A Curation-Based Search for the Enhancer Locus that Increases Tolerance of a Loss of Chloroplast Translation in *Arabidopsis thaliana*

Kayla Cook and David Meinke

Department of Plant Biology, Ecology, and Evolution; Oklahoma State University

Abstract

Arabidopsis thaliana accessions show variation in their response to a loss of chloroplast translation. A duplicated locus (*ACC2)* encoding homomeric acetyl-coenzyme A carboxylase, an enzyme that functions in fatty acid biosynthesis, was previously identified in the Meinke laboratory as an essential factor in promoting tolerance of a loss of plastid translation. Other unknown loci seem to play a secondary role to *ACC2* in enhancing tolerance. One such locus, termed the enhancer due to its proposed role in enhancing ACC2 function, has been mapped to a defined region of *A. thaliana* chromosome 5 and shows tight linkage to *EMB3137* (*At5g14320*). In this study, we manually curated the region adjacent to *EMB3137* and captured detailed information for each locus. We identified enhancer candidates based on their protein function and evaluated how well each candidate fit our models of enhancer function. Our results underscore the range of diverse protein functions encoded by this small region of the *A. thaliana* genome and point to several loci of potential interest for future studies.

Introduction

Natural accessions of the model organism *Arabidopsis thaliana* show considerable variation in response to a loss of chloroplast translation (Parker et al., 2014). Tolerance of spectinomycin, an antibiotic that inhibits chloroplast translation, is due in part to partial rescue by the protein product of *ACC2.* This nuclear gene is the result of a tandem duplication of *ACC1*,

and in normal conditions, *acc2* knockout mutants do not display a phenotype (Babiychuk et al., 2011). Both *ACC1* and *ACC2* encode a homomeric acetyl-coenzyme A carboxylase (ACCase), yet only ACC2 is targeted to the plastid (Babiychuk et al., 2011). The plastid also contains a heteromeric ACCase that is encoded in part by the plastid genome. Both the heteromeric and homomeric ACCases catalyze essential steps in fatty acid biosynthesis.

Spectinomycin-induced inhibition of plastid translation prevents seedlings from forming heteromeric ACCase because the accD subunit of this enzyme is encoded by the plastid genome. Functional ACC2, which is targeted to the plastid, can partially compensate for the loss of heteromeric ACCase when seedlings are grown on spectinomycin (Parker et al., 2014). Although rescue of the seedling is reliant on functional *ACC2*, several other unlinked loci seem to enhance tolerance, including a locus defined as the enhancer, which maps to a region on chromosome 5 (Parker et al., 2014).

Genetic analysis indicated that the enhancer is linked to *EMB3137* (Parker et al., 2014). *OEP80* and *TOC34* were two early candidates for the enhancer due to their known role in plastid protein import and their relative proximity to *EMB3137* on chromosome 5. One hundred and ten plants produced from crosses between the tolerant Tsu-0 accession and the *emb3137* mutant in a sensitive ("Nossen") background were genotyped for *OEP80* and *TOC34* and phenotyped for the presence of the enhancer (Parker et al., 2014). This analysis revealed 10 recombinants between the enhancer and *TOC34* as well as 10 recombinants between the enhancer and *OEP80* (Parker et al., 2014). Thus, neither of these genes represents the enhancer. No recombinants were found between *EMB3137* and the enhancer, indicating that the enhancer is tightly linked to *EMB3137*. As the distance between these two loci is likely to be less than 1 cM, we reasoned that manual curation of the region closest to *EMB3137* might reveal enhancer candidates. The purpose of this

research study was to manually curate the region closest to *EMB3137* and identify candidate enhancers that might increase tolerance of a loss of plastid translation.

Materials and Methods

A 662 Kb region flanking EMB3137 was manually curated and evaluated for enhancer candidates

Due to limited time and resources, exploring the entire region between *TOC34* and *OEP80* was not feasible. Previous studies had indicated that the enhancer locus is tightly linked to *EMB3137* (located approximately midway between *TOC34* and *OEP80*). Thus, the most logical place to begin searching for the enhancer was the region closest to this locus. Our initial goal was to manually curate approximately 100 loci upstream and downstream of *EMB3137.* We later decided on a Kb-based approach to ensure that approximately the same distance surrounding *EMB3137* was curated. Our final results encompassed a manually-curated 331 Kb upstream region (1A) and 331 Kb downstream region (1B), which included 104 and 101 loci, respectively (Table I; Figure 1). We also quickly scanned an additional 388 Kb upstream region (2A) and 325 Kb downstream region (2B) for obvious enhancer candidates (Table I; Figure 1).

A list of potential enhancer functions was constructed

To facilitate our search for the enhancer, we first needed to consider possible mechanisms of enhancer-promoted tolerance of a loss of chloroplast translation (Table II). Previous studies revealed that a functional copy of *ACC2* is required for the enhancer to increase tolerance. With this in mind, we reasoned that the simplest model for the enhancer might involve improved ACC2 stability, function or import into the plastid. Other models acknowledged that the enhancer might not directly interact with ACC2 or impact its role in fatty acid biosynthesis, but

might instead alter some other process or pathway that becomes rate limiting in the presence of functional ACC2. Alternatively, the enhancer might function in the chloroplast-nucleus signaling pathway responsible for signaling the nucleus when plastid translation has been inhibited. These models were kept in mind when searching through loci within our region of interest.

A protocol was developed to facilitate manual curation

A protocol was created to ensure that the same method was used to capture information on each locus within the region of interest and to ensure that appropriate online sources were reviewed. To begin the curation efforts, we searched The Arabidopsis Information Resource database (TAIR; www.arabidopsis.org) for information regarding each locus within the 1A – 1B region. Relevant TAIR information was recorded, including locus number, gene symbol, protein function and predicted localization taken from the Subcellular Localisation Database for Arabidopsis Proteins (SUBA; www.suba.live).

Literature publication searches were conducted for regions 1A and 1B

For regions 1A and 1B, literature publication searches were conducted to provide additional information on protein function. Determining protein function for each locus was essential because we used information regarding protein function to identify potential enhancer candidates. A protocol was constructed for publication searches to ensure the same method was used throughout this study. For loci studied elsewhere, the TAIR database provides links to relevant PubMed (www.ncbi.nlm.nih.gov/pubmed) articles. PubMed articles found via TAIR were searched first for pertinent information on protein function and localization. The PubMed IDs for informative articles were recorded, as were notes regarding relevant information for each locus. If TAIR-listed articles failed to produce the desired information on a given locus or its protein product, a PubMed search was conducted at NCBI – starting with locus number and then

gene symbol as needed. If the above efforts failed to produce information on protein function, a Google scholar (www.scholar.google.com) search was conducted using the locus number.

Loci within the 1A and 1B regions were sorted by protein function

To facilitate our search for the enhancer, we used information gathered from our literature publication searches to sort loci based on the function of their protein product. This allowed us to analyze the functions of similar proteins as a group and identify loci whose protein product might act as the enhancer. The system developed by Lloyd and Meinke (2012) grouped proteins into twelve general categories based on biological functions. We modified this system to include several noncoding RNAs encoded by loci within our manually curated region (Supplemental Table 2) and used the modified Lloyd and Meinke (2012) system when subjectively sorting each locus by protein function (Supplemental Table S1B). Additionally, we edited protein function details obtained from TAIR and our PubMed literature searches and summarized this information in the final two columns of our master list (Supplemental Table S1A).

Regions 2A and 2B were scanned for obvious enhancer candidates

For regions 2A and 2B, TAIR was searched for the same information gathered for regions 1A and 1B, including locus number, gene symbol, and protein function. We quickly scanned these regions for loci encoding proteins that might play an obvious role as the enhancer. We did not conduct PubMed searches or edit TAIR information on protein function for loci in regions 2A and 2B unless the protein product localized to either the chloroplast or mitochondria based on SUBA localization data. These intracellular regions were emphasized because ACC2 is localized to the plastid and some mitochondrial proteins are co-localized to plastids. The

PubMed article search protocol was followed as described above for loci whose protein products were predicted to localize to the plastid or mitochondria.

Results

The limited genomic region analyzed encompasses a wide range of diverse loci

The 662 Kb region (1A and 1B) that was manually curated encompasses 205 loci (Supplemental Table S1). The most common protein functions include transcription factors, membrane transporters, and proteins involved in signaling and regulatory pathways (Table III). The full range of protein functions encoded by loci within this region is wide – including metabolism, electron transport, signaling pathways, transcription regulation, protein degradation and modification, RNA and protein synthesis, and more. The diversity is not limited simply to loci encoding proteins. This region also includes loci for several non-coding RNA molecules such as tRNAs, miRNAs and antisense RNAs, as well as retrotransposons and pseudogenes. Loci with protein products of unknown function constitute 18% of the genes identified within this region (Table III).

Enhancer candidates were identified and ranked according to how well they fit enhancer models

From the 205 loci manually curated, we identified five loci in region 1A (Figure 2) whose protein functions fit one of our enhancer models. Two loci in the 2A and 2B region (Figure 2) also fit one of our models and were added to the list of candidates. We did not identify any loci of interest within the 1B region (Figure 2). To quickly recollect the position of each enhancer candidate relative to *EMB3137*, we devised a simple system that documents by how many genes above (upstream) or below (downstream) of *EMB3137* each enhancer candidate is located. For

example, Ag156 refers to the 156th gene above *EMB3137* and Bg100 refers to the 100th gene below *EMB3137*. To assist in the analysis of each candidate enhancer, we ranked the seven candidates based on how well their protein products fit our models of enhancer function (Table IV). Our three most promising candidates include a plastid integral membrane protein, a chaperone protein, and a protein involved in retrograde nucleus-plastid signaling. The remaining candidates are involved in protein folding, protein aggregation, and fatty acid homeostasis.

The most promising candidates are consistent with existing models of enhancer function

***At5g13630* (Ag76):**

At5g13630 (*GUN5/CHLH*) encodes a subunit of magnesium chelatase, an enzyme involved in tetrapyrrole biosynthesis (Due et al., 2012). CHLH has several roles in the cell, including chelating magnesium to protoporphyrin IX, mediating plastid-nucleus retrograde signaling, and regulating ABA signaling (Du et al., 2012). Plastid-nucleus retrograde signaling allows for communication between the plastid and nucleus when the plastid is under stress. The current model for this pathway is that a signal generated when plastid function is disrupted results in the repression of nuclear genes encoding plastid-localized proteins (Pogson et al., 2008). The role of CHLH in nucleus-plastid retrograde signaling makes this protein a potential candidate for the enhancer (Table IV). Under this model, the tolerant version of the enhancer limits activation of a signal that is normally initiated upon inhibition of plastid gene expression. The tolerant enhancer might increase the threshold for activation of the signal or alter successful transmission of the signal to the nucleus. In either case, inhibition of this plastid-nucleus signaling pathway would result in continued expression of nuclear genes encoding plastid-localized proteins and, potentially, limited compensation for the loss of plastid-encoded proteins.

***At5g13850 (*Ag51):**

At5g13850 (*NACA3*) encodes a putative alpha subunit of the Nascent Polypeptide Associated Complex (NAC). The NAC acts as a chaperone molecule that facilitates targeting of newly-synthesized peptides to specific organelles in the cell. Recent research has shown that a yeast gene (*EGD2*) encoding the alpha subunit of the NAC is involved in targeting of nascent proteins destined for the mitochondria (Ponce-Rojas et al., 2017). Mutants lacking functional subunits of the NAC heterodimer are defective in mitochondrial protein import (Ponce-Rojas et al., 2017). WU-BLAST (www.arabidopsis.org/wublast/index2.jsp) analysis between yeast EGD2 and the *Arabidopsis* NACA3 subunit encoded by *At5g13850* showed a high level of amino acid sequence similarity. Given this information, *NACA3* was chosen as a candidate for the enhancer (Table IV). Under this model, the enhancer might function as a subunit of a chaperone protein that helps direct ACC2 from the cytosol to the plastid. The tolerant enhancer might be more efficient in guiding ACC2 to the plastid, where it can partially compensate for loss of heteromeric ACCase. This proposed function of NACA3 fits directly with one of our models, though to our knowledge this *Arabidopsis* protein has not been studied in detail. Although yeast EGD2 showed amino acid sequence similarity to NACA3, this is not conclusive evidence that NACA3 functions as a subunit of the NAC complex in plants or targets specific proteins to the plastid. In addition, one model for yeast mitochondrial protein import involves binding of the mRNA encoding the mitochondrial-localized protein to the outside surface of mitochondria (Ponce-Rojas et al., 2017), and this model of protein import is not thought to occur in plastids.

***At5g13390* (Ag101):**

At5g13390 (*NEF1*) encodes a plastid integral membrane protein. This gene is named for the phenotype of *nef1* mutants, which fail to form exine in pollen. These mutants display altered

plastid development in pollen and leaves, as well as altered lipid accumulation in pollen and flower buds (Ariizumi et al., 2004). Although the specific protein function is not currently understood, NEF1 has been shown to be a plastid-targeted protein with similarity to integral membrane and transporter proteins (Ariizumi et al., 2004). The proposed functions of NEF1 implicate this protein in maintaining plastid membrane integrity relative to importation and exportation of fatty acids (Ariizumi et al., 2004). TAIR expression data (AtGenExpress Visualization Tool) for *NEF1* indicate that this gene is widely expressed throughout the plant, including both the embryo and seedling stages of development. The proposed role for NEF1 fits a model for enhancer function where the enhancer helps export fatty acids synthesized by ACC2 into the cytosol to be used for plant growth (Table IV). When fatty acid biosynthesis becomes limited due to the inhibition of plastid translation, a more efficient (tolerant) enhancer might allow slightly more fatty acids to be exported into the cytosol compared to plants with the sensitive enhancer. This difference in efficiency would only become phenotypically relevant when fatty acid biosynthesis is severely constrained.

Other enhancer candidates are involved in protein folding, protein aggregation and fatty acid homeostasis

***At5g13640* (Ag75):**

At5g13640 (*PDAT1*) encodes an enzyme involved in triacylglycerol (TAG) biosynthesis. This enzyme catalyzes the final addition of a fatty acid chain (synthesized by the plastid) to create triacylglycerol (Fan et al., 2013). PDAT1 has been associated with regulation of TAG biosynthesis because knockout *pdat1* mutants show increased levels of free fatty acids (Fan et al., 2013). PDAT1 was identified as a potential candidate because of its role in the integration of

fatty acids into triacylglycerols (Table IV). With this model, the tolerant enhancer might facilitate better utilization of fatty acids synthesized by ACC2.

***At5g15450* (Bg122):**

At5g15450 (*CLPB3/APG6*) encodes CLPB3 – a Clp protein targeted to the plastid. Clp proteins are chaperone proteins that disassemble protein complexes or aggregates (Lee et al., 2006). Knockout *clpb3* mutants are seedling lethal and fail to develop proper chloroplasts (Lee et al., 2006). This protein was identified as a candidate for the enhancer due to its plastid localization and identity as a chaperone protein (Table IV). With this model, the tolerant enhancer might recognize improperly folded ACC2 inside the plastid and disassemble these complexes, allowing ACC2 to dimerize properly.

***At5g13410* (Ag99):**

At5g13410 encodes a plastid localized FKBP-like protein (also known as an immunophilin). Although their physiological significance in plants is not well understood, these proteins have peptidyl prolyl cis-trans isomerase potential that is believed to be involved in protein folding (He et al., 2004). The *A. thaliana* genome encodes multiple immunophilins and many of them are targeted to the plastid, where they presumably participate in protein folding or other protein interactions (He et al., 2004). We chose this locus as an enhancer candidate because of this presumed role in protein folding (Table IV). Under this model, the tolerant enhancer might improve ACC2 folding or dimerization inside the plastid. One potential conflict of this model is that the subcellular localization of this protein is in the thylakoid lumen, a region distinct from where ACC2 functions in the plastid.

Less promising but potentially relevant genes are varied in protein function

At5g12860 (*OMT1;*Ag156*)* encodes a protein that acts as an oxaloacetate–malate transporter essential for the malate valve and a 2-oxoglutarate–malate transporter involved in the nitrogen assimilation pathway (Kinoshita et al., 2011). This locus provides one example of how the enhancer might function in another rate-limiting pathway once ACC2 partially restores fatty acid biosynthesis (Table IV). This model for enhancer function suggests that the tolerant enhancer increases tolerance separately from ACC2 function via its role in nitrogen assimilation.

Other potentially relevant genes included *At5g13690, At5g15250,* and *At5g16070.* *At5g13690* (*NAGLU/CYL1*) encodes an N-Acetylglucosaminidase essential for seed development (Ronceret et al., 2008). *At5g15250* (*FTSH6*) encodes a FtsH metalloprotease (Sedaghatmehr et al., 2016). *At5g16070* encodes a putative TCP1 chaperone protein but is located outside of the 1A and 1B region, and thus was not manually curated. These loci were initially identified as potential candidates but further analysis is required to evaluate how well they fit the enhancer models. Although not considered an enhancer candidate, another locus worth noting was *At5g13270,* which encodes a PPR protein that edits *accD* transcripts inside the chloroplast.

Discussion

Functional diversity in the genomic region analyzed complicated our search for the enhancer

Our search for the enhancer was challenging because we have multiple models for how the enhancer might work, and because regions 1A and 1B contain loci encoding proteins with a wide variety of biological functions. Examples include replication, transcription, translation, protein modification or degradation, transcriptional regulation, metabolism and cell structure. SUBA-predicted subcellular localizations for these proteins covered virtually every compartment

within the cell, including the cell membrane, nucleus, Golgi complex, peroxisome, plastid, mitochondria, vacuole, cytosol and endoplasmic reticulum. Our enhancer models helped focus the search to loci encoding proteins involved in certain biological functions and proteins localized to certain cell compartments of interest, including the cytosol, plastid, and mitochondria.

Proteins of unknown function and transcription factors remain enhancer candidates

One complication encountered during this curation process was the large number of loci encoding transcription factors (18 genes) and proteins with unknown functions (38 genes). Previous studies have indicated that the enhancer is unlikely to be a transcription factor that increases expression of *ACC2* because both tolerant and sensitive accessions have similar transcript levels (Parker et al., 2014). However, the enhancer could potentially encode a transcription factor that affects expression of another locus whose protein product fits one of the enhancer models. Though the seven candidate loci described here represent the most logical candidates for enhancer function, there remains a distinct possibility that the enhancer could be a protein of unknown function or transcription factor encoded within this region.

This research provides a foundation for efforts to determine the cellular function of the enhancer

Although the exact cellular function and locus number of the enhancer remain unresolved, this study identified seven loci whose protein products might function as the enhancer. Our results provide a list of potential enhancers that could be expanded upon in future efforts to further analyze this region. While the 1A and 1B regions were the most logical places to begin searching for the enhancer, future studies might also need to consider the neighboring regions.

The next step in identifying the enhancer will involve genotyping each of the 20 recombinant plants already identified using additional markers to determine whether the enhancer is located upstream or downstream of *EMB3137*. Once genetic analysis identifies a smaller region of interest, further manual curation of this region may ultimately help to uncover the identity of the enhancer and lead to a better understanding of how *Arabidopsis* accessions differ in their response to a loss of chloroplast translation.

References

Ariizumi T, Hatakeyama K, Hinata K, Inatsugi R, Nishida I, Sato S, Kato T, Tabata S, Toriyama K (2004) Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility tn *Arabidopsis thaliana*. Plant J **39:** 170–181

Babiychuk E, Vandepoele K, Wissing J, Garcia-Diaz M, De Rycke R, Akbari H, Joubès J, Beeckman T, Jänsch L, Frentzen M, et al (2011) Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. Proc Natl Acad Sci USA **108:** 6674–6679

Du S, Zhang X, Lu Z, Xin Q, Wu Z, Jiang T, Lu Y, Wang X, Zhang D (2012) Roles of the different components of magnesium chelatase in abscisic acid signal transduction. Plant Mol Biol **80:** 519–537

Fan J, Yan C, Xu C (2013) Phospholipid:diacylglycerol acyltransferase-mediated triacylglycerol biosynthesis is crucial for protection against fatty acid-induced cell death in growing tissues of Arabidopsis. Plant J **76**: 930–942

He Z, Li L, Luan S (2004) Immunophilins and parvulins. Superfamily of peptidyl prolyl isomerases in Arabidopsis. Plant Physiol **134:** 1248–1267

Kinoshita H, Nagasaki J, Yoshikawa N, Yamamoto A, Takito S, Kawasaki M, Sugiyama T, Miyake H, Weber A, Taniguchi M (2011) The chloroplastic 2-oxoglutarate/malate transporter has dual Function as the malate valve and in carbon/nitrogen metabolism. Plant J **65:** 15–26

Lee U, Rioflorido I, Hong S, Larkindale J, Waters E, Vierling E (2006) The Arabidopsis

ClpB/Hsp100 family of proteins: chaperones for stress and chloroplast development. Plant J **49:** 115–127

Lloyd J, Meinke D (2012) A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. Plant Physiol **158:** 1115–1129

Parker N, Wang Y, Meinke D (2014) Natural variation in sensitivity to a loss of chloroplast translation in arabidopsis. Plant Physiol **166:** 2013–2027

Pogson B, Woo N, Forster B, Small I (2008) Plastid signaling to the nucleus and beyond. Trends Plant Sci **13:** 602–609

Ponce-Rojas J, Avendaño-Monsalve M, Yañez-Falcón A, Jaimes-Miranda F, Garay E, Torres-Quiroz F, DeLuna A, Funes S (2017) αβ′-NAC cooperates with Sam37 to mediate early stages of mitochondrial protein import. FEBS J **284:** 814–830

Ronceret A, Gadea-Vacas J, Guilleminot J, Devic M (2008) The alpha-N-acetyl-glucosaminidase gene is transcriptionally activated in male and female gametes prior to fertilization and is essential for seed development in Arabidopsis. J Exp B **59:** 3649–3659

Sedaghatmehr M, Mueller-Roeber B, Balazadeh S (2016) The plastid metalloprotease FtsH6 and small heat shock protein HSP21 jointly regulate thermomemory in Arabidopsis. Nat Commun **7:** 1–14

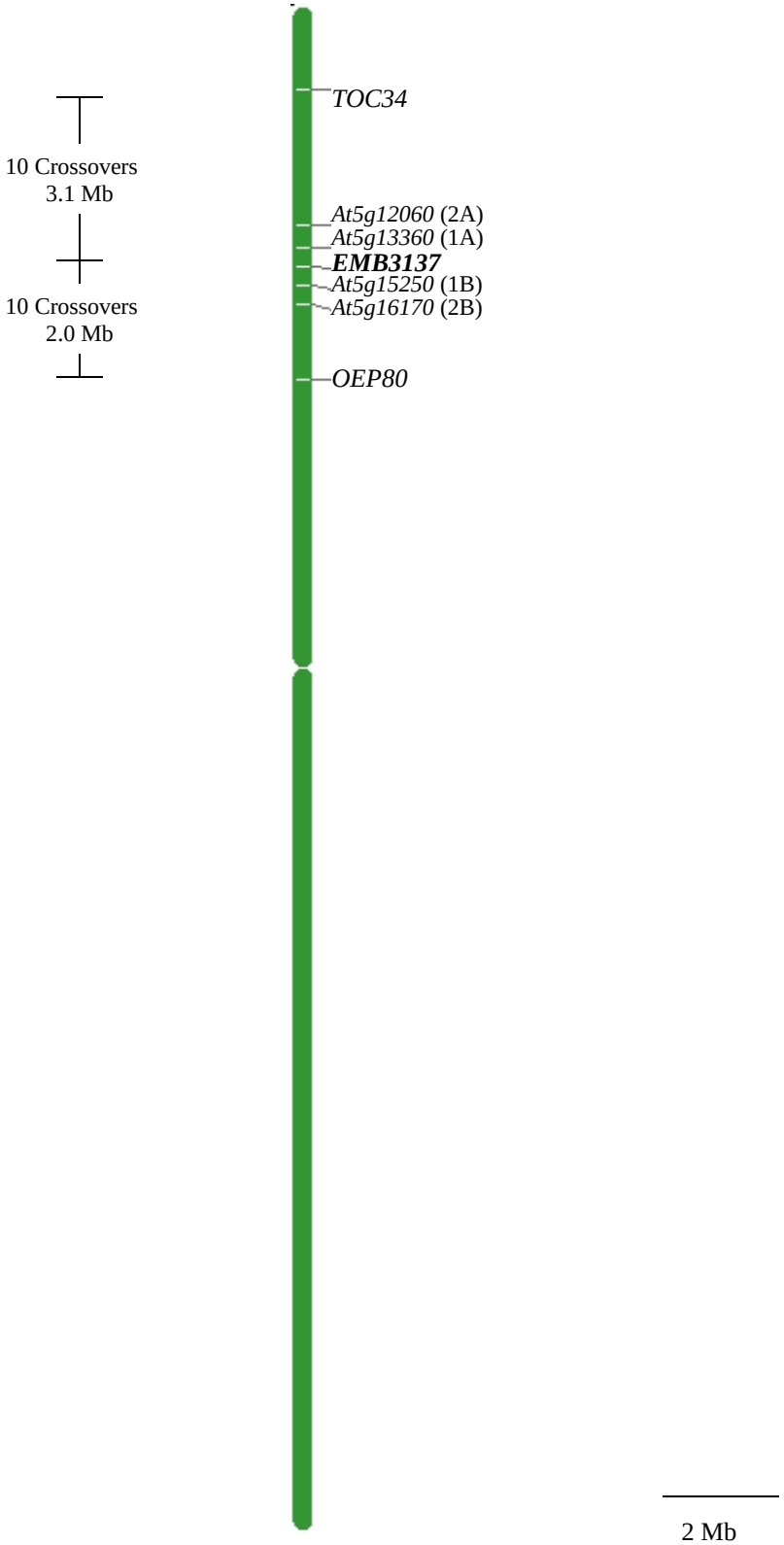

Figure 1. Map of *Arabidopsis thaliana* chromosome 5 created using the TAIR Chromosome Map Tool (www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp). *At5g12060* and *At5g13360* are the upper boundaries of regions 2A and 1A, respectively. *At5g15250* and *At5g16170* are the lower boundaries of regions 1B and 2B, respectively.

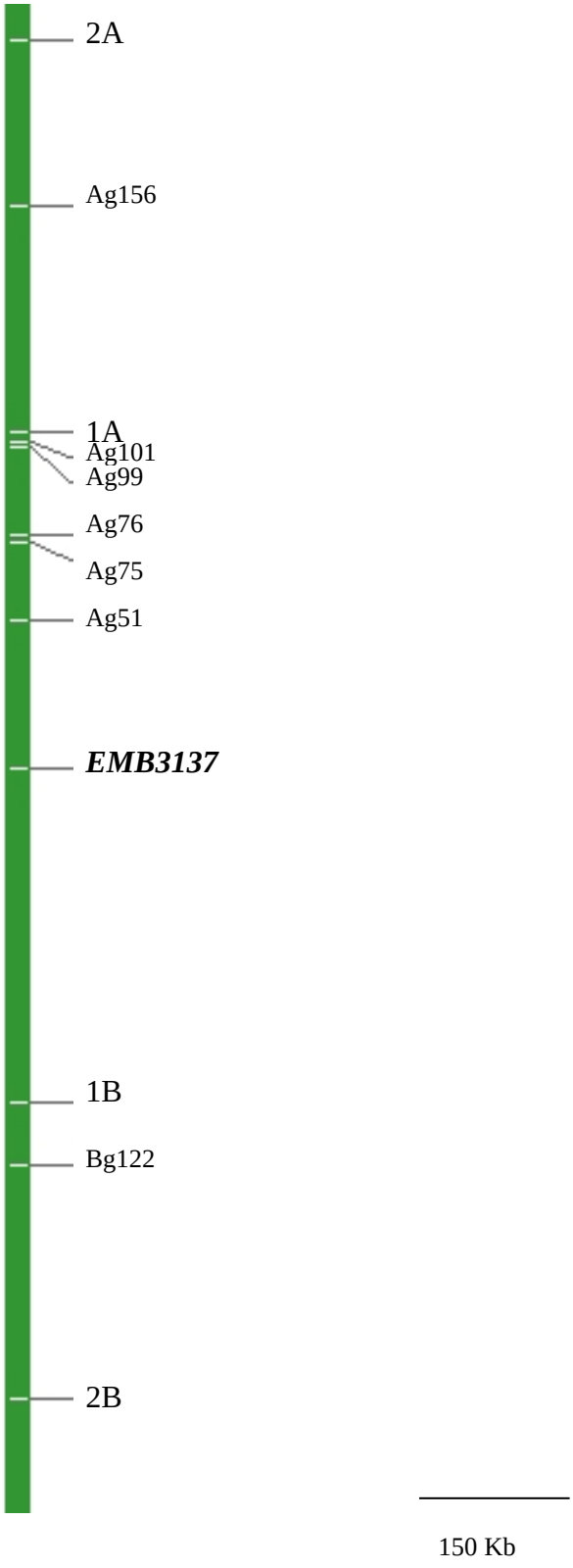

Figure 2. Map of potential enhancer candidates on *Arabidopsis thaliana* chromosome 5. 2A, *At5g12060*; 1A, *At5g13360*; 1B, *At5g15250*; 2B, *At5g16190*. Ag, gene number above (upstream of) *EMB3137*; Bg, gene number below (downstream of) *EMB3137* (e.g. Ag156 refers to the 156th gene above *EMB3137).*

Table I. *Distance, total genes and identified enhancer candidates in regions 2A – 2B*

Region	Distance (Kb)	Total Genes	Enhancer Candidates
2A	388	100	1
1A	331	104	5
1B	331	101	0
2B	325	104	1

Table II. *Models for enhancer function in the absence of chloroplast translation*

1. Enhances Function, Abundance, or Localization of ACC2 Protein
    a. Improves translational efficiency of *ACC2* mRNA
    b. Improves targeting of ACC2 to plastid via chaperone molecule
    c. Improves import of ACC2 through plastid membrane
    d. Improves ACC2 folding and dimerization inside plastid
2. Improves Fatty Acid Biosynthesis in Plastid
    a. Increases efficiency of upstream/downstream reactions
    b. Improves export of ACC2-synthesized fatty acids
3. Compensates for Loss of Ycf1, Ycf2, ClpP1 Functions in Plastid
4. Impacts Chloroplast-Nucleus Retrograde Signaling Pathways
5. Improves Other Rate-Limiting Metabolic Pathways in Plastid

Table III. *Protein functions assigned to loci within the 1A and 1B regions flanking EMB3137*

Genes	Protein Function[a]
1	**DNA Synthesis/Repair**
12	**RNA Synthesis/Modification**
1	Transcription Machinery
3	PPR/RNA Binding Proteins
8	RNA Splicing/Modification/Degradation
11	**Protein Synthesis**
5	Ribosomal Proteins
0	Aminoacyl-tRNA Synthetases
2	Translation Machinery
4	tRNAs
11	**Protein Modification/Transport**
9	Modification; Chaperones
2	Protein Transport/Import
10	**Protein Degradation**
6	**Chromosome Dynamics**
5	Chromosome Structure/Modification
1	Chromosome Mechanics; Cell Cycle
18	**Transcription Regulation**
16	**Signaling and Regulatory Pathways**
3	**Energy; Electron Transport**
26	**Metabolism**
3	Biosynthesis of Amino Acids, Vitamins, Nucleotides, Fatty Acids
2	Biosynthesis of Chlorophyll, Carotenoids, Terpenoids
4	Biosynthesis of Lipds; Modifcation of Fatty Acids and Lipids
3	Biosynthesis/Modification of Complex Carbohydrates
14	Other Metabolic Pathways/Enzymes
28	**Cell Structure; Membrane Function/Trafficking**
4	Cytoskeleton; Cell Wall; Organelle Biogenesis and Division; Other Structural Proteins
21	Membrane Transporters
3	Vesicle/Membrane Trafficking/Secretion
15	**Remain to Be Classified/Uncertain**
38	**Unknown**
11	**Other DNA**
3	Pseudogenes
4	Transposons
4	miRNAs and Antisense RNAs
206[b]	**Total**

[a] Modified protein function categories (Lloyd and Meinke, 2012).
[b] Total includes 205 loci within Regions 1A and 1B that were manually curated, including *EMB3137*.

Table IV. *Enhancer candidates identified in the region flanking EMB3137 on chromosome 5*

Rank[a]	Gene Location[b]	Region	Locus Number	Gene Symbol	Edited Function[c]	Edited Function Details[c]
A	Ag101	1A	At5g13390	*NEF1*	Plastid Integral Membrane Protein	Required for pollen exine formation; Proposed roles in plastid membrane integrity and fatty acid export
A	Ag76	1A	At5g13630	*ABAR; CCH; CHLH; GUN5*	Magnesium Chelatase	Plastid to nucleus retrograde signal transduction
A	Ag51	1A	At5g13850	*NACA3*	Nascent Polypeptide Associated Complex Subunit Alpha-Like Protein 3	Potential role in translocation of nascent polypeptides into chloroplasts
A/B	Ag99	1A	At5g13410		Plastid-Localized FKBP-Like Protein; Immunophilin	Potential role in protein folding
A/B	Ag75	1A	At5g13640	*PDAT1*	Phospholipid: Diacylglycerol Acyltransferase	TAG biosynthesis; Fatty acid and membrane lipid homeostasis
A/B	Bg122	1B	At5g15450	*CLPB3; APG6*	Plastid-Localized ClpB Homologue; Chaperone	Remodeling of protein aggregates
B	Ag156	2A	At5g12860	*DIT1; OMT1*	Plastid Dicarboxylate Transporter	Integration of carbon, nitrogen metabolism

[a] System used to subjectively rank each candidate. Rank A, most likely; Rank A/B, promising; Rank B, possible.
[b] Naming system used to denote location by counting how many genes above or below *EMB3137* each locus is positioned.
[c] Based on information from TAIR and relevant publications.

Supplemental Table S2. Edited protein function classification system based on Lloyd and Meinke (2012)

1. DNA Synthesis/Repair

2. RNA Synthesis/Modification

    2.1 Transcription Machinery
    2.2 PPR/RNA Binding Proteins
    2.3 RNA Splicing/Modification/Degradation

3. Protein Synthesis

    1.1 Ribosomal Proteins
    1.2 Aminoacyl-tRNA Synthetases
    1.3 Translation Machinery
    1.4 tRNAs

4. Protein Modification/Transport

    1.1 Modification; Chaperones
    1.2 Protein Transport/Import

5. Protein Degradation

6. Chromosome Dynamics

    6.1 Chromosome Structure/Modification
    6.2 Chromosome Mechanics; Cell Cycle

7. Transcription Regulation

8. Signaling and Regulatory Pathways

9. Energy; Electron Transport

10. Metabolism

    10.1 Biosynthesis of Amino Acids, Vitamins, Nucleotides, Fatty Acids
    10.2 Biosynthesis of Chlorophyll, Carotenoids, Terpenoids
    10.3 Biosynthesis of Lipids; Modification of Fatty Acids and Lipids
    10.4 Biosynthesis/Modification of Complex Carbohydrates
    10.5 Other Metabolic Pathways/Enzymes

11. Cell Structure; Membrane Function/Trafficking

11.1 Cytoskeleton; Cell Wall; Organelle Biogenesis and Division; Other Structural Proteins
11.2 Membrane Transporters
11.3 Vesicle/Membrane Trafficking/Secretion

12. Remain to be Classified/Uncertain

13. Unknown

14. Other DNA

14.1 Pseudogenes
14.2 Transposons
14.3 miRNAs and Antisense RNAs