The quest for tolerant varieties: the importance of integrating "omics" techniques to phenotyping

Michel Zivy [1], Stefanie Wienkoop [2], Jenny Renaut [3], Carla Pinheiro [4,5], Estelle Goulas [6] and Sebastien Carpentier [7,8]*

[1] Department Génétique Quantitative et Évolution, Le Moulon INRA, CNRS, AgroParisTech, Plateforme PAPPSO, Université Paris-Sud, Gif-sur-Yvette, France, [2] Department of Ecogenomics and Systems Biology, University of Vienna, Vienna, Austria, [3] Department of Environmental Research and Innovation, Luxembourg Institute of Science and Technology, Belvaux, Luxembourg, [4] Instituto de Tecnologia Química e Biológica, New University of Lisbon, Oeiras, Portugal, [5] Faculdade de Ciências e Tecnologia, New University of Lisbon, Caparica, Portugal, [6] Department of Sciences et Technologies, CNRS/Université Lille, Villeneuve d'Ascq, France, [7] Department of Biosystems, University of Leuven, Leuven, Belgium, [8] SYBIOMA, University of Leuven, Leuven, Belgium

The primary objective of crop breeding is to improve yield and/or harvest quality while minimizing inputs. Global climate change and the increase in world population are significant challenges for agriculture and call for further improvements to crops and the development of new tools for research. Significant progress has been made in the molecular and genetic analysis of model plants. However, is science generating false expectations? Are 'omic techniques generating valuable information that can be translated into the field? The exploration of crop biodiversity and the correlation of cellular responses to stress tolerance at the plant level is currently a challenge. This viewpoint reviews concisely the problems one encounters when working on a crop and provides an outline of possible workflows when initiating cellular phenotyping via "-omic" techniques (transcriptomics, proteomics, metabolomics).

Keywords: proteomics, data integration and computational methods, phenotype, omics-technologies, crop improvement

Introduction

The need for higher yields with lower inputs is widely recognized as necessary to meet the challenge of feeding 9 billion people in 2050 (Godfray et al., 2010). Agricultural management is being challenged by erratic climates and the occurrence of extreme stress events that have the potential to destroy crop production in many geographical regions. Stress is complex and involves timing, duration and severity (Blum, 2014). Moreover, plant stress in agriculture is a phenomenon that is correlated with the genotype, environment and management (G × E × M). Breeding toward stress tolerance is limited in many crops and the current commercially grown varieties have mainly been selected for production and excellent post-harvest qualities, with less attention to other features (e.g., drought tolerance, nutrient use efficiency, durable pest and disease resistance, environmental repercussions etc.). The effects of these factors have been/are mitigated by the use of treatments such as irrigation, pesticides and fertilizers. These management processes have reduced the sense of urgency and resulted in the use of existing plant genetic resources in overcoming crop limitations. Consequently, several ancient varieties and landraces or even wild crop relatives containing useful sources of resistance or tolerance are now underutilized. However, the use of pesticides, fertilizer and water must be reduced and agriculture must become more sustainable. Recently, many governments have

commenced initiatives to promote plant germplasm collections which increase the range of material that can be explored in search of genotypes less affected by stress[1,2]. Reliable identification of tolerant varieties and the understanding of their genetic diversity are urgently needed. The knowledge gap is a strong propeller for the generation of biological knowledge (both fundamental and applied) and provides the plant biology community with the opportunity to establish different experimental models (genomics, transcriptomics, proteomics, metabolomics, phenomics) for different crops. Phenotyping is an emerging field that characterizes plant behavior and quantify features, such as growth and yield, in a way that allows linking to genetic control. However, the evaluation of genetic biodiversity is a research bottleneck and there is still a significant gap between the lab and the field. Is science generating false expectations? Are 'omic techniques, such as proteomics and metabolomics, generating valuable information that can be translated into practice? Is proteomics better than metabolomics, transcriptomics or genomics? Recently, a European network was created to help tackle this issue and develop new workflows to integrate the different 'omic techniques: COST action FA1306 "The quest for tolerant varieties—Phenotyping at plant and cellular level[3]." The aim of the Action is the improvement and exchange of scientific knowledge in plant phenotyping through the creation of a network between European interdisciplinary scientists and to use this network to: map valuable gene bank collections and breeding programs in Europe, train breeders and physiologists in screening techniques and data interpretation, get insight into the genetic basis of tolerance, to characterize current biodiversity and rank it according to tolerance levels and to apply the knowledge for agricultural management. The Action started in May 2014 and 28 countries have currently joined. This viewpoint embodies the vision of this COST action and describes concisely the problems one encounters when phenotyping the diversity found in crops. It specifically provides an outline of the problems encountered when initiating "cellular phenotyping" through 'omics techniques, highlighting proteomics.

Understanding Gene function

Understanding gene function can be approached via several techniques: genomics, transcriptomics (messenger, structural and regulatory RNA's), proteomics (proteins and their putative post-translational modifications (PTM) and peptides) and metabolomics (primary and secondary metabolites). In prokaryotes, gene finding is essentially a matter of identifying open reading frames. As genomes get larger, it becomes increasingly complicated. Several sophisticated software algorithms have been designed to handle gene prediction in eukaryotic genomes. Despite considerable progress, gene prediction entirely based on DNA analysis is cumbersome and requires support from "functional genomics," i.e., transcriptomics, proteomics, and metabolomics. Indeed, genomics focuses on the static aspects of genome information. Gene prediction and annotation in a reference

[1]http://www.croptrust.org/
[2]http://www.cgiar.org/
[3]http://costfa1306.eu/

variety is the initial step for every crop, but this is not sufficient to get complete insight into the phenotypic plasticity and the agricultural potential of the biodiversity. Functional genomics deals with dynamic aspects, reflecting environmental adaptations and allows the description of gene functions as well as the interactions between gene products that may provide a view of the agricultural potential of a variety/genotype.

Transcriptomics

Probably the easiest way to study changes on a genome-wide scale is through transcriptomics. The structure of RNA is homogenous and relatively simple and therefore the analysis is the most straightforward when compared to protein and metabolite analyses.

Serial analysis of gene expression (SAGE), developed by Velculescu et al. (1995), is based on the generation of 15 bp tags from a defined position in each transcript, which are then concatenated, cloned into a plasmid vector and ultimately sequenced (Velculescu et al., 1995). Massively parallel signature sequencing is a more advanced technique based on sequencing of tags. It generates 17 bp tags that are sequenced using a fluorescence-based signature sequencing method on microbeads and was applied on multiple model organisms (Brenner et al., 2000; Reinartz et al., 2002). To analyze the abundance of a transcript, one simply calculates the number of times that a certain tag was found. Though no sequence data needs to be identified *a priori*, the tag needs to be identified as belonging to a gene to convey its biological meaning and this step can be difficult in plants with limited genetic resources. DNA sequences are not as well conserved as amino acid sequences and therefore a cross-species identification based on a short tag is problematic. Matsumura et al. (2003) developed superSAGE in rice, which utilizes longer tags (26 bp). However, the generation of longer tags still resulted in the SAGE-approach for un-sequenced non-model crops challenging, as illustrated for banana (Coemans et al., 2005; Carpentier et al., 2008a). At the same time microarrays, which were significantly cheaper and was a more high-throughput technology, were also developed for transcriptomic studies. Microarrays use known probes that will hybridize with the labeled sample and based on the intensity of these dyes, transcript levels are estimated. This however, implies that sequence information exists before generation of the microarray and this is a serious limitation when applied to non-model crops. The limited sequence availability in non-models can be overcome by the use of microarrays of closely related species or by the generation of a species-specific microarray based on known expressed sequence tag (EST) data for instance, but these analyses will be less informative (Davey et al., 2009; Pariset et al., 2009). Microarray analysis is hampered by a high background noise due to cross-hybridization as well as saturation of signals. Microarrays therefore have a limited sensitivity and dynamic range. Furthermore, microarrays are closed platforms as unknown transcripts cannot be detected. An alternative for the standard gene expression microarray is the tiling microarray. These are high-density arrays composed of oligonucleotide probes that span the entire genome of an organism (Yazaki et al., 2007).

Whole-genome tiling arrays may provide part of the solution toward the detection of new gene transcripts in a sequenced organism but are more expensive and still suffer from the general drawbacks of a microarray approach (Valdés et al., 2013). With the availability of next generation sequencing (NGS) technologies, the possibility to directly sequence mRNA at relatively reduced costs became available. This technique, termed RNA-seq, has clear advantages over the other transcriptomics methods: a higher sensitivity and dynamic range can be achieved (Wang et al., 2009) and no previous sequence knowledge is *per se* required. Reads can be mapped to a known reference genome or *de novo* assembled. *De novo* assembly of reads into contigs increases the use of this technique for crops whose genome has not been sequenced, as was demonstrated for wheat, agave and horse gram (Bhardwaj et al., 2013; Gross et al., 2013; Oono et al., 2013). However, a reference genome is highly recommended to assure the correct assembly of the reads and to deal with paralogs and allelic variants. Moreover, most read-mapping software has been written to analyze diploid genomes (Langmead and Salzberg, 2012) and is unsuited for polyploid organisms (Page et al., 2013a). Read mapping is a fundamental part of next-generation genomic research but is complicated by genome duplication in many plants. When a reference genome is already available, RNA-seq can provide additional information necessary to identify previously unknown gene coding sequences. Categorizing DNA sequence reads into their respective genomes enables current methods to analyze polyploid genomes as if they were diploid. Page et al. (2013a,b) developed software for SNP detection in cotton, which is an allotetraploid. Using SNP-tolerant mapping, the software uses the SNPs between genomes to categorize reads according to their respective genomes. Furthermore RNA-seq data can also be used to improve existing annotations both in identifying actual intron-exon structures as well as in identifying different splice variants as was shown in maize (Kakumanu et al., 2012) or identifying homeologs. Genome-wide quantification of homeolog expression ratios was technically hindered because of the high homology between homeologous gene pairs. Additionally, in contrast to the high background noise caused by cross-hybridization in microarrays, most RNA-seq reads can be unambiguously mapped to a region of the reference genome. This makes RNA-seq in combination with reference genomes an excellent tool to differentiate between isoforms of a gene family, which are a widespread phenomenon in complex crop genomes. On the other hand, the alignment of short sequence reads that are shared between several loci and therefore align to several locations on the genome is still complicated. One solution is to assign these reads proportionally to the number of unique splice reads at these loci (Mortazavi et al., 2008). Moreover, aside from being relatively unbiased toward previous sequence knowledge, RNA-seq is also more sensitive. This sensitivity comes at a price. To detect rare transcripts, coverage and therefore sequencing depth is required, which increases the sequencing cost. Lastly, the dynamic range of RNA-seq is also substantially higher, at about five orders of magnitude compared to several hundred-fold for microarrays (Wang et al., 2009; Zhao et al., 2014).

With the introduction of NGS, RNA-seq appears to be the transcriptomic tool for the future, especially in crops. At the moment, the costs associated with RNA-seq prevent large scale analysis of many varieties in different conditions and multiple biological replicates. However, as the NGS technique keeps evolving, costs are likely to drop and may no longer be a limiting factor in the future. As more and more genomes are sequenced, alignments to reference genomes should become standard practice, which will also significantly reduce the analysis time required for *de novo* assembly.

Proteomics

In contrast to genomics and transcriptomics, proteomics is often regarded as a slow and cumbersome art. The discovery of soft ionization techniques for mass spectrometry (MS) by Nobel Prize winners Fenn and Tanaka, the coupling of MS to liquid chromatography and the genomic and computational advances, have made the high throughput large scale analysis of proteins feasible (Karas and Hillenkamp, 1988; Fenn et al., 1989; Henzel et al., 1993; McCormack et al., 1997). Thus, after a significant lag phase, high throughput proteomics has become an important research tool for model organisms and is currently finding its way to crop species.

Two approaches are generally distinguished in the field of proteome analysis: a protein based approach (in general, referred to as gel based) and a peptide based approach (in general referred to as gel free or shotgun). In the gel based approach, proteins are separated and quantified via gel electrophoresis. The proteins of interest are then picked from the gel, digested and the resulting peptides identified via MS by comparing experimental *versus* theoretical masses present in various databases. This technique has the advantage that protein separation and analysis via (two-dimensional) electrophoresis prior to MS analysis ensures physical connectivity between the peptides and the protein and significantly reduces complexity (Carpentier et al., 2008b). Currently, it is still the most widely used approach in crop proteomics. Unfortunately the technique has some major drawbacks, i.e., it has a very poor performance when analyzing hydrophobic and basic proteins and can be quite limited with respect to throughput.

In the gel free approach, protein digestion precedes the separation and quantification of peptides. Gel free differential proteomics provides a broader coverage of the proteome and also enables the identification of membrane proteins. However, a major disadvantage of this approach lies in the disconnection between the protein and its peptides (Carpentier and America, 2014). In general, most approaches use a bottom-up strategy where proteins are first digested with a proteolytic enzyme. Yates and colleagues were one of the early pioneers to explore the use of liquid chromatography coupled to electrospray ionization tandem mass spectrometry (LC/MS/MS) and realize the potential of automated high throughput proteomics (McCormack et al., 1997; Ducret et al., 1998; Link et al., 1999). However, proteolytic digests of a higher eukaryotic proteomes, like crops, exceed the analytical capacity of most MS. During recent years, MS have been developed with high mass accuracies, resolving power, sensitivity, scan speed, reproducibility and lower detection limits (Domon and Aebersold, 2006; Mann and Kelleher, 2008). For

example, the use of Fourier transform ion cyclotron resonance based spectrometry (Yang and Yen, 2002), hybrid Linear Trap Quadrupole-Orbitrap devices (Makarov et al., 2006; Olsen et al., 2009), high energy C-trap dissociation (Olsen et al., 2007), parallel reaction monitoring (Peterson et al., 2012), the coupling of a quadrupole mass filter to an Orbitrap analyser (Michalski et al., 2011; Kelstrup et al., 2012), Ultra Performance Liquid Chromatography (UPLC) combined with moist static energy (MSE; Plumb et al., 2006), combining quadrupole, Orbitrap and ion trap mass analysis (Lebedev et al., 2014), and hybrid quadrupole time-of-flight MS (Andrews et al., 2011), have all contributed to improvements in proteomic experiments, and in particular toward better peptide identifications and quantification. Despite the development of new MS, a protein sample from a crop species is still challenging to analyze and contains several thousand proteins. This might lead to both identification and quantification problems, especially in the case of crops with complex polyploid genomes and large protein families. Peptides shared between several proteins do not contribute to the conclusive identification of a particular protein. This is the so-called protein inference problem (Nesvizhskii and Aebersold, 2005). Unique peptides need to be measured and identified for final protein identification and quantification. So a gel free approach is only applicable for crops, in practice, once a reference genome is available or when substantial EST libraries become available (Vertommen et al., 2011b). Typically, a gel free analysis starts with an MS survey scan where peptide precursor masses are measured, followed by an MS/MS scan for fragmentation of the selected precursor ion. This is called data-dependent acquisition (DDA). The serial nature of the MS and MS/MS cycles and the complexity of the proteome in crops remain a challenge. There is a bias toward the more abundant peptides and no MS scan can be obtained while fragmentation is being performed in the MS/MS scan in most current MS. Moreover, co-eluting peaks can lead to chimeric spectra, reduced reproducibility and loss of information about less abundant peptides. To increase the chance of identifying specific tryptic peptides, it is important to ensure a good peptide separation and to keep the mixture of co-ionizing peptides as simple as possible even in fast modern MS. Vertommen et al. (2011a) proposed a workflow for banana samples where this was achieved by using a two-dimensional RP-RP chromatography system coupled to a high accuracy MS. To identify the peptides in a gel free approach for a crop, it is important to build a species specific in-house database and to search this database in consecutive steps: a non-error tolerant manner, subsequently an error tolerant and finally *de novo* sequencing. This *de novo* approach is a crucial step, since it allows the identification of variety specific peptides via a homology search of sequences instead of a search based on m/z values (Vertommen et al., 2011a).

To identify and quantify peptides in a rapid, consistent, reproducible, accurate and sensitive way, data independent acquisition (DIA) protocols have been developed for label-free shotgun proteomics as an alternative to DDA (Plumb et al., 2006). To increase the identification rate of label free DIA experiments for samples from the apple variety Braeburn, a new workflow was developed by Buts et al. (2014) where a DDA database was

constructed and linked to the DIA data. A ten-fold increase in peptides was identified from a single DIA run and proteins correlated to the storage quality of apples were found.

While in sequenced crops, two-dimensional electrophoresis (2DE) may currently no longer be the tool of choice for high-throughput differential proteomics, it is still very effective in identifying and quantifying protein species as a result of genetic variations, alternative splicing and/or PTM. As an example, by using combined 2DE and 2D DIGE with *de novo* MS/MS sequencing, Carpentier et al. (2011) were able to identify inter- and intra-cultivar protein polymorphisms in banana correlated to drought tolerance. Using an 2D-DIGE LC MS/MS approach Vanhove et al. (2015) were able to characterize the complex protein family of HSP70s and identify an osmotic stress specific isoform. Likewise, the molecular mechanisms for rapid metabolic responses to stress remain largely unknown and to fill this gap, the role of PTMs needs to be investigated. As an example, in response to cold stress, which involves quick adjustments to the photosynthetic machinery, many cold-acclimation-related proteins are putatively regulated by PTMs, as has been recently highlighted in pea by using 2D-DIGE analysis (Grimaud et al., 2013).

Metabolomics

In the 'omics field, metabolomics generates large datasets for the identification and quantification of small molecules. Usually, the approach is undertaken for the high throughput detection of secondary (flavonoids, sugar-phosphates, phytohormones, phytoalexins, etc.) and primary metabolites (sugars, organic- and amino acids, etc.). Complementary MS-based LC and GC approaches are required to adequately profile this metabolic diversity (Scherling et al., 2010). Although it is possible to gather tens of thousands of metabolic variables, their accurate identification remains the major bottleneck in metabolomics studies to date. Like for proteins, due to the large dynamic range and the difference in physico-chemical properties, detecting the entire "metabolome" is not possible. In contrast to proteomics, for metabolomics analyses, functional identification is not dependent on the availability of genome sequence information. However, the availability of purified standard substances and/or spectral libraries/databases is necessary. Of the various MS methods, two profiling strategies can be typically distinguished: a targeted and a non-targeted approach (Wienkoop et al., 2008, 2010). The non-targeted technique allows for the quantitative evaluation of unknown, yet unidentified, variables. Nevertheless, metabolomics databases and the availability of MS-derived metabolite fragment information are increasing. Furthermore, novel strategies for pathway and structural assignments of untargeted high-throughput metabolomics data are being developed. For example, an algorithm termed mzGroupAnalyzer was developed to investigate metabolite transformations caused by biochemical or chemical modifications, and the approach led to the identification of novel molecule structures (Doerfler et al., 2014). Specifically, it resulted in the detection of 15 unknown, putative cold and light stress regulated metabolites of the flavonoid-pathway in *Arabidopsis thaliana*. Metabolomics

is also playing an increased role in stress marker detection and contributing to improved stress tolerance in crops, as previously reviewed (Hirayama and Shinozaki, 2010; Weckwerth, 2011; Martinez-Gomez et al., 2012; Rasmussen et al., 2012). Metabolomics has also contributed to the detection of putative marker(s) induced by the pathogen *Rhizoctonia solani* in potatoes (Aliferis and Jabaji, 2012) and soybean (Aliferis et al., 2014) as well as bacterial blight-resistance in rice (Wu et al., 2012).

QTL Analysis

Most traits of interest for crop breeding are controlled by multiple quantitative trait loci (QTL), and the major objective of using 'omics in this context is the characterization of these QTLs. Analyzing the genetic variations of transcripts, proteins and metabolites at a large-scale allows the search for causal relationships between molecular and phenotypic variations with a global approach and without *a priori* knowledge. The study of natural genetic variations is not only interesting for breeding purposes, but also to decipher the biological processes involved in the genotype to phenotype relationship. Indeed, it is not uncommon that loss of function mutations have no clear phenotype, and the analysis of small disturbances caused by QTLs may allow a better understanding of metabolic pathways and regulatory networks involved in the variations of the phenotypic trait (Sulpice et al., 2010).

Genotyping is the information on which all breeding programs are based to map the QTLs, to measure kinship between genotypes or populations, to analyze changes in allele frequency during selection, etc. In recent years, considerable advances have been made in genotyping and hundreds of thousands of SNP markers (single nucleotide polymorphism) are available today, at relatively low cost for crops whose genome has been sequenced. This information is subsequently used to identify candidate genes or proteins. Overall, the strategy consists of mapping QTLs of transcript expression (eQTLs) or abundance proteins (PQLs or pQTLs), and looking for co-localization between them and QTLs of the traits of interest. The underlying idea is that QTL/(eQTL or PQL) co-localizations may be due to a single polymorphism that causes quantitative or qualitative (amino acid polymorphism) variation of the transcript and/or protein, that in turn would be the cause of the variation of the trait of interest. When these co-locations also co-localize with the gene encoding the transcript or protein, then the causal polymorphism is likely located within the gene, including the promoter region (*cis*-QTLs). This gives breeders the opportunity to select the favorable allele itself, which is far more efficient than using QTL flanking markers.

When QTL/(eQTL or PQL) co-localizations are found outside the region of the gene (*trans*-QTLs), the QTL can be any sequence that influences the protein or transcript abundance. In this case, it is not precisely identified; nevertheless the co-localization indicates a possible involvement of the gene/protein in the genetic variation of the phenotypic trait. This information is interesting both for the breeder and for understanding the mechanisms involved in the variation of the trait. As co-localizations can also be found due to chance, candidate genes or proteins are selected according to *a priori* knowledge on gene function or regulation

that support the involvement of the gene in the studied trait. After validation, the information brought by these co-localizations can be used in breeding programs by selecting alleles that influence the abundance of the transcript or protein, or by transgenesis to over- or under-express these genes.

As can be deduced from the discussions above, each 'omics approach offers advantages and drawbacks. Transcripts are direct gene products, and thus the link between genomic information and expression data is quite simple. On the other hand, the link between transcript abundance and the phenotype is loose, because of the multiple steps from transcripts to protein, from protein abundance to activity and metabolite concentrations, and from metabolites to cellular, physiological and plant phenotypes. Quite a number of eQTL analyses have been performed in plants, including in crops. For example, Li et al. (2013b) mapped a total of 30,774 eQTLs for 22,242 genes by RNA sequencing a maize population of intermated recombinant lines. Thirty-seven percent were *cis*-eQTLs, while the other 63% were *trans*-eQTLs. The latter were often grouped in hotspots. In many of these hotspots, the effect of alleles from the same parent affected gene expression in the same direction. The genes controlled by hotspots were often enriched in a particular functional category. The last two observations suggested that hotspots contain upstream regulators controlling cellular processes. *Cis*- and *trans*-eQTLs have been observed in various proportions according to the species and the study, and hotspots showing some kind of functional specialization are also often observed (Kliebenstein, 2009). Several eQTL studies have been performed with the aim of identifying candidate genes. For example, Li et al. (2013a) identified in 2013 74 loci highly associated with maize oil concentrations or composition by genome-wide association study (GWAS). The expression of all 41 genes at these loci was controlled by *cis*-eQTLs and for 32 it was correlated to the targeted or related traits. Sequencing five of them in a collection of genotypes allowed the identification of polymorphisms in their UTR or promoter regions, which was likely the cause of the variation of their expression and for the variation of kernel oil content and composition. Most eQTL studies have been carried out by analyzing segregating populations, but when the objective was to target a particular QTL, near isogenic lines (NILs) were also used (Bolon et al., 2010; Kugler et al., 2013). Bulk segregant analysis, where different genotypes are mixed according to the genotype in the QTL region, were also used (Chen et al., 2011).

Metabolite QTLs (mQTLs) have also been mapped both in crops and in model organisms (Kliebenstein, 2009). Metabolites are products of biochemical reactions catalyzed by proteins (enzymes), and as such they are closer to the phenotype than transcripts. On the other hand, the relation to genomic sequences is tenuous, because metabolite amounts may depend on many other biochemical reactions than those directly involved in their synthesis or degradation: many enzymes can be directly or indirectly responsible for the variation of a single metabolite. Many of the mQTL studies or analyses of the natural genetic variability of metabolites were performed with the goal of linking metabolite variations to other biochemical, physiological or phenotypic traits. Kerwin et al. (2011) compared *Arabidopsis* mQTLs to eQTLs of genes and showed a single peak of expression

per day. The results, combined with the analysis of mutants, allowed them to conclude that variations in glucosinolate content can influence the internal circadian clock. Sulpice et al. (2013) analyzed the relationships between metabolism and biomass in a panel of 97 *Arabidopsis* accessions and observed that correlation-based networks were very specific to growth conditions. Desnoues et al. (2014) analyzed the correlations between sugars and enzymatic activities in peach fruits in a progeny of 106 genotypes. Interestingly the variations in sugar content was only poorly explained by the variations of enzymatic capacities of the enzymes directly involved in their synthesis or degradation, suggesting that their variations could be related to changes in other components of sugar metabolism.

In several mQTL studies, parallel mapping of eQTL or analysis of gene expression were performed to identify candidate genes. For example, Brotman et al. (2011) identified a gene encoding a putative fumarase in the region of a fumarate mQTL in *Arabidopsis*. They showed that the fumarate content was greatly reduced in mutants and that the expression of the candidate gene was 16 times more highly expressed in the parent that showed the highest level of fumarate. These results suggest that there is a causal relationship between the genetic variation of the expression of this gene at this locus and the natural variation in fumarate content. A similar strategy was followed by Zorrilla-Fontanesi et al. (2012) to identify a candidate gene for the production of mesifurane by the strawberry fruit.

Some proteins are in-between transcripts and metabolites in the genome-to-phenotype relationship, since they are effectors of biological processes and other proteins are real end products influencing the phenotype directly. In proteomics studies, the link between genes and proteins can be ambiguous in particular because of peptides shared between members of gene families. On the other hand, as proteins are the results of transcription, transcript turnover and translatability, post-translational events and protein turnover, their amount represents the integration of many processes that lead to the cellular and *in fine* plant phenotype. Consequently, correlations between protein and transcripts amounts are relatively low (Gygi et al., 1999; Carpentier et al., 2008a; Schwanhausser et al., 2011) which supports the necessity to analyze proteome variations. Wilhelm et al. (2014) have calculated the protein/mRNA ratio for many protein and transcripts in a certain tissues and claim that knowing the translation rate constant, it becomes possible to predict protein abundances with good accuracy from the measured mRNA abundance. This is because the translation rate constant is a dominant factor in determining protein abundance (Wilhelm et al., 2014). Some further confirmatory experiments will need to be undertaken to confirm whether this hypothesis is adequately supported.

The first PQLs were mapped in 1994 (Damerval et al., 1994) and the candidate protein strategy was developed in 1999 (de Vienne et al., 1999). It allowed the identification of the ZmASR1 candidate protein for tolerance to drought, whose effect on yield was then confirmed by over-expression (Virlouvet et al., 2011). The ASR protein was also found to be relevant in other species, e.g., tomato, grape, lily and banana (Maskin et al., 2001; Cakir et al., 2003; Wang et al., 2005; Henry et al., 2011). Segregating populations

were also used to map seed and leaf protein PQLs in pea and wheat (Amiour et al., 2003; Bourgeois et al., 2011; Legrand et al., 2013). More often, proteomics has been used to search for candidate proteins in relation to a particular QTL, by using NILs (Hajduch et al., 2007; Torabi et al., 2009; Bernardo et al., 2012; Gunnaiah et al., 2012; Lesage et al., 2012; Li et al., 2014). The analysis of bulked genotypes grouped according to their phenotype can also be useful. It has helped in identifying candidate genes in potato as a complement to a study based on association genetics (Fischer et al., 2013).

To our knowledge, no study on the genetic diversity of PTMs has been performed, although it is well known that protein phosphorylation can play an important role in plant response to biotic and abiotic stresses (Bonhomme et al., 2012; Rampitsch and Bykova, 2012). Although the quantification of PTMs is more difficult than the analysis of protein abundance, and that even identifying a correlation between a modifying enzyme and its target is challenging, it is likely that "candidate PTMs" would be of great interest for QTL characterizations. The analysis of large numbers of genotypes is necessary for PQL mapping or genome wise genetics association studies. The constant progression of the performances of MS-based quantitative proteomics will allow those types of analysis to be performed in the near future, to achieve in plants what has already begun in yeast (Foss et al., 2007; Picotti et al., 2013; Skelly et al., 2013) and in humans (Wu et al., 2013).

Data Integration

Prior to the development of high throughput methods for extensive metabolome, proteome, transcriptome, genome and recently phenome research, scientists dreamed about the integration of different datasets to gain a deeper insight. Data integration, as a tool to enable in-depth insights into processes, can be performed at two levels. On the one hand, one can integrate homogeneous data derived from measurements on the same entity made in the same experimental set-up. On the other hand, meta-analysis can be applied to integrate heterogeneous data derived from different experiments (Dupae et al., 2014). The latter approach poses a challenge as the integration of data performed using different standards and methods are difficult. Thus, we will focus on the integration of homogenous data derived from the same experimental set-up that has been used to get a clearer view on plant (dys)function. Over the past few years, questions on how to unravel the existing links between the metabolome, proteome, transcriptome, genome and the phenome from a particular entity have been discussed. As well as how this knowledge can provide better insights in mechanisms like abiotic stress tolerance or biotic stress resistance. Statistical multivariate methods have been refined to uncover those links and are now used to integrate two or even more datasets based on predefined models of one-by-one dataset relationships (Le Cao et al., 2011; Gunther et al., 2014; Tenenhaus et al., 2014). This not only makes it possible to integrate multiple 'omic levels, but even enables scientists to "supervise" and direct such analyses toward meaningful relationships between the 'omics datasets (Gunther et al., 2014).

A major advantage in integrating metabolomics with proteomics is gaining spatial and temporal information on the end products upon environmental constrains. Since specific protein isoforms can target a specific protein to particular tissues and/or compartment, the use of integrative subcellular fractionation and localization strategies will allow the detection of dynamic distributions within the cell. The temporal plasticity of metabolism constitutes various phases of adjustment. In order to capture the interaction between the metabolome and the proteome, it is necessary to investigate the system along a period of time. Correlative network- as well as Granger-Causality (time-series correlation) analyses have been demonstrated to be effective tools in obtaining information on pathway interplay and reprogramming cues (Doerfler et al., 2014). Linking 'omics data with such mathematical approaches facilitates the interpretation of time dependent chronological processes and the identification of variables being controlled by time-lagged values of other variables.

The next step is to link the transcriptome/proteome/metabolome to plant performance (phenome), i.e., linking cellular phenotyping to plant phenotyping. The detailed characterisation (physiological, cytological, biochemical, gene expression, protein and metabolite profiles) of different plant species represent a real and difficult challenge. The necessary resources and or knowledge are most likely not present in one research group or country. This is where a European network such as the COST project can make the difference. An integration of these data into regulation/interaction networks will represent new and important advances for understanding plant responses. A multivariate analysis approach becomes of interest because the algorithms are designed to explain as much of the variability at the plant level as well as the variability at the cellular level. This method is especially useful to assess polygenic traits such as drought tolerance, because one can specifically look for cellular and plant level variations that explain the variation in genotypes and treatments used. In this way candidate genes related to important tolerance mechanisms at the plant level can be retained for further analysis, a procedure called feature selection. Retained candidate genes can further be related to the kyoto encyclopedia of genes and genomes (KEGG), gene ontology (GO), protein domain, protein family, protein function, and more categorical databases, as a form of heterogeneous data integration (Gomez-Cabrero et al., 2014). Identification of key genes, pathways, regulation networks of metabolism and stress responses, in association with physiological stress-related data, represent crucial information that have to be integrated and presented in web databases. This way the physiological context of the genes of interest can be compared to the literature to confirm the obtained results.

'Omics and Breeding: From Lab to Field

The genetic gain obtained via breeding programs supports the yield gain of new crop varieties. In this way it is relevant to increase the genetic base available for selection, i.e., the genetic diversity, and also to characterize the molecular and phenotypic consequences of such diversity. Germplasm

characterization through 'omics allows one to perform the molecular characterization of genotypes, providing a list of candidate genes/gene products that are highly valuable for breeding and/or engineering stress-tolerant crops with novel traits (Mir et al., 2012; Pinheiro et al., 2014). The characterization of genetic diversity in germplasm collections is typically performed through DNA based markers. However, the use of non-DNA markers (as transcripts, proteins and metabolites) has the advantage to provide information on the molecular interactions and networks operating in a given genotype; these have the potential to help evaluate the potential of the different genotypes. As indicated above, each technique focuses on a subset of the biological interaction network and each technique has its strong and weak points. At present, blind high throughput proteomics and metabolomics studies are often regarded as being descriptive. This might be due to the fact that not many of the putative markers that have been proposed have been evaluated and transferred to the productive sector. The reason might be that high throughput crop proteomics and metabolomics are still emerging 'omic approaches with limited resources and high associated costs. Due to cost, setup, training, requirement for labs /greenhouses etc, most studies are limited in size and number of biological replicates. In many cases these numbers are far too low. Currently the number of studies in crops exploring diversity via different 'omic techniques is limited. Germplasm
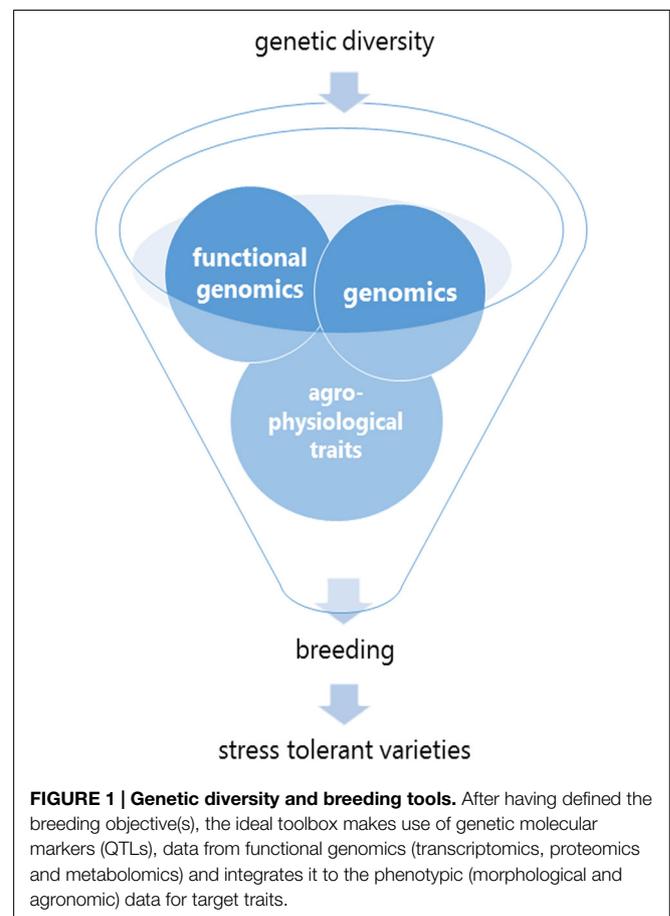
FIGURE 1 | Genetic diversity and breeding tools. After having defined the breeding objective(s), the ideal toolbox makes use of genetic molecular markers (QTLs), data from functional genomics (transcriptomics, proteomics and metabolomics) and integrates it to the phenotypic (morphological and agronomic) data for target traits.

screening and/or discrimination between genotypes of *Phaseolus vulgaris* (Mensack et al., 2010), *Oryza sativa* (Heuberger et al., 2010), *Miscanthus* (Straub et al., 2013), and *Hordeum vulgare* (Heuberger et al., 2014) are some examples. Technologies keep evolving and become more powerful and cheaper. Proteomics and metabolomics are powerful tools when combined with a good experimental design and when the candidates are validated under realistic conditions. It is essential to identify potential traits under lab conditions that are responsible for superior yields under realistic field conditions in different environments.

The current breeder's toolbox makes use of genetic molecular markers, QTLs, gene expression and biochemistry and phenotypic (morphological and agronomic) data for target traits. The use of these tools is dependent on the breeding objective (phenology, yield, yield components, quality, disease, adaptation) and on the crop. While available tools can differ from crop to crop, the crop growing habit (cool vs warm season crops) and the agricultural system also needs to be considered. For each crop, and for each climatic zone, major constraint(s) need to be identified and goals established. The strategy of "one size fits all" is not well suitable, and regional and local goals need to be addressed (climate, soil, social). Another important issue is the availability and the costs of each technology. Langridge and Fleury (2011) present a summary of 'omic resources available for some crops.

In the 21st century, the research and agricultural community face a challenge to deliver new stress tolerant and productive crops. Does science and 'omic techniques generate false expectations? No, 'omics technology is powerful and promising when combined with relevant experimental design and subsequently validated in a realistic environment. The use of the 'omics generated knowledge is a promising tool in the breeders tool box (**Figure 1**). A dialog and consultation between breeders, farmers, researchers from natural and social sciences and politicians is required. This is where the COST project "The quest for tolerant varieties—Phenotyping at plant and cellular level" needs to contribute. Through a strong dialog between stakeholders and their cooperative activity it will be possible to deliver better agricultural products that utilize less inputs, have lower environmental costs and provide higher levels of social well-being.

References

Aliferis, K. A., Faubert, D., and Jabaji, S. (2014). A metabolic profiling strategy for the dissection of plant defense against fungal pathogens. *PLoS ONE* 9:e111930. doi: 10.1371/journal.pone.0111930

Aliferis, K. A., and Jabaji, S. (2012). FT-ICR/MS and GC-EI/MS metabolomics networking unravels global potato sprout's responses to *Rhizoctonia solani* infection. *PLoS ONE* 7:e42576. doi: 10.1371/journal.pone.0042576

Amiour, N., Merlino, M., Leroy, P., and Branlard, G. (2003). Chromosome mapping and identification of amphiphilic proteins of hexaploid wheat kernels. *Theor. Appl. Genet.* 108, 62–72. doi: 10.1007/s00122-003-1411-0

Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., and Muddiman, D. C. (2011). Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* 83, 5442–5446. doi: 10.1021/ac200812d

Bernardo, L., Prinsi, B., Negri, A. S., Cattivelli, L., Espen, L., and Vale, G. (2012). Proteomic characterization of the Rph15 barley resistance gene-mediated defence responses to leaf rust. *BMC Genomics* 13:642. doi: 10.1186/1471-2164-13-642

Bhardwaj, J., Chauhan, R., Swarnkar, M. K., Chahota, R. K., Singh, A. K., Shankar, R., et al. (2013). Comprehensive transcriptomic study on horse gram (*Macrotyloma uniflorum*): *de novo* assembly, functional characterization and comparative analysis in relation to drought stress. *BMC Genomics* 14:647. doi: 10.1186/1471-2164-14-647

Blum, A. (2014). Genomics for drought resistance—getting down to earth. *Funct. Plant Biol.* 41, 1191–1198. doi: 10.1071/FP14018

Bolon, Y. T., Joseph, B., Cannon, S. B., Graham, M. A., Diers, B. W., Farmer, A. D., et al. (2010). Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. *BMC Plant Biol.* 10:41. doi: 10.1186/1471-2229-10-41

Bonhomme, L., Valot, B., Tardieu, F., and Zivy, M. (2012). Phosphoproteome dynamics upon changes in plant water status reveal early events associated with rapid growth adjustment in maize leaves. *Mol. Cell. Proteom.* 11, 957–972. doi: 10.1074/mcp.M111.015867

Bourgeois, M., Jacquin, F., Cassecuelle, F., Savois, V., Belghazi, M., Aubert, G., et al. (2011). A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *Proteomics* 11, 1581–1594. doi: 10.1002/pmic.201000687

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 18, 630–634. doi: 10.1038/76469

Brotman, Y., Riewe, D., Lisec, J., Meyer, R. C., Willmitzer, L., and Altmann, T. (2011). Identification of enzymatic and regulatory genes of plant metabolism through QTL analysis in *Arabidopsis*. *J. Plant Physiol.* 168, 1387–1394. doi: 10.1016/j.jplph.2011.03.008

Buts, K., Michielssens, S., Hertog, M. L., Hayakawa, E., Cordewener, J., America, A. H., et al. (2014). Improving the identification rate of data independent label-free quantitative proteomics experiments on non-model crops: a case study on apple fruit. *J. Proteom.* 105, 31–45. doi: 10.1016/j.jprot.2014.02.015

Cakir, B., Agasse, A., Gaillard, C., Saumonneau, A., Delrot, S., and Atanassova, R. (2003). A Grape ASR protein involved in sugar and abscisic acid signaling. *Plant Cell* 15, 2165–2180. doi: 10.1105/tpc.013854

Carpentier, S. C., and America, T. (2014). Proteome analysis of orphan plant species, fact or fiction? *Methods M-ol. Biol.* 1072, 333–346. doi: 10.1007/978-1-62703-631-3_24

Carpentier, S. C., Coemans, B., Podevin, N., Laukens, K., Witters, E., Matsumura, H., et al. (2008a). Functional genomics in a non-model crop: transcriptomics or proteomics? *Physiol. Plant* 133, 117–130. doi: 10.1111/j.1399-3054.2008.01069.x

Carpentier, S. C., Panis, B., Renaut, J., Samyn, B., Vertommen, A., Vanhove, A. C., et al. (2011). The use of 2D-electrophoresis and *de novo* sequencing to characterize inter- and intra-cultivar protein polymorphisms in an allopolyploid crop. *Phytochemistry* 72, 1243–1250. doi: 10.1016/j.phytochem.2010.10.016

Carpentier, S. C., Panis, B., Vertommen, A., Swennen, R., Sergeant, K., Renaut, J., et al. (2008b). Proteome analysis of non-model plants: a challenging but powerful approach. *Mass Spectrom. Rev.* 27, 354–377. doi: 10.1002/mas.20170

Chen, X., Hedley, P. E., Morris, J., Liu, H., Niks, R. E., and Waugh, R. (2011). Combining genetical genomics and bulked segregant analysis-based differential expression: an approach to gene localization. *Theor. Appl. Genet.* 122, 1375–1383. doi: 10.1007/s00122-011-1538-3

Coemans, B., Matsumura, H., Terauchi, R., Remy, S., Swennen, R., and Sagi, L. (2005). SuperSAGE combined with PCR walking allows global gene expression profiling of banana (*Musa acuminata*), a non-model organism. *Theor. Appl. Genet.* 111, 1118–1126. doi: 10.1007/s00122-005-0039-7

Damerval, C., Maurice, A., Josse, J. M., and De Vienne, D. (1994). Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* 137, 289–301.

Davey, M. W., Graham, N. S., Vanholme, B., Swennen, R., May, S. T., and Keulemans, J. (2009). Heterologous oligonucleotide microarrays for transcriptomics in a non-model species; a proof-of-concept study of drought stress in Musa. *BMC Genomics* 10:436. doi: 10.1186/1471-2164-10-436

de Vienne, D., Leonardi, A., Damerval, C., and Zivy, M. (1999). Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. *J. Exp. Bot.* 50, 303–309. doi: 10.1093/jxb/50.332.303

Desnoues, E., Gibon, Y., Baldazzi, V., Signoret, V., Genard, M., and Quilot-Turion, B. (2014). Profiling sugar metabolism during fruit development in a peach progeny with different fructose-to-glucose ratios. *BMC Plant Biol.* 14:336. doi: 10.1186/s12870-014-0336-x

Doerfler, H., Sun, X., Wang, L., Engelmeier, D., Lyon, D., and Weckwerth, W. (2014). mzGroupAnalyzer–predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data. *PLoS ONE* 9:e96188. doi: 10.1371/journal.pone.0096188

Domon, B., and Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science* 312, 212–217. doi: 10.1126/science.1124619

Ducret, A., Van Oostveen, I., Eng, J. K., Yates, J. R., and Aebersold, R. (1998). High throughput protein characterization by automated reverse-phase chromatography electrospray tandem mass spectrometry. *Prot. Sci.* 7, 706–719. doi: 10.1002/pro.5560070320

Dupae, J., Bohler, S., Noben, J. P., Carpentier, S., Vangronsveld, J., and Cuypers, A. (2014). Problems inherent to a meta-analysis of proteomics data: a case study on the plants' response to Cd in different cultivation conditions. *J. Proteom.* 108, 30–54. doi: 10.1016/j.jprot.2014.04.029

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass-spectrometry of large biomolecules. *Science* 246, 64–71. doi: 10.1126/science.2675315

Fischer, M., Schreiber, L., Colby, T., Kuckenberg, M., Tacke, E., Hofferbert, H. R., et al. (2013). Novel candidate genes influencing natural variation in potato tuber cold sweetening identified by comparative proteomics and association mapping. *BMC Plant Biol.* 13:113. doi: 10.1186/1471-2229-13-113

Foss, E. J., Radulovic, D., Shaffer, S. A., Ruderfer, D. M., Bedalov, A., Goodlett, D. R., et al. (2007). Genetic basis of proteome variation in yeast. *Nat. Genet.* 39, 1369–1375. doi: 10.1038/ng.2007.22

Godfray, H. C., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327, 812–818. doi: 10.1126/science.1185383

Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., et al. (2014). Data integration in the era of omics: current and future challenges. *BMC Syst. Biol.* 8(Suppl. 2):I1. doi: 10.1186/1752-0509-8-S2-I1

Grimaud, F., Renaut, J., Dumont, E., Sergeant, K., Lucau-Danila, A., Blervacq, A. S., et al. (2013). Exploring chloroplastic changes related to chilling and freezing tolerance during cold acclimation of pea (*Pisum sativum* L.). *J. Proteom.* 80, 145–159. doi: 10.1016/j.jprot.2012.12.030

Gross, S. M., Martin, J. A., Simpson, J., Abraham-Juarez, M. J., Wang, Z., and Visel, A. (2013). *De novo* transcriptome assembly of drought tolerant CAM plants, Agave deserti and *Agave tequilana*. *BMC Genomics* 14:563. doi: 10.1186/1471-2164-14-563

Gunnaiah, R., Kushalappa, A. C., Duggavathi, R., Fox, S., and Somers, D. J. (2012). Integrated metabolo-proteomic approach to decipher the mechanisms by which wheat QTL (Fhb1) contributes to resistance against *Fusarium graminearum*. *PLoS ONE* 7:e40695. doi: 10.1371/journal.pone.0040695

Gunther, O. P., Shin, H., Ng, R. T., Mcmaster, W. R., Mcmanus, B. M., Keown, P. A., et al. (2014). Novel multivariate methods for integration of genomics and proteomics data: applications in a kidney transplant rejection study. *OMICS* 18, 682–695. doi: 10.1089/omi.2014.0062

Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720–1730.

Hajduch, M., Casteel, J. E., Tang, S., Hearne, L. B., Knapp, S., and Thelen, J. J. (2007). Proteomic analysis of near-isogenic sunflower varieties differing in seed oil traits. *J. Proteome Res.* 6, 3232–3241. doi: 10.1021/pr070149a

Henry, I. M., Carpentier, S. C., Pampurova, S., Van Hoylandt, A., Panis, B., Swennen, R., et al. (2011). Structure and regulation of the Asr gene family in banana. *Planta* 234, 785–798. doi: 10.1007/s00425-011-1421-0

Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U.S.A.* 90, 5011–5015. doi: 10.1073/pnas.90.11.5011

Heuberger, A. L., Broeckling, C. D., Kirkpatrick, K. R., and Prenni, J. E. (2014). Application of nontargeted metabolite profiling to discover novel markers of quality traits in an advanced population of malting barley. *Plant Biotechnol. J.* 12, 147–160. doi: 10.1111/pbi.12122

Heuberger, A. L., Lewis, M. R., Chen, M. H., Brick, M. A., Leach, J. E., and Ryan, E. P. (2010). Metabolomic and functional genomic analyses reveal varietal differences in bioactive compounds of cooked rice. *PLoS ONE* 5:e12915. doi: 10.1371/journal.pone.0012915

Hirayama, T., and Shinozaki, K. (2010). Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant J.* 61, 1041–1052. doi: 10.1111/j.1365-313X.2010.04124.x

Kakumanu, A., Ambavaram, M. M., Klumas, C., Krishnan, A., Batlang, U., Myers, E., et al. (2012). Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiol.* 160, 846–867. doi: 10.1104/pp.112.200444

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 Daltons. *Anal. Chem.* 60, 2299–2301. doi: 10.1021/ac00171a028

Kelstrup, C. D., Young, C., Lavallee, R., Nielsen, M. L., and Olsen, J. V. (2012). Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* 11, 3487–3497. doi: 10.1021/pr3000249

Kerwin, R. E., Jimenez-Gomez, J. M., Fulop, D., Harmer, S. L., Maloof, J. N., and Kliebenstein, D. J. (2011). Network quantitative trait loci mapping of circadian clock outputs identifies metabolic pathway-to-clock linkages in *Arabidopsis*. *Plant Cell* 23, 471–485. doi: 10.1105/tpc.110.082065

Kliebenstein, D. (2009). Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Annu. Rev. Plant Biol.* 60, 93–114. doi: 10.1146/annurev.arplant.043008.092114

Kugler, K. G., Siegwart, G., Nussbaumer, T., Ametz, C., Spannagl, M., Steiner, B., et al. (2013). Quantitative trait loci-dependent analysis of a gene co-expression network associated with *Fusarium* head blight resistance in bread wheat (*Triticum aestivum* L.). *BMC Genomics* 14:728. doi: 10.1186/1471-2164-14-728

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Langridge, P., and Fleury, D. (2011). Making the most of 'omics' for crop breeding. *Trends Biotechnol.* 29, 33–40. doi: 10.1016/j.tibtech.2010.09.006

Lebedev, A. T., Damoc, E., Makarov, A. A., and Samgina, T. Y. (2014). Discrimination of leucine and isoleucine in peptides sequencing with Orbitrap Fusion mass spectrometer. *Anal. Chem.* 86, 7017–7022. doi: 10.1021/ac501200h

Le Cao, K. A., Boitard, S., and Besse, P. (2011). Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinform.* 12:253. doi: 10.1186/1471-2105-12-253

Legrand, S., Marque, G., Blassiau, C., Bluteau, A., Canoy, A. S., Fontaine, V., et al. (2013). Combining gene expression and genetic analyses to identify candidate genes involved in cold responses in pea. *J. Plant Physiol.* 170, 1148–1157. doi: 10.1016/j.jplph.2013.03.014

Lesage, V. S., Merlino, M., Chambon, C., Bouchet, B., Marion, D., and Branlard, G. (2012). Proteomes of hard and soft near-isogenic wheat lines reveal that kernel hardness is related to the amplification of a stress response during endosperm development. *J. Exp. Bot.* 63, 1001–1011. doi: 10.1093/jxb/err330

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2013a). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484

Li, L., Petsch, K., Shimizu, R., Liu, S., Xu, W. W., Ying, K., et al. (2013b). Mendelian and non-Mendelian regulation of gene expression in maize. *PLoS Genet.* 9:e1003202. doi: 10.1371/journal.pgen.1003202

Li, Y., Nie, Y., Zhang, Z., Ye, Z., Zou, X., Zhang, L., et al. (2014). Comparative proteomic analysis of methyl jasmonate-induced defense responses in different rice cultivars. *Proteomics* 14, 1088–1101. doi: 10.1002/pmic.201300104

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., et al. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 17, 676–682. doi: 10.1038/10890

Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., et al. (2006). Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* 78, 2113–2120. doi: 10.1021/ac0518811

Mann, M., and Kelleher, N. L. (2008). Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18132–18138. doi: 10.1073/pnas.0800788105

Martinez-Gomez, P., Sanchez-Perez, R., and Rubio, M. (2012). Clarifying omics concepts, challenges, and opportunities for Prunus breeding in the postgenomic era. *OMICS* 16, 268–283. doi: 10.1089/omi.2011.0133

Maskin, L., Gudesblat, G. E., Moreno, J. E., Carrari, F. O., Frankel, N., Sambade, A., et al. (2001). Differential expression of the members of the Asr gene family in tomato (*Lycopersicon esculentum*). *Plant Sci.* 161, 739–746. doi: 10.1016/S0168-9452(01)00464-2

Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., et al. (2003). Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15718–15723. doi: 10.1073/pnas.2536670100

McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S., Barnes, G., Drubin, D., et al. (1997). Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* 69, 767–776. doi: 10.1021/ac960799q

Mensack, M. M., Fitzgerald, V. K., Ryan, E. P., Lewis, M. R., Thompson, H. J., and Brick, M. A. (2010). Evaluation of diversity among common beans (*Phaseolus vulgaris* L.) from two centers of domestication using 'omics' technologies. *BMC Genomics* 11:686. doi: 10.1186/1471-2164-11-686

Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., et al. (2011). Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteom.* 10, M111 011015. doi: 10.1074/mcp.m111.011015

Mir, R. R., Zaman-Allah, M., Sreenivasulu, N., Trethowan, R., and Varshney, R. K. (2012). Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. *Theor. Appl. Genet.* 125, 625–645. doi: 10.1007/s00122-012-1904-9

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Nesvizhskii, A. I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteom.* 4, 1419–1440. doi: 10.1074/mcp.R500012-MCP200

Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 4, 709–712. doi: 10.1038/nmeth1060

Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., et al. (2009). A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteom.* 8, 2759–2769. doi: 10.1074/mcp.M900375-MCP200

Oono, Y., Kobayashi, F., Kawahara, Y., Yazawa, T., Handa, H., Itoh, T., et al. (2013). Characterisation of the wheat (*Triticum aestivum* L.) transcriptome by *de novo* assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. *BMC Genomics* 14:77. doi: 10.1186/1471-2164-14-77

Page, J. T., Gingle, A. R., and Udall, J. A. (2013a). PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. *G3 (Bethesda)* 3, 517–525. doi: 10.1534/g3.112.005298

Page, J. T., Huynh, M. D., Liechty, Z. S., Grupp, K., Stelly, D., Hulse, A. M., et al. (2013b). Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *G3 (Bethesda)* 3, 1809–1818. doi: 10.1534/g3.113.007229

Pariset, L., Chillemi, G., Bongiorni, S., Romano Spica, V., and Valentini, A. (2009). Microarrays and high-throughput transcriptomic analysis in species with incomplete availability of genomic sequences. *N. Biotechnol.* 25, 272–279. doi: 10.1016/j.nbt.2009.03.013

Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., and Coon, J. J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteom.* 11, 1475–1488. doi: 10.1074/mcp.O112.020131

Picotti, P., Clement-Ziza, M., Lam, H., Campbell, D. S., Schmidt, A., Deutsch, E. W., et al. (2013). A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* 494, 266–270. doi: 10.1038/nature11835

Pinheiro, C., Guerra-Guimarães, L., David, T. S., and Vieira, A. (2014). Proteomics: state of the art to study Mediterranean woody species under stress. *Environ. Exp. Bot.* 103, 117–127. doi: 10.1016/j.envexpbot.2014.01.010

Plumb, R. S., Johnson, K. A., Rainville, P., Smith, B. W., Wilson, I. D., Castro-Perez, J. M., et al. (2006). UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* 20, 1989–1994. doi: 10.1002/rcm.2550

Rampitsch, C., and Bykova, N. V. (2012). The beginnings of crop phosphoproteomics: exploring early warning systems of stress. *Front. Plant Sci.* 3:144. doi: 10.3389/fpls.2012.00144

Rasmussen, S., Parsons, A. J., and Jones, C. S. (2012). Metabolomics of forage plants: a review. *Ann. Bot.* 110, 1281–1290. doi: 10.1093/aob/mcs023

Reinartz, J., Bruyns, E., Lin, J. Z., Burcham, T., Brenner, S., Bowen, B., et al. (2002). Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Brief. Funct. Genomic. Proteomic.* 1, 95–104. doi: 10.1093/bfgp/1.1.95

Scherling, C., Roscher, C., Giavalisco, P., Schulze, E. D., and Weckwerth, W. (2010). Metabolomics unravel contrasting effects of biodiversity on the performance of individual plant species. *PLoS ONE* 5:e12569. doi: 10.1371/journal.pone.0012569

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098

Skelly, D. A., Merrihew, G. E., Riffle, M., Connelly, C. F., Kerr, E. O., Johansson, M., et al. (2013). Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res.* 23, 1496–1504. doi: 10.1101/gr.155762.113

Straub, D., Yang, H., Liu, Y., and Ludewig, U. (2013). Transcriptomic and proteomic comparison of two *Miscanthus* genotypes: high biomass correlates with investment in primary carbon assimilation and decreased secondary metabolism. *Plant Soil* 372, 151–165. doi: 10.1007/s11104-013-1693-1

Sulpice, R., Nikoloski, Z., Tschoep, H., Antonio, C., Kleessen, S., Larhlimi, A., et al. (2013). Impact of the Carbon and Nitrogen Supply on Relationships and Connectivity between Metabolism and Biomass in a Broad Panel of *Arabidopsis* Accessions[1][W][OA]. *Plant Physiol.* 162, 347–363. doi: 10.1104/pp.112.210104

Sulpice, R., Trenkamp, S., Steinfath, M., Usadel, B., Gibon, Y., Witucka-Wall, H., et al. (2010). Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of *Arabidopsis* accessions. *Plant Cell* 22, 2872–2893. doi: 10.1105/tpc.110.076653

Tenenhaus, A., Philippe, C., Guillemot, V., Le Cao, K. A., Grill, J., and Frouin, V. (2014). Variable selection for generalized canonical correlation analysis. *Biostatistics* 15, 569–583. doi: 10.1093/biostatistics/kxu001

Torabi, S., Wissuwa, M., Heidari, M., Naghavi, M. R., Gilany, K., Hajirezaei, M. R., et al. (2009). A comparative proteome approach to decipher the mechanism of rice adaptation to phosphorous deficiency. *Proteomics* 9, 159–170. doi: 10.1002/pmic.200800350

Valdés, A., Ibáñez, C., Simó, C., and García-Cañas, V. (2013). Recent transcriptomics advances and emerging applications in food science. *TrAC Trends Anal. Chem.* 52, 142–154. doi: 10.1016/j.trac.2013.06.014

Vanhove, A. C., Vermaelen, W., Swennen, R., and Carpentier, S. C. (2015). A look behind the screens: characterization of the HSP70 family during osmotic stress in a non-model crop. *J. Proteom.* 119C, 10–20. doi: 10.1016/j.jprot.2015.01.014

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* 270, 484–487. doi: 10.1126/science.270.5235.484

Vertommen, A., Moller, A. L., Cordewener, J. H., Swennen, R., Panis, B., Finnie, C., et al. (2011a). A workflow for peptide-based proteomics in a poorly sequenced plant: a case study on the plasma membrane proteome of banana. *J. Proteom.* 74, 1218–1229. doi: 10.1016/j.jprot.2011.02.008

Vertommen, A., Panis, B., Swennen, R., and Carpentier, S. C. (2011b). Challenges and solutions for the identification of membrane proteins in non-model plants. *J. Proteom.* 74, 1165–1181. doi: 10.1016/j.jprot.2011.02.016

Virlouvet, L. C., Jacquemot, M.-P., Gerentes, D., Corti, H., Bouton, S., Gilard, F., et al. (2011). The ZmASR1 protein influences branched-chain amino acid biosynthesis and maintains kernel yield in maize under water-limited conditions. *Plant Physiol.* 157, 917–936. doi: 10.1104/pp.111.176818

Wang, H. J., Hsu, C. M., Jauh, G. Y., and Wang, C. S. (2005). A lily pollen ASR protein localizes to both cytoplasm and nuclei requiring a nuclear localization signal. *Physiol. Plant.* 123, 314–320. doi: 10.1111/j.1399-3054.2005.00454.x

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Weckwerth, W. (2011). Unpredictability of metabolism–the key role of metabolomics science in combination with next-generation genome

sequencing. *Anal. Bioanal. Chem.* 400, 1967–1978. doi: 10.1007/s00216-011-4948-9

Wienkoop, S., Larrainzar, E., Glinski, M., Gonzalez, E. M., Arrese-Igor, C., and Weckwerth, W. (2008). Absolute quantification of *Medicago truncatula* sucrose synthase isoforms and N-metabolism enzymes in symbiotic root nodules and the detection of novel nodule phosphoproteins by mass spectrometry. *J. Exp. Bot.* 59, 3307–3315. doi: 10.1093/jxb/ern182

Wienkoop, S., Weiss, J., May, P., Kempa, S., Irgang, S., Recuenco-Munoz, L., et al. (2010). Targeted proteomics for *Chlamydomonas reinhardtii* combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. *Mol. Biosyst.* 6, 1018–1031. doi: 10.1039/b920913a

Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., et al. (2014). Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582–587. doi: 10.1038/nature13319

Wu, J., Yu, H., Dai, H., Mei, W., Huang, X., Zhu, S., et al. (2012). Metabolite profiles of rice cultivars containing bacterial blight-resistant genes are distinctive from susceptible rice. *Acta Biochim. Biophys. Sin. (Shanghai)* 44, 650–659. doi: 10.1093/abbs/gms043

Wu, L., Candille, S. I., Choi, Y., Xie, D., Jiang, L., Li-Pook-Than, J., et al. (2013). Variation and genetic control of protein abundance in humans. *Nature* 499, 79–82. doi: 10.1038/nature12223

Yang, J., and Yen, H. E. (2002). Early salt stress effects on the changes in chemical composition in leaves of ice plant and *Arabidopsis*. A Fourier transform infrared spectroscopy study. *Plant Physiol.* 130, 1032–1042. doi: 10.1104/pp.004325

Yazaki, J., Gregory, B. D., and Ecker, J. R. (2007). Mapping the genome landscape using tiling array technology. *Curr. Opin. Plant Biol* 10, 534–542. doi: 10.1016/j.pbi.2007.07.006

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE* 9:e78644. doi: 10.1371/journal.pone.0078644

Zorrilla-Fontanesi, Y., Rambla, J. L., Cabeza, A., Medina, J. J., Sanchez-Sevilla, J. F., Valpuesta, V., et al. (2012). Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesifurane content. *Plant Physiol.* 159, 851–870. doi: 10.1104/pp.111.188318

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.