Plant Physiology and Biochemistry xxx (2013) 1–9

Contents lists available at SciVerse ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Research article

Selection of candidate genes for grape proanthocyanidin pathway by an integrative approach

Carrier Grégory [a,c,1], Huang Yung Fen [a,b,2], Le Cunff Loïc [a,c], Fournier-Level Alexandre [a], Vialet Sandrine [b], Souquet Jean-Marc [b], Cheynier Véronique [b], Terrier Nancy [b,*], This Patrice [a]

[a] UMR AGAP, INRA-Montpellier SupAgro-CIRAD, 2 Place Pierre Viala, F-34060 Montpellier, France
[b] UMR1083 SPO, INRA-Montpellier SupAgro-UM1, 2 Place Pierre Viala, F-34060 Montpellier, France
[c] UMT Geno-Vigne, 2 Place Viala, F-34060 Montpellier, France

ARTICLE INFO

Article history:
Received 28 November 2012
Accepted 19 April 2013
Available online xxx

Keywords:
Proanthocyanidin biosynthesis
Candidate genes
QTLs
Association genetics
Vitis vinifera L.
Transcriptomics

ABSTRACT

Proanthocyanidins (PA) play a major role in plant protection against biotic and abiotic stresses. Moreover these molecules are known to be beneficial for human health and are responsible for astringency of foods and beverages such as wine and thus have a great impact on the final quality of the product. Genes playing a role in the PA pathway are only partially known. The amount of available transcriptomic and genetic data to select candidate genes without *a priori* knowledge from orthologous function increases every day. However, the methods used so far generate so many candidate genes that it is impossible to validate all of them. In this study, we used an integrative strategy based on different screening methods to select a reduced list of candidate genes. We have crossed results from different screening methods including QTL mapping and three transcriptomic studies to select 20 candidate genes, located in QTL intervals and fulfilling at least two transcriptomic screenings. This list includes three glucosyltransferases, already suspected to have a role in the PA biosynthetic pathway. Among the 17 remaining genes, we selected three genes to perform further analysis by association genetic studies. For each of these genes, we found a polymorphism linked to PA variation. The three genes (*VvMybC2-L1*, *VvGAT-like* and *VvCob-like*), not previously known to play a role in PA synthesis, are promising candidates for further molecular physiology studies.

1. Introduction

Proanthocyanidins (PA) are secondary metabolites belonging to flavonoids and involved in plant protection against biotic and abiotic stresses [1]. From a human nutrition viewpoint, they contribute to the organoleptic properties of food products (i.e. bitterness and astringency) and their health benefit is increasingly described in the literature [2]. PAs are oligomers and polymers of flavan-3-ols. In grape, they are composed of four major subunits (catechin, epicatechin, epigallocatechin and epicatechin-3-*O*-gallate) that differ according to their 2,3 stereochemistry (catechin/epicatechin), their level of hydroxylation on the B-ring (epicatechin/epigallocatechin), and their acylation by gallic acid (epicatechin/epicatechin-3-*O*-gallate). The properties of PAs are tightly linked to their chemical structure: gallic esters of flavan-3-ols monomers and oligomers show higher antibacterial, antiviral and antioxidant activity than without the galloyl group [3–5]. Astringency properties increase with the chain length and with galloylation level [6].

PA are mostly synthesized during the early stages of grape berry development both in skin and seed, and the amount of PA on a per berry basis hardly changes between the onset of ripening and maturity [7]. Three specific PA structural genes have been isolated in grapevine (*VvLAR-1*, *VvLAR-2* and *VvANR* [8]). Genes that regulate PA synthesis have also recently been reported [9–12]. The first transcription factors whose function was validated in grape are two Myb-type proteins, VvMYBPA1 and VvMYBPA2 [9,10]. They were isolated by cloning cDNA exhibiting sequence homology with *AtTT2*, the main PA regulator in Arabidopsis [13], which specifically

* Corresponding author. Tel.: +33 (0)4 99 61 24 60; fax: +33 (0)4 99 61 28 57.
E-mail addresses: gregory.carrier@supagro.inra.fr (C. Grégory), huangyungfen@gmail.com (H.Y. Fen), loic.lecunff@supagro.inra.fr (L.C. Loïc), alexandre_fournier-level@brown.edu (F.-L. Alexandre), vialet@supagro.inra.fr (V. Sandrine), jean-marc.souquet@supagro.inra.fr (S. Jean-Marc), cheynier@supagro.inra.fr (C. Véronique), terrier@supagro.inra.fr (T. Nancy), patrice.this@supagro.inra.fr (T. Patrice).
[1] Present address: IFREMER Laboratory of PBA, rue de l'Ile d'Yeu, 44311 Nantes, France.
[2] Present address: Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, 960 Carling Ave, Ottawa, Ontario K1A 0C6, Canada.

0981-9428/$ – see front matter © 2013 Published by Elsevier Masson SAS.
http://dx.doi.org/10.1016/j.plaphy.2013.04.014

induces the expression of PA structural genes. However, several biosynthesis and regulation actors are still unknown: mechanisms involved in PA galloylation, PA transport and storage inside the cell or PA formation (i.e. polymerization) and determination of chain length still remain to be elucidated, even in model plants.

Considerable effort has been invested in identifying and characterizing genes responsible for phenotypic variation of agronomical traits in order to develop selection programs using genetic markers. Candidate genes have been identified using different strategies. *A priori* strategies rely on functional information available for model species to investigate orthologous genes in the organism of interest [14,15]. However the situation where genes with known functions are conserved between species is not the general rule in plants, and as stated above, some steps are still lacking even in model plants. In absence of *a priori* strategies, approaches used to identify candidate genes are either based on genetic mapping (quantitative trait loci (QTL) mapping), transcriptomics (RNA microarrays), proteomics or epigenetics (differential methylation hybridization or bisulfite conversion) [16,17]. However, all these methods yield so many putative candidate genes that it is impossible to validate the function for all of them.

Among strategies that do not require *a priori* knowledge of gene function, QTL mapping enables identification of the genetic intervals correlated to the phenotype variation in segregating populations. We have recently published the first QTL study for grape PA composition [18] where we have identified 46 QTLs for 13 berry skin PA traits in grapevine. However, QTL intervals are generally large in plants and can contain up to a few hundreds genes. For instance, 10 cM intervals in maize (*Zea mays*) and grapevine (*Vitis vinifera* L.) represents on average 10 and 5 Mb, respectively. Each of these intervals corresponds to approximately 790 and 270 genes [19,20]. In most cases, only one or a few genes in the QTL intervals are responsible for the observed phenotypic variation [21]. Map-based positional cloning approaches have been successfully used to identify the causal allele of phenotypic variation in several species such as barley or rice [22,23] but are time-consuming in species with a long biological cycle [21]. It is also possible to refine the QTL results through association mapping in natural populations [24]. This enables precise identification of potential causal polymorphisms by exploring the extent of genetic diversity at candidate gene loci [25,26]. Once a few consistent candidate genes are identified, functional validation becomes more straightforward. Combining QTL detection and association mapping was successful in assessing the effect of putative causal polymorphisms on anthocyanin composition and “Muscat” flavor in a broad grapevine germplasm collection [27,28].

On the other hand, transcriptomic approaches can also help in the identification of candidate genes. Terrier et al. have identified genes differentially expressed between grapevine hairy roots over-expressing *VvMybPA1* or *VvMybPA2* and wild type hairy roots [10]. However, any genes exhibiting differential expression profiles between samples can be either directly or indirectly involved in phenotypic variation. It is therefore difficult to reduce thousands of genes to a few relevant candidates for functional validation. Nevertheless, this strategy has been successfully used in several studies [29–31].

Today QTL and transcriptomic data sets are exponentially accumulating. Resources have been developed to cross-validate the findings of different strategies in order to identify sets of candidate genes involved in a particular trait [17]. In this study, we used an integrative strategy based on different methods without *a priori* knowledge to select the most promising candidate genes involved in PA biosynthesis in grape berry. We combined genetic mapping and transcriptomic approaches to unravel these candidates. We focused on PA traits in berry skin that play an important role in wine astringency: i) the total quantities of PA, ii) the extent of PA galloylation (% of galloylated units) and iii) the mean degree of polymerization of PA. Among the 20 candidate genes that fulfilled the screening criteria, three genes were investigated by association genetics and appeared as excellent candidates for further functional analyses.

2. Results

We focused on the genetic control of four grape skin PA variables which are particularly involved in wine astringency: the amount of PA in mg/berry (concB) and in mg/kg berry (concK), the percentage of galloylated units (Gal) and the mean degree of polymerization (mDP). These four PA-related variables varied considerably in both F1 progeny and diversity panel [18].

2.1. Identification of candidate genes

2.1.1. QTLs selection

Among QTLs detected in Huang et al. [18], nine QTL intervals were linked with the four selected PA traits: four for PA total content (one for concB and three for concK), two for mDP and three for Gal (Fig. 1). The QTL sizes varied from 6 to 48.1 cM. Five out of the nine QTLs displayed a marked individual effect on PA trait variation ($R^2 > 10\%$, Fig. 1). These nine QTLs represented a total length of 36.2 Mb containing 2405 positional candidate genes. The number of genes located in the QTL confidence intervals was proportional to their size (Pearson correlation coefficient: 0.8), and a similar gene density was observed for the nine QTLs.

2.1.2. Transcriptomic data

Part of the transcriptomic data used in this study was extracted from Terrier et al. [10] between control hairy-roots and those over-expressing PA pathway-specific transcription factors (*VvMybPA1* and *VvMybPA2*). The other new data set consisted of genes differentially expressed in skin during berry development (list of genes, expression values and magnitude of variation are available in Supplementary Table 1). These three transcriptomic analyses allowed the identification of a total of 1647 differentially expressed genes (509 from *VvMybPA1* screening [10], 370 from *VvMybPA2* screening [10] and 1020 from development stage screening).

We then compared these three transcriptomic data sets: 72 genes were differentially expressed both after over-expression of *VvMybPA1* and *VvMybPA2*; 78 genes were differentially expressed both after over-expression of *VvMybPA1* and during berry skin development; 66 genes were differentially expressed both after over-expression of *VvMybPA2* and during berry skin development; and 18 genes were differentially expressed in the three screenings (Supplementary Fig. 1).

2.2. Selection of candidate genes

Among the 2405 genes under the nine QTLs, 41 were differentially expressed after over-expression of *VvMybPA1*, 29 were differentially expressed after over-expression of *VvMybPA2* and 126 genes during berry skin development (Supplementary Table 1). We observed that transcriptomic candidate genes were more often present in QTL than in other regions of the genome (Chi-square test; *P*-value = 6.4 E^{-6}). Thereafter, we focused on potential candidate genes located in a QTL interval and fulfilling at least two transcriptomic screening conditions. Based on this criterion, 20 candidate genes were selected (Table 1): 2 genes were differentially expressed in skin during berry development and after over-expression of *VvMybPA1* and 9 other genes during skin berry development and after over-expression of *VvMybPA2*; 6 genes were differentially expressed after over-expression of *VvMybPA2* and *VvMybPA1*. Only 3 genes appeared in all four lists of candidate genes (Fig. 2).

Please cite this article in press as: C. Grégory, et al., Selection of candidate genes for grape proanthocyanidin pathway by an integrative approach, Plant Physiology and Biochemistry (2013), http://dx.doi.org/10.1016/j.plaphy.2013.04.014