



Contents lists available at ScienceDirect

Plant Science

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



1 Review

2 Genomics-based strategies for the use of natural variation in the
3 improvement of crop metabolism

4 **Q1** Federico Scossa^{a,b}, Yariv Brotman^a, Francisco de Abreu e Lima^a, Lothar Willmitzer^a,
5 Zoran Nikoloski^a, Takayuki Tohge^a, Alisdair R. Fernie^{a,*}

6 ^a Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Potsdam, Germany

7 ^b Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per la Frutticoltura, Via di Fioranello 52, 00134 Rome, Italy

90 A R T I C L E I N F O

11 *Article history:*
12 Available online xxx

13 *Keywords:*
14 Genomics
15 Crop wild relative
16 Hybridization
17 Natural variation
18 Metabolomics

A B S T R A C T

Next-generation genomics holds great potential in the study of plant phenotypic variation. With several crop reference genomes now available, the affordable costs of de novo genome assembly or target resequencing offer the opportunity to mine the enormous amount of genetic diversity hidden in crop wild relatives. Wide introgressions from these wild ancestors species or land races represent a possible strategy to improve cultivated varieties. In this review, we discuss the mechanisms underlying metabolic diversity within plant species and the possible strategies (and barriers) to introgress novel metabolic traits into cultivated varieties. We show how deep genomic surveys uncover various types of structural variants from extended gene pools of major crops and highlight how this variation may be used for the improvement of crop metabolism.

© 2015 Published by Elsevier Ireland Ltd.

21 Contents

22	1. Introduction.....	00
23	2. Genetic basis of metabolic diversity.....	00
24	3. Next-generation sequencing approaches to mine genomic diversity.....	00
25	4. Wide introgression strategies.....	00
26	5. Metabolic quantitative trait loci (mQTL) and metabolite-based genome-wide association studies (mGWAS).....	00
27	6. Transposable elements as presence-absence variants.....	00
28	7. Other cases of structural variants.....	00
29	8. Conclusions.....	00
30	Acknowledgement.....	00
31	Appendix A. Supplementary data.....	00
32	References.....	00

Abbreviations: CCD, carotenoid cleavage dioxygenase; CGH, comparative genome hybridization; CNV, copy number variant/variation; CWR, crop wild relative; eQTL, expression quantitative trait locus; InDels, insertions/deletions; LD, linkage disequilibrium; mGWAS, metabolic genome-wide association study; mQTL, metabolic quantitative trait locus; NAHR, non-allelic homologous recombination; NGS, next-generation sequencing; PAV, presence-absence variant; RNAseq, RNA sequencing; RIL, recombinant inbred line; siRNA, small-interfering RNA; SNP, single-nucleotide polymorphism; SNV, single-nucleotide variant; SV, structural variant/variation; TCM, trans-chromosomal methylation; TSS, transcriptional start site; WGD, whole genome duplication; WGS, whole genome sequencing.

* Corresponding author.

E-mail addresses: scossa@mpimp-golm.mpg.de (F. Scossa), brotman@mpimp-golm.mpg.de (Y. Brotman), Lima@mpimp-golm.mpg.de (F. de Abreu e Lima), willmitzer@mpimp-golm.mpg.de (L. Willmitzer), nikoloski@mpimp-golm.mpg.de (Z. Nikoloski), tohge@mpimp-golm.mpg.de (T. Tohge), fernie@mpimp-golm.mpg.de (A.R. Fernie).

<http://dx.doi.org/10.1016/j.plantsci.2015.05.021>

0168-9452/© 2015 Published by Elsevier Ireland Ltd.

Please cite this article in press as: F. Scossa, et al., Genomics-based strategies for the use of natural variation in the improvement of crop metabolism, Plant Sci. (2015), <http://dx.doi.org/10.1016/j.plantsci.2015.05.021>

1. Introduction

A number of grand challenges await plant researchers in the next decade. The recent document from the Plant Science Research Summit has gathered consensus views from a wide research community [1], with the objective to define those areas of investigation in plant biology deserving highest priorities on the next decade. Coordinated research efforts, on a global scale, are of course driven by the increased pressures we put on the environment: the estimated rise in world population, accompanied by the reduction in arable land, is seriously challenging our capacity to secure food, combat malnourishment, and preserve natural resources [2,3].

Perhaps the common priority emerging from the specific grand challenges of the decadal vision is to achieve a deeper understanding of the genetic basis of phenotypic variation. The range of phenotypic traits exhibited today by cultivated plants is the result of complex interactions between genotypes and environmental factors; essentially, standing natural genetic variation in modern crops (both intraspecific and interspecific) is the result of spontaneous mutations and subsequent selection during the evolution from plant ancestral forms. During domestication, the limited and selective sampling of wild populations brought into cultivation has led to a reduction of genetic variation of cultivated crops. Perennial plant species, in general, have retained a higher degree of genetic variation with respect to annuals, mostly due to lower number of sexual cycles imposed by their long juvenile phases [4]. Crop wild relatives (CWRs), locally adapted varieties (land races), and, in general, unadapted germplasm accessions, having not undergone domestication bottlenecks and intense breeding pressure, represent a fundamental resource to re-enrich cultivated forms with desirable traits [5]. Wild relatives of modern breeds are still found today in their natural habitats, where they display, for example, superior resilience to changing climates and ecological adaptations to specific threats (saline soils, pest resistance, seasonal heat, etc.). Most often, these same threats are also affecting today the majority of the agricultural crops worldwide [6]. Understanding genetics of phenotypic variation is thus important not only to retrace history of plant domestication, but contributes to define heritabilities of simple and complex traits and design successful introgression strategies from diverse germplasm resources. In its simplest form, crop improvement relies on the identification of natural genetic variants and on the subsequent assembly of beneficial alleles in a unique ideotype/individual, minimizing the yield penalties often associated with linkage drag [7].

Recent progresses in crop genomics holds great promise for addressing the analysis of natural variation. Since the release of the first plant genome, *Arabidopsis thaliana*, in 2000 [8], decreasing costs associated with the advent of next-generation sequencing (NGS) technologies have permitted the analysis of more than 100 genome sequences [9]. The genome online database (GOLD, web: <https://gold.jgi-psf.org/index>) [10] aims to keep track of completed and ongoing genome sequencing projects. At the time of writing this review, it contained around 1900 entries related to projects of de novo assembly, resequencing and transcriptome analysis of land plants (Table S1). At the beginning of the plant genomics era, the initial focus was on few, representative model and major crop species (*Arabidopsis*, rice, grape, the moss *Physcomitrella patens*), while more recently, the focus shifted toward the analysis of larger genomes, characterized by high degree of heterozygosity (e.g., sweet orange [11]), ploidy (e.g. hexaploid wheat [12]) and content of repetitive sequences (e.g. pepper [13,14]).

Although the available reference crop genomes represent several families of land plants (covering the most important staple crops), to fulfill the ambitions stated above, the focus of current genomics efforts should now turn to the analysis of CWRs. A limited amount of genomic information is currently available from WRs:

a few groups have been the subject of deep genomic surveys (e.g. rice, maize, soybean); for the remaining crops, for example, when data are available, they usually derive from target sampling of a few wild accessions or land races (Table 1). The development of genomic resources from a wider collection of ancestral forms, or, in general, from the natural variants representing the extended gene pools of reference crops (e.g. land races and closely related interfertile species) will afford us a map of the full spectrum of genetic variations within the pan-genome [15]. Extant genetic variation within the pan-genome can then be used to design rational-wide introgression strategies in order to develop improved cultivated crops [16].

The goal of this review is to survey recent advances of genomics in uncovering the natural genetic variation existing in the extended gene pools of current crops. Our focus will be mainly on plant metabolism, of both herbaceous and perennial species, and we hope to illustrate the various types of structural genetic variations affecting metabolic and regulatory genes and their consequences on the plant metabolome.

We will first briefly survey the genetic mechanisms underlying metabolic diversity, covering the possible current next-generation strategies for analysis of natural variation within pan-genomes; later in this review, we will additionally examine basis of hybrid vigor, compatibility and phenotypic stability, since these latter aspects are fundamental during the design and evaluation of progenies from wide introgression strategies. In the rest of this review, we will present case studies where various targeted sequencing or NGS approaches have been combined with metabolic profiling to identify the causal structural variants in natural populations, illustrating how this variation has been used to enrich or diversify metabolic traits in hybrids or progenies from wide crosses.

2. Genetic basis of metabolic diversity

Plants display an enormous variation in their metabolic repertoires. This diversity is mainly represented by secondary metabolites, which are broadly defined as those involved in plant interactions with the environment (e.g. pigments, hormones, volatiles). Current estimates are in the range of over 50,000 different chemical species (with a defined structure), considering only the main classes of terpenoids, alkaloids and phenylpropanoids [17]. These estimates are of course provisional, given that new natural compounds are continuously isolated and their structure defined [18]. Secondary metabolites are widely different across plant lineages, with biosynthetic pathways usually restricted to specific plant genera; this staggering chemical diversity emerged during the colonization and radiation of plant species in terrestrial habitats, most probably from simple ancestral metabolism [19]. Primary metabolites (those involved in energy metabolism, e.g. carbohydrates, amino acids, organic acids, vitamins, etc.), on the other hand, do not display such high diversity and their number and structures are highly conserved through the plant kingdom.

Plant chemodiversity can be explained on the basis of various types of genome rearrangements occurring during evolution of land plants. Essentially, the view is that various mechanisms of local and whole-genome duplication (WGD) may account for the diversity of secondary metabolites, as was demonstrated in the case of glucosinolates [20]. For example, transposable elements (TEs), which may account up to 80% of angiosperm genomes [21], may generate duplicates and various genome rearrangements as a result of their transpositional activities [22,23]. TEs have also been recently suggested to play a role in the genesis of metabolic clusters. In *Arabidopsis*, for example, the clusters for marneral and thalianol biosynthesis occur in regions with high density of TEs,

Download English Version:

<https://daneshyari.com/en/article/8357270>

Download Persian Version:

<https://daneshyari.com/article/8357270>

[Daneshyari.com](https://daneshyari.com)