genetics/automated-plant-phenotyping

[ii]www.psb.ugent.be/phenotyping/phenovision

[iii]www.npec.nl/

[iv]www.youtube.com/watch?v=BOGbWw58YJ0

[v]www.danforthcenter.org/our-work/core-facilities/phenotyping/

[vi]www.plantphenomics.org.au/about-us/#about-the-appf

[vii]https://pprcen.njau.edu.cn/PPRC/About_PPRC.htm

References

1. Yang, W. *et al.* (2020) Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214
2. Asaari, M.S.M. *et al.* (2018) Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform. *ISPRS J. Photogramm. Remote Sens.* 138, 121–138
3. Pieruschka, R. and Schurr, U. (2019) Plant phenotyping: past, present, and future. *Plant Phenomics* 2019, 7507131
4. Roitsch, T. *et al.* (2019) Review: New sensors and data-driven approaches—a path to next generation phenomics. *Plant Sci.* 282, 2–10
5. Zhao, C. *et al.* (2019) Crop phenomics: current status and perspectives. *Front. Plant Sci.* 10, 714
6. Singh, A. *et al.* (2021) Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends Plant Sci.* 26, 53–69
7. Mishra, P. *et al.* (2020) Close-range hyperspectral imaging of whole plants for digital phenotyping: recent applications and illumination correction approaches. *Comput. Electron. Agric.* 178, 105780
8. Nguyen, V.D. *et al.* (2022) Noninvasive imaging technologies in plant phenotyping. *Trends Plant Sci.* 27, 316–317
9. Mahlein, A.-K. *et al.* (2018) Hyperspectral sensors and imaging technologies in phytopathology: state of the art. *Annu. Rev. Phytopathol.* 56, 535–558
10. Alseekh, S. *et al.* (2021) Domestication of crop metabolomes: desired and unintended consequences. *Trends Plant Sci.* 26, 650–661
11. Naikoo, M.I. *et al.* (2019) Role and regulation of plants phenolics in abiotic stress tolerance: an overview. In *Plant Signalling Molecules* (Khan, M.I.R. *et al.*, eds), pp. 157–168, Elsevier/Woodhead Publishing
12. Patti, G.J. *et al.* (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13, 263–269

Outstanding questions

The key outstanding questions relate to the need for feasibility studies regarding already available instrumentation which is yet untested for the types of application proposed here.

Are the existing mobile GC-MS units already appropriate for application? Do the sensitivity and measurement speed match the release rate of marker VOCs?

Similarly, tissue sampling robots have already been designed for seedling genotyping – can these already be implemented in a HTP metabo-phenotyping facility, or do they require significant redesign?

Additional research and data analysis tool development will also be needed before we can fully integrate all data types coming in from spectral, chemical, imaging, and potentially other especially, environmental data sources in what must become a true systems biology approach to phenotypic analysis. How can we ensure that experts from the different disciplines find a common goal to work towards?

How can the new artificial intelligence tools such as deep learning support the understanding and integration of multimodel data generated with diverse sensors integrated in a HTP metabo-phenotyping facility?

Plant breeders are accustomed to work with (DNA) markers of unknown composition and could potentially apply metabolite 'unknowns' in a similar way. However, chemical annotation would greatly advance our capacity for biological interpretation. How can we improve our metabolite identification abilities to meet the desires of the biologist?

13. Alseekh, S. *et al.* (2021) Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat. Methods* 18, 747–756
14. Baker, M. (2011) Metabolomics: from small molecules to big ideas. *Nat. Methods* 8, 117–121
15. Eick, K. and Pohnert, G. (2015) Simplifying complexity in metabolomics. *Chem. Biol.* 22, 567–568
16. Barupal, D.K. *et al.* (2018) Integrating bioinformatics approaches for a comprehensive interpretation of metabolomics datasets. *Curr. Opin. Biotechnol.* 54, 1–9
17. Blaženović, I. *et al.* (2018) Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites* 8, 31
18. Tsugawa, H. (2018) Advances in computational metabolomics and databases deepen the understanding of metabolisms. *Curr. Opin. Biotechnol.* 54, 10–17
19. Schorn, M.A. *et al.* (2021) A community resource for paired genomic and metabolomic data mining. *Nat. Chem. Biol.* 17, 363–368
20. Lu, Y. *et al.* (2021) A rapid liquid chromatography-electrospray ionization-ion mobility spectrometry method for monitoring nine representative metabolites in the seedlings of cucumber and wheat. *J. Sep. Sci.* 44, 709–716
21. Masike, K. *et al.* (2021) Recent applications of ion mobility spectrometry in natural product research. *J. Pharm. Biomed. Anal.* 195, 113846
22. Wojcik, R. *et al.* (2019) SLIM ultrahigh resolution ion mobility spectrometry separations of isotopologues and isotopomers reveal mobility shifts due to mass distribution changes. *Anal. Chem.* 91, 11952–11962
23. Emwas, A.H. *et al.* (2015) Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. *Metabolomics* 11, 872–894
24. Spicer, R.A. *et al.* (2017) Comment: A decade after the metabolomics standards initiative it's time for a revision. *Sci. Data* 4, 2–4
25. Dhondt, S. *et al.* (2013) Cell to whole-plant phenotyping: the best is yet to come. *Trends Plant Sci.* 18, 428–439
26. Rahaman, M.M. *et al.* (2019) Data-mining techniques for image-based plant phenotypic traits identification and classification. *Sci. Rep.* 9, 19526
27. Awada, L. *et al.* (2018) The adoption of automated phenotyping by plant breeders. *Euphytica* 214, 148
28. Li, D. *et al.* (2021) High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. *Front. Bioeng. Biotechnol.* 8, 623705
29. Prado, S.A. *et al.* (2018) Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. *Plant Cell Environ.* 41, 314–326
30. Mishra, P. *et al.* (2019) Homogenising and segmenting hyperspectral images of plants and testing chemicals in a high-throughput plant phenotyping setup. In *10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*
31. Mishra, P. *et al.* (2019) Early detection of drought stress in *Arabidopsis thaliana* utilsing a portable hyperspectral imaging setup. In *Proceedings of 10th Workshop on Hyperspectral Imaging and Signal Processing:Evolution in Remote Sensing (WHISPERS)*, The Netherlands, Amsterdam
32. Asaari, M.S.M. *et al.* (2019) Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Comput. Electron. Agric.* 162, 749–758
33. Awlia, M. *et al.* (2016) High-throughput non-destructive phenotyping of traits that contribute to salinity tolerance in *Arabidopsis thaliana*. *Front. Plant Sci.* 7, 1414
34. Polder, G. *et al.* (2019) Potato virus Y detection in seed potatoes using deep learning on hyperspectral images. *Front. Plant Sci.* 10, 209
35. Mishra, P. *et al.* (2020) Close range spectral imaging for disease detection in plants using autonomous platforms: a review on recent studies. *Curr. Robot. Reports* 1, 43–48
36. Mishra, P. *et al.* (2021) A generic workflow combining deep learning and chemometrics for processing close-range spectral images to detect drought stress in Arabidopsis thaliana to support digital phenotyping. *Chemom. Intell. Lab. Syst.* 216, 104373
37. Gao, G. *et al.* (2020) The use of high-throughput phenotyping for assessment of heat stress-induced changes in arabidopsis. *Plant Phenomics* 2020, 3723916
38. Mazis, A. *et al.* (2020) Application of high-throughput plant phenotyping for assessing biophysical traits and drought response in two oak species under controlled environment. *For. Ecol. Manag.* 465, 118101
39. Mishra, P. *et al.* (2017) Close-range hyperspectral imaging of plants: a review. *Biosyst. Eng.* 164, 49–67
40. Mahlein, A.-K. *et al.* (2019) Comparison and combination of thermal, fluorescence, and hyperspectral imaging for monitoring *Fusrium* head blight of wheat on spikelet scale. *Sensors* 19, 2281
41. Couture, J.J. *et al.* (2016) Spectroscopic determination of ecologically relevant plant secondary metabolites. *Methods Ecol. Evol.* 7, 1402–1412
42. Maurya, A.K. (2020) Application of plant volatile organic compounds (VOCs) in agriculture. In *New Frontiers in Stress Management for Durable Agriculture* (Rakshit, A. *et al.*, eds), pp. 369–388, Springer, Singapore
43. Lubes, G. and Goodarzi, M. (2017) Analysis of volatile compounds by advanced analytical techniques and multivariate chemometrics. *Chem. Rev.* 117, 6399–6422
44. Ninkovic, V. *et al.* (2021) Plant volatiles as cues and signals in plant communication. *Plant Cell Environ.* 44, 1030–1043
45. Weisskopf, L. *et al.* (2021) Microbial volatile organic compounds in intra-kingdom and inter-kingdom interactions. *Nat. Rev. Microbiol.* 19, 391–404
46. Korankye, E.A. *et al.* (2017) Plant senescence: the role of volatile terpene compounds (VTCs). *Am. J. Plant Sci.* 8, 3120–3139
47. Cui, S. *et al.* (2019) Development of fast e-nose system for early-stage diagnosis of aphid-stressed tomato plants. *Sensors* 19, 3480
48. Jud, W. *et al.* (2018) Volatilomics: a non-invasive technique for screening plant phenotypic traits. *Plant Methods* 14, 109
49. Majchrzak, T. *et al.* (2020) Real-time volatilomics: a novel approach for analyzing biological samples. *Trends Plant Sci.* 25, 302–312
50. Mozaffar, A. *et al.* (2017) Methanol emissions from maize: ontogenetic dependence to varying light conditions and guttation as an additional factor constraining the flux. *Atmos. Environ.* 152, 405–417
51. Mozaffar, A. *et al.* (2018) Biogenic volatile organic compound emissions from senescent maize leaves and a comparison with other leaf developmental stages. *Atmos. Environ.* 176, 71–81
52. Materic, D. *et al.* (2015) Monoterpene separation by coupling proton transfer reaction time-of-flight mass spectrometry with fastGC. *Anal. Bioanal. Chem.* 407, 7757–7763
53. Meyer, R.C. *et al.* (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4759–4764
54. Riedelsheimer, C. *et al.* (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44, 217–220
55. Westhues, M. *et al.* (2017) Omics-based hybrid prediction in maize. *Theor. Appl. Genet.* 130, 1927–1939
56. Yadav, A.K. *et al.* (2019) Wheat drought tolerance in the field is predicted by amino acid responses to glasshouse-imposed drought. *J. Exp. Bot.* 70, 4931–4947
57. Rincent, R. *et al.* (2018) Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar. *G3 Genes Genomes Genet.* 8, 3961–3972
58. Villate, A. *et al.* (2021) Review: Metabolomics as a prediction tool for plants performance under environmental stress. *Plant Sci.* 303, 110789
59. Fekete, S. *et al.* (2014) Current and future trends in UHPLC. *TrAC Trends Anal. Chem.* 63, 2–13
60. Fernandez, O. *et al.* (2021) Plant metabolomics and breeding. *Adv. Bot. Res.* 98, 207–235

61. Gibon, Y. *et al.* (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in *Arabidopsis* rosettes. *Genome Biol.* 7, R76
62. Hannemann, J. *et al.* (2009) Xeml Lab: a tool that supports the design of experiments at a graphical interface and generates computer-readable metadata files, which capture information about genotypes, growth conditions, environmental perturbations and sampling strategy. *Plant Cell Environ.* 32, 1185–1200
63. van Herck, L. *et al.* (2020) Crop design for improved robotic harvesting: a case study of sweet pepper harvesting. *Biosyst. Eng.* 192, 294–308
64. Foix, S. *et al.* (2018) Task-driven active sensing framework applied to leaf probing. *Comput. Electron. Agric.* 147, 166–175
65. Kim, H.K. and Verpoorte, R. (2010) Sample preparation for plant metabolomics. *Phytochem. Anal.* 21, 4–13
66. Gomez, L.D. *et al.* (2010) Automated saccharification assay for determination of digestibility in plant materials. *Biotechnol. Biofuels* 3, 23
67. Bénard, C. *et al.* (2018) Untargeted analysis of semipolar compounds by LC-MS and targeted analysis of fatty acids by GC-MS/GC-FID: from plant cultivation to extract preparation. In *Methods in Molecular Biology: Plant Metabolomics* (1778) (Antonio, C., ed.), pp. 101–124, Springer, Berlin Heidelberg
68. Sarvin, B. *et al.* (2020) Fast and sensitive flow-injection mass spectrometry metabolomics by analyzing sample-specific ion distributions. *Nat. Commun.* 11, 3186
69. Vasconcelos Soares Maciel, E. *et al.* (2020) Miniaturized liquid chromatography focusing on analytical columns and mass spectrometry: a review. *Anal. Chim. Acta* 1103, 11–31
70. Dührkop, K. *et al.* (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16, 299–302
71. Shen, X. *et al.* (2019) Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. *Nat. Commun.* 10, 1516
72. Emami Khoonsari, P. *et al.* (2019) Interoperable and scalable data analysis with microservices: applications in metabolomics. *Bioinformatics* 35, 3752–3760
73. Peters, K. *et al.* (2019) PhenoMeNal: processing and analysis of metabolomics data in the cloud. *Gigascience* 8, giy149
74. Izumi, Y. *et al.* (2019) Inter-laboratory comparison of metabolite measurements for metabolomics data integration. *Metabolites* 9, 257
75. Cavill, R. *et al.* (2016) Transcriptomic and metabolomic data integration. *Brief. Bioinform.* 17, 891–901
76. Fabres, P.J. *et al.* (2017) A concise review on multi-omics data integration for terroir analysis in *Vitis vinifera*. *Front. Plant Sci.* 8, 1065
77. Perez de Souza, L. *et al.* (2019) Multi-tissue integration of transcriptomic and specialized metabolite profiling provides tools for assessing the common bean (*Phaseolus vulgaris*) metabolome. *Plant J.* 97, 1132–1153
78. Glaubitz, U. *et al.* (2017) Integrated analysis of rice transcriptomic and metabolomic responses to elevated night temperatures identifies sensitivity- and tolerance-related profiles. *Plant Cell Environ.* 40, 121–137
79. Subramanian, I. *et al.* (2020) Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights* 14, 7–9
80. Gu, Z. *et al.* (2021) Statistical integration of two omics datasets using GO2PLS. *BMC Bioinforma.* 22, 1–18
81. el Bouhaddani, S. *et al.* (2018) Integrating omics datasets with the OmicsPLS package. *BMC Bioinforma.* 19, 371
82. Szymanski, J. *et al.* (2014) Linking gene expression and membrane lipid composition of Arabidopsis. *Plant Cell* 26, 915–928
83. Copley, T.R. *et al.* (2017) An integrated RNAseq-1H NMR metabolomics approach to understand soybean primary metabolism regulation in response to Rhizoctonia foliar blight disease. *BMC Plant Biol.* 17, 84
84. Kang, W. *et al.* (2018) Transcriptomic and metabolomic analyses reveal that bacteria promote plant defense during infection of soybean cyst nematode in soybean. *BMC Plant Biol.* 18, 86
85. Zhang, B. *et al.* (2018) RaMP: a comprehensive relational database of metabolomics pathways for pathway enrichment analysis of genes and metabolites. *Metabolites* 8, 16
86. Mishra, P. *et al.* (2021) Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC Trends Anal. Chem.* 137, 116206
87. Reel, P.S. *et al.* (2021) Using machine learning approaches for multi-omics data analysis: a review. *Biotechnol. Adv.* 49, 107739
88. Picard, M. *et al.* (2021) Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* 19, 3735–3746
89. Toubiana, D. *et al.* (2019) Combined network analysis and machine learning allows the prediction of metabolic pathways from tomato metabolomics data. *Commun. Biol.* 2, 214
90. Mariette, J. and Villa-Vialaneix, N. (2018) Unsupervised multiple kernel learning for heterogeneous data integration. *Bioinformatics* 34, 1009–1015
91. Knoch, D. *et al.* (2021) Multi-omics-based prediction of hybrid performance in canola. *Theor. Appl. Genet.* 134, 1147–1165
92. Zhu, X.G. *et al.* (2016) Plants in silico: why, why now and what?-an integrative platform for plant systems biology research. *Plant Cell Environ.* 39, 1049–1057
93. Topfer, N. *et al.* (2018) Integration of plant metabolomics data with metabolic networks: progresses and challenges. In *Plant Metabolomics* (Antonio, C., ed.), pp. 297–310, Springer, Berlin Heidelberg
94. Seaver, S.M.D. *et al.* (2018) PlantSEED enables automated annotation and reconstruction of plant primary metabolism with improved compartmentalization and comparative consistency. *Plant J.* 95, 1102–1113
95. Heirendt, L. *et al.* (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* 14, 639–702
96. Cottret, L. *et al.* (2018) MetExplore: collaborative edition and exploration of metabolic networks. *Nucleic Acids Res.* 46, W495–W502
97. Rowe, E. *et al.* (2018) Escher-FBA: a web application for interactive flux balance analysis. *BMC Syst. Biol.* 12, 84
98. Zhou, G. *et al.* (2020) Network-based approaches for multi-omics integration. In *Computational Methods and Data Analysis for Metabolomics* (Li, S., ed.), pp. 469–487, Springer, Berlin Heidelberg
99. Jamil, I.N. *et al.* (2020) Systematic multi-omics integration (MOI) approach in plant systems biology. *Front. Plant Sci.* 11, 944
100. Perez De Souza, L. *et al.* (2020) Network-based strategies in metabolomics data analysis and interpretation: from molecular networking to biological interpretation. *Expert Rev. Proteomics* 17, 243–255
101. Cho, K. *et al.* (2016) Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. *J. Exp. Bot.* 67, 1519–1533
102. Lin, W. *et al.* (2020) Combined analysis of the metabolome and transcriptome identified candidate genes involved in phenolic acid biosynthesis in the leaves of *Cyclocarya paliurus*. *Int. J. Mol. Sci.* 21, 1337
103. Szymański, J. *et al.* (2020) Analysis of wild tomato introgression lines elucidates the genetic basis of transcriptome and metabolome variation underlying fruit traits and pathogen response. *Nat. Genet.* 52, 1111–1121
104. Siddiqui, J.K. *et al.* (2018) IntLIM: integration using linear models of metabolomics and gene expression data. *BMC Bioinforma.* 19, 81
105. Lê Cao, K.A. *et al.* (2009) IntegrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* 25, 2855–2856
106. Rohart, F. *et al.* (2017) mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13, e1005752
107. Feussner, K. and Feussner, I. (2019) Comprehensive LC-MS-based metabolite fingerprinting approach for plant and fungal-derived samples. In *High-Throughput Metabolomics. Methods in Molecular Biology* (vol 1978) (D'Alessandro, A., ed.), pp. 167–185, Humana New York
108. Villa-Vialaneix, N. (2017) Stochastic self-organizing map variants with the R package SOMbrero. In *2017 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM)*, pp. 1–7

109. Wenke, K. *et al.* (2019) Volatiles of rhizobacteria *Serratia* and *Stenotrophomonas* alter growth and metabolite composition of *Arabidopsis thaliana*. *Plant Biol. J.* 21, 109–119
110. Meng, C. *et al.* (2014) A multivariate approach to the integration of multi-omics datasets. *BMC Bioinforma.* 15, 162
111. Uppal, K. *et al.* (2018) XMWAS: a data-driven integration and differential network analysis tool. *Bioinformatics* 34, 701–702
112. Tran, V.D.T. *et al.* (2019) Condition-specific series of metabolic sub-networks and its application for gene set enrichment analysis. *Bioinformatics* 35, 2258–2266
113. Canzler, S. and Hackermüller, J. (2020) multiGSEA: a GSEA-based pathway enrichment analysis for multi-omics data. *BMC Bioinforma.* 21, 561
114. Picart-Armada, S. *et al.* (2018) FELLA: an R package to enrich metabolomics data. *BMC Bioinforma.* 19, 538
115. McIntyre, L.M. *et al.* (2020) GAIT-GM: galaxy tools for modeling metabolite changes as a function of gene expression. *bioRxiv* Published online December 27, 2020. https://doi.org/10.1101/2020.12.25.424407
116. Pang, Z. *et al.* (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* 49, W388–W396
117. Schwacke, R. *et al.* (2019) MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant* 12, 879–892
118. Hernández-De-Diego, R. *et al.* (2018) PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Res.* 46, W503–W509
119. Zhou, G. *et al.* (2021) OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res.* 49, W476–W482
120. Zhou, G. and Xia, J. (2018) OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic Acids Res.* 46, W514–W522
121. Zoppi, J. *et al.* (2021) MiBiOmics: an interactive web application for multi-omics data exploration and integration. *BMC Bioinforma.* 22, 6
122. Tal, O. *et al.* (2020) Netmet: a network-based tool for predicting metabolic capacities of microbial species and their interactions. *Microorganisms* 8, 1–11
123. Hinshaw, S.J. *et al.* (2018) MetaBridge: enabling network-based integrative analysis via direct protein interactors of metabolites. *Bioinformatics* 34, 3225–3227
124. Blimkie, T. *et al.* (2020) MetaBridge: an integrative multi-omics tool for metabolite- enzyme mapping. *Curr. Protoc. Bioinforma.* 70, 1–22
125. Mohamed, A. *et al.* (2014) NetPathMiner: R/Bioconductor package for network path mining through gene expression. *Bioinformatics* 30, 3139–3141
126. Harbig, T. *et al.* (2015) OmicsTIDE: interactive exploration of trends in multi-omics data. *bioRxiv* Published online February 2, 2021. https://doi.org/10.1101/2021.02.01.428836
127. Hernandez-Ferrer, C. *et al.* (2017) MultiDataSet: an R package for encapsulating multiple data sets with application to omic data integration. *BMC Bioinforma.* 18, 1–7