Journal of Plant Physiology 168 (2011) 1387–1394

Contents lists available at ScienceDirect

Journal of Plant Physiology

journal homepage: www.elsevier.de/jplph

Review

Identification of enzymatic and regulatory genes of plant metabolism through QTL analysis in *Arabidopsis*

Yariv Brotman[b,1], David Riewe[a,1], Jan Lisec[b], Rhonda C. Meyer[a], Lothar Willmitzer[b], Thomas Altmann[a,*]

[a] Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466 Gatersleben, Germany
[b] Department of Molecular Physiology, Max-Planck-Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam-Golm, Germany

ARTICLE INFO

Article history:
Received 1 October 2010
Received in revised form 20 March 2011
Accepted 21 March 2011

Keywords:
IL
mQTL
Metabolomics
RIL
Fumarase

ABSTRACT

The biochemical diversity in the plant kingdom is estimated to well exceed 100,000 distinct compounds (Weckwerth, 2003) and 4000 to 20,000 metabolites per species seem likely (Fernie et al., 2004). In recent years extensive progress has been made towards the identification of enzymes and regulatory genes working in a complex network to generate this large arsenal of metabolites. Genetic loci influencing quantitative traits, e.g. metabolites or biomass, may be mapped to associated molecular markers, a method called quantitative trait locus mapping (QTL mapping), which may facilitate the identification of novel genes in biochemical pathways. *Arabidopsis thaliana*, as a model organism for seed plants, is a suitable target for metabolic QTL (mQTL) studies due to the availability of highly developed molecular and genetic tools, and the extensive knowledge accumulated on the metabolite profile. While intensely studied, in particular since the availability of its complete sequence, the genome of *Arabidopsis* still comprises a large proportion of genes with only tentative function based on sequence homology. From a total number of 33,518 genes currently listed (TAIR 9, http://www.arabidopsis.org), only about 25% have direct experimental evidence for their molecular function and biological process, while for more than 30% no biological data are available. Modern metabolomics approaches together with continually extended genomic resources will facilitate the task of assigning functions to those genes. In our previous study we reported on the identification of mQTL (Lisec et al., 2008). In this paper, we summarize the current status of mQTL analyses and causal gene identification in *Arabidopsis* and present evidence that a candidate gene located within the confidence interval of a fumarate mQTL (*AT5G50950*) encoding a putative fumarase is likely to be the causal gene of this QTL. The total number of genes molecularly identified based on mQTL studies is still limited, but the advent of multi-parallel analysis techniques for measurement of gene expression, as well as protein and metabolite abundances and for rapid gene identification will assist in the important task of assigning enzymes and regulatory genes to the growing network of known metabolic reactions.

Contents

* Corresponding author. Tel.: +49 39482 5 214; fax: +49 39482 5 758.
E-mail address: altmann@ipk-gatersleben.de (T. Altmann).
[1] These authors contributed equally to this work.

0176-1617/$ – see front matter © 2011 Elsevier GmbH. All rights reserved.
doi:10.1016/j.jplph.2011.03.008

Introduction

Over the past decade, metabolic profiling was the focus of an increasing number of studies. Novel methods used for metabolite detection and analysis that offer robust, accurate and sensitive analysis of several hundreds of compounds have been established (Kopka et al., 2004; Lisec et al., 2006). Mapping those metabolites in segregating populations in order to obtain metabolic QTL (mQTL), will eventually lead to the identification of novel enzymatic and regulatory genes controlling diverse biochemical pathways. In *Arabidopsis* several studies focused on the identification of mQTL, mainly mapping of glucosinolate metabolites where the synthetic pathway is well known (Table 1).

Similar to mapping of metabolite QTL, levels of transcript and protein abundance were mapped to identify genomic loci controlling the observed variation in mRNA and protein levels, generating expression QTL (eQTL) and protein QTL (pQTL) (Schadt et al., 2003; Keurentjes et al., 2007; Wentzell et al., 2007; Fu et al., 2009).

Plant metabolite QTL analyses were in the focus of previous reviews (Fernie and Schauer, 2009; Keurentjes, 2009; Kliebenstein, 2009a,b). In the present overview we focus on several examples for cloning of enzymatic and regulatory factors by mQTL analysis in *Arabidopsis*. In addition, we present an example for validation of a mQTL candidate gene using reverse genetic tools for the gene *AT5G50950* which encodes a putative fumarase on chromosome V and which is located within the confidence interval of a fumarate mQTL reported in a previous study (Lisec et al., 2008).

The use of immortal populations and natural variation in plant quantitative genetics

In plants, the identification and characterization of individual genes causal for a specific phenotype (i.e. plant disease resistance genes, *R*-genes) using map-based cloning approaches became almost a routine, but laborious procedure (Jander et al., 2002). However, quantitative traits, such as biomass and growth, which show a continuous distribution of trait values throughout a population, are often controlled by more than one and potentially up to hundreds of genetic loci. Contributions of single loci may therefore be small and their identification requires precise estimation of genotypic values that can be achieved using large segregating populations. Therefore, recombinant inbred lines (RILs) and introgression lines (ILs; also termed near isogenic lines, NILs) are two types of populations, which are frequently used to investigate such quantitative traits in order to identify the underlying genes (Salvi and Tuberosa, 2005). Both are developed from a cross of two parental accessions (P1, P2) which are preferably homozygous and genetically distinct to ensure the presence of alternate alleles at loci potentially influencing the trait of interest. Due to their "immortal" nature, RIL and IL populations are suitable for monitoring metabolite level in replicated experiments under the same or different environmental conditions.

The current method of choice for the required determination of the genetic composition of these lines is the typing of single nucleotide polymorphisms (SNPs) due to their widespread occurrence, their high stability (Kruglyak, 2008), and their amenability to cost-efficient high-throughput analyses (Wang et al., 1998). In *Arabidopsis*, a re-sequencing effort led to the prediction of over one million SNPs (Clark et al., 2007). The progress in SNP detection technologies facilitated the generation of new genotyping tools (Ganal et al., 2009; Maresso and Broeckel, 2008) and the genotyping of large RIL and IL populations.

RIL and IL populations are available for several plant species like tomato, rice, maize (Burr et al., 1988; Eshed and Zamir, 1995; Li et al., 1995; Alonso-Blanco et al., 1998). For *Arabidopsis*, numerous RIL populations and some IL populations have been constructed using a wide range of diverse parental lines including Columbia-0 (Col-0) and C24 (Törjék et al., 2006, 2008) and are available to the research community (see: http://www.inra.fr/internet/Produits/vast/RILs.htm).

Segregating populations, such as RILs and ILs, derived from pairwise or multiple crosses of accessions (Paulo et al., 2008) show some inherent limitations regarding the identification and cloning of quantitative trait loci such as genetic interactions between genomic loci, low recombination frequencies and a low rate of polymorphisms between genotypes. With the advances in sequencing and genotyping technology, genome-wide association studies (GWAS) recently became available as an alternative method to classical mapping approaches. GWAS make use of the natural genotypic variation and allow the analysis of associations between hundreds of thousands of single nucleotide polymorphisms and a specific phenotype (Yu and Buckler, 2006). GWAS are based on linkage disequilibrium (LD) in accessions. In *Arabidopsis*, the LD decays rapidly (<10 kb) in a sample of accessions selected for high genetic diversity, possibly allowing near-gene-level resolution in mapping approaches (Kim et al., 2007; Atwell et al., 2010). However, spurious associations, which can arise due to an underlying population structure have to be taken into account. Currently, re-sequencing efforts are on the way for a large number of *Arabidopsis* accessions (Ossowski et al., 2008; Weigel and Mott, 2009), thus further facilitating future GWAS.

Metabolome analyses

Similar to integrative traits like biomass, growth and resistance, metabolite levels can be measured in the genetically defined populations described in the previous section and, therefore, subjected to QTL analysis. The metabolome is defined as the entire set of low molecular weight compounds of an organism and its composition is tightly linked to many traits like those mentioned above. While methods for the measurement of individual metabolites by spectrophotometric assays or simple chromatographic separation have been used for a long time, the analysis of several hundreds to thousands of compounds only started to become feasible with the hyphenation of separation methods to various detection systems (Fernie et al., 2004). The separation methods which are commonly applied include gas chromatography (GC), liquid chromatography