Fortune telling: metabolic markers of plant performance

Olivier Fernandez, Maria Urrutia, Stéphane Bernillon, Catherine Giauffret, Francois F. Tardieu, Jacques Le Gouis, Nicolas Langlade, Alain A. Charcosset, Annick Moing, Yves Y. Gibon

HAL Id: hal-02637624
https://hal.inrae.fr/hal-02637624

Submitted on 27 May 2020

CrossMark

REVIEW ARTICLE

Fortune telling: metabolic markers of plant performance

Olivier Fernandez[1] · Maria Urrutia[1] · Stéphane Bernillon[1,2] · Catherine Giauffret[3] ·
François Tardieu[4] · Jacques Le Gouis[5] · Nicolas Langlade[6] · Alain Charcosset[7] ·
Annick Moing[1,2] · Yves Gibon[1,2]

Abstract

Background In the last decade, metabolomics has emerged as a powerful diagnostic and predictive tool in many branches of science. Researchers in microbes, animal, food, medical and plant science have generated a large number of targeted or non-targeted metabolic profiles by using a vast array of analytical methods (GC–MS, LC–MS, ^{1}H-NMR….). Comprehensive analysis of such profiles using adapted statistical methods and modeling has opened up the possibility of using single or combinations of metabolites as markers. Metabolic markers have been proposed as proxy, diagnostic or predictors of key traits in a range of model species and accurate predictions of disease outbreak frequency, developmental stages, food sensory evaluation and crop yield have been obtained.

Aim of review (i) To provide a definition of plant performance and metabolic markers, (ii) to highlight recent key applications involving metabolic markers as tools for monitoring or predicting plant performance, and (iii) to propose a workable and cost-efficient pipeline to generate and use metabolic markers with a special focus on plant breeding.

Key message Using examples in other models and domains, the review proposes that metabolic markers are tending to complement and possibly replace traditional molecular markers in plant science as efficient estimators of performance.

Keywords Breeding · Metabolic marker · Metabolomics · Plant performance · Prediction

✉ Olivier Fernandez
olivier.fernandez@bordeaux.inra.fr

1 UMR 1332 Biologie du Fruit et Pathologie, INRA, Centre INRA de Bordeaux, 71 av Edouard Bourlaux, 33140 Villenave d'Ornon, France

2 Plateforme Métabolome Bordeaux, CGFB, MetaboHUB-PHENOME, 33140 Villenave d'Ornon, France

3 UMR AgroImpact, INRA, Estrées-Mons, 80203 Péronne, France

4 UMR LEPSE, INRA, Montpellier SupAgro, 34000 Montpellier, France

5 UMR GDEC, INRA, UBP, 63039 Clermont-Ferrand, France

6 UMR LIPM, INRA, CNRS, Université de Toulouse, 31326 Castanet-Tolosan, France

7 UMR GQE, INRA, CNRS, Université Paris Sud, AgroParisTech, Ferme du Moulon, 91190 Gif-Sur-Yvette, France

1 Introduction

Forecasting the future is as old as the hills. How odd it might sound today but animals' entrails, palm-reading and coffee grounds have been used in the past as a source of information by leaders and decision-makers. In modern society, we still need to anticipate. Proxy, diagnosis or estimation remain helpful for many human activities including scientific domains.

Metabolomics has recently taken a quantum leap forward. Using a combination of approaches such as proton nuclear magnetic resonance (^{1}H-NMR), liquid or gas chromatography coupled with mass spectrometry (GC–MS, LC–MS) as well as robotized spectrometric and fluorimetric assays, it is now possible to measure thousands of analytes in thousands of samples whether of microbial, plant or animal origin (Gibon et al. 2012; Nicholson et al. 2007), even in non-model species. Metabolomics has a

wide range of applications in an impressive list of organisms. For example, several 'silent' mutations in *Saccharomyces cerevisiae* bearing no overt phenotypes have been revealed by measuring metabolite concentrations (Raamsdonk et al. 2001). Metabolomics has also led to considerable progress in understanding the regulation of cellular metabolism in *Escherichia coli* (Nöh et al. 2007). In animal science, it has been used for studying the responses to adverse conditions in nematode and fruit fly (Coquin et al. 2008; Hughes et al. 2009; Malmendal et al. 2006) and for classifying the stages of embryogenesis in zebra fish by using fingerprints of highly correlated metabolites (Hayashi et al. 2009, 2011). Metabolomics is also widely used in edible products for predicting geographical origin, terroir and varietal effect, e.g. for wine (Cynkar et al. 2010; Tarr et al. 2013), green tea (Lee et al. 2015) and orange (Díaz et al. 2014), for assessing the legal requirements for oil, coffee, honey (Cubero-Leon et al. 2014) and for profiling the sensory qualities of wine and meat (Schmidtke et al. 2013; Straadt et al. 2014). Readers are referred to recent reviews on this subject (Cubero-Leon et al. 2014; Oms-Oliu et al. 2013; Putri et al. 2013; Sumner et al. 2015) for a more comprehensive view of these applications. The spread of metabolomics has been supported by increased computational power, which facilitates statistical analyses of large datasets and raises the possibility of applying correlative methods and finding metabolites associated with a given state or condition (Gibon et al. 2012; Wolfender et al. 2013). These so-called biomarkers can also be referred to as metabolic markers when constructed with metabolite concentrations.

Medical science has been precursor in the use of metabolic markers. Indian physicians around 1500 BC noted that the sugar-enriched urine of patients with diabetes attracted ants (Zajac et al. 2010). Nowadays, body fluid analyses offer numerous opportunities to profile metabolites and correlate them with a diagnosis and/or prediction of disease susceptibility. This is illustrated by the emergence of patient stratification and personalized medicine (Lindon and Nicholson 2014; Nicholson et al. 2012). Urine metabolic profiling led to the identification of metabolic markers of symptomatic gout (Liu et al. 2012) and preeclampsia (Austdal et al. 2015) and blood profiling has been used to estimate the risk of bacteremic sepsis in emergency rescue situations (Kauppi et al. 2016). Another promising application of metabolite analysis in medical science is the prediction of cancer risk (Lee et al. 2014; McDunn et al. 2013; Truong et al. 2013) or the evaluation of the putative effect of cancer treatments (Hou et al. 2014; Jiang et al. 2014; Wei et al. 2013).

Metabolic markers are also used in plant science. Early examples include diagnostic methods such as Jubil® and N-tester®. They have both been used to proxy the nitrogen status in plants for the sustainable fertilization of wheat, barley and maize (Justes et al. 1997; Uddling et al. 2007) through measurements of nitrate in stem fluids or chlorophyll in leaves respectively. Because plant scientists and breeders are eager to improve crop performances in challenging conditions for human food security and to find varieties selected for more complex traits, metabolic markers are also becoming popular in plant science and breeding (Herrmann and Schauer 2013; Zabotina 2013). However, the use of metabolic markers is not straightforward. Metabolite levels belong to the phenotype, which means that they can be associated with the genotype, the environment, the developmental stage and the interactions between them, as any other trait. This might be why metabolic markers were first proposed as a tool for searching for metabolite quantitative trait loci (mQTLs) and finding the related genes (Fridman et al. 2000), which were subsequently used for selection. Nevertheless, metabolic markers can be used as direct predictors when associated with plant performance criteria. They can also contribute to understanding how plant physiology processes are co-ordinated in various growth conditions [e.g. as detailed for water deficit by (Tardieu et al. 2011)], although this may not be the primary objective, especially when using metabolic markers in breeding.

The aim of this paper is to define plant performance and metabolic markers and to explain why and when they can be used as a tool for monitoring or predicting such performance. Finally, we describe a cost-efficient pipeline using metabolic markers as putative predictors of performance, with notable applications in plant-breeding.

2 What is plant performance?

The definition of crop performance is often limited to the yield of the harvested part of the plant bearing the added value. Yield is indubitably an important trait of performance and its pattern under various growth conditions may allow the simple comparison of genotypes. However practical, this definition of performance is partly inadequate. Performance traits can be qualitative such as behavior in a series of environmental scenarios (high temperature, water deficit or biotic stresses), crop subtypes (afila in pea, bearded wheat) or the association of traits that are desirable for a given crop. Additionally, crop performance can be related to an industrial procedure through which the crop has to be processed. We propose here a general definition of plant performance as being **an association of several traits that need to be monitored with regard to the plant life cycle or improved through a breeding process**. We propose the following non-exhaustive list of traits:

- Grain or tissue yield
- Stability and consistency of yield over various natural environments, meteorological conditions or stresses
- Plant morphology (number of leaves, stems, flowers per bunch, plant height…) or phenology (duration of a particular stage of development)
- Storage properties such as fruit shelf-life or grain stability
- Yield of a specific compound or metabolite (to increase its concentration or to eliminate it)
- Technological properties (e.g. malting properties for barley, protein quantity and quality for breadmaking in wheat, cooking properties for potato, etc.)
- Sensory quality such as the presence of aromas or aroma precursors
- Nutritional attributes such as absence or low content of anti-nutritional compounds, or presence of vitamins, glycemic index, saturated lipid content
- Tolerance to a specific adverse condition, biotic or abiotic stress (extreme temperatures, salinity…)
- Efficiency of water and nutrient use.

Several of these criteria are now included in large crop-breeding projects such as the French aMaizING (maize, www.amaizing.fr), BreedWheat (wheat, www.breedwheat.fr) and SUNRISE (sunflower, www.sunrise-project.fr) projects, which address a variety of agronomical objectives (e.g., tolerance to water stress, chilling, low nitrogen or sulphur availability) and include precise phenotyping. Biochemical or metabolic phenotyping are tentatively integrated into the breeding process, notably in order to establish more precise estimations of plant performance and access the underlying mechanisms.

3 Definition of a metabolic marker

The term biomarker (or biological marker) originates from the field of medicine. It has been defined as 'a characteristic that is objectively measured and evaluated as an indication of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention' (NIH Definitions Working Group, 2000). In plants, the concept of biomarker is often associated with plant performance and could be defined as **a characteristic that is objectively measured or evaluated as a predictor of plant performance**.

Biomarkers can be genotypic (e.g., nucleotide polymorphisms such as single-nucleotide polymorphisms or SNPs generally) or phenotypic (e.g., transcript levels, protein levels, enzyme activities, metabolite levels, images in different wavelengths). In addition to being predictive,

biomarkers are preferably easy and cheap to score (Aronson 2005). This is probably why the use of molecular and biochemical markers, which proved to be excellent predictors and are relatively easy to measure in high-throughput conditions, became widespread in medicine (Menard et al. 2013; Robinette et al. 2013).

Metabolic markers are a sub-category of biomarkers that are involved in metabolism. Importantly, unlike DNA sequences, most metabolic traits vary during plant development, potentially with diurnal patterns, between tissue/organ and in response to environmental cues. Therefore, their use as biomarkers has to take into account developmental stage, position on the plant, time of day and growth scenario. Three types of metabolic markers can be distinguished:

- Traits of agricultural importance. An obvious strategy is to screen germplasm with direct measurements of such molecules or their precursors. Such traits can be desirable, like vitamin C or aromas (Ruiz-García et al. 2014; Pissard et al. 2013), or undesirable (e.g., toxins such as cyanogenic glucosides in cassava, anti-nutrients such as erucic acid in rapeseed).
- Diagnostic markers. In plants, single metabolic markers have been proposed to estimate the intensity of a given stress, for example proline, which accumulates in many species experiencing drought (Dib et al. 1994; Hayat et al. 2012). More recently, the idea that combinations of metabolic variables could be used to diagnose stress damage or resistance has emerged and the use of transcripts (Tamaoki et al. 2004), enzymes (Gibon et al. 2004) or metabolites (Korn et al. 2010, 2008; Roessner et al. 2000) has been proposed.
- Markers of genotype performance. In 2007, metabolic profiles were used for the first time to estimate biomass production in the model plant species *Arabidopsis thaliana*, with a coefficient of correlation of 0.58 (Table 1). This pioneering study paved the way for several others where associations between performance traits and metabolic markers were found, as summarized in Table 1. It also opened up new possibilities for plant breeding in which metabolic markers would be used to search for combinations of alleles that provide higher plant performance (Meyer et al. 2007). Ultimately, this would consist in searching for associations (e.g. with correlation, regression or classification methods), in a given set of genotypes, between metabolite data obtained for a given organ, developmental stage and environment combinations and plant performance, and then assuming that these associations remain valid for any genotypes grown subsequently in other environmental conditions.

Table 1 List of associations of metabolic markers and plant performance in recent literature

Plant	Trait	Numbers of markers	Statistical approach	Association	Statistical validation	Reference
Arabidopsis	Biomass	181	CCA/PLS	corr = 0.73/0.58	Random permutation/train-test sets	Meyer et al. 2007
	Dry weight	181	OLS/PLS	Q2Y = 0.11 (C24) and 0.11 (Columbia)/Q2Y = 0.12 (C24) and 0.23 (Columbia)	Random permutation/ random permutation	Steinfath et al. 2010a
	Dry weight	9 (Columbia) 13 (C24)	PLS	Q2Y = 0.26/0.38	Random permutation	
Barley	Several malt quality traits	216	O2PLS	Q2Y = 0.17 to 0.79	–	Heuberger et al. 2014
Maize	Several performance traits	130	RR-BLUP	$r(\hat{g},g)$ = 0.61 to 0.79	Fivefold cross validation	Riedelsheimer et al. 2012a
	Dry matter yield	7	Pearson correlation	corr = −0.35 to 0.12	p value < 0.05	Riedelsheimer et al. 2012b
	Lignin content	7	Pearson correlation	corr = −0.20 to 0.15	p value < 0.01	
	Plant height	5	Pearson correlation	corr = −0.23 to 0.16	p value < 0.008	
	GCA for several performance traits	1/563	Pearson correlation/RR-BLUP	corr = −0.54 to 0.48/ $r(\hat{g},g)$ = 0.47 to 0.78	Fivefold cross validation	Riedelsheimer et al. 2013
	Grain yield under drought stress	5	Pearson correlation	corr = −0.47 to −0.54	p value < 0.01	Obata et al. 2015
Pine	Plant height	11	Pearson correlation	corr = 0.13 to 0.35	p value < 0.05	Kang et al. 2015
	Stem dry mass	11	Pearson correlation	corr = 0.15 to 0.34	p value < 0.05	
Potato	Chip property	2	PLS and VIP selection	0.66 to 0.75	Random permutation	Steinfath et al. 2010b
	Susceptibility to blackspotedness	5	PLS and VIP selection	0.53 to 0.82	Random permutation	
Rice	Tolerance to mild salinity stress	2	t-test(foldchange)/ PLS-DA	Delta log2(FCh) > 1/ Q2Y = 0.49	p value < 0.05/ random permutation	Nam et al. 2015
	Yield under drought stress	16	Pearson correlation	corr = −0.71 to 0.53	p value < 0.05	Degenkolbe et al.2013
	yield under drought stress	5	Pearson correlation	corr = −0.72 to 0.45	p value < 0.05	
Tomato	Firmess and shelf life	2	Correlation network	–	p value < 0.001	López et al. 2015
	TYLCV resistance	120	FCH/Correlation network	0.88 to 1.43/mean r2 = 0.62	p value < 0.05/–	Sade et al. 2015
Wheat	Fusarium graminearum resistance	60	FCH/Correlation network	–	–	Cuperlovic-Culf et al. 2016
Grape	Esca disease sensitivity	34	PCA	–	–	Lima et al. 2010

CCA canonical correlation analysis, *corr* correlation, *FCH* fold change, *GCA* general combining ability, *O2PLS* orthogonal partial least squares projections to latent structures, *OLS* ordinary least squares, *PCA* principal component analysis, *PLS* partial least squares to latent structures, *Q2Y* cumulative predictive explained variation, $r(\hat{g},g)$ correlation between predicted and unobserved true values, *RR-BLUP* ridge regression-best linear unbiased prediction, *VIP* variables importance in the projection

4 Why use metabolic markers?

Measuring metabolites implies destructive sampling and sometimes costly and labor-intensive analytics. Furthermore, the use of molecular markers such as single nucleotide polymorphisms (SNPs), which are cheap, independent of the environment, amenable to high-throughput and are now commonplace in the research departments of breeding companies, is becoming the standard for breeders. So what would metabolic markers be good for?

4.1 When metabolite levels are the trait of performance

Some metabolic traits are important *per se*. A famous example is zero-erucic-acid rapeseed oil, which is suitable for human nutrition. It was obtained with a strategy involving the non-destructive sampling of single cotyledons (to guarantee seedling survival to form the next generation) and quantification via gas liquid chromatography (Downey and Harvey 1963). Cyanogenic glucoside content in cassava, an important food source in tropical regions, could be measured by a low-cost spectroscopic method after acid hydrolysis (Bradbury et al. 1991) and then used in classical breeding programs aiming at reducing toxin levels (Nambisan 2011). Similarly, low phytic acid content in maize kernels is of interest for food and animal feed (Hazebroek et al. 2007). The screening of desirable metabolites is also possible, e.g. nutritional compounds such as vitamin C (Pissard et al. 2013) or aroma precursors such as rose oxide, which highly correlate with the "Muscat Aroma" in the grape cultivar (Ruiz-García et al. 2014). The role of metabolomics in improving the nutritional values of crops has already been underlined in rice (Fitzgerald et al. 2009) and these approaches could be a way to ensure that plant breeding programs place more emphasis on nutritional optimization (Anonymous 2016b).

4.2 When metabolites provide condensed information

So far, most of the molecular marker–trait associations found in academic programs that have been transferred to commercial breeding programs involve traits with simple genetic determinism (Heffner et al. 2009; Xu and Crouch 2008). This is probably due to the fact that the number of molecular markers was initially low in most cases. Additionally, qualitative traits (disease resistance mostly) are overrepresented (Gupta et al. 2010). Furthermore, pyramiding beneficial alleles associated with traits resulting from complex interactions such as epistasis and genotype by environment interactions is still considered as very challenging (Furbank and Tester 2011).

In 2012, Riedelsheimer et al. (2012a) compared the predictive power of metabolic and molecular markers. Although the precision was slightly lower for metabolites with correlations ranging from 0.60 to 0.80 (Table 1) compared to 0.72 to 0.81, the authors underlined the fact that 130 metabolites were almost as good predictors as 38,000 SNPs. They concluded that metabolites provide condensed information and could be especially interesting when dealing with highly polygenic traits.

Two further studies in maize used a similar approach. The lipid profiling of maize leaves revealed high correlations with several agronomical traits [Riedelsheimer et al. (2013), including dry matter yield (0.47) and flowering time (0.78); Table 1]. A tempting follow-up would be to identify highly efficient hybrids in test-crosses via lipidomics. Caffeic- and p-coumaric acid also showed significant correlations with dry matter yield [−0.28 and 0.12 respectively; Table 1; Riedelsheimer et al. (2012b)], suggesting that a low-cost strategy targeting these metabolites could be developed to screen thousands of hybrids for selection purposes. In these examples, there is little difference in dealing with metabolic markers compared to molecular markers. Associations between metabolic markers and performance criteria would nevertheless have to be generated with adequate statistical methods that take into account potential interactions, e.g., between genotype and environment.

4.3 When metabolites open the way to mechanistic insights

The fact that metabolic markers provide biological information that can narrow down the genotype-phenotype gap opens the door for mechanistic insights, starting with the detection of SNPs or candidate genes via mQTL mapping strategies. Riedelsheimer et al. (2012b) detected several mQTL for lignin precursors such as p-coumaric acid and caffeic acid, which they found to be good predictors of a range of plant performance criteria (e.g., plant height and dry matter yield; Table 1). The corresponding region harbors a key enzyme in monolignol synthesis (cinnamoyl-CoA reductase) and has been proposed as a good target for improving the quality of lignocellulosic biomass. In addition, candidate gene allelic variability (natural or induced) could be explored to evaluate changes in lignocellulosic quality. The use of metabolic markers to gain mechanistic knowledge can also be illustrated by the negative correlation of starch with biomass (Sulpice et al. 2009). This led the authors to conclude that starch is an integrator of plant growth, reflecting a fine balance between carbon supply and growth.

Such findings highlight the usefulness of metabolic markers for estimating agronomical traits and revealing biological mechanisms underlying phenotypes.

4.4 When metabolites can be a diagnostic tool in crop processing

An original application of metabolic markers is the evaluation of crop performance in an industrial or commercial process. One of the first publications to mention such a possibility was focused on potato susceptibility to black spot bruising (induced by collisions during transport and storage) and undesirable 'browning while frying'. Five amino acids (tyrosine, threonine, valine, serine and glutamine) and two sugars (glucose and fructose) were detected as the best metabolic markers (VIP in a PLS analysis; Table 1) for these traits, respectively (Steinfath et al. 2010b). To validate these markers, a model was entrained to compare measured and predicted traits in an independent location bearing significant correlation (ranging 0.53 to 0.82 and 0.66 to 0.75 respectively for susceptibility to blackspottedness and chip property; Table 1). Another example of metabolites linked to industrial properties is the association of a profile of 216 features (Table 1) to malting quality in barley (Heuberger et al. 2014).

Fresh fruit marketability is linked to shelf-life, which is affected by firmness. Both traits have been shown to be associated with malate content in tomato (López et al. 2015) through a neural network approach (self-organizing maps; Table 1). In the same study, another important commercial trait (fruit morphology) showed to be associated strongly with aspartate, glutamate and 2-oxoglutarate (López et al. 2015).

4.5 When assessing diversity of crop core collections or other genetic resources

A recent application of plant metabolomics that has already been implemented in biotechnology and seed companies is the assessment of metabolic diversity within their crop core population or genetic lineage. This has been done for instance by Monsanto® in soybean (Kusano et al. 2015; Harrigan et al. 2015) and maize (Venkatesh et al. 2016) as well as by Pioneer® in the latter species (Baniasadi et al. 2014; Zeng et al. 2014; Asiago et al. 2012). Authors underline the potential of metabolomics to separate genetic and environmental effects on crop diversity (Venkatesh et al. 2016; Baniasadi et al. 2014) or for substantial equivalence studies of genetically modified (GM) genotypes (Harrigan et al. 2015; Baniasadi et al. 2014; Asiago et al. 2012). These results could be used to improve acceptance of GMOs and might also be used for regulatory purposes (Zeng et al. 2014). These companies have all the necessary tools in house to use metabolic data for breeding. Indeed several of their publications have already shown associations of key performance criteria with metabolites, for instance for yield in soybean (Kusano et al. 2015) or plant and ear height in maize (Venkatesh et al. 2016).

4.6 When working on impact of abiotic and biotic stress

Metabolites can also be used as markers to estimate plant performance under stress conditions (Feussner and Polle 2015; Fraire-Velázquez and Balderas-Hernández 2013). Obata et al. (2015) found that myo-inositol accumulated in young leaves was constitutively and negatively associated with grain yield under at least some drought stress scenarios in maize (−0.54; Table 1) In rice, Quistián-Martínez et al. (2011) identified trehalose as a putative inducible marker in drought-tolerant rice genotypes, while Degenkolbe et al. (2013) reported eight metabolites that were positively accumulated in drought-tolerant varieties (including allantoin, galactaric and gluconic acid, glucose and salicylic acid glucopyranoside; Table 1). Interestingly, allantoin was also associated with salt-stress tolerance in rice (Table 1; Nam et al. 2015). Although 'constitutive' metabolic markers, e.g. those measured in plant material obtained under standard conditions and at young developmental stages (Riedelsheimer et al. 2012b; Riedelsheimer et al. 2012a), might be of great interest when stress resistance can be estimated, it is likely that 'inducible' metabolic markers will be needed to evaluate tolerance in stressed conditions and to train the prediction models of resistance. For this, the combined use of phenotyping platforms (Tisne et al. 2013) providing reproductive and relevant stress scenarios combined with pertinent metabolic analyses could be very valuable. However, such a strategy involving ecophysiologists, biochemists and geneticists still requires sustained exploratory efforts.

Regarding biotic stress, metabolomics has recently emerged as a tool for studying plant immunity, especially for deciphering the role of small molecules involved in plant–microbe interactions (Feussner and Polle 2015). Diagnostic-like strategies separating diseased from healthy plants with metabolic markers have been proposed using ^{1}H-NMR in ornamental periwinkle and grapevine (Table 1; Choi et al. 2004; Lima et al. 2010). Finally, metabolic markers have been associated with tolerance to yellow leaf curl virus in tomato (Sade et al. 2015) and to fusarium in wheat (Cuperlovic-Culf et al. 2016). Of particular interest in the tomato study, the authors highlighted a more coordinated response of the primary metabolism in resistant cultivars (Sade et al. 2015).

5 What pipeline to work with metabolic markers of plant performance?

The major challenge when using metabolic markers will be to establish combinations of growth scenarios, sampling strategies and metabolic marker measurements that provide estimations of plant performance that are consistent with the 'real' world. As mentioned above, it is indeed known that QTL associated with plant performance can have positive effects under given growth scenarios and negative effects under others (Tardieu 2011), and that there is a priori no reason why this would not be the case for such estimations. Vast numbers of metabolic fingerprints can be generated by profiling diverse organs or tissues at different stages and under various growth conditions. The fact that this diversity is challenging when looking for metabolic markers of performance implies that several steps listed below have to be taken into account.

5.1 Growth scenarios: reproducible and crop-adapted to reveal diversity

Metabolite levels and fluxes are sensitive to growth conditions, especially to temperature which modifies enzymatic activities independently (Strand et al. 1999; Parent et al. 2010). They are also subject to large changes throughout plant and organ development and even throughout night and day cycles. Simulating the diversity of scenarios that any crop would face in the field is not a realistic option. Therefore, careful implementation of reproducible growth scenarios seems necessary to find the best metabolic markers, especially if the studied performance criterion is tolerance towards adverse conditions.

These scenarios should be designed in order to reveal genotype diversity for a given plant performance criterion. They can be seen as a proxy of the growth conditions of the crop with the additional constraint of reproducibility in order to generate robust markers. Academic (Cabrera-Bosquet et al. 2016; Kumar et al. 2015) and private robotized phenotyping facilities offer solutions for programming such scenarios and for phenotyping crops while limiting costs compared to field trials (Humplík et al. 2015). These facilities, which so far tend to focus on growth and architecture, could be used to perform metabolic studies, eventually identify metabolic markers and ultimately deepen our knowledge about how metabolism and plant performance are integrated. It is likely that this will require large experimental (e.g., what should be harvested, at what developmental stage, at what time of the day, what should be measured) and technological (e.g., cost-efficient sample collection) efforts.

In association with this type of facilities, data and metadata management solutions (Hannemann et al. 2009) would be of great help. Indeed, the extensive follow-up of experimental conditions (detailed scoring of all environmental and developmental factors that may impact metabolism…) from growth scenarios to sample handling and metabolomics data, would greatly facilitate the integration of such factors with plant performance and help in generating accurate metabolic markers.

5.2 Sampling procedure: easy to harvest and process

Wen et al. (2015) studied the predictive power of metabolomic data obtained from different organs/stages for agronomical traits in a maize population (leaves at seedling and reproductive stages and kernels at 15 days after pollination). Only 33 of the 79 identified metabolites were commonly detected between these organs/stages and the evaluated agronomical traits were predicted by different combinations of metabolites depending on the sampling matrix. Metabolic marker selection might therefore be conditioned by both the organ/tissue and the developmental stage at sampling time, and also largely depend on the trait studied. Pragmatically, metabolic markers would be sought at young developmental stages first in order to reduce screening costs, and in leaves, which are easy to collect, handle and analyze. Furthermore, it seems logical that the later the samples are taken during development, the greater the chances of finding good correlations between metabolite levels and traits of interest. Thus, taking samples as early as possible in plant development would result in robust prediction and metabolic markers. Finally, the best option for each case needs to be carefully evaluated and pondered considering the expected results and required investment.

5.3 Number of metabolic markers vs sample size: finding the right balance for cost efficiency

Although targeted metabolite profiling by electrospray ionization tandem mass spectrometry allows hundreds of metabolites to be measured in thousands of samples for human Genome-Wide Association Studies (Gieger et al. 2008), in depth metabolomics remains too costly for the analysis of very large numbers of plant genotypes (ranging from 30 to 300 € per sample; Gibon et al. 2012). In other words, when looking for associations with plant performance, 'metabotyping' every genotype appears to be impossible at a reasonable cost so subpanels have to be designed. Subpanel selection is rarely described in detail. One possibility is to maximize diversity based on

phenotypic or molecular data (Rincent et al. 2014). The constitution of bulks of extreme genotypes has been widely used for genomics (Zou et al. 2016) and has been successfully tested for metabolic data (Zhang et al. 2010). Numerous sampling survey methods exist (Singh and Singh Mangat 1996) but their adaptability to plant metabolomics data is uncertain and has received little attention to date. We foresee two possible non-mutually exclusive options for in depth metabolomics analysis:

- Untargeted metabolic phenotyping in diversity subpanels

Subpanels of highly diverse genotypes and/or given growth scenarios could be investigated first by using non-targeted analytical approaches and identifying the best markers, thus keeping costs acceptable by reducing the sample number. The number of potential metabolic markers generated via untargeted analysis could then be reduced by selecting those that provide good discrimination between genotypes, environments and their interactions, on the one hand, and which are easily amenable to high-throughput on the other. Targeted methods would then be developed to characterize the full panel and/or the full set of growth conditions. If the metabolic marker has been generated through LC–MS technology, the development of a targeted method requires accurate annotation of the compound. Readers are referred to (Wolfender et al. 2015) as a guideline for annotation in complex extracts.

- Targeted measurements

Such measurements should enable high numbers of samples to be processed at low costs, thus enabling screens of large populations and/or complex experimental setups (diverse growth scenarios, developmental stages, etc.). For example, LC–MS targeted profiles could be generated automatically at moderate cost (50–100 € per sample; Heuberger et al. 2014). Sample preparation and equipment investment still account for a large part of LC–MS analysis costs and they can both be improved by automation and increase in throughput (de Raad et al. 2016; Novakova 2013). The cost of data handling, curation and analysis also has to be taken into account (Anonymous 2016a).

High-throughput spectrophotometric analysis of major sugars and organic acids, which are respectively powerful predictors of potato quality (Steinfath et al. 2010b) and tomato (López et al. 2015), could be easily implemented in facilities using robotized microplate measurements (Ménard et al. 2014) and for less than 20 € per sample. However, for many volatile compounds and secondary metabolites, there will still be certain limitations to reducing costs by methodologic adaptations (Kallenbach et al. 2014), although future developments may offer new possibilities.
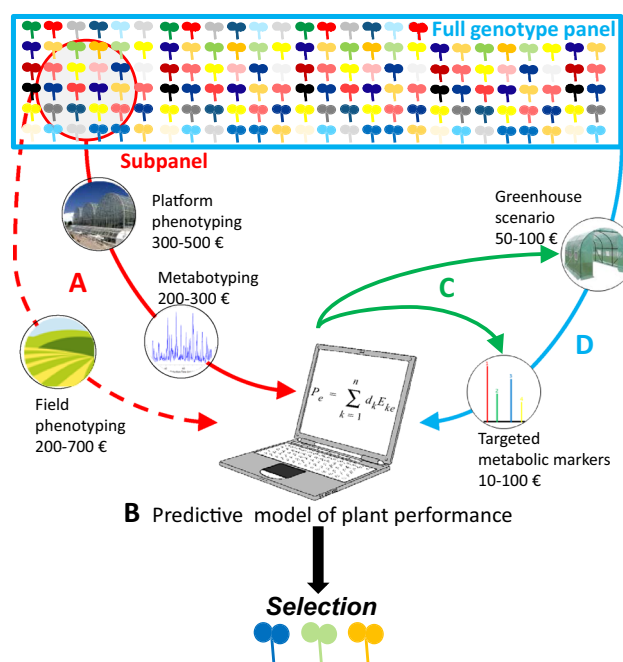

Fig. 1 Strategy combining phenotyping, metabotyping and modeling for selection in order to find a few performing genotypes from a full panel of genotypes for a given criterion. Metabolic marker may optimize cost and speed of the process by (*A*) "metabotyping" and precision phenotyping of a diversity subpanel in a series of representative environmental conditions, (*B*) using collected data to model genotype performance. The model would generate a workable combination of (*C*) adapted growth scenario, sampling procedure and a small cost-efficient set of metabolic markers which would be used for (*D*) validation on the full panel of genotypes or for a further selection program. For the purpose of estimating costs, we consider 1 sample per genotype as a pool of 5 plants

5.4 Data analysis for modeling plant performance: custom-made solutions

Detection of markers is linked to the idea of associating explanatory variables (X, markers) and response variables (Y, targeted phenotype). Therefore, an appropriate statistical method estimating such an association between metabolites or metabolite signatures and phenotypic variables and its significance is necessary.

In the simplest scenario where one metabolite is highly correlated to the targeted phenotypic trait, a pair-wise Pearson's correlation might be sufficient to detect an appropriate marker. However, a more likely situation is that more than one metabolite is needed to build a predictive model. In such cases, some commonly applied statistical methods are used to maximize the correlation between X and Y. Among them, canonical correlation analysis (CCA) estimates the maximum correlation between linear combinations of X and Y matrices, while stepwise regression and best subset regression aim at maximizing the correlation by selecting a minimum
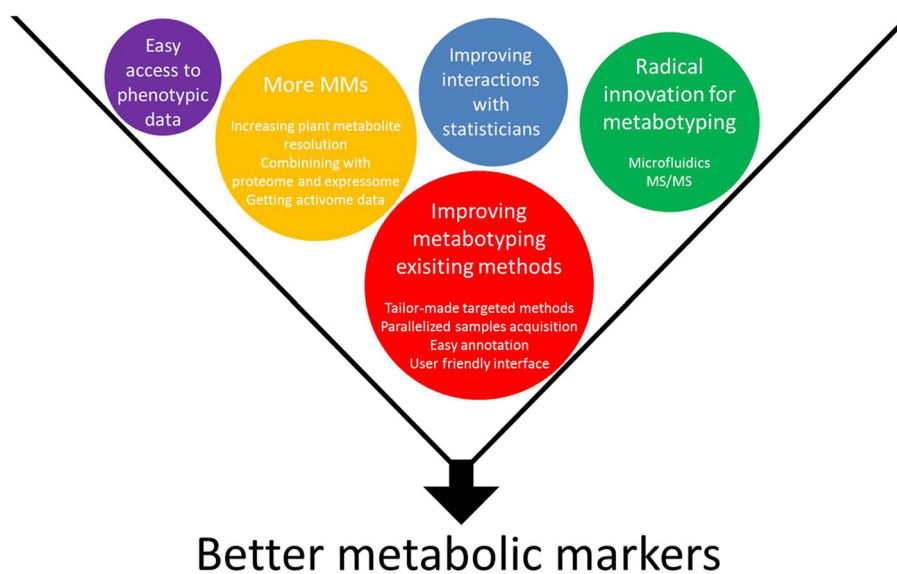
Fig. 2 Key milestones for improving and developing the use of metabolic markers

number of variables in X that predict Y (Song et al. 2016). Other very widespread methods are used to maximize covariance. If genotypes can be easily grouped in a few clusters based on their agronomical performance(s), these groups can be used to search for biomarkers using discriminant analysis. Partial Least Square Discriminant Analysis (PLS-DA) maximizes covariance between X and Y, thereby reducing the explanatory variables to a set of PLS components whose optimal number is selected by cross-validation. PLS methods have the advantage of handling highly collinear and noisy datasets (Wold et al. 2001), as is the case for most metabolomics data sets. A variant of PLS, Orthogonal Partial Least Squares (OPLS), reduces the noise effect by splitting variation in X matrix between correlated (predictive) and uncorrelated (orthogonal) to Y. This orthogonal signal correction aims at maximizing the explained covariance between X and Y on the first OPLS component while the subsequent components explain the uncorrelated variance to Y (Trygg and Wold 2002). (O)PLS statistical validation is performed by random permutation of labels and by dividing the samples into two random groups, one of them aiming to fit a model and the other to estimate its predictive power or quality. In addition, (O)PLS allows variable selection among X variables through several statistics, variable importance in projection (VIP) being the most commonly known but not the only one (Galindo-Prieto et al. 2014; Mehmood et al. 2012). Although these are very popular methods in metabolomics, there are other appropriate alternatives like principal component-discriminant function analysis, support vector machines and random forest (Gromski et al. 2015). All the above multivariate methods are prone to

overfitting, so validation with a different dataset from the one used to fit the model is mandatory.

A possibility is to begin a metabolic marker search process using the following workflow. Normalization has to be done first according to data scale and heteroscedasticity (van den Berg et al. 2006). Log 2 normalization is often preferred for univariate analysis, whereas Z-score or Pareto normalization is done before multivariate analysis. The data matrix is first analyzed with a univariate method (e.g. one or two-factor ANOVA, possibly genotype and treatment) to obtain the most significant metabolites affected by each factor and to check whether genotype x treatment interactions are present. Some highly correlated variables may also be removed at this stage to improve further modeling. Multivariate unsupervised analyses (PCA) are generally performed to give a global snapshot of the data and check for outlier samples. Finally, supervised methods such as PLS-DA and OPLS-DA are carried out. They provide VIP values that can be used to select potential candidates for metabolic markers. In parallel, machine learning methods (random forest, neural network…) might be applied but their use is still limited in plant metabolomics. Note that this analytic procedure is given as a basic guideline and should be adapted for each target and type of data matrix, then complemented with other statistical methods.

5.5 The example of plant breeding

To illustrate and summarize the search for and use of metabolic markers, an example of a pipeline for plant breeding is given in Fig. 1: (1) 'Metabotyping' of smaller

representative subpanels of genotypes [see for instance Rincent et al. (2012) for discussion on panel sampling in a predictive context] is performed in parallel with acquisition of other phenotypic variables of interest in the field or on phenotypic platforms (Fig. 1A). (2) These data are used to train models estimating traits of interest (Fig. 1B) and aiming at optimizing growth conditions and sampling, and if possible, at reducing the number of metabolic markers (Fig. 1C). (3) With such optimization, a small set of metabolic markers (10–20 markers) can be measured at a cost of 10–100 € per sample in a breeding pipeline (as shown in Fig. 1; e.g. a pool of 5 individuals from the same genotype), making it possible to use them for full diversity panels (Fig. 1D). The estimated cost for use of a molecular marker is between 10 and 30 € per sample and they will continue to be improved thanks to sequencing technologies (Next-generation sequencing, Genotyping by Sequencing). Nevertheless, if the proposed pipeline is carefully followed, metabolic markers would be able to compete with molecular markers based on relevance rather than just on cost in certain situations.

6 Conclusion

Metabolites have a great potential as markers of plant performance because they contain more information in certain scenarios and give a more realistic picture of 'real' plant performance than molecular markers. Indeed, leading biotech companies have already or are in the process of integrating these tools in their crop selection projects (Venkatesh et al. 2016; Baniasadi et al. 2014).

However, if metabolic markers are to express their full potential, several technological breakthroughs will be needed (Fig. 2). Available analytical methods have to be democratized and made more user-friendly, especially the possibility of parallelizing sample flow and data acquisition (Deng et al. 2002). Furthermore, solvent quantities need to be reduced by using micro-fluidic devices (Gao et al. 2013) and tailor-made targeted methods able to measure 10–20 metabolic markers simultaneously need to be developed. Dedicated new methods with metabolite sensors using microfluidics could be used for plant samples, as is already the case in human health (Tharakan et al. 2015). In addition to the development of methods for the parallel measurement of individual small molecules such as ELAKCA (a sandwich-type enzyme-linked assay), breeding would benefit from a tunable platform in which such assays could be easily adapted to each specific marker (Chovelon et al. 2016). Methods targeting other types of metabolic markers such as transcripts or proteins could also be implemented. Thus, enzymatic activities could well prove to be efficient markers as well since they correlate poorly with

metabolites (Sulpice et al. 2010) and would therefore add a new layer of information for modeling plant performance. Closer collaboration between statisticians and bioinformaticians is required and plant scientists need to become more familiar with advanced statistical methods.

Finally, phenotypic data on existing genotypes should be made more accessible because they offer a great potential for correlating or associating putative markers with known genotype performance. This is clearly the goal of the DivSeek consortium (Anonymous 2015) but other initiatives, be they public or private, should be fostered.

Compliances with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical approval This study does not involve the use of animal or human samples.

References

Anonymous. (2015). Growing access to phenotype data. [Editorial]. *Nature Genetics, 47*(2), 99–99. doi:10.1038/ng.3213.

Anonymous. (2016a). FAIR principles for data stewardship. [Editorial]. *Nature Genetics, 48*(4), 343–343. doi:10.1038/ng.3544.

Anonymous. (2016b). Purple plants. [Editorial]. *Nature Genetics, 48*(6), 587–587. doi:10.1038/ng.3585.

Aronson, J. K. (2005). Biomarkers and surrogate endpoints. *British Journal of Clinical Pharmacology, 59*(5), 491–494. doi:10.1111/j.1365-2125.2005.02435.x.

Asiago, V. M., Hazebroek, J., Harp, T., & Zhong, C. (2012). Effects of genetics and environment on the metabolome of commercial maize hybrids: A multisite study. *Journal of Agricultural and Food Chemistry, 60*(46), 11498–11508. doi:10.1021/jf303873a.

Austdal, M., Tangeras, L. H., Skrastad, R. B., Salvesen, K., Austgulen, R., Iversen, A. C., et al. (2015). First trimester urine and serum metabolomics for prediction of preeclampsia and gestational hypertension: A prospective screening study. *International Journal of Molecular Sciences, 16*(9), 21520–21538. doi:10.3390/ijms160921520.

Baniasadi, H., Vlahakis, C., Hazebroek, J., Zhong, C., & Asiago, V. (2014). Effect of environment and genotype on commercial maize hybrids using LC/MS-based metabolomics. *Journal of

Agricultural and Food Chemistry, 62(6), 1412–1422. doi:10.1021/jf404702g.

Bradbury, J. H., Egan, S. V., & Lynch, M. J. (1991). Analysis of cyanide in cassava using acid hydrolysis of cyanogenic glucosides. *Journal of the Science of Food and Agriculture, 55*(2), 277–290. doi:10.1002/jsfa.2740550213.

Cabrera-Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., & Tardieu, F. (2016). High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytologist, n/a-n/a,.* doi:10.1111/nph.14027.

Choi, Y. H., Tapias, E. C., Kim, H. K., Lefeber, A. W., Erkelens, C., Verhoeven, J. T., et al. (2004). Metabolic discrimination of *Catharanthus roseus* leaves infected by phytoplasma using 1H-NMR spectroscopy and multivariate data analysis. *Plant Physiology, 135*(4), 2398–2410. doi:10.1104/pp.104.041012.

Chovelon, B., Durand, G., Dausse, E., Toulmé, J.-J., Faure, P., Peyrin, E., et al. (2016). ELAKCA: Enzyme-linked aptamer kissing complex Assay as a small molecule sensing platform. *Analytical Chemistry, 88*(5), 2570–2575. doi:10.1021/acs.analchem.5b04575.

Coquin, L., Feala, J. D., McCulloch, A. D., & Paternostro, G. (2008). Metabolomic and flux-balance analysis of age-related decline of hypoxia tolerance in *Drosophila muscle* tissue. *Molecular Systems Biology, 4*, 233. doi:10.1038/msb.2008.71.

Cubero-Leon, E., Peñalver, R., & Maquet, A. (2014). Review on metabolomics for food authentication. *Food Research International, 60*, 95–107. doi:10.1016/j.foodres.2013.11.041.

Cuperlovic-Culf, M., Wang, L., Forseille, L., Boyle, K., Merkley, N., Burton, I., et al. (2016). Metabolic biomarker panels of response to fusarium head blight infection in different wheat varieties. *PLoS One, 11*(4), e0153642. doi:10.1371/journal.pone.0153642.

Cynkar, W., Dambergs, R., Smith, P., & Cozzolino, D. (2010). Classification of Tempranillo wines according to geographic origin: Combination of mass spectrometry based electronic nose and chemometrics. *Analytica Chimica Acta, 660*(1–2), 227–231. doi:10.1016/j.aca.2009.09.030.

de Raad, M., Fischer, C. R., & Northen, T. R. (2016). High-throughput platforms for metabolomics. *Current Opinion in Chemical Biology, 30*, 7–13. doi:10.1016/j.cbpa.2015.10.012.

Degenkolbe, T., Do, P. T., Kopka, J., Zuther, E., Hincha, D. K., & Köhl, K. I. (2013). Identification of drought tolerance markers in a diverse population of rice cultivars by expression and metabolite profiling. *PLoS One, 8*(5), e63637. doi:10.1371/journal.pone.0063637.

Deng, Y., Wu, J.-T., Lloyd, T. L., Chi, C. L., Olah, T. V., & Unger, S. E. (2002). High-speed gradient parallel liquid chromatography/tandem mass spectrometry with fully automated sample preparation for bioanalysis: 30 seconds per sample from plasma. *Rapid Communications in Mass Spectrometry, 16*(11), 1116–1123. doi:10.1002/rcm.688.

Díaz, R., Pozo, O. J., Sancho, J. V., & Hernández, F. (2014). Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *Food Chemistry, 157*, 84–93. doi:10.1016/j.foodchem.2014.02.009.

Dib, T. A., Monneveux, P., Acevedo, E., & Nachit, M. M. (1994). Evaluation of proline analysis and chlorophyll fluorescence quenching measurements as drought tolerance indicators in durum wheat (*Triticum turgidum* L. var. durum). *Euphytica, 79*(1), 65–73. doi:10.1007/bf00023577.

Downey, R. K., & Harvey, B. L. (1963). Methods of breeding for oil quality in rape. *Canadian Journal of Plant Science, 43*(3), 271–275. doi:10.4141/cjps63-054.

Feussner, I., & Polle, A. (2015). What the transcriptome does not tell—Proteomics and metabolomics are closer to the plants'

patho-phenotype. *Current Opinion in Plant Biology, 26*, 26–31. doi:10.1016/j.pbi.2015.05.023.

Fitzgerald, M. A., McCouch, S. R., & Hall, R. D. (2009). Not just a grain of rice: The quest for quality. *Trends in Plant Science, 14*(3), 133–139. doi:10.1016/j.tplants.2008.12.004.

Fraire-Velázquez, S. L., & Balderas-Hernández, V. E. (2013). Abiotic stress in plants and metabolic responses. In K. Vahdati & C. Leslie (Eds.), *Abiotic stress—Plant responses and applications in agriculture* (pp. 25–46). Rijeka: INTECH.

Fridman, E., Pleban, T., & Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proceedings of the National Academy of Sciences, 97*(9), 4718–4723.

Furbank, R. T., & Tester, M. (2011). Phenomics—Technologies to relieve the phenotyping bottleneck. *Trends in Plant Science, 16*(12), 635–644. doi:10.1016/j.tplants.2011.09.005.

Galindo-Prieto, B., Eriksson, L., & Trygg, J. (2014). Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *Journal of Chemometrics, 28*(8), 623–632. doi:10.1002/cem.2627.

Gao, D., Liu, H., Jiang, Y., & Lin, J.-M. (2013). Recent advances in microfluidics combined with mass spectrometry: Technologies and applications. *Lab on a Chip, 13*(17), 3309–3322. doi:10.1039/C3LC50449B.

Gibon, Y., Blaesing, O. E., Hannemann, J., Carillo, P., Hohne, M., Hendriks, J. H., et al. (2004). A Robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: Comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. *The Plant Cell, 16*(12), 3304–3325. doi:10.1105/tpc.104.025973.

Gibon, Y., Rolin, D., Deborde, C., Bernillon, S., & Moing, A. (2012). New opportunities in metabolomics and biochemical phenotyping for plant systems biology. In D. U. Roessner (Ed.), *Metabolomics* (p. 374). Rijeka: INTECH.

Gieger, C., Geistlinger, L., Altmaier, E., Hrabe de Angelis, M., Kronenberg, F., Meitinger, T., et al. (2008). Genetics meets metabolomics: A genome-wide association study of metabolite profiles in human serum. *PLoS Genetics, 4*(11), e1000282. doi:10.1371/journal.pgen.1000282.

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., et al. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis—A marriage of convenience or a shotgun wedding. *Analytica Chimica Acta, 879*, 10–23. doi:10.1016/j.aca.2015.02.012.

Gupta, P. K., Langridge, P., & Mir, R. R. (2010). Marker-assisted wheat breeding: Present status and future possibilities. *Molecular Breeding, 26*(2), 145–161. doi:10.1007/s11032-009-9359-7.

Hannemann, J., Poorter, H., Usadel, B., Blasing, O. E., Finck, A., Tardieu, F., 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 and Environment, 32*(9), 1185–1200. doi:10.1111/j.1365-3040.2009.01964.x.

Harrigan, G. G., Skogerson, K., MacIsaac, S., Bickel, A., Perez, T., & Li, X. (2015). Application of (1)h NMR profiling to assess seed metabolomic diversity. A case study on a soybean era population. *Journal of Agricultural and Food Chemistry, 63*(18), 4690–4697. doi:10.1021/acs.jafc.5b01069.

Hayashi, S., Akiyama, S., Tamaru, Y., Takeda, Y., Fujiwara, T., Inoue, K., et al. (2009). A novel application of metabolomics in vertebrate development. *Biochemical and Biophysical Research Communications, 386*(1), 268–272. doi:10.1016/j.bbrc.2009.06.041.

Hayashi, S., Yoshida, M., Fujiwara, T., Maegawa, S., & Fukusaki, E. (2011). Single-embryo metabolomics and systematic prediction

of developmental stage in zebrafish. *Zeitschrift fur Naturforschung. C. Journal of Biosciences, 66*(3–4), 191–198.

Hayat, S., Hayat, Q., Alyemeni, M. N., Wani, A. S., Pichtel, J., & Ahmad, A. (2012). Role of proline under changing environments. *Plant Signaling & Behavior, 7*(11), 1456–1466. doi:10.4161/psb.21949.

Hazebroek, J., Harp, T., Shi, J., & Wang, H. (2007). Metabolomic analysis of low phytic acid maize kernels. In B. J. Nikolau & E. S. Wurtele (Eds.), *Concepts in plant metabolomics* (pp. 221–238). Dordrecht: Springer.

Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Science, 49*(1), 1–12. doi:10.2135/cropsci2008.08.0512.

Herrmann, A., & Schauer, N. (2013). Metabolomics-assisted plant breeding. In *The handbook of plant metabolomics* (pp. 245–254). New York: Wiley, KGaA.

Heuberger, A. L., Broeckling, C. D., Kirkpatrick, K. R., & Prenni, J. E. (2014). Application of nontargeted metabolite profiling to discover novel markers of quality traits in an advanced population of malting barley. *Plant Biotechnology Journal, 12*(2), 147–160. doi:10.1111/pbi.12122.

Hou, Y., Yin, M., Sun, F., Zhang, T., Zhou, X., Li, H., et al. (2014). A metabolomics approach for predicting the response to neoadjuvant chemotherapy in cervical cancer patients. *Molecular BioSystems, 10*(8), 2126–2133. doi:10.1039/c4mb00054d.

Hughes, S. L., Bundy, J. G., Want, E. J., Kille, P., & Stürzenbaum, S. R. (2009). The metabolomic responses of caenorhabditis elegans to cadmium are largely independent of metallothionein status, but dominated by changes in cystathionine and phytochelatins. *Journal of Proteome Research, 8*(7), 3512–3519. doi:10.1021/pr9001806.

Humplík, J. F., Lazár, D., Husičková, A., & Spíchal, L. (2015). Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses—A review. *Plant Methods, 11*(1), 1–10. doi:10.1186/s13007-015-0072-8.

Jiang, Y., Djuric, Z., Sen, A., Ren, J., Kuklev, D., Waters, I., et al. (2014). Biomarkers for personalizing omega-3 fatty acid dosing. *Cancer Prevention Research (Philadelphia, Pa.), 7*(10), 1011–1022. doi:10.1158/1940-6207.capr-14-0134.

Kang, J. W., Kim, H.-T., Lee, W. Y., Choi, M. N., Park, E.-J. (2015) Identification of a potential metabolic marker, inositol, for the inherently fast growth trait by stems of via a retrospective approach. *Canadian Journal of Forest Research, 45*(6), 770–775.

Justes, E., Meynard, J. M., Mary, B., & Plénet, D. (1997). Diagnosis using stem base extract: JUBIL method. In G. Lemaire (Ed.), *Diagnosis of the nitrogen status in crops* (pp. 163–187). Berlin: Springer.

Kallenbach, M., Oh, Y., Eilers, E. J., Veit, D., Baldwin, I. T., & Schuman, M. C. (2014). A robust, simple, high-throughput technique for time-resolved plant volatile analysis in field experiments. *The Plant Journal, 78*(6), 1060–1072. doi:10.1111/tpj.12523.

Kauppi, A. M., Edin, A., Ziegler, I., Mölling, P., Sjöstedt, A., Gylfe, Å., et al. (2016). Metabolites in blood for prediction of bacteremic sepsis in the emergency room. *PLoS One, 11*(1), e0147670. doi:10.1371/journal.pone.0147670.

Korn, M., Gartner, T., Erban, A., Kopka, J., Selbig, J., & Hincha, D. K. (2010). Predicting Arabidopsis freezing tolerance and heterosis in freezing tolerance from metabolite composition. *Molecular Plant, 3*(1), 224–235. doi:10.1093/mp/ssp105.

Korn, M., Peterek, S., Mock, H. P., Heyer, A. G., & Hincha, D. K. (2008). Heterosis in the freezing tolerance, and sugar and flavonoid contents of crosses between *Arabidopsis thaliana* accessions of widely varying freezing tolerance. *Plant, Cell and Environment, 31*(6), 813–827. doi:10.1111/j.1365-3040.2008.01800.x.

Kumar, J., Pratap, A., & Kumar, S. (2015). Plant phenomics: An overview. In J. Kumar, A. Pratap, & S. Kumar (Eds.), *Phenomics in crop plants: Trends, options and limitations* (pp. 1–10). New Delhi: Springer.

Kusano, M., Baxter, I., Fukushima, A., Oikawa, A., Okazaki, Y., Nakabayashi, R., et al. (2015). Assessing metabolomic and chemical diversity of a soybean lineage representing 35 years of breeding. *Metabolomics, 11*(2), 261–270. doi:10.1007/s11306-014-0702-6.

Lee, J.-E., Lee, B.-J., Chung, J.-O., Kim, H.-N., Kim, E.-H., Jung, S., et al. (2015). Metabolomic unveiling of a diverse range of green tea (*Camellia sinensis*) metabolites dependent on geography. *Food Chemistry, 174*, 452–459. doi:10.1016/j.foodchem.2014.11.086.

Lee, S. C., Tan, H. T., & Chung, M. C. M. (2014). Prognostic biomarkers for prediction of recurrence of hepatocellular carcinoma: Current status and future prospects. *World Journal of Gastroenterology, 20*(12), 3112–3124. doi:10.3748/wjg.v20.i12.3112.

Lima, M. R., Felgueiras, M. L., Graca, G., Rodrigues, J. E., Barros, A., Gil, A. M., et al. (2010). NMR metabolomics of esca disease-affected *Vitis vinifera* cv. Alvarinho leaves. *Journal of Experimental Botany, 61*(14), 4033–4042. doi:10.1093/jxb/erq214.

Lindon, J. C., & Nicholson, J. K. (2014). The emergent role of metabolic phenotyping in dynamic patient stratification. *Expert Opinion on Drug Metabolism & Toxicology, 10*(7), 915–919. doi:10.1517/17425255.2014.922954.

Liu, Y., Yu, P., Sun, X., & Di, D. (2012). Metabolite target analysis of human urine combined with pattern recognition techniques for the study of symptomatic gout. *Molecular BioSystems, 8*(11), 2956–2963. doi:10.1039/c2mb25227a.

López, M. G., Zanor, M. I., Pratta, G. R., Stegmayer, G., Boggio, S. B., Conte, M., et al. (2015). Metabolic analyses of interspecific tomato recombinant inbred lines for fruit quality improvement. *Metabolomics, 11*(5), 1416–1431. doi:10.1007/s11306-015-0798-3.

Malmendal, A., Overgaard, J., Bundy, J. G., Sørensen, J. G., Nielsen, N. C., Loeschcke, V., et al. (2006). Metabolomic profiling of heat stress: hardening and recovery of homeostasis in Drosophila. *American Journal of Physiology—Regulatory, Integrative and Comparative Physiology, 291*(1), R205–R212. doi:10.1152/ajpregu.00867.2005.

McDunn, J. E., Li, Z., Adam, K.-P., Neri, B. P., Wolfert, R. L., Milburn, M. V., et al. (2013). Metabolomic signatures of aggressive prostate cancer. *The Prostate, 73*(14), 1547–1560. doi:10.1002/pros.22704.

Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 118*, 62–69. doi:10.1016/j.chemolab.2012.07.010.

Ménard, G., Biais, B., Prodhomme, D., Ballias, P., & Gibon, Y. (2014). Analysis of enzyme activities. In M. Dieuaide-Noubhani & P. A. Alonso (Eds.), *Plant metabolic flux analysis: Methods and protocols* (pp. 249–259). Totowa, NJ: Humana Press.

Menard, G. E., Grant, P. J., Cohn, S. L., & Smetana, G. W. (2013). Update in perioperative medicine 2012. *Hospital Practice (1995), 41*(2), 85–92. doi:10.3810/hp.2013.04.1050.

Meyer, R. C., Steinfath, M., Lisec, J., Becher, M., Witucka-Wall, H., Törjék, O., et al. (2007). The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, 104*(11), 4759–4764.

Nam, H. M., Bang, E., Kwon, Y. T., Kim, Y., Kim, H. E., Cho, K., et al. (2015). Metabolite profiling of diverse rice germplasm and identification of conserved metabolic markers of rice roots in

response to long-term mild salinity stress. *International Journal of Molecular Sciences, 16*(9), 21959–21974. doi:10.3390/ijms160921959.

Nambisan, B. (2011). Strategies for elimination of cyanogens from cassava for reducing toxicity and improving food safety. *Food and Chemical Toxicology, 49*(3), 690–693. doi:10.1016/j.fct.2010.10.035.

Nicholson, J. K., Holmes, E., Kinross, J. M., Darzi, A. W., Takats, Z., & Lindon, J. C. (2012). Metabolic phenotyping in clinical and surgical environments. *Nature, 491*(7424), 384–392. doi:10.1038/nature11708.

Nicholson, J. K., Holmes, E., & Lindon, J. C. (2007). Chapter 1— Metabonomics and metabolomics techniques and their applications in mammalian systems. In *The handbook of metabonomics and metabolomics* (pp. 1–33). Amsterdam: Elsevier.

Nöh, K., Grönke, K., Luo, B., Takors, R., Oldiges, M., & Wiechert, W. (2007). Metabolic flux analysis at ultra short time scale: Isotopically non-stationary 13C labeling experiments. *Journal of Biotechnology, 129*(2), 249–267. doi:10.1016/j.jbiotec.2006.11.015.

Novakova, L. (2013). Challenges in the development of bioanalytical liquid chromatography-mass spectrometry method with emphasis on fast analysis. *Journal of Chromatography A, 1292*, 25–37. doi:10.1016/j.chroma.2012.08.087.

Obata, T., Witt, S., Lisec, J., Palacios-Rojas, N., Florez-Sarasa, I., Yousfi, S., et al. (2015). Metabolite profiles of maize leaves in drought, heat, and combined stress field trials reveal the relationship between metabolism and grain yield. *Plant Physiology, 169*(4), 2665–2683. doi:10.1104/pp.15.01164.

Oms-Oliu, G., Odriozola-Serrano, I., & Martín-Belloso, O. (2013). Metabolomics for assessing safety and quality of plant-derived food. *Food Research International, 54*(1), 1172–1183. doi:10.1016/j.foodres.2013.04.005.

Parent, B., Turc, O., Gibon, Y., Stitt, M., & Tardieu, F. (2010). Modelling temperature-compensated physiological rates, based on the coordination of responses to temperature of developmental processes. *Journal of Experimental Botany, 61*(8), 2057–2069.

Pissard, A., Fernández Pierna, J. A., Baeten, V., Sinnaeve, G., Lognay, G., Mouteau, A., et al. (2013). Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture, 93*(2), 238–244. doi:10.1002/jsfa.5779.

Putri, S. P., Nakayama, Y., Matsuda, F., Uchikata, T., Kobayashi, S., Matsubara, A., et al. (2013). Current metabolomics: Practical applications. *Journal of Bioscience and Bioengineering, 115*(6), 579–589. doi:10.1016/j.jbiosc.2012.12.007.

Quistián-Martínez, D., Estrada-Luna, A. A., Altamirano-Hernández, J., Peña-Cabriales, J. J., Oca-Luna, R. M., & Cabrera-Ponce, J. L. (2011). Use of trehalose metabolism as a biochemical marker in rice breeding. *Molecular Breeding, 30*(1), 469–477. doi:10.1007/s11032-011-9636-0.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., et al. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology, 19*(1), 45–50. doi:10.1038/83496.

Riedelsheimer, C., Brotman, Y., Méret, M., Melchinger, A. E., & Willmitzer, L. (2013). The maize leaf lipidome shows multilevel genetic control and high predictive value for agronomic traits. *Scientific Reports, 3*, 2479. doi:10.1038/srep02479.

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., & Sulpice, R., et al. (2012a). Genomic and metabolic prediction of complex heterotic traits in hybrid maize.

Nature Genetics, 44(2), 217–220. http://www.nature.com/ng/journal/v44/n2/abs/ng.1033.html#supplementary-information.

Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., et al. (2012b). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proceedings of the National Academy of Sciences of the USA, 109*(23), 8872–8877. doi:10.1073/pnas.1120813109.

Rincent, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., et al. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize Inbreds (*Zea mays* L.). *Genetics, 192*(2), 715–728. doi:10.1534/genetics.112.141473.

Rincent, R., Nicolas, S., Bouchet, S., Altmann, T., Brunel, D., Revilla, P., et al. (2014). Dent and Flint maize diversity panels reveal important genetic potential for increasing biomass production. *Theoretical and Applied Genetics, 127*(11), 2313–2331. doi:10.1007/s00122-014-2379-7.

Robinette, S. L., Lindon, J. C., & Nicholson, J. K. (2013). Statistical spectroscopic tools for biomarker discovery and systems medicine. *Analytical Chemistry, 85*(11), 5297–5303. doi:10.1021/ac4007254.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N., & Willmitzer, L. (2000). Technical advance: Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *The Plant Journal, 23*(1), 131–142. doi:10.1046/j.1365-313x.2000.00774.x.

Ruiz-García, L., Hellín, P., Flores, P., & Fenoll, J. (2014). Prediction of Muscat aroma in table grape by analysis of rose oxide. *Food Chemistry, 154*, 151–157. doi:10.1016/j.foodchem.2014.01.005.

Sade, D., Shriki, O., Cuadros-Inostroza, A., Tohge, T., Semel, Y., Haviv, Y., et al. (2015). Comparative metabolomics and transcriptomics of plant response to Tomato yellow leaf curl virus infection in resistant and susceptible tomato cultivars. *Metabolomics, 11*(1), 81–97. doi:10.1007/s11306-014-0670-x.

Schmidtke, L. M., Blackman, J. W., Clark, A. C., & Grant-Preece, P. (2013). Wine metabolomics: Objective measures of sensory properties of semillon from GC-MS profiles. *Journal of Agricultural and Food Chemistry, 61*(49), 11957–11967. doi:10.1021/jf403504p.

Singh, R., & Singh Mangat, N. (1996). *Elements of survey sampling* (1 edn., Texts in the Mathematical Sciences, Vol. 15). Dordrecht: Springer.

Song, Y., Schreier, P. J., Ramírez, D., & Hasija, T. (2016). Canonical correlation analysis of high-dimensional data with very small sample support. *Signal Processing, 128*, 449–458. doi:10.1016/j.sigpro.2016.05.020.

Steinfath, M., Gärtner, T., Lisec, J., Meyer, R. C., Altmann, T., Willmitzer, L. et al. (2010a). Prediction of hybrid biomass in Arabidopsis thaliana by selected parental SNP and metabolic markers. *Theoretical and Applied Genetics, 120*(2), 239–247. doi:10.1007/s00122-009-1191-2.

Steinfath, M., Strehmel, N., Peters, R., Schauer, N., Groth, D., Hummel, J., et al. (2010b). Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. *Plant Biotechnology Journal, 8*(8), 900–911. doi:10.1111/j.1467-7652.2010.00516.x.

Straadt, I. K., Aaslyng, M. D., & Bertram, H. C. (2014). An NMR-based metabolomics study of pork from different crossbreeds and relation to sensory perception. *Meat Science Part A, 96*(2), 719–728. doi:10.1016/j.meatsci.2013.10.006.

Strand, A., Hurry, V., Henkes, S., Huner, N., Gustafsson, P., Gardestrom, P., et al. (1999). Acclimation of Arabidopsis leaves developing at low temperatures. Increasing cytoplasmic volume accompanies increased activities of enzymes in the Calvin cycle

and in the sucrose-biosynthesis pathway. *Plant Physiology, 119*(4), 1387–1398.

Sulpice, R., Pyl, E.-T., Ishihara, H., Trenkamp, S., Steinfath, M., Witucka-Wall, H., et al. (2009). Starch as a major integrator in the regulation of plant growth. *Proceedings of the National Academy of Sciences of the USA, 106*(25), 10348–10353.

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. *The Plant Cell, 22*(8), 2872–2893. doi:10.1105/tpc.110.076653.

Sumner, L. W., Lei, Z., Nikolau, B. J., & Saito, K. (2015). Modern plant metabolomics: Advanced natural product gene discoveries, improved technologies, and future prospects. *Natural Product Reports, 32*(2), 212–229. doi:10.1039/C4NP00072B.

Tamaoki, M., Matsuyama, T., Nakajima, N., Aono, M., Kubo, A., & Saji, H. (2004). A method for diagnosis of plant environmental stresses by gene expression profiling using a cDNA macroarray. *Environmental Pollution, 131*(1), 137–145. doi:10.1016/j.envpol.2004.01.008.

Tardieu, F. (2011). Any trait or trait-related allele can confer drought tolerance: Just design the right drought scenario. *Journal of Experimental Botany, 63*(1), 25–31. doi:10.1093/jxb/err269.

Tardieu, F., Granier, C., & Muller, B. (2011). Water deficit and growth. Co-ordinating processes without an orchestrator? *Current Opinion in Plant Biology, 14*(3), 283–289. doi:10.1016/j.pbi.2011.02.002.

Tarr, P. T., Dreyer, M. L., Athanas, M., Shahgholi, M., Saarloos, K., & Second, T. P. (2013). A metabolomics based approach for understanding the influence of terroir in Vitis Vinifera L. *Metabolomics, 9*(1), 170–177. doi:10.1007/s11306-013-0497-x.

Tharakan, R., Tao, D., Ubaida-Mohien, C., Dinglasan, R. R., & Graham, D. R. (2015). Integrated microfluidic chip and online SCX separation allows untargeted nanoscale metabolomic and peptidomic profiling. *Journal of Proteome Research, 14*(3), 1621–1626. doi:10.1021/pr5011422.

Tisne, S., Serrand, Y., Bach, L., Gilbault, E., Ben Ameur, R., Balasse, H., et al. (2013). Phenoscope: An automated large-scale phenotyping platform offering high spatial homogeneity. *The Plant Journal, 74*(3), 534–544. doi:10.1111/tpj.12131.

Truong, M., Yang, B., & Jarrard, D. F. (2013). Toward the detection of prostate cancer in urine: A critical analysis. *The Journal of Urology, 189*(2), 422–429. doi:10.1016/j.juro.2012.04.143.

Trygg, J., & Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics, 16*(3), 119–128. doi:10.1002/cem.695.

Uddling, J., Gelang-Alfredsson, J., Piikki, K., & Pleijel, H. (2007). Evaluating the relationship between leaf chlorophyll concentration and SPAD-502 chlorophyll meter readings. *Photosynthesis Research, 91*(1), 37–46. doi:10.1007/s11120-006-9077-5.

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics, 7*(1), 1–15. doi:10.1186/1471-2164-7-142.

Venkatesh, T. V., Chassy, A. W., Fiehn, O., Flint-Garcia, S., Zeng, Q., Skogerson, K., et al. (2016). Metabolomic assessment of key maize resources: GC-MS and NMR profiling of grain from B73 Hybrids of the nested association mapping (NAM) founders and of geographically diverse landraces. *Journal of Agricultural and Food Chemistry, 64*(10), 2162–2172. doi:10.1021/acs.jafc.5b04901.

Wei, S., Liu, L., Zhang, J., Bowers, J., Gowda, G. A. N., Seeger, H., et al. (2013). Metabolomics approach for predicting response to neoadjuvant chemotherapy for breast cancer. *Molecular Oncology, 7*(3), 297–307. doi:10.1016/j.molonc.2012.10.003.

Wen, W., Li, K., Alseekh, S., Omranian, N., Zhao, L., Zhou, Y., et al. (2015). Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *The Plant Cell, 27*(7), 1839–1856. doi:10.1105/tpc.15.00208.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109–130. doi:10.1016/S0169-7439(01)00155-1.

Wolfender, J.-L., Marti, G., Thomas, A., & Bertrand, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. *Journal of Chromatography A, 1382*, 136–164. doi:10.1016/j.chroma.2014.10.091.

Wolfender, J. L., Rudaz, S., Choi, Y. H., & Kim, H. K. (2013). Plant metabolomics: From holistic data to relevant biomarkers. *Current Medicinal Chemistry, 20*(8), 1056–1090. doi:10.2174/0929867311320080009.

Xu, Y., & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: From publications to practice. *Crop Science, 48*, 391–407. doi:10.2135/cropsci2007.04.0191.

Zabotina, O. A. (2013). Metabolite-based biomarkers for plant genetics and breeding. In T. Lübberstedt & K. R. Varshney (Eds.), *Diagnostics in plant breeding* (pp. 281–309). Dordrecht: Springer.

Zajac, J., Shrestha, A., Patel, P., & Poretsky, L. (2010). The main events in the history of diabetes mellitus. In L. Poretsky (Ed.), *Principles of diabetes mellitus* (pp. 3–16). Boston, MA: Springer.

Zeng, W., Hazebroek, J., Beatty, M., Hayes, K., Ponte, C., Maxwell, C., et al. (2014). Analytical method evaluation and discovery of variation within maize varieties in the context of food safety: Transcript profiling and metabolomics. *Journal of Agricultural and Food Chemistry, 62*(13), 2997–3009. doi:10.1021/jf405652j.

Zhang, N., Gur, A., Gibon, Y., Sulpice, R., Flint-Garcia, S., McMullen, M. D., et al. (2010). Genetic analysis of central carbon metabolism unveils an amino acid substitution that alters maize NAD-Dependent isocitrate dehydrogenase activity. *PLoS One, 5*(4), e9991. doi:10.1371/journal.pone.0009991.

Zou, C., Wang, P., & Xu, Y. (2016). Bulked sample analysis in genetics, genomics and crop improvement. *Plant Biotechnology Journal,*. doi:10.1111/pbi.12559.