Post-transcriptional regulation of photosynthetic genes is a key driver of C₄ leaf ontogeny

Nicklaus Fankhauser¹,* and Sylvain Aubry²,*,†

¹ Clinical Trials Unit, University of Bern, Finkenhübelweg 11, 3012 Bern, Switzerland
² Institute of Plant and Microbial Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

* These authors contributed equally to this work.
† Correspondence: sylvain.aubry@botinst.uzh.ch

Received 2 June 2016; Accepted 26 September 2016

Abstract

C₄ photosynthesis allows highly efficient carbon fixation that originates from tightly regulated anatomical and biochemical modifications of leaf architecture. Recent studies showed that leaf transcriptome modifications during leaf ontogeny of closely related C₃ (Tarenaya hassleriana) and C₄ (Gynandropsis gynandra) species within the Cleomaceae family existed but they did not identify any dedicated transcriptional networks or factors specifically driving C₄ leaf ontogeny. RNAseq analysis provides a steady-state quantification of whole-cell mRNAs but does not allow any discrimination between transcriptional and post-transcriptional processes that may occur simultaneously during leaf ontogeny. Here we use exon–intron split analysis (EISA) to determine the extent to which transcriptional and post-transcriptional processes are involved in the regulation of gene expression between young and expanded leaves in both species. C₄-specific changes in post-transcriptional regulation were observed for genes involved in the Calvin–Benson cycle and some photosystem components but not for C₄ core-cycle genes. Overall, this study provides an unbiased genome-wide insight into the post-transcriptional mechanisms that regulate gene expression through the control of mRNA levels and could be central to the onset of C₄ photosynthesis. This mechanism is cytosolic which implies cell-specific modifications of mRNA stability. Understanding this mechanism may be crucial when aiming to transform C₃ crops into C₄ crops.

Key words: C₄, Cleomaceae, EISA, leaf ontogeny, photosynthesis, post-transcription.

Introduction

C₄ photosynthesis is one of the most efficient carbon-fixing reactions on Earth (Hatch and Slack, 1966). It is widespread in many ecosystems and has existed since the late Miocene (Ehleringer et al., 1997). Based on ancestral C₃ photosynthesis, it is a fantastic example of convergent evolution in response to variable atmospheric CO₂ in angiosperms (Aubry et al., 2011; Williams et al., 2012). Interestingly, the independent evolution of at least 18 C₄ plant families seem to be constrained to a recurrent sequence of small evolutionary steps that progressively increased the carbon concentration in the vicinity of the leaf vasculature (Heckmann et al., 2013; Williams et al., 2013). Briefly, in order to favour the ribulose 1,5-bisphosphatase carboxylase/oxygenase (RuBisCO) carboxylation reaction over its undesired oxygenation reaction, C₄ plants concentrate CO₂ around the enzyme and thereby drastically limit photorespiration (and therefore increase

© The Author 2016. Published by Oxford University Press on behalf of the Society for Experimental Biology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
photosynthetic activity). In order to allow carbon concentration, the carbon cycle is segregated into two compartments, each containing one of the two carboxylases involved: the phosphoenolpyruvate carboxylase (PPC) in mesophyll (M) cells and RuBisCO in bundle-sheath (BS) cells. After diffusing into the leaf through stomata that have acquired the ability specifically to regulate their opening in otherwise unfavourable CO₂ and humidity conditions (Aubry et al., 2016), CO₂ is primarily fixed by the O₂-insensitive PPC. It is subsequently exported as a four carbon compound to the BS cells where a decarboxylase charges the Calvin-Benson cycle (CBC) to feed the final carboxylation by RuBisCO. The biochemistry of various subtypes of C₄ photosynthesis is classified by one of the three different decarboxylases recruited (Furbank, 2011).

Several complementary studies using high-throughput transcriptomics approaches performed multiple comparisons of steady-state transcript levels between C₃ and C₄ leaves as well as during their ontogeny (Brautigam et al., 2010; Gowik et al., 2011; Aubry et al., 2014a; Külahoglu et al., 2014). These ‘omics’ approaches are aimed at identifying new elements of transcriptional networks that regulate the C₄ pathway (reviewed in Burgess and Hibberd, 2015). However, they also confirmed the assumption that multiple levels of regulation—transcriptional (Brown et al., 2011), post-transcriptional (Patel and Berry, 2008), and post-translational (Pick et al., 2011)—contribute to the onset and maintenance of the C₄ metabolic system (Hibberd and Covshoff, 2010).

Modifications in gene expression are thought to be associated with many of the C₄-specific anatomical or biochemical features, such as increased vein density, changes in stomata regulation, chloroplast dimorphism, chloroplast positioning, BS suberization, and endoreduplication (Li et al., 2010; Majeran et al., 2010; Gowik et al., 2011; Pick et al., 2011; Aubry et al., 2014a, 2016; Külahoglu et al., 2014). Further studies focused on differences in the transcriptome between specialized leaf cell types—bundle-sheath (BS) and mesophyll (M) cells—and seem to confirm the function of transcriptional regulation for gene expression tuning in C₄ cycle regulation but, so far, without identifying any dedicated transcriptional networks (Li et al., 2010; Chang et al., 2012; John et al., 2014; Aubry et al., 2014a; Burgess and Hibberd, 2015). It is noteworthy that, in C₄ leaves, cell-specificity is not limited to the core C₄ cycle (i.e. from the initial HCO₃⁻ fixation until the RuBisCO carboxylation step) but applies to other parts of the photosynthetic apparatus. In Flaveria and Cleome, transcriptional signatures show a clear segregation of photosystem I (PSI) and photosystem II (PSII) elements between BS and M cells (Gowik et al., 2011; Aubry et al., 2014a). In C₄ species, from the NADP malic enzyme (NADPME)-subtype (like Flaveria and maize) BS chloroplasts are mostly agranal and there is a strong transcriptional investment in ATP production by cyclic electron flow that includes genes encoding proteins of PSI, cytochrome b/f and NDH (NADH dehydrogenase) complexes (Nakamura et al., 2013; Aubry et al., 2014a). RuBisCO transcript levels (and most other CBC transcripts) are largely reduced in C₃ compared with C₄ leaves (Ku et al., 1979; Bräutigam et al., 2010). This reduction in RuBisCO is thought to be one of the major advantages of C₄ nitrogen use efficiency (Oaks, 1994; Ghanoum et al., 2005). Mechanisms that restrain RuBisCO from accumulating in BS are complex and involve a mixture of transcriptional and post-transcriptional features (Berry et al., 2016). In both monocots and dicots, BS-specific expression of RuBisCO is under robust post-transcriptional control. Thus, during leaf development, expression of RbcS and RbcL have been shown to be uncoupled from translation (Patel and Berry, 2008) and RbcS untranslated regions (UTR) are suggested to play a role in RNA stability during the later stages of leaf development. Post-transcriptional regulation of gene expression may not be restricted to RuBisCO but may also be a more widely used strategy for other (photosynthesis-related) gene regulation, especially during C₄ leaf development.

In order to assess the involvement of post-transcriptional regulation of gene expression, we took advantage of comparative experiments on closely related species during leaf ontogeny in both C₃ and C₄ leaves (Kühlahoglu et al., 2014) using a computational approach called exon–intron split analysis (EISA). EISA takes advantage of the presence of both mature and intron-containing pre-mRNAs in RNAseq samples to discriminate transcriptional from post-transcriptional effects on gene expression (Gaidatzis et al., 2015). While the abundance of transcripts may vary between two experimental conditions (such as here during leaf development), a gene that is exclusively subjected to transcriptional regulation should show tightly correlated proportions of reads mapping to introns and exons. By contrast, a gene that is subjected to post-transcriptional regulation is expected to exhibit a decreased amount of reads mapping to exons compared with reads mapping to introns, as only mature intron-free cytosolic mRNA should be impacted. This has been shown to be a possible proxy to predict gene regulation status (Gaidatzis et al., 2015). For example, a significant increase in intronic RNA expression without an increase of corresponding exonic RNA has been observed in some animal systems after antioxidant treatment, even if the physiological significance of such mechanisms remains unclear (Menon et al., 2016). We applied EISA to leaf development transcriptome series from Tarenaya hassleriana (C₃) and Gynandropsis gynandra (C₄), two closely related species within the Cleomaceae (Kühlahoglu et al., 2014). Despite some anatomical and biochemical modifications, gene expression profiles of the two species appeared relatively well conserved during leaf ontogeny (Aubry et al., 2014a; Külahoglu et al., 2014). Nevertheless genes encoding photosynthetic components and CBC enzymes appear to be subject to post-transcriptional regulation at the level of mRNA accumulation. This process is likely to be a key driver of C₄ leaf ontogeny.

Materials and methods

Data access

RNAseq data were obtained from NCBI (accessions: SRP036637 and SRP036837, for G. gynandra and T. hassleriana, respectively). The reference genome for Arabidopsis was accessed from Phytozone (http://www.phytozone.net) with the most up-to-date
gene model. The genome sequences from *G. gynandra* and *T. hassleriana* were accessed as described previously (Aubry et al., 2016; Cheng et al., 2013).

**Processing of the reads and quantification of exonic and intronic expression levels**

The EISA method was implemented to obtain expression values for reads mapping to introns or exons as published (Gaidatzis et al., 2015) with the following criteria used. Only transcripts that map to a unique position in the genome were considered. Only reads that map inside the body of the gene (no UTRs) were taken into account and only introns enclosed by exons were considered. Exon/intron counts were quantified using qCount from QuasR (Lun et al., 2016). A threshold requiring at least two reads for every exon and intron was applied. Normalization based on the total number of reads for each library was performed separately for exons and introns. Overlapping genes were not considered. ΔExon and ΔIntron were defined as the difference in log₂ of exonic or intronic expression levels between respective compartments. The model for statistical significance of differential post-transcriptional regulation was implemented in edgeR (Lun et al., 2016). Generally, the lack of reliable genome annotation, especially a definition of exon–intron boundaries, can limit the interpretation of our data (Weber, 2015).

We validated our approach using publically available transcriptome data from the *ago1-27* mutant (accession number: GSE77211, see Supplementary Table S1 at JXB online). This analysis showed that suppression of ARGONAUTE1 (AGO1) protein, a core component of the gene-silencing machinery (Morel et al., 2002; Vaucheret et al., 2004) modifies the extent of post-transcriptional regulation of 216 genes, especially RNA-metabolic genes (GO term enriched: RNA splicing, RNA processing, and RNA metabolism, Supplementary Table S1). It should be noted that the number of biological replicates and the coverage of the sequencing are crucial when intending to use EISA for a comprehensive evaluation of post-transcriptional regulation of gene expression.

**Analysis of differential gene expression and gene ontology enrichment**

Reads from RNAseq libraries of young and expanded leaves were aligned to protein-coding transcripts from the respective genomes of *G. gynandra* and *T. hassleriana* (Cheng et al., 2013; Aubry et al., 2016). Transcript expression was quantified using RSEM version 1.2.30 (Li and Dewey, 2011). Posterior probability of differential expression (PPDE) between the two conditions was estimated using the empirical Bayesian approach implemented in EBSeq version 1.1.6 (Leng et al., 2013). Gene ontology enrichment was performed using a corrected Benjamini–Hochberg enrichment score implemented in Pageman (Usadel et al., 2006).

**Results and discussion**

**Application of exon–intron split analysis to plant datasets**

During standard RNAseq library preparation, the total extraction of mRNAs results in a mixture of mature cytosolic RNA and nuclear nascent or unspliced RNAs (Fig. 1A). Even if the vast majority of the reads analysed originate from exons, a fraction of the reads maps as introns (Fig. 1B). These might originate from genomic contamination, wrong gene models (mis-annotated exons) or from premature RNAs. Levels of intron-specific reads have been proposed to correlate with transcriptional activity and to be used as a proxy for post-transcriptional regulation of photosynthetic genes.

**Fig. 1.** The exon–intron split analysis (EISA) detects the extent of gene expression regulation under post-transcriptional control between two experimental conditions. EISA was applied during leaf ontogeny between young and expanded leaves of *G. gynandra* and *T. hassleriana*. (A) PremRNA comprising introns (red) and exons (blue) are transcribed in the nucleus, then spliced and translocated to the cytosol where the mature mRNA will be translated. EISA counts the differences in intronic (ΔIntron) and exonic (ΔExons) levels between the experimental conditions and compares them as a measure of post-transcriptional controls. The bottom part shows plots expressing ΔExon as a function of ΔIntron to show the possible patterns that are representative of the predominant mechanisms regulating gene expression (mostly transcriptional, post-transcriptional or a mixture of the two). The schematic representation was adapted from Gaidatzis et al. (2015). (B) RNA extraction results in a majority of reads mapping to exons but also reads originating from intronic segments.
analysing transcription (Zeisel et al., 2011; Hendriks et al., 2014). Intronic read levels often, but not always, correlate with exonic reads when compared across experimental conditions. In the case of post-transcriptionally regulated genes that are for example under miRNA control, the relative level of mature transcripts declines as the silencing machinery exclusively acts on cytosolic mature RNAs. It is of note that there are fewer intronic reads in the preparation when mRNA is selected by poly-A purification compared with total RNA extractions by ribosomal RNA depletion (Trapnell et al., 2010). However, data that are derived from both methods are still suitable for EISA, given a sufficient coverage of the transcriptome (Gaidatzis et al., 2015).

The EISA method, by quantifying reads mapping exclusively to either introns or exons (see Materials and methods for details on filtering), aims to identify transcripts for which the stoichiometry between intron and exon is lost when comparing two experimental conditions (Fig. 1A). In the present study, it was possible successfully to apply EISA to recent congeneric Cleomaceae whole leaf gradient data (Külahoglu et al., 2014) but cell-specific data sets were not sequenced deep enough to provide a statistically robust result (data not shown; Li et al., 2010; Chang et al., 2012; Aubry et al., 2014a; John et al., 2014).

**Table 1.** Genes that were detected and that have been filtered (minimal coverage of two reads per exon or intron) by the EISA pipeline and the number of genes differentially transcribed between young and expanded leaves (FDR ≤0.05) in the two Cleomaceae species

<table>
<thead>
<tr>
<th>Differential transcription (EISA)</th>
<th>G. gynandra</th>
<th>T. hassleriana</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. of genes</td>
<td>23 340</td>
<td>22 227</td>
</tr>
<tr>
<td>Genes selected for EISA</td>
<td>4 769</td>
<td>5 720</td>
</tr>
<tr>
<td>Post-transcriptionally regulated (FDR ≤0.05)</td>
<td>678 (15%)</td>
<td>263 (5%)</td>
</tr>
<tr>
<td>Differential expression (EBSeq)</td>
<td>No of genes DE (FC ≥2, PPDE ≥0.95)</td>
<td>8 855</td>
</tr>
<tr>
<td></td>
<td>Up in young leaves</td>
<td>3 949</td>
</tr>
<tr>
<td></td>
<td>Up in expanded leaves</td>
<td>4 906</td>
</tr>
</tbody>
</table>

Genome-scale prediction of transcriptional and post-transcriptional regulation in Cleomaceae leaf gradients

Leaf ontogeny variations between C_{3} and C_{4} leaves observed from comparisons between congeneric species have been well documented for *Cleome*, *Flaveria*, and *Alloteropsis* species (Gowik et al., 2011; Christin et al., 2012; Külahoglu et al., 2014). Here we used publically available transcriptomic data from congeneric Cleomaceae leaf gradients (Külahoglu et al., 2014). For each species, we selected the extreme stages of the leaf gradient (stage 0: ‘Young’ and stage 5: ‘Expanded’) analysed in the study (Külahoglu et al., 2014). 8 855 (37% of all genes detected) and 6 672 genes (28%) were differentially expressed between these two conditions in *G. gynandra* and *T. hassleriana*, respectively. EBSeq was applied using a posterior probability of being differentially expressed ≥0.95 and a minimal fold change of 2 between the two conditions. The results are summarized in Table 1 and Supplementary Table S2.

The EISA method was used on all expressed genes to identify post-transcriptional events occurring between young and expanded leaves in both C_{3} *T. hassleriana* and C_{4} *G. gynandra* (Fig. 2A, B). This method requires complete coverage in exons as well as introns. In order to limit false positives, a stringent minimal read cut-off was applied (see Materials and methods for details). This filtering left a quarter of the genes for both species: 4 769 and 5 720 in *G. gynandra* and *T. hassleriana*, respectively (Table 1). Of these genes, 678 in *G. gynandra* and 263 genes in *T. hassleriana* were most likely post-transcriptionally regulated (EISA positive) between young and expanded leaves with a false discovery rate FDR ≤0.05 (Table 1). Finally, 49 post-transcriptionally regulated genes were common to both species, including, for example, the gene encoding phosphoribulokinase (PRK) (Fig. 2C; Supplementary Table S3). This is consistent with a global rewiring of gene expression at the post-transcriptional level specific to the C_{4} leaf. A majority of the predicted post-transcriptionally regulated genes: 349 (51%) for *G. gynandra* and 116 (44%) for *T. hassleriana* were also significantly over-expressed during leaf ontogeny (Fig. 2D). Gene Ontology (GO) terms associated with photosystems and the CBC were significantly over-represented in genes that were post-transcriptionally regulated between young and expanded leaves from *G. gynandra*.

Expression of genes encoding proteins of the C_{4} cycle is up-regulated during leaf development in both species, but to a much larger extent in the C_{4} species (see Supplementary Fig. S1A at *JXB* online). Increased expression of C_{4} cycle genes in the C_{3} species during leaf development has previously been shown in many other dicotyledonous systems. This suggests that C_{4} gene regulatory networks pre-existing in C_{3} species and were associated with their ancestral anaplerotic function (Aubry et al., 2011; Wang et al., 2013). The EISA model tries to identify genes that show expression changes at the exon level that are not accounted for at the intron level.
Post-transcriptional regulation of photosynthetic genes

Genes encoding C₄ core-cycle proteins are not post-transcriptionally regulated: variations of the number of reads that map introns and exons are correlated for most of these genes except for two, GgPPT1 (phosphoenolpyruvate/phosphate translocator) and ThPPC2 (phosphoenolpyruvate carboxylase 2) (Fig. 3A). This shows that, in both C₃ and C₄ species, C₄ cycle genes are subjected to an increase in amplitude as the leaf matures but the magnitude of this amplitude is much higher in C₄ leaves.

Interestingly, this increase in expression does not seem to be regulated at the post-transcriptional level according to EISA. This is consistent with data previously published suggesting a strong involvement of the transcriptional machinery to the C₄ cycle (Bräutigam et al., 2010; Aubry et al., 2014a). However, regulation of M specificity of C₄ cycle genes like GgPPDK (pyruvate-orthophosphate dikinase) or GgCA4 (β-carboxy anhydrase 4) is regulated by post-transcriptional mechanisms involving untranslated regions (Williams et al., 2016). This suggests a complex interplay between transcriptional and post-transcriptional regulation of gene expression for these genes.

Calvin–Benson cycle and photosystem genes are post-transcriptionally regulated in the C₄ leaf gradient

Genes encoding proteins involved in the CBC are up-regulated during leaf development in both C₃ and C₄ species (Supplementary Fig. S1). In contrast to the core C₄ cycle, there is a tendency towards a lower expression of CBC genes in C₄ leaves (Bräutigam et al., 2010). In G. gynandra, most genes (10 out of 13 detected by EISA) showed a significant (FDR ≤0.05) drop in ΔExon compared with ΔIntron (Fig. 3B). In most C₄ species, the CBC is split between BS and M cells (Majeran et al., 2005, 2010; Aubry et al., 2014a; John et al., 2014). Interestingly, all genes involved in the CBC known to be specifically expressed in BS cells are predicted as being post-transcriptionally regulated (FDR ≤0.05, Fig. 3B). Genes that are preferentially expressed in M, like triose phosphate isomerase (TPI) or the GAPDH subunit B (GAP-B) are also post-transcriptionally regulated, suggesting this mechanism is not exclusive to BS genes. These genes, nonetheless, have their expression level increased at the tissue level, proportional to the leaf age (Supplementary Fig. S1). We postulate that the significant decrease in ΔExon during leaf development is a mixture of transcriptional and post-transcriptional regulation. None of the T. hassleriana CBC genes, except phosphoribulokinase (PRK), were predicted to be under post-transcriptional regulation during leaf development (blue dots in Fig. 3B). This suggests a post-transcriptional regulation mechanism specific to the C₄ leaf. It is known that most CBC genes are expressed cell-specifically in G. gynandra leaves. Therefore, we propose a model of post-transcriptional regulation that operates between BS and M cells on a specific subset of transcripts. For example, sedoheptulose-bisphosphatase (SBPA) is mostly expressed in BS cells and the ΔExon/ΔIntron correlation is lost during leaf development (Fig. 3B). This can be interpreted as an increase in expression at the transcriptional level in BS cells, combined with a mesophyll-specific degradation of exon-only mature transcripts. In a similar manner, a large proportion, i.e. 18 and 23 transcripts encoding for PSI and PSII proteins, respectively, were post-transcriptionally regulated (decrease in ΔExon, Fig. 3C, D). These results indicate that regulation of C₄ core genes is somehow distinct from photosynthetic genes during C₄ leaf ontogeny and that this mechanism could be the basis for the actual cell-specific regulation of photosynthesis in C₄ leaves.

Fig. 2. Changes in intronic and exonic levels during leaf development for (A) G. gynandra and (B) T. hassleriana. Significantly up- or down-regulated genes after EISA analysis (FDR ≤0.05) are shown in red. (C) Overlap between the genes that were post-transcriptionally regulated and identified as potential orthologues by BLAST analysis across the two species. (D) Among post-transcriptionally regulated genes, a majority of genes were differentially expressed between young and expanded leaves in each species (differential analysis using EBSeq with a PPDE ≥0.95 and a minimal fold change of 2). PT, post-transcriptionally regulated.
Post-transcriptional regulation of RbcS and cell-specificity

RuBisCO is a hexadecameric protein, composed of eight nuclear-encoded subunits (RBCS) and eight chloroplast-encoded subunits (RBCL). In C₄ dicots like Flaveria and amaranths, RbcL and RbcS expression are tightly regulated and repressed in the mesophyll (Patel and Berry, 2008; Berry et al., 2016). RbcS expression at the translational level appears to be central to the early stage of development, whereas mRNA stability mediates cell-specificity at a later stage (Patel and Berry, 2008). Consistent with this, ectopic expression of RbcS under a constitutive promoter in maize failed to accumulate transcripts in the mesophyll compartment (Wostrikoff et al., 2012).

Our data imply that, despite an overall increase in RbcS expression on the whole leaf scale (Supplementary Fig. S1), at least part of the RbcS1a expression is under post-transcriptional regulation in G. gynandra. Given the high specificity of RbcS transcripts in the BS compartment at a steady-state level in G. gynandra (Aubry et al., 2014a), the mesophyll compartment is the most obvious place for this to happen. The cytosolic degradation of mature transcripts translates to a decrease in ΔExon during leaf development. The potential influence of cytosol-based regulation of RbcS transcript levels on chloroplast expression of RbcL remains to be shown, as gene expression of both subunits is tightly linked (Wostrikoff et al., 2012).

Models for post-transcriptional regulation of photosynthesis in C₄ leaves

We have shown here that post-transcriptional control of RNA stability is likely to play a key function during C₄ leaf ontogeny, but the molecular basis of the mechanism involved in C₄ species remains unclear. Our results confirm a large body of data indicating that post-transcriptional regulation is crucial for photosynthesis regulation as, for example, in the case of ferredoxin-1 mRNA stability controlled by light in tobacco (Petracek et al., 1998). In plants, multiple levels of control can regulate cytosolic mRNA stability: basal RNA decay
mRNA

miRNA-mediated degradation that is known to repress gene expression in diverse developmental processes (phase transition, leaf shape, floral organ identity; Chen, 2009) could be a possible mechanism to explain the observed post-transcriptional regulation. Unfortunately, no cell-specific miRNA profiling of C₄ leaves has been published to our knowledge.

Non-sense-mediated decay (NMD)

Cytosolic aberrant mRNAs may be degraded via the non-sense-mediated mRNA decay (NMD) pathways, considered to be one of the main post-transcriptional regulation processes in eukaryotes (Siwaszek et al., 2014; Dai et al., 2016). However, genes involved in the NMD pathway in plants such as upframeshift protein 1 and 3 (UPF1 and 3) and exonubonucleases (XRN4, 5, 6) were not differentially expressed between BS and M cells in mature G. gynandra leaves and no cell-specific polymorphism could be observed between the transcripts from the two cell types (Aubry et al., 2014a). This mechanism is therefore unlikely to be recruited for transcriptional regulation of photosynthetic genes during C₄ leaf ontogeny.

Translation

Interaction with the translational machinery is another way of controlling mRNA levels by translational repression (Roy and Jacobson, 2013). The relationship between mRNA decay and translation efficiency are, however, complex and differ on a gene-to-gene basis. More work is necessary to show whether ribosomal footprints of mRNA is actually different in genes predicted to be post-transcriptionally regulated using EISA modelling. Interestingly, a complex regulation of translation of RbcS transcripts observed in Arabidopsis (assessed by monitoring the association of transcripts to various polysomal fractions) may indicate that already in C₃ ancestors RuBisCO expression is under a specific regime of transcriptional and post-transcriptional regulation (Piques et al., 2009). Therefore, further experiments analysing the association of photosynthesis genes with ribosomes on a cell-specific basis are required. Moreover, applying recently developed ribosomal footprints to such leaf material (Ingolia et al., 2009; Juntaawong et al., 2014) could shine new light on transcription–translation interactions in a C₄ context.

Splicing

Differences in mRNA splicing at the pathway level that would lead to a mis-interpretation of the variations in intronic signals are unlikely. Indeed, a large proportion of genes predicted to be post-transcriptionally regulated and involved in the same pathway also present very distinct intron/exon organization (Gaidatzis et al., 2015). The median number of exons and introns of post-transcriptionally (EISA FDR ≤0.05) or transcriptionally regulated (EISA FDR >0.05) genes were compared in each species and showed no significant differences based on Wilcoxon–ann–Whitney test. This indicates that the EISA output is not biased by the structures of the genes.

Post-transcriptional co-ordination of gene expression

In G. gynandra, M specificity of two C₄ genes, PPDK and CA4, is mediated by 5’ and 3’ UTR regions (Kajala et al., 2012; Williams et al., 2012, 2016). The mechanism co-ordinating the cell-specific expression involves a cis-element (MEM2) that specifically increases translation in the M compartment (Williams et al., 2016). Other examples of gene expression co-ordination of functionally related genes have been described (Keene, 2007; Goodarzi et al., 2012). Here we postulate that mechanisms similar to those co-ordinating the expression of photosynthesis genes have been recruited during C₄ leaf development. More work is necessary to identify which of the 200 predicted RNA-binding proteins from six different families in Arabidopsis might regulate mRNA stability (Fedoroff, 2002; Ambrosone et al., 2012).

Finally, a model involving RNA binding proteins that regulate the mRNA stability of a given set of photosynthesis-related genes during leaf ontogeny of C₄ might be the simplest scenario that explains our observations (Williams et al., 2012). Distinct families of RNA binding proteins could be involved (Silverman et al., 2013), such as PUF proteins that often recognize 3’ UTRs and are dependent on secondary RNA structure (Francischini and Quaggio, 2009; Lu et al., 2009) or sequence-specific PENTATRICOPEPTIDE REPEAT (PPR) that are mostly but not exclusively localized in mitochondria and chloroplasts (Colcomb et al., 2013). More work is required to understand the extent of its complexity (e.g. timing of the regulation in both cells) and diversity (e.g. how that regulation accommodates multiple types of Kranz anatomy) across other C₄ species. Interestingly, this model would provide two advantages at the pathway level, (i) a reasonable control over transcripts originating from M cells in M cells themselves, and, (ii) would avoid the potential accumulation of transcripts that could have leaked from BS cells (or other cells) into M cells and is, therefore an efficient way to control translation tightly in each respective compartment.

EISA limitations and potentials for the C₄ photosynthesis field

Evidence of the presence of unspliced pre-mature RNAs in total RNA samples has already been reported when comparing transcriptome and polyribosomal translatome fractions from Arabidopsis leaves (Aubry et al., 2014b; Zhang et al.,...
When the fraction of RNA that binds ribosomes (referred to as the ‘translatome’) was immunopurified, most of the intron-containing pre-mRNAs were eliminated. This also implies that the EISA method cannot be applied to data deriving from TRAPseq experiments recently used in one of the few cell-specific expression-profiling studies of C₄ BS cells from Arabidopsis (Aubry et al., 2014b).

Using data from entire leaf tissue might not be sufficient to obtain a precise cell-specific understanding of the post-transcriptional mechanisms, but does provide a genome-wide insight of their involvement in the pathway regulation. Further experiments using cell-specific fractions with deep sequencing coverage, and taking into account the untranslated regions of the genes for the EISA, are needed to refine the EISA analysis.

Conclusions

Relatively few genes are known to be post-transcriptionally regulated in C₄ leaves and the extent to which this level of regulation is involved in cellular fate remains unclear. Some level of post-transcriptional regulation has been suggested to occur in a single gene-based manner, such as for RbcS, PPDK, and RbcS that are specific to C₄ leaves of multiple genes encoding photosynthetic machinery components. The model proposed here is reminiscent of the theory that conceptualizes high-order regulation of functionally related mRNAs (Keene, 2007; Tenenbaum et al., 2011). More work is needed to clarify the exact significance of this post-transcriptional signature in each cell type, in particular, the mesophyll compartment of C₄ leaves and how this mechanism interacts with predicted transcriptional modules and translation (Williams et al., 2012; Aubry et al., 2014a; Külahgolu et al., 2014). Perhaps not surprisingly, the stability of chloroplast-encoded transcripts is also very variable among the developmental stages (Klaff and Gruissem, 1991). The fact that a large majority of the nuclear-encoded proteins in the C₄ leaf under post-transcriptional regulation are eventually targeted to the chloroplast might suggest an involvement of some plastid-to-nucleus retrograde signalling in this process. Unfortunately, all high-throughput C₄ cell-specific data available (Li et al., 2010; Chang et al., 2012; Aubry et al., 2014a; John et al., 2014) do not have enough coverage to use the EISA method robustly (data not shown). Therefore, in the context of C₄ engineering, more research is required to identify relevant RNA binding, RNA processing (helicases) proteins, and non-coding regulatory RNA. Together with a careful monitoring of mRNA half-lives, this could set the stage for the engineering of an efficient C₄ cycle in a C₃ species such as rice. Our unbiased genome-wide approach shows a massive rewiring of C₄ leaves at the post-transcriptional level compared with C₃ leaves. It may be crucial to understand this mechanism before trying to transform C₃ into C₄ crops.

Supplementary data

Supplementary data can be found at JXB online.

Table S1. EISA analysis on a control dataset from the WT and an ago1-27 Arabidopsis mutant.

Table S2. List of transcripts that are differentially expressed between young and expanded leaves in G. gynandra and T. hassleriana.

Table S3. List of transcripts that are common to both species after EISA analysis during leaf development (described by their AGI reference number).

Table S4. List of transcripts that are predicted to be post-transcriptionally regulated by EISA between young and expanded leaves in G. gynandra and T. hassleriana.

Acknowledgements

SA is supported by the University of Zurich and the Plant Fellows program of the Zürich–Basel Plant Science Centre. NF is supported by the Clinical Trial Unit of the University of Bern. We would like to thank Stefan Höftenstein and Enrico Martinoina for material support, and Julian Hibberd, Ivan Reyna-Llorens and Ben Williams for helpful discussion.

References


