



Conserved peptide upstream open reading frames are associated with regulatory genes in angiosperms

Richard A. Jorgensen^{1*} and Ana E. Dorantes-Acosta²

¹ Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, México

² Instituto de Biotecnología y Ecología Aplicada, Universidad Veracruzana, Xalapa, Veracruz, México

Edited by:

Anireddy S. N. Reddy, Colorado State University, USA

Reviewed by:

Bernie Carroll, The University of Queensland, Australia

Joseph Rothnagel, University of Queensland, Australia

Georg Weiller, Australian National University, Australia

*Correspondence:

Richard A. Jorgensen, Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional, Km 9.6 Libramiento Norte Carretera León, 36821 Irapuato, Guanajuato, México. e-mail: rajorgensen@langebio.cinvestav.mx

Upstream open reading frames (uORFs) are common in eukaryotic transcripts, but those that encode conserved peptides occur in less than 1% of transcripts. The peptides encoded by three plant conserved peptide uORF (CPuORF) families are known to control translation of the downstream ORF in response to a small signal molecule (sucrose, polyamines, and phosphocholine). In flowering plants, transcription factors are statistically over-represented among genes that possess CPuORFs, and in general it appeared that many CPuORF genes also had other regulatory functions, though the significance of this suggestion was uncertain (Hayden and Jorgensen, 2007). Five years later the literature provides much more information on the functions of many CPuORF genes. Here we reassess the functions of 27 known CPuORF gene families and find that 22 of these families play a variety of different regulatory roles, from transcriptional control to protein turnover, and from small signal molecules to signal transduction kinases. Clearly then, there is indeed a strong association of CPuORFs with regulatory genes. In addition, 16 of these families play key roles in a variety of different biological processes. Most strikingly, the core sucrose response network includes three different CPuORFs, creating the potential for sophisticated balancing of the network in response to three different molecular inputs. We propose that the function of most CPuORFs is to modulate translation of a downstream major ORF (mORF) in response to a signal molecule recognized by the conserved peptide and that because the mORFs of CPuORF genes generally encode regulatory proteins, many of them centrally important in the biology of plants, CPuORFs play key roles in balancing such regulatory networks.

Keywords: translational control, gene regulation, peptoswitch, sucrose signaling, polyamine signaling, trehalose signaling, regulatory networks, dicistronic transcripts

INTRODUCTION

Upstream open reading frames (uORFs) occur in 20–50% of eukaryotic transcripts (Kochetov, 2008). However, those encoding conserved peptides uORFs (CPuORFs) are relatively rare, occurring in less than 1% of angiosperm (Hayden and Jorgensen, 2007; Tran et al., 2008), dipteran (Hayden and Bosco, 2008), and mammalian (Crowe et al., 2006) transcripts. The function of CPuORFs has been studied in only a few cases, but in each case the CPuORF has been shown to control translation of a downstream ORF that is modulated by a small signal molecule. Three such examples are known in plants: *S*-adenosylmethionine decarboxylase, which is translationally regulated by polyamines (Ivanov et al., 2010), S1-group bZIP transcription factors, translationally regulated by sucrose (Rahmani et al., 2009), and phosphoethanolamine *N*-methyltransferase, translationally regulated by phosphocholine (Alatorre-Cobos et al., 2012).

Twenty-six different CPuORFs were identified in plants by comparison of rice and *Arabidopsis* full-length cDNAs (Hayden and Jorgensen, 2007). Many CPuORF genes appeared to have regulatory functions, and transcription factors were statistically over-represented among these genes. Here we reevaluate and extend the association of CPuORF genes with regulatory functions through

detailed analysis of the literature and consider the hypothesis that the function of the CPuORFs is to “fine tune” translation of downstream ORFs according to a signal to which the CPuORF peptide responds and further that CPuORF genes encode key regulatory proteins that mediate important responses to the environment as well as in growth and development. CPuORF peptides may be thought of as “peptoswitches,” by analogy to riboswitches, which are RNA structures that bind small molecules and thereby attenuate transcription and translation in prokaryotes and regulate splicing in eukaryotes according to the concentration of the small molecule.

THREE CLASSES OF CANDIDATE CPuORFs

Hayden and Jorgensen (2007) developed a program called *uORFinder* to pair the two most similar full-length cDNA sequences between two distantly related genomes and to identify candidate CPuORFs upstream of the major ORF (mORF). By comparison of *Arabidopsis* and rice cDNAs, *uORFinder* detected two classes of candidate CPuORF-carrying transcripts: (1) alternatively spliced transcripts that encode ORFs with alternate N-termini, where one alternative transcript encodes a candidate CPuORF that in the other alternative transcript is fused to the

major downstream ORF (mORF), and (2) transcripts that are not alternatively spliced. The latter were considered to represent putatively “true” CPuORFs. We suggest that a third class also exists, namely, dicistronic transcripts, which encode two transacting proteins, and that this class should be considered separately. Although dicistronic transcripts are relatively rare in plants, some other taxa, such as dipterans, possess hundreds of such genes (Lin et al., 2011). Thus, while dicistronic genes are at most a minor “contaminant” in CPuORF detection in plants, they are expected to be a major class in some other taxa. (Note: We do not exclude the possibility that the uORF of dicistronic transcripts might function also in regulating translation of the downstream ORF).

By reanalysis of more recently available *Arabidopsis* and rice cDNAs with *uORFinder*, we have also identified three new “true” CPuORF families (HG27–29) and one new dicistronic transcript (HG30). The putative CPuORF of HG30 encodes a cdc26 homolog. Among the original 26 angiosperm CPuORF families of Hayden and Jorgensen (2007), HG8’s CPuORF peptide is the only one that encodes a known domain, a CHCH domain that appears to belong to the twin CX₉C class of CHCH proteins (Cavallero, 2010). Cavallero (2010) suggested that twin CX₉C proteins play a scaffolding role in mitochondrial structure and function, although HG8’s CPuORF is a member of twin CX₉C cluster 21, which lacks any known function. HG8’s CX₉C protein is conserved among plants, animals, and fungi, yet is associated with different mORFs in different taxa, and even exists free of any mORF in some taxa (Hayden and Jorgensen, 2007). Thus, we suggest that HG8 should be considered as a probable dicistronic transcript.

All detected CPuORF families and dicistronic genes are shown in **Table 1**.

CPuORFs ARE ASSOCIATED WITH REGULATORY GENES IN ANGIOSPERMS

Hayden and Jorgensen (2007) reported that transcription factors were significantly over-represented among CPuORF families and noted also that most CPuORFs were associated with regulatory genes. Since that time, a great deal of new information has been published in the literature regarding functions in *Arabidopsis* of the mORFs associated with many of the identified CPuORF genes. By means of a comprehensive literature and database analysis, we have been able to identify known or inferred functions for the mORFs encoded by 22 of the 27 putatively “true” CPuORF families, summarized in **Table 1**. Remarkably, all 22 of these mORFs with a known or inferred function appear to play a regulatory role:

(1) Seven (26%) CPuORF families encode a variety of different types of transcription factor (**Table 1**), including bZIP (HG1), bHLH (HG2 and HG15), MADS (HG4), HD-ZIP (HG14), AP2/ERF (HG21), and an HSF-like activator (HG18). [Note: One of these (HG4) was not identified as a transcription factor by Hayden and Jorgensen (2007) because it had not been annotated as such in public database; however, the rice member of this family has been reported to be a MADS transcription factor by Arora et al. (2007) and Gramzow and Theissen (2010).

- (2) Nine (33%) CPuORF families are involved in signaling processes (**Table 1**): (a) Five of these families encode signal transduction protein kinases, including three MAP kinases (HG10, HG16, and HG23), a calcium response kinase (HG25), and a CBL-interacting protein kinase (HG27). (b) Four families are involved in the biosynthesis or catabolism of small molecule signals, including polyamines (HG3 and HG6), trehalose (HG11), and phosphocholine (HG13).
- (3) Four (15%) CPuORF families are involved in control of translation or protein degradation (**Table 1**). One family encodes translation initiation factor eIF5, whereas three other CPuORF families are predicted to play roles in the regulation of protein degradation: The HG22 mORF encodes a ubiquitin-specific protease, proteins which are implicated in regulation of ubiquitin-dependent protein degradation (Reyes-Turcu et al., 2009). HG26’s mORF encodes an E3 ubiquitin ligase, a protein that mediates degradation of a specific target protein(s) (Aguilar-Hernández et al., 2011). The mORF of HG28 encodes an F-box domain protein; F-boxes function in substrate recognition by SCF ubiquitin ligases, determining specificity.
- (4) CPuORF family HG5 encodes ANK6, an ankyrin repeat protein that is proposed to play a central role in male-female gamete recognition by regulating mitochondrial gene expression (Yu et al., 2010). It interacts in mitochondria with the non-conserved N-terminal region of SIG5, a mitochondrial sigma factor. By analogy to bacterial sigma factor 70, Yu et al. (2010) propose that by interacting with the inhibitory N-terminal sequence of SIG5, ANK6 uncovers SIG5’s C-terminal domain which then binds both to the mitochondrial RNA polymerase core to form a holoenzyme and to a mitochondrial promoter, initiating transcription of gamete-specific genes necessary for gamete recognition.
- (5) Finally, CPuORF family HG29 encodes DIC1, a mitochondrial dicarboxylate carrier which imports dicarboxylates into mitochondria to replenish the TCA cycle and is proposed to function as a malate/oxaloacetate shuttle providing other cellular compartments with reducing equivalents. When this shuttle is coupled with the chloroplastic malate valve, the plant acquires the ability to balance cellular energy supply and control redox poise (Palmieri et al., 2008).

Of the five genes of unknown function (19%), three have domains that are not sufficient to suggest whether these proteins perform a regulatory function:

- (1) HG9 encodes a cysteine-rich protein in which we detected CX_{4/7}CX₁₀CX₂HX₅ tandem repeats that to our knowledge have never been described in the literature or databases. Conceivably, it could be a novel zinc finger domain, but its function is unknown.
- (2) HG19 encodes a histidine-rich SAUR-like protein that binds metal ions *in vitro* (Hara et al., 2010). Conceivably, it could be involved in metal tolerance or homeostasis, but its function is not known.
- (3) HG24 encodes a domain of unknown function, DUF506 which has been shown to be a member of the PD-(D/E)XK

Table 1 | CPuORF genes: mORF molecular function and biological role.

Homology group	Arabidopsis locus ID	Arabidopsis gene name	mORF: known or inferred molecular function/domain	Known or inferred biological role	Comments	Citations
CONTROL OF TRANSCRIPTION						
HG1	At4g34590	<i>bZIP11 (GBF6)</i>	bZIP transcription factor	Transcriptional control	S1-group bZIPs	Rahmani et al. (2009)
	At2g18160	<i>bZIP2 (GBF5)</i>		Controls amino acid metabolism in response to sucrose status	Sucrose modulates translation via CPuORF	
	At3g62420	<i>bZIP53</i>				
	At5g49450	<i>bZIP1</i>				
	At1g75390	<i>bZIP44</i>				
HG2	At2g27230	<i>LHW (BHLH156)</i>	bHLH transcription factor	Transcriptional control	SACL3 (HG15) binds LHW	Parizot et al. (2008), Ohashi-Ito and Bergmann (2007)
	At2g31280	<i>BHLH155</i>		Root stele determination and lateral root initiation		Arora et al. (2007), Gramzow and Theissen (2010)
	At1g06150	<i>EMB1444</i>		Transcriptional control	Annotation based on rice ortholog Os02g01360 (OsMADS60)	Yao et al. (2011)
HG4	At5g52550		MADS-box, MIKC ^C type transcription factor			
	At4g25670					
	At4g25690					
HG14	At3g01470	<i>HD-ZIP1 (AtHB-1)</i>	HD-Zip class I transcription factor	Transcriptional control	Highly connected hub in light acclimation	
HG15	At5g64340	<i>SAC51 (BHLH142)</i>	bHLH transcription factor	Photosynthetic light acclimation	Thermospermine activates translation of SAC51; SACL3 binds LHW (HG2)	Imai et al. (2006), Kakehi et al. (2008), Takahashi and Kakehi (2009)
	At5g09460	<i>SACL1 (BHLH143)</i>		Transcriptional control		
	At1g29950	<i>SACL3 (BHLH144)</i>		Vascular (xylem) development		
	At5g50010	<i>SACL2 (BHLH145)</i>				
HG18	At4g36990	<i>TBF1 (HsfB1; HSF4)</i>	HSF-like transcription factor	Transcriptional control	Central regulator in pathogen response	Pajerowska-Mukhtar et al. (2012)
HG21	At1g25470	<i>ERF116 (CRF12)</i>	AP2 transcription factor;	Growth-to-defense transition	Cytokinin responsive	Rashotte and Goertzen (2010)
	At1g68550	<i>ERF118 (CRF10)</i>	cytokinin response factor	Transcriptional control		
CONTROL OF TRANSLATION AND PROTEIN DEGRADATION						
HG7	At1g36730	<i>eIF5-1</i>	Translation initiation factor eIF5	Regulates general translation in response to amino acid starvation	Start codon selection; GDI and GAP activity	Jennings and Pavitt (2010)
HG22	At1g78880		Ubiquitin-specific protease family C19-related	Regulation of ubiquitin-dependent protein degradation		Reyes-Turcu et al. (2009)
HG26	At5g05280	<i>ATL73</i>	RING finger protein; E3	Mediator of specific protein degradation		Aguiar-Hernández et al. (2011)
	At3g10910	<i>ATL72</i>	ubiquitin ligase			Xu et al. (2009)
HG28	At1g67480		F-box/keich-repeat protein	F-box determines substrate recognition by SCF ubiquitin ligases	Similar to SKIP6 (SKIP1-interacting partner 6)	
SIGNALING – SMALL MOLECULE SIGNALS						
HG3	At3g02470	<i>AdoMetDC1</i>	S-adenosylmethionine decarboxylase	Polyamine biosynthesis	Regulates growth, development and stress responses	Ivanov et al. (2010)
	At5g15950	<i>AdoMetDC2</i>				
	At3g25570	<i>AdoMetDC3</i>				
HG6	At2g43020	<i>PAO2</i>	Polyamine oxidase	Catabolizes spermidine, norspermine, thermospermine	Growth, development and stress responses	Takahashi et al. (2010), Fincato et al. (2011)
	At3g59050	<i>PAO3</i>				

(Continued)

Table 1 | Continued

Homology group	Arabidopsis locus ID	Arabidopsis gene name	mORF: known or inferred molecular function/domain	Known or inferred biological role	Comments	Citations
HG11	At4g12430 At4g22590	<i>TPP5 (TPPF)</i> <i>TPP6 (TPPG)</i>	Trehalose-6-phosphate phosphatase	T6P inhibits SNRK1 indicator of sucrose status, regulates sucrose utilization and carbon partitioning	TPP6 lacks synthase domain	Ponnu et al. (2011), Ma et al. (2011), Valluru and Van den Ende (2011)
HG13	At3g18000 At1g48600 At1g73600	<i>XPL1 (NMT1)</i> <i>NMT2</i> <i>NMT3</i>	Phosphoethanolamine N-methyltransferase	Phosphatidic acid and phosphocholine biosynthesis; root development	Phosphocholine regulates translation of XPL1 via CPuORF	Alatorre-Cobos et al. (2012)
SIGNALING – SIGNAL TRANSDUCTION KINASES						
HG10	At4g19110		MAP kinase, PPC family 4.5.1	Signal transduction; <i>Chlamydomonas</i> ortholog LF4 regulates flagellar length	Pollen and sperm cell-specific expression	Berman et al. (2003)
HG16	At5g45430		MAP kinase, PPC family 4.1.5	Signal transduction		
HG23	At1g64630 At5g41990	<i>WNK5 (ZIK1)</i> <i>WNK8 (ZIK6)</i> <i>WNK10 (ZIK10)</i>	MAP kinase, PPC family 4.1.5	Signal transduction promotes flowering	Phosphorylates EDM2 and V-ATPase	Tsuchiya and Eulgem (2010)
HG25	At5g60550	<i>GRIK1 (SNAK2)</i>	Calcium response protein kinase	Ca ⁺⁺ /CaM—dependent signal transduction; sugar signaling	Phosphorylates SNRK1	Shen and Hanley-Bowdoin (2006)
HG27	At3g45240 At4g30960	<i>GRIK2 (SNAK1)</i> <i>CIPK6 (SIP2)</i>	CBL-interacting protein kinase	Signal transduction salt and drought tolerance	Activates K ⁺ transporter AKT2	Tsou et al. (2012), Dreyer and Uozumi (2011)
OTHER REGULATORY FUNCTION						
HG5	At5g61230	<i>ANK6</i>	Ankyrin repeats – protein–protein interaction	Male-female gamete recognition	Interacts with mitochondrial sigma factor SIG5	Yu et al. (2010)
HG29	At2g22500	<i>DIC1</i>	Dicarboxylate carrier protein dicarboxylic acid transporter	Redox connection between mitochondria and cytosol		Palmieri et al. (2008)
OTHER FUNCTION OR UNKNOWN FUNCTION						
HG9	At5g64550 At5g09670		Cysteine-rich protein: CX ₄₇ -CX ₁₀ CX ₂ HX ₅ tandem repeats	Unknown	Novel zinc finger?	
HG12	At1g64140 At1g23150 At1g70780		Unknown	Unknown		
HG19	At5g53590	<i>HIRP1</i>	Histidine-rich SAUR-like protein	Unknown	Metal binding: metal homeostasis or tolerance?	Hara et al. (2010)
HG20	At2g37480		Unknown	Unknown		
HG24	At3g53670 At3g22970 At4g14620		DUF506; PD-(D/E)XK nuclease	Unknown		Knizewski et al. (2007)
PUTATIVE DIGESTRONIC TRANSCRIPTS (NOT “TRUE” CPuORFs?)						
HG8	At3g12010		TGF-beta superfamily; Mic-1 (GDF15) putative ortholog	Cell cycle regulation, S phase arrest, and starvation for DNA precursors (in animals)	CPuORF conserved across eucaryotes; “free” of mORF in many taxa	Agarwal et al. (2006), Cavallero (2010)
HG30	At2g11890		Adenylate cyclase	cAMP biosynthesis	CPuORF is cdc26	

nuclease superfamily that has many functions, including restriction endonucleases, transposases, and roles in DNA degradation, recombination, and repair, and RNA processing (Knizewski et al., 2007). The DUF506 family has 18 members in rice and 23 in *Arabidopsis*, but none have any known function.

The distribution of functional classes is diagrammed in **Figure 1**.

Although the two putative dicistronic families, HG8 and HG30, have not been included in this summary, our conclusion would not be weakened by their inclusion because both appear to encode regulatory genes. The HG8 mORF encodes a member of the TGF-beta superfamily, an ortholog of the Mic-1/GDF15 gene, which is involved in cell cycle regulation in response to starvation for DNA precursors in animals. HG30's mORF encodes adenylate cyclase, which is involved in cyclic AMP biosynthesis; cyclic AMP is a small signaling molecule. Thus, inclusion of HG8 and HG30 would only reinforce our conclusion that CPuORFs are associated with regulatory genes.

It is not possible to calculate the statistical significance of the conclusion that CPuORFs are associated with regulatory genes because the genome has not been annotated at the same level of detail as we have re-annotated CPuORF genes. However, with the exception of small signal molecules, biosynthetic, and other "housekeeping" enzymes responsible for primary and secondary metabolism are strikingly absent, such as biosynthesis of carbohydrates, amino acids, fatty acids, nucleotides, enzyme cofactors, sterols, chlorophylls, carotenoids, terpenoids, alkaloids, flavonoids, isoflavonoids, and anthocyanins, as well as polymers such as cellulose, polysaccharides, waxes, lignins, and tannins. Even leaving out small molecule signal genes, the association of CPuORFs with

regulatory genes (e.g., transcription factors and signal transduction kinases comprise 12 of the 27 CPuORF families) is very striking.

FUNCTIONS ASSOCIATED WITH GENES ENCODING ALTERNATIVE N-TERMINI DUE TO ALTERNATIVE SPLICING

Some alternatively spliced genes that encode proteins with different N-termini due to fusion of the uORF and mORF of one transcript into a single ORF in the alternatively spliced transcript, i.e., the N-terminus of the second alternative transcript is the same as the N-terminus of the uORF of the first transcript. Hayden and Jorgensen (2007) did not report the identities of genes in the alternative N-termini (aNT) class, so we repeated the analysis of *Arabidopsis* and rice full-length cDNAs with *uORFinder* and detected 25 cases of N-terminal alternative splicing (**Table A1** in Appendix). (Note: Based on alternative transcripts now available in public databases, we have reclassified family HG17 as aNT family aNT25.) Alternative transcripts encoding similar proteins with different N-termini were produced by several different types of alternative splicing: intron retention, alternative acceptor site, and alternative splice site selection.

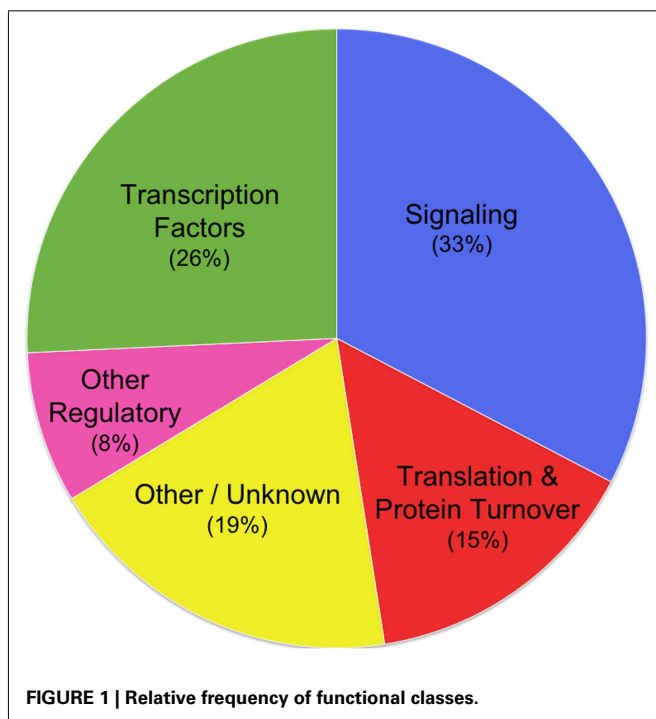
The spectrum of functions of the 25 aNT genes (**Table A1** in Appendix) is strikingly different than the spectrum of functions found in CPuORF genes (**Table 1**). Although a number of these genes play regulatory roles, what is most striking is the abundance of proteins associated with membranes or organelles: nine are localized to the plasma membrane and endomembrane system seven in chloroplasts (four of these in chloroplast membranes), six in the nucleus, one to the cytoskeleton, one to the cytosol, and one unknown location. Alternative splicing to produce aNT has been shown in some cases to be responsible for alternative localization of proteins to different compartments in plant cells (Silva-Filho, 2003), and this may also be the case with the alternatively spliced, aNT transcripts identified by *uORFinder*. These aNT genes function in diverse processes, including glycolysis, cell wall biogenesis, membrane fusion, microtubule organization, spliceosome assembly, apoptosis, light harvesting, membrane targeting and transport, and biosynthesis of chlorophyll and phyloquinone in the chloroplast. The striking difference in the spectrum of functions for the aNT class as compared to the spectrum of functions identified for the CPuORF class lends support to the observation that CPuORFs are associated with regulatory genes in plants.

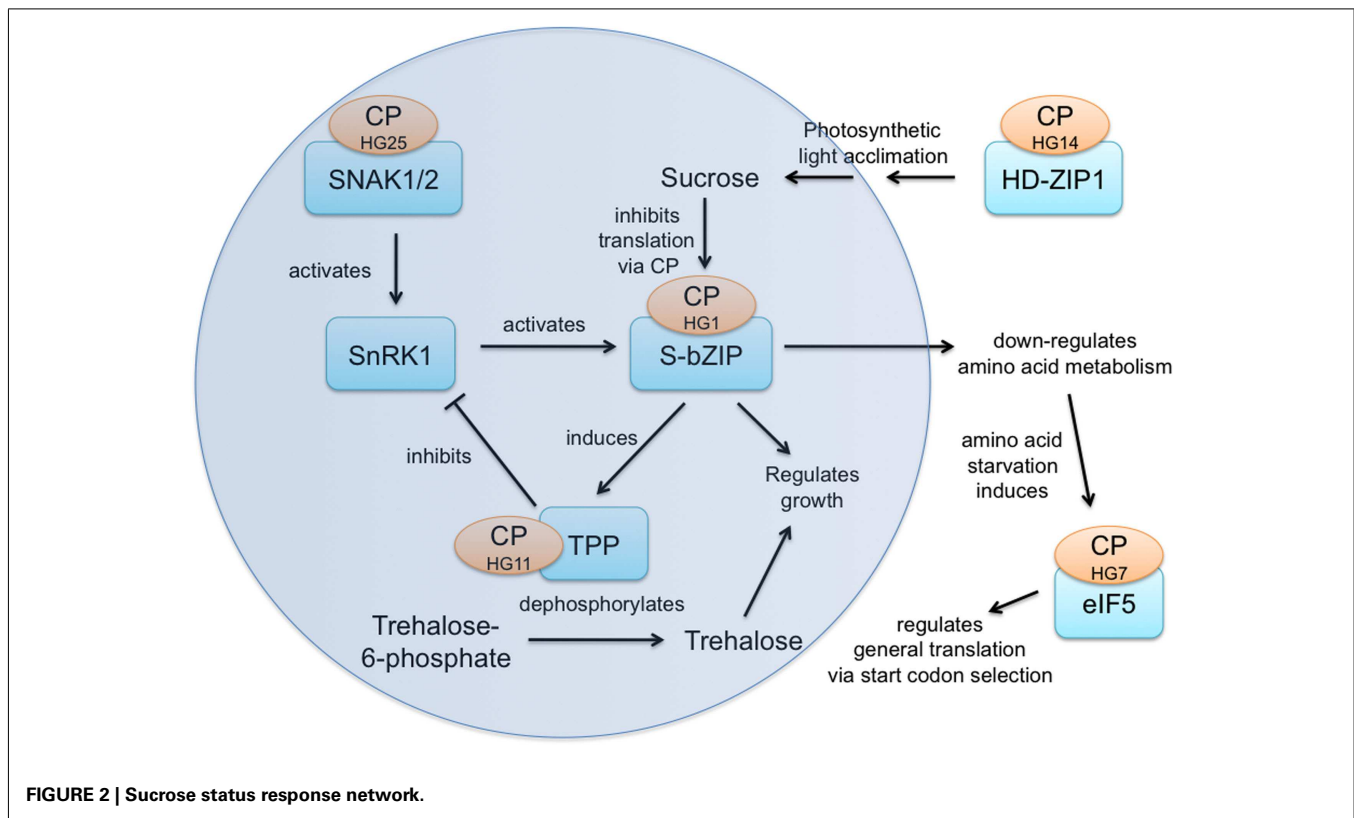
DIVERSE TYPES OF REGULATORY PROCESSES ARE ASSOCIATED WITH CPuORF GENES

Extensive review of the experimental research literature showed that 16 CPuORF families are involved in key regulatory processes in growth, development, and physiology (**Table 1**), as follows.

THE SUCROSE STATUS RESPONSE NETWORK INVOLVES FIVE CPuORF FAMILIES

Sucrose is a signal molecule in plants and has been shown to modulate translation of the S1-group of bZIP transcription factors in order to adjust metabolism to altered sucrose status (Rahmani et al., 2009). The core SnRK1/T6P/S-bZIP sucrose status response network (**Figure 2**) includes three CPuORF families:





- (1) SnRK1 activates SNAK1/2 (Shen and Hanley-Bowdoin, 2006), which have the HG25 CPuORF.
- (2) SnRK1 acts through S1-bZIP and the other S1-group bZIP transcription factors (HG1 family) whose translation is modulated by sucrose via the HG1 CPuORF (Rahmani et al., 2009).
- (3) S1-bZIP induces transcription of TPP5 and TPP6 (trehalose-6-phosphate phosphatases) whose translation may be modulated via the CPuORF of HG11. Trehalose-6-phosphate (T6P) is known to inhibit SnRK1 (Zhang et al., 2009).

Homeobox leucine zipper transcription factor HD-ZIP1 (HG14) is a highly connected central hub in the long-term photosynthetic light acclimation response whose expression is induced by a signal of the oxidized state of the plastoquinone pool (Yao et al., 2011), thereby influencing sucrose levels and the sucrose status response network. In addition, the S1-bZIP transcription factors regulate growth, in part by down-regulating amino acid metabolism in response to sucrose acting through CPuORF family HG1. Amino acid starvation induces translation initiation factor eIF5 (HG7), which regulates general translation by altering start codon selection (Jennings and Pavitt, 2010), potentially influencing translation of many transcripts possessing upstream ATGs. Thus, five CPuORF families participate in, influence, or respond to this energy sensing network, diagrammed in **Figure 2**. Only the HG1 CPuORF has an experimentally determined function in regulating mORF translation. Clearly, it would be very interesting to determine the functions of the other four CPuORF families and the nature of any signals to which they might respond.

THREE CPuORF FAMILIES ARE INVOLVED IN POLYAMINE METABOLISM OR ACTION

Polyamines are small signal molecules that play important roles in both vascular development and abiotic stress tolerance (Takahashi and Kakehi, 2009; Alcázar et al., 2010; Vera-Sirera et al., 2010). Three CPuORF families involve polyamines:

- (1) HG3 encodes S-adenosylmethionine decarboxylase, which is required for polyamine biosynthesis (Ivanov et al., 2010).
- (2) HG6 encodes polyamine oxidases that catabolize spermidine, norspermidine, and thermospermine (Takahashi et al., 2010; Fincato et al., 2011).
- (3) Mutation of the HG15 CPuORF suppresses thermospermine synthase mutations, and thermospermine is thought to regulate translation of SAC51 via the HG15 CPuORF (Takahashi and Kakehi, 2009; Vera-Sirera et al., 2010). SAC51 encodes a bHLH transcription factor necessary for xylem development (Imai et al., 2006).

OTHER REGULATORS OF PLANT METABOLISM AND RESPONSES TO THE ENVIRONMENT

Three additional CPuORF families play key roles in the acclimation of physiology and metabolism to changing conditions:

- (1) TBF1 (HG18) encodes a transcription factor that is a “master molecular switch” in the growth-to-defense transition, activating, and repressing more than 1,000 genes that carry a TBF1 binding site in their promoters (Pajerowska-Mukhtar et al., 2012). TBF1 adjusts physiology in response to pathogen

attack via its uORF, which is rich in phenylalanine and regulates translation in response phenylalanine starvation. Thus, TBF1 senses metabolic changes during pathogen attack and triggers specific transcriptional reprogramming. However, the phenylalanine richness of this CPuORF does not explain the highly conserved 22 amino acid segment of the CP. The possibility should be considered that this CP recognizes some signal molecule that regulates translation via the interaction.

- (2) DIC1 (HG29), as mentioned above, has been proposed to balance cellular energy supply and control redox poise (Palmieri et al., 2008).
- (3) CIPK6 (HG27) is a serine/threonine protein kinase whose up-regulation results in enhanced drought and salt tolerance (Tsou et al., 2012).

REGULATION OF DEVELOPMENT

Six other CPuORF families play important roles in the regulation of development:

- (1) LONESOME HIGHWAY (LHW; HG2) encodes a bHLH transcription factor necessary for root stele determination and lateral root initiation (Ohashi-Ito and Bergmann, 2007).
- (2) As was mentioned in the discussion of polyamines, SAC51 (HG15) encodes a bHLH transcription factor necessary for xylem development (Imai et al., 2006).
- (3) ANK6 (HG5), as was mentioned above, is believed to regulate transcription of gamete-specific genes in the mitochondrial genome, thereby controlling male-female gamete recognition (Yu et al., 2010).
- (4) XPL1 (HG13) is required for phosphocholine synthesis in plants. Phosphocholine concentration regulates the proliferative capacity of root meristems, the proper elongation of root cells, and the viability of the elongating root epidermal cells. Phosphocholine regulates XPL1 translation via its CPuORF (Alatorre-Cobos et al., 2012).
- (5) WNK8 (HG23) acts upstream of EDM2 in an FLC-dependent mechanism that controls the timing of the floral transition and counteracts the promoting effect of EDM2 on this developmental process, i.e., it is a key component of a regulatory module that resembles the autonomous floral promotion pathway (Tsuchiya and Eulgem, 2010).
- (6) Family HG10 encodes a MAP kinase that is highly conserved between plants and animals. The HG10 ortholog in *Chlamydomonas*, LF4, regulates flagellar length (Berman et al., 2003). LF4 has no role in assembly, but rather maintains a specific length in assembled flagella, apparently by shortening. Interestingly, the plant HG10 family is expressed specifically in pollen and sperm cells where it may serve a similar function.

BIOLOGICAL PROCESS UNKNOWN

Six families whose mORFs encode domains inferred to have a regulatory function have not been associated experimentally with a specific biological process. Neither have the five families whose mORF functions are unknown been associated experimentally with a biological process. Given the important roles of the 16 better-characterized CPuORF families, it is interesting to consider

that some of these other 11 families might play similarly important roles in the biology of plants.

DISCUSSION

THE HYPOTHESIS FOR MODULATING mORF TRANSLATION VIA CPuORFs

The three families whose CPuORFs have been investigated experimentally regulate mORF translation via the CPuORF in response to a small molecule signal: translation of bZIP11 (HG1) is regulated by sucrose via its CPuORF; translation of *S*-adenosyl methionine decarboxylase (HG2) is regulated by polyamines via its CPuORF; and translation of phosphoethanolamine *N*-methyltransferase is regulated by phosphocholine via the HG13 CPuORF. In each case, the CPuORF has been shown to modulate, or “fine tune,” translation of its downstream mORF in response to a particular small molecule signal. Thus, an attractive hypothesis is that other CPuORF families also exhibit small molecule regulation of mORF translation. In other words, CPuORFs act as “peptoswitches,” by analogy to riboswitches which are RNA structures that bind small molecules and attenuate transcription and translation in prokaryotes and regulate splicing in eukaryotes.

Our observation that many CPuORF genes are associated with key regulatory processes suggests an expanded hypothesis, i.e., that CPuORFs play a major role in the biology plants by sensitively modulating translation of key regulatory proteins in response to specific signal molecules, thereby optimizing metabolic, physiological, and developmental networks for particular conditions.

Given the central importance of many CPuORF genes in the biology of plants, we would suggest that analysis of all CPuORF families should be a priority for experimental research designed to ascertain whether such signal molecules exist and regulate mORF translation, and if they do, to determine their molecular natures and whether they fine tune gene expression networks. This is important at two levels. First, in the 16 CPuORF families with known regulatory roles it would be interesting to investigate the potential role of the CPuORF in regulating translation of its downstream mORF, as well as the nature of any signal molecule that might act through the CPuORF. Second, in the 12 families lacking an experimentally determined role in regulation, it could be very interesting to learn the role played in the biology of plants because these genes may also play central roles in the biology of plants. Given the apparently key regulatory roles that many CPuORF genes play, we feel that such experiments should be given a high priority relative to many other genes whose function also remains unknown or poorly understood.

SELECTIVE REGULATION OF TRANSLATION INITIATION IN PLANTS

Selective translation of mRNAs has been shown to play an important role in some plant acclimation responses. For instance, cellular oxygen deprivation requires plants to acclimatize in order to survive. It has been shown that the conservation of energy and the transition to anaerobic metabolism that occur during such hypoxia stress is coordinated by selective mRNA translation, resulting in the inhibition of translation of ~70% of cellular mRNAs (Branco-Price et al., 2008). CPuORF genes represent a specific class of “selectively” translated genes that are relatively rare. CPuORFs

were estimated by Hayden and Jorgensen (2007) to occur in only ~0.5% of *Arabidopsis* genes.

FURTHER CONSIDERATIONS OF THREE CPuORF GENES

eIF5 (HG7)

The presence of one or more uORFs is a near universal feature of eukaryotic eIF5 mRNAs (Loughran et al., 2011). Experiments in human cells have shown that eIF5 mRNA autoregulates its own translation via inhibitory uORFs. eIF5 protein is known to increase the stringency of start codon selection, whereas at high eIF5 levels it induces translation initiation at AUGs in poor contexts. Because the uORF AUG(s) of eIF5 mRNAs have poor initiation contexts and translation of these uORFs is inhibitory to translation of the eIF5 mORF, high levels of eIF5 reduce translation of the eIF5 mORF, providing negative feedback (Loughran et al., 2011). Given that plant eIF5 mRNAs encode a CPuORF (HG7), we suggest that modulation of eIF5 translation by an unknown signal molecule triggers a general shift in mRNA translation in favor of AUGs in “poor” initiation contexts at the expense of those in “good” contexts, thereby mediating a global change in translation rates of many thousands of mRNAs and shifting the relative abundances of many proteins. Alternatively, the conserved peptide might function by directly interacting with ribosomes or other initiation factors. Distinguishing these possibilities will require direct experimentation.

TBF1 (HG18)

TBF1 has been shown to be a key regulator of the growth-to-defense transition (Pajeroska-Mukhtar et al., 2012). The CPuORF of TBF1 is the second of two uORFs in the 5' UTR that repress TBF1 translation. Both uORF peptide sequences contain four phenylalanine residues, and phenylalanine starvation was shown to alleviate translational repression by the uORFs. Pathogen challenge increases uncharged tRNA^{Phe} and phosphorylation of translation initiation factor eIF2alpha and releases the inhibitory effects of the uORFs on TBF1 translation. This is similar to GCN4 in yeast whose translation is derepressed by amino acid starvation via accumulation of uncharged tRNAs that induce

phosphorylation of eIF2alpha, allowing translation to reinitiate at the GCN4 start codon. Translational control allows rapid response to pathogen detection, via the rapid, dramatic increase in uncharged tRNA^{Phe}, by producing TBF1 protein from transcripts already present, thereby activating the immune response (Pajeroska-Mukhtar et al., 2012). An important question that remains about TBF1 is why the peptide sequence of its second uORF is conserved. One possibility is that the peptide interacts with the ribosome to influence whether it is able to reinitiate at the TBF1 start codon, though it should be noted that none of the GCN4 uORFs is conserved. The other possibility that should be considered is that the HG18 conserved peptide interacts with a signal molecule that could influence translation.

CIPK6 (HG27)

CIPK6 is a serine/threonine kinase that enhances drought and salt tolerance when upregulated. It appears to do so by modulating the activity of potassium channel AKT2 and mediates AKT2 translocation to the plasma membrane (Held et al., 2011). Phosphorylation and desphosphorylation of AKT2 channels control switching between two gating modes, one mode opening channels at voltage more negative than -100 mV and the other mode being voltage independent. Toggling between modes assists in energizing transmembrane transport processes, predominantly in phloem tissues where K⁺ ions serve as decentralized energy storage (Dreyer and Uozumi, 2011). CIPK6 is only one of many CIPKs in plants, and presumably it regulates AKT2 in response to a specific condition or signal that it transmits to AKT2. Conceivably, translational regulation of CIPK6 could be a key aspect of that response, perhaps leading to more rapid switching of AKT2 channel gating than might be achieved by only a transcriptional response, and possibly modulating the response quantitatively in response to the degree of drought or salt stress.

ACKNOWLEDGMENTS

The authors would like to thank Cassie Heinle for assisting with updating the uORFinder searches of rice and *Arabidopsis* cDNA sequences. CONACyT provided financial support to Richard A. Jorgensen (133990) and Ana E. Dorantes-Acosta (158561).

REFERENCES

- Agarwal, M. K., Hastak, K., Jackson, M. W., Breit, S. N., Stark, G. R., and Agarwal, M. L. (2006). Macrophage inhibitory cytokine 1 mediates a p53-dependent protective arrest in S phase in response to starvation for DNA precursors. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16278–16283.
- Aguilar-Hernández, V., Aguilar-Henonin, L., and Guzmán, P. (2011). Diversity in the architecture of ATLs, a family of plant ubiquitin-ligases, leads to recognition and targeting of substrates in different cellular environments. *PLoS ONE* 6, e23934. doi:10.1371/journal.pone.0023934
- Alatorre-Cobos, F., Cruz-Ramírez, A., Hayden, C. A., Pérez-Torres, C.-A., Chauvin, A., Ibarra-Laclette, E., Alva-Cortés, E., Jorgensen, R. A., and Herrera-Estrella, L. (2012). Translational regulation of *Arabidopsis* XIPOTL1 is modulated by phosphocholine levels via the conserved upstream open reading frame 30. *J. Exp. Bot.* doi:10.1093/jxb/ers180
- Alcázar, R., Altabella, T., Marco, F., Bor-tolotti, C., Reymond, M., Koncz, C., Carrasco, P., and Tiburcio, A. F. (2010). Polyamines: molecules with regulatory functions in plant abiotic stress tolerance. *Planta* 231, 1237–1249.
- Arora, R., Agarwal, P., Ray, S., Singh, A. K., Singh, V. P., Tyagi, A. K., and Kapoor, S. (2007). MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* 8, 242. doi:10.1186/1471-2164-8-242
- Berman, S. A., Wilson, N. F., Haas, N. A., and Lefebvre, P. (2003). A novel MAP kinase regulates flagellar length in *Chlamydomonas*. *Cell* 13, 1145–1149.
- Branco-Price, C., Kayla, A., Kaiser, K.A., Jang, C. J. H., Larive, C. K., and Bailey-Serres, J. (2008). Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*. *Plant J.* 56, 743–755.
- Cavallero, G. (2010). Genome-wide analysis of eukaryotic twin CX9C proteins. *Mol. Biosyst.* 6, 2459–2470.
- Crowe, M. L., Wang, X. Q., and Rothnagel, J. A. (2006). Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides. *BMC Genomics* 7, 16. doi:10.1186/1471-2164-7-16
- Dreyer, I., and Uozumi, N. (2011). Potassium channels in plant cells. *FEBS J.* 278, 4293–4303.
- Fincato, P., Moschou, P. N., Spedaletti, V., Tavazza, R., Angelini, R., Federico, R., Roubelakis-Angelakis, K. A., and Tavladoraki, P. (2011). Functional diversity inside the *Arabidopsis* polyamine oxidase gene family. *J. Exp. Bot.* 62, 1155–1168.
- Gramzow, L., and Theissen, G. (2010). A Hitchhiker's guide to the MADS world of plants. *Genome Biol.* 11, 214.
- Hara, M., Kashima, D., Horiike, T., and Toru Kuboi, T. (2010). Metal-binding characteristics of the protein which shows the highest histidine content in the *Arabidopsis* genome. *Plant Biotechnol.* 27, 475–480.

- Hayden, C. A., and Bosco, G. (2008). Comparative genomic analysis of novel conserved peptide upstream open reading frames in *Drosophila melanogaster* and other dipteran species. *BMC Genomics* 9:61. doi:10.1186/1471-2164-9-61
- Hayden, C. A., and Jorgensen, R. A. (2007). Identification of novel conserved peptide uORF homology groups in *Arabidopsis* and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. *BMC Biol.* 5, 32. doi:10.1186/1741-7007-5-32
- Held, K., Pascaud, F., Eckert, C., Gajdanowicz, P., Hashimoto, K., Corratge-Faillie, C., Offenborn, J. N., Lacombe, B., Dreyer, I., Thibaud, J. B., and Kudla, J. (2011). Calcium-dependent modulation and plasma membrane targeting of the AKT2 potassium channel by the CBL4/CIPK6 calcium sensor/protein kinase complex. *Cell Res.* 21, 1116–1130.
- Imai, A., Hanzawa, Y., Komura, M., Yamamoto, K. T., Komeda, Y., and Takahashi, T. (2006). The dwarf phenotype of the *Arabidopsis* ac5 mutant is suppressed by a mutation in an upstream ORF of a bHLH gene. *Development* 133, 3575–3585.
- Ivanov, I. P., Atkins, J. F., and Michael, A. J. (2010). A profusion of upstream open reading frame mechanisms in polyamine-responsive translational regulation. *Nucleic Acids Res.* 38, 353–359.
- Jennings, M. D., and Pavitt, G. D. (2010). eIF5 is a dual function GAP and GDI for eukaryotic translational control. *Small GTPases* 1, 1–6.
- Takehi, J., Kuwashiro, Y., Niitsu, M., and Takahashi, T. (2008). Thermospermine is required for stem elongation in *Arabidopsis thaliana*. *Plant Cell Physiol.* 49, 1342–1349.
- Knizewski, L., Kinch, L. N., Grishin, N. V., Rychlewski, L., and Ginalski, K. (2007). Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.* 7, 40. doi:10.1186/1472-6807-7-40
- Kochetov, A. V. (2008). Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *Bioessays* 30, 683–691.
- Lin, M. F., Carlson, J. W., Crosby, M. A., Matthews, B. B., Yu, C., Park, S., Wan, K. H., Schroeder, A. J., Gramates, L. S., St Pierre, S. E., Roark, M., Wiley, K. L. Jr., Kulathinal, R. J., Zhang, P., Myrick, K. V., Antone, J. V., Celniker, S. E., Gelbart, W. M., and Kellis, M. (2011). Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* 17, 1823–1836.
- Loughran, G., Sachs, M. S., Atkins, J. F., and Ivanov, I. P. (2011). Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Res.* 40, 2898–2906.
- Ma, J., Hanssen, M., Lundgren, K., Hernández, L., Delatte, T., Ehlert, A., Liu, C.-M., Schlupepmann, H., Dröge-Laser, W., Moritz, T., Smeekens, S., and Hanson, J. (2011). The sucrose-regulated *Arabidopsis* transcription factor bZIP11 reprograms metabolism and regulates trehalose metabolism. *New Phytol.* 191, 733–745.
- Ohashi-Ito, K., and Bergmann, D. C. (2007). Regulation of the *Arabidopsis* root vascular initial population by LONESOME HIGHWAY. *Development* 134, 2959–2968.
- Pajerowska-Mukhtar, K. M., Wang, W., Tada, Y., Oka, N., Tucker, C. L., Fonseca, J. P., and Dong, X. (2012). The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Curr. Biol.* 22, 103–112.
- Palmieri, L., Picault, N., Arrigoni, R., Besin, E., Palmieri, F., and Hodges, M. (2008). Molecular identification of three *Arabidopsis thaliana* mitochondrial dicarboxylate carrier isoforms: organ distribution, bacterial expression, reconstitution into liposomes and functional characterization. *Biochem. J.* 410, 621–629.
- Parizot, B., Laplaze, L., Ricaud, L., Boucheron-Dubuisson, E., Bayle, V., Bonke, M., De Smet, I., Poethig, S. R., Helariutta, Y., Haseloff, J., Chriqui, D., Beekman, T., and Nussaume, L. (2008). Diarch symmetry of the vascular bundle in *Arabidopsis* root encompasses the pericycle and is reflected in distich lateral root initiation. *Plant Physiol.* 146, 140–148.
- Ponnu, J., Wahl, V., and Schmid, M. (2011). Trehalose-6-phosphate: connecting plant metabolism and development. *Front. Plant Sci.* 2:70. doi:10.3389/fpls.2011.00070
- Rahmani, F., Hummel, M., Schurmans, J., Wiese-Klinkenberg, A., Smeekens, S., and Hanson, J. (2009). Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol.* 150, 1356–1367.
- Rashotte, A. M., and Goertzen, L. R. (2010). The CRF domain defines Cytokinin Response Factor proteins in plants. *BMC Plant Biol.* 10, 74. doi:10.1186/1471-2229-10-74
- Reyes-Turcu, F. E., Ventii, K. H., and Wilkinson, K. D. (2009). Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu. Rev. Biochem.* 78, 363–397.
- Shen, W., and Hanley-Bowdoin, L. (2006). Geminivirus infection up-regulates the expression of two *Arabidopsis* protein kinases related to yeast SNF1 and mammalian AMPK-activating kinases. *Plant Physiol.* 142, 1642–1655.
- Silva-Filho, M. C. (2003). One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. *Curr. Opin. Plant Biol.* 6, 589–595.
- Takahashi, T., Cong, R., Sagor, G. H., Niitsu, M., Berberich, T., and Kusano, T. (2010). Characterization of five polyamine oxidase isoforms in *Arabidopsis thaliana*. *Plant Cell Rep.* 29, 955–965.
- Takahashi, T., and Takehi, J.-I. (2009). Polyamines: ubiquitous polycations with unique roles in growth and stress responses. *Ann. Bot.* 105, 1–6.
- Tran, M. K., Carolyn, J., Schultz, C. J., and Ute Baumann, U. (2008). Conserved upstream open reading frames in higher plants. *BMC Genomics* 9, 361. doi:10.1186/1471-2164-9-361
- Tsou, P.-L., Lee, S. Y., en Allen, N. S., Winter-Sederoff, H., and Robertson, D. (2012). An ER-targeted calcium-binding peptide confers salt and drought tolerance mediated by CIPK6 in *Arabidopsis*. *Planta* 235, 539–552.
- Tsuchiya, T., and Eulgem, T. (2010). Co-option of EDM2 to distinct regulatory modules in *Arabidopsis thaliana* development. *BMC Plant Biol.* 10, 203. doi:10.1186/1471-2229-10-203
- Valluru, R., and Van den Ende, W. (2011). Myo-inositol and beyond-emerging networks under stress. *Plant Sci.* 181, 397–400.
- Vera-Sirera, F., Minguet, E. G., Singh, S. K., Ljung, K., Tuominen, H., Blázquez, M. A., and Carbonell, J. (2010). Role of polyamines in plant vascular development. *Plant Physiol. Biochem.* 48, 534–539.
- Xu, G., Ma, H., Nei, M., and Kong, H. (2009). Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification. *Proc. Natl. Acad. Sci. U.S.A.* 106, 835–840.
- Yao, C. W., Hsu, B. D., and Chen, B. S. (2011). Constructing gene regulatory networks for long term photosynthetic light acclimation in *Arabidopsis thaliana*. *BMC Bioinformatics* 12, 335. doi:10.1186/1471-2105-12-335
- Yu, F., Shi, J., Zhou, J., Gu, J., Chen, Q., Li, J., Cheng, W., Mao, D., Tian, L., Buchanan, B. B., Li, L., Chen, L., Li, D., and Shen, L. (2010). ANK6, a mitochondrial ankyrin repeat protein, is required for male-female gamete recognition in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22332–22337.
- Zhang, Y., Primavesi, L. F., Jhurreea, D., Andralojc, P. J., Mitchell, R. A. C., Powers, S. J., Schlupepmann, H., Delatte, T., Wingler, A., and Paul, M. J. (2009). Inhibition of SNF1-related protein kinase1 activity and regulation of metabolic pathways by trehalose-6-phosphate. *Plant Physiol.* 149, 1860–1871.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 June 2012; accepted: 04 August 2012; published online: 24 August 2012.

Citation: Jorgensen RA and Dorantes-Acosta AE (2012) Conserved peptide upstream open reading frames are associated with regulatory genes in angiosperms. *Front. Plant Sci.* 3:191. doi:10.3389/fpls.2012.00191

This article was submitted to *Frontiers in Plant Genetics and Genomics*, a specialty of *Frontiers in Plant Science*.

Copyright © 2012 Jorgensen and Dorantes-Acosta. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

APPENDIX

Table A1 | Alternative splicing fuses uORF to mORF.

Alt NT group	TAIR locus ID	<i>Arabidopsis</i> gene name	Inferred molecular function/domain(s)	Inferred biological role/process and cellular location (TAIR)	Alt splice type (<i>Arabidopsis</i>)
LOCALIZED TO PLASMA MEMBRANE AND ENDOMEMBRANE SYSTEM					
aNT1	At1g02145	<i>EBS4/ALG12</i>	alpha-1,6-mannosyltransferase; Alg9-like mannosyltransferase	Protein catabolic process Located in endomembrane system; intrinsic to ER membrane	Intron retention
aNT3	At1g08350		Endomembrane protein 70 family Non-aspanin	Cellular membrane fusion Integral to membrane, phragmoplast	Intron retention
aNT5	At1g29890	<i>REDUCED WALL ACETYLATION (RWA4)</i>	<i>O</i> -acetyltransferase; Cas1p-like	Secondary wall biogenesis Located in Golgi apparatus	Intron retention
aNT13	At3g45960	<i>EXPANSIN-LIKE A3 (EXLA3)</i>	Expansin; endoglucanase	Cell growth, cell wall loosening; located in cytosol, cell wall, membrane, plasmodesma	Intron retention
aNT15	At3g52990	PK90	Pyruvate kinase	Glycolysis Located in cytosol, membrane	Intron retention
aNT18	At5g22140		FAD/NAD(P)-binding oxidoreductase family protein; electron carrier	Oxidation-reduction process Located in plasma membrane, plasmodesma	Alt splice site selection
aNT19	At5g41610	<i>CATION/H+ EXCHANGER 18 (CHX18)</i>	Na ⁺ /H ⁺ antiporter	Cation transport Located in late endosome	Intron retention
aNT22	At5g61520	<i>STP3</i>	Sugar:hydrogen symporter	Carbohydrate transmembrane transport Integral to membrane	Intron retention
aNT25	At5g01710		Methyltransferase domains: type 11; FkbM	Metabolism Located in endomembrane system	Intron retention (alt NT based on paralog model At5g03190.2)
LOCALIZED IN CHLOROPLAST					
aNT4	At1g23360	<i>MENG</i>	2-phytyl-1,4-naphthoquinone methyltransferase	Phylloquinone (vitamin K1) biosynthesis Located in chloroplast	Intron retention (3)
aNT6	At1g61520	<i>LIGHT HARVESTING COMPLEX A3 (LHCA3)</i>	PSI type III chlorophyll a/b-binding protein	Photosynthesis, light harvesting; located in chloroplast thylakoid membrane	Intron retention (2)
aNT10	At3g01500	<i>CARBONIC ANHYDRASE 1 (CA1)</i>	Beta-carbonic anhydrase	Carbon utilization; CO ₂ -controlled stomatal movements in guard cells Located in chloroplast envelope, stroma, thylakoid membranes	Alt splice site selection
aNT12	At3g14590	<i>NTMC2T6.2</i>	C ₂ calcium-dependent membrane targeting domain	Membrane targeting; located in chloroplast	Alt splice site selection; intron retention
aNT17	At4g35860	<i>GTP-BINDING 2 (GB2)</i>	Small GTP-binding protein	Protein transport, signal transduction Located in chloroplast	Intron retention
aNT20	At5g54190	<i>PROTOCHLOROPHYLLIDE OXIDOREDUCTASE A (PORA)</i>	Light-dependent NADPH:protochlorophyllide oxidoreductase A	Chlorophyll biosynthesis Located in chloroplast envelope and thylakoid	Intron retention
aNT21	At5g59950		RNA-binding (RRM/RBD/RNP motifs)	Process unknown Located in chloroplast	Intron retention
LOCALIZED TO CYTOSKELETON					
aNT16	At4g12790	<i>None</i>	ATP/GTP-binding site motif A (P-loop)	Paralog QQT1 participates in the organization of microtubules during cell division	uORF of At4g12790 is N-term of paralogs

(Continued)

Table A1 | Continued

Alt NT group	TAIR locus ID	<i>Arabidopsis</i> gene name	Inferred molecular function/domain(s)	Inferred biological role/process and cellular location (TAIR)	Alt splice type (<i>Arabidopsis</i>)
	At5g22370	<i>QUATRE-QUART1 (QQT1)</i>		Paralog QQT1 localized to mitotic cytoskeleton	QQT1 & QQT2
	At4g21800	<i>QUATRE-QUART1 (QQT2)</i>			
LOCATED IN NUCLEUS					
aNT2	At1g04870	<i>PROTEIN ARGININE METHYLTRANSFERASE 10 (PRMT10)</i>	histone-arginine N-methyltransferase	Histone-arginine methylation Located in nucleus	Alt acceptor site
aNT7	At1g68130	<i>INDETERMINATE DOMAIN14 (IDD14)/IDD16</i>	Zinc finger, C2H2-type transcription factor	Regulation of transcription, starch metabolism Located in nucleus	Intron retention
aNT8	At2g35990	<i>LONELY GUY2 (LOG2)</i>	Cytokinin riboside 5'-monophosphate phosphoribohydrolase	Cytokinin activating enzyme Located in cytosol and nucleus	Alt splice site selection
aNT9	At2g37340	<i>RSZ33</i>	Arginine-serine-rich Zinc knuckle protein	Spliceosome assembly Located in nuclear speck	Alt splice site selection; intron retention
aNT14	At3g46130	<i>MYB48</i>	MYB transcription factor	Regulation of transcription in flavonol biosynthesis; located in nucleus	Alt splice site selection; intron retention
aNT24	At5g64960	<i>CYCLIN-DEPENDENT KINASE C2 (CDKC2)</i>	Cyclin-dependent kinase	Involved in development, viral reproduction. colocalizes with microtubule, spliceosomal complex (nucleus)	Intron retention
LOCATED IN CYTOSOL, OR UNKNOWN LOCATION					
aNT11	At3g13225		WW/Rsp5/WWP Domain	Protein binding domain Cellular location unknown	Intron retention
aNT23	At5g64830		Programmed cell death protein 2, C-terminal domain	Apoptosis Located in cytoplasm	Alt splice site selection