Plant Science 223 (2014) 146–152

Contents lists available at ScienceDirect

Plant Science

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

Review

Predicting transcriptional circuitry underlying seed coat development

Deirdre Khan, Ainsley Chan, Jenna L. Millar, Ian J. Girard, Mark F. Belmonte*

Department of Biological Sciences, University of Manitoba, Winnipeg R3T 2N2, Canada

ARTICLE INFO

Article history:
Received 9 December 2013
Received in revised form 16 March 2014
Accepted 17 March 2014
Available online 25 March 2014

Keywords:
Arabidopsis
Chalazal seed coat
Glycine max
Oilseed
Seed coat
Transcriptional circuit

ABSTRACT

Filling, protection, and dispersal of angiosperm seeds are largely dependent on the development of the maternally derived seed coat. The development of the seed coat in plants such as *Arabidopsis thaliana* and *Glycine max* (soybean) is regulated by a complex network of genes and gene products responsible for the establishment and identity of this multicellular structure. Recent studies support the hypothesis that the structure, development, and function of the seed coat are under the control of transcriptional regulators that are specified in space and time. Furthermore, these transcriptional regulators can act in combination to orchestrate the expression of large gene sets. We discuss the underlying transcriptional circuits of the seed coat sub-regions through the interrogation of large-scale datasets, and also provide some ideas on how the identification and analysis of these datasets can be further improved in these two model oilseed systems.

Contents

1. Introduction

The evolution of the seed has largely accounted for the success of flowering plants since it allows the next plant generation to inhabit new environments away from the parent, and delays germination until conditions are favorable for seedling growth. The young embryo is protected by the maternal seed coat, which encloses the internal embryo and endosperm, and ensures the viability of the next sporophyte generation. Traditional methods of investigating the seed coat have included anatomical and forward genetics studies. While these approaches were useful in the initial characterization of the seed coat, they limit research to the study of genes with obviously screenable phenotypes.

Recent advances in transcriptomics have allowed for the study of seed development on a more global scale [1]. By identifying patterns of gene expression and analyzing smaller subsets of co-expressed genes, we can discover genetic regulators such as

Abbreviations: CZSC, chalazal seed coat; GO, gene ontology; TF, transcription factor.

* Corresponding author. Tel.: +1 204 474 8556.
E-mail addresses: Mark.Belmonte@umanitoba.ca, belmonte@cc.umanitoba.ca (M.F. Belmonte).

http://dx.doi.org/10.1016/j.plantsci.2014.03.016
0168-9452/

transcription factors (TFs) that are masked by redundancy, or which produce only subtle phenotypes [2]. Whether through the analysis of large or small datasets, association of transcriptional regulators through gene regulatory network analysis may provide the tools necessary to uncover these genes.

The effort to discover all of the gene regulatory networks underlying the development of each cell type in the seed coat will likely require the combination and integration of a number of techniques and methodologies needed to detect, quantify, and experimentally test each transcriptional circuit. This review is focused on current methods used to generate transcriptional modules using large-scale mRNA datasets, enrichment analysis, and protein–DNA interactions. We discuss current tools available to plant scientists to predict transcriptional circuits within cell and tissue layers of the seed. We focus on the development of the seed coat in two important model oilseed plants: *Arabidopsis thaliana* (Fig. 1A) and *Glycine max* (soybean) (Fig. 1B), representatives of the Brassicaceae and Leguminosae families, respectively. We also present the consequences of technical advancements in transcriptome research on TF detection and transcriptional module construction, and ideas for future investigations.

2. Seed coat sub-regions in *Arabidopsis* and soybean

2.1. The distal seed coat sub-region

The distal seed coat sub-region of *Arabidopsis* (Fig. 1A, navy blue) is composed of two integument layers. Both of these layers are multiple cell layers thick, with each layer undergoing dramatic morphological changes throughout seed development. Many excellent reviews have already described the anatomy of the different cell layers in *Arabidopsis* and how they transform from the time of ovule fertilization to the time of seed maturity [3–5]. However, due to the small size of the *Arabidopsis* seed, -omics-based studies have not been completed at the level of the individual cell layers, and more cutting edge technologies must be applied to access these cells. The implications of integrating these technologies into regulatory network-based investigations will be discussed further in this review.

Like *Arabidopsis*, the seed coat of soybean is also separated into an inner and an outer integument (Fig. 1B, navy and light blues), each of which contains different cell layers that change throughout development [6–8]. Soybean seed coat structure has also been the subject of extensive study and review [6–8]. The use of laser microdissection technology has allowed the separate capture of the inner and outer integuments of the soybean seed coat, though individual cell layers have yet to be isolated (https://www.seedgenenetwork.net).

2.2. The chalazal sub-region

The chalazal seed coat sub-region (CZSC, Fig. 1A, orange) – the portion of the seed most proximal to the funiculus – is morphologically distinct from its distal counterpart. In addition to structural differences at the mature ovule stage [9], throughout seed development, the CZSC in *Arabidopsis* exhibits different patterns of cell division and pigment accumulation compared to the distal seed coat [3,10]. The study of proanthocyanidin biosynthesis has led to some description and genetic analysis involving the CZSC [10], but the CZSC has not otherwise been the main experimental focus until recently [13]. Because of these structural differences, it was predicted that changes in transcriptome activity could also be observed [11–13].

A "sub-hilum" region – an area of the seed coat proximal to the funiculus – has been identified in soybean [11], and appears to be analogous to the CZSC (Fig. 1B, orange). Similar to the CZSC in *Arabidopsis*, there are unique features in the hilum that are not found in the rest of the seed coat [11,14,15]. The chalazal region of the soybean seed coat has different anatomy and developmental patterns from the rest of the seed, and deserves further investigation.

3. Predicting transcriptional circuits underlying seed coat development

To fully understand how each cell and cell layer of the seed coat contributes to the making of a seed, one must be able to harvest and analyze the underlying genetic molecules required for seed coat development. Initiatives through the Seed Gene Network (https://www.seedgenenetwork.net) have undertaken an ambitious genomics study to investigate the organization and regulation of the *Arabidopsis* and soybean genomes [11,12,16]. While the initial goal was to isolate, characterize, and study all of the genes required to make a seed, the recent addition of information about small RNAs and DNA methylation will increase the resolution of possible regulatory mechanisms underlying cell identity and seed development as a whole. A systems biology approach to studying the seed coat integrates our knowledge of plant molecular biology, mathematics, statistics, and computer science to better understand and model, or predict the gene regulatory networks underlying plant development [17], metabolism [18] and adaptation [19].

Like *Arabidopsis*, the soybean genome has been fully sequenced [20], and public resources for the study of this agriculturally important crop are rapidly improving. The Seed Gene Network has compiled GeneChip, next generation RNA sequencing, and methylome analyses of both *Arabidopsis* and soybean, and maintains a collection of RNAi lines targeting TFs implicated in the regulation of seed development for functional analysis of soybean. Likewise, the *Arabidopsis* Information Resource (http://www.arabidopsis.org) maintains a database of all genes expressed in *Arabidopsis*, as well as access to T-DNA insertion lines that can be used for functional analysis of almost every gene.

Predictive gene regulatory networks are useful for studying transcriptome datasets as they consolidate a large amount of information into manageable modules or circuits. In order to construct informative transcriptional modules, one must have access to: (i) large-scale datasets or a list of co-expressed genes; (ii) enrichment of biological information like Gene Ontology (GO) terms, active DNA sequence motifs, TF families, and (iv) lists of TFs known or predicted to bind to the DNA sequence motifs to complete the circuit.

3.1. Generation of large-scale seed coat-specific datasets

Various -omic based studies on specific tissues or cell types rely on contamination-free isolation of the cells of interest. Laser microdissection is one of the few technologies capable of this precise dissection. Unlike other isolation methods, laser microdissection does not require cell-specific molecular markers, and is capable of harvesting tissues and cells visible with compound microscopy [21]. The utility of the system has proven successful in the study of cells and tissues of a number of plant systems including rice [21], *Arabidopsis* [22], maize [23], *Medicago* [24], cucumber [25], and canola [26]. With regard to the seed coat specifically, laser microdissection methods were used to isolate different tissues of the *Arabidopsis* seed – including the distal and CZSC – for microarray studies [11,12]. The inner and outer integument of the soybean seed has also been isolated and transcriptionally profiled using laser microdissection [11]. Moreover, this technique would be an