

Original Article

Replication protein A subunit 3 and the iron efficiency response in soybean

Sarah E. Atwood^{1†}, Jamie A. O'Rourke^{2†}, Gregory A. Peiffer^{1†}, Tengfei Yin³, Mahbulul Majumder³, Chunquan Zhang⁴, Silvia R. Cianzio⁵, John H. Hill⁴, Dianne Cook³, Steven A. Whitham⁴, Randy C. Shoemaker^{5,6} & Michelle A. Graham^{5,6}

¹Interdepartmental Genetics Program, Iowa State University, Ames, IA 50011, USA, ²Plant Science Research Unit, USDA-Agricultural Research Service, Saint Paul, MN 55108, USA, ³Department of Statistics, Iowa State University, Ames, IA 50011, USA, ⁴Department of Plant Pathology & Microbiology, Iowa State University, Ames, IA 50011, USA, ⁵Department of Agronomy, Iowa State University, Ames, IA 50011, USA and ⁶Corn Insects and Crop Genetics Research Unit, USDA-Agricultural Research Service, Ames, IA 50011, USA

ABSTRACT

In soybean [*Glycine max* (L.) Merr.], iron deficiency results in interveinal chlorosis and decreased photosynthetic capacity, leading to stunting and yield loss. In this study, gene expression analyses investigated the role of soybean replication protein A (RPA) subunits during iron stress. Nine RPA homologs were significantly differentially expressed in response to iron stress in the near isogenic lines (NILs) Clark (iron efficient) and IsoClark (iron inefficient). RPA homologs exhibited opposing expression patterns in the two NILs, with RPA expression significantly repressed during iron deficiency in Clark but induced in IsoClark. We used virus induced gene silencing (VIGS) to repress *GmRPA3* expression in the iron inefficient line IsoClark and mirror expression in Clark. *GmRPA3*-silenced plants had improved IDC symptoms and chlorophyll content under iron deficient conditions and also displayed stunted growth regardless of iron availability. RNA-Seq comparing gene expression between *GmRPA3*-silenced and empty vector plants revealed massive transcriptional reprogramming with differential expression of genes associated with defense, immunity, aging, death, protein modification, protein synthesis, photosynthesis and iron uptake and transport genes. Our findings suggest the iron efficient genotype Clark is able to induce energy controlling pathways, possibly regulated by SnRK1/TOR, to promote nutrient recycling and stress responses in iron deficient conditions.

Key-word: iron; virus induced gene silencing.

INTRODUCTION

Iron is an essential micronutrient required for photosynthesis, respiration and other metabolic processes in plants. However, an overabundance of iron is toxic to cells, as free iron can cause reactions that damage DNA, proteins and lipids (Winterbourn 1995). Essential for proper growth and development, iron

homeostasis is dependent on the tightly regulated uptake, transport and storage of iron (Guerinot & Yi 1994).

Iron content in soybean [*Glycine max* (L.) Merr.] has both nutritional and agricultural importance. It is estimated that nearly 25% of the global population is anaemic, with the highest percentage in pregnant women and young children in developing countries (McLean *et al.* 2008). Biofortification of crops is considered to be the best solution for solving iron deficiency in the developing world, where diets are mainly plant-based (Mayer, Pfeiffer & Beyer 2008). Agriculturally, yield losses from iron deficiency in soybean can be quite large. In 2004, the estimated loss from iron deficiency chlorosis (IDC) of soybeans in the United States was over \$120 million (Hansen *et al.* 2004). Furthering our knowledge of the uptake, transport and regulation of iron in plants is essential to improving both human nutrition and preventing detrimental yield losses for farmers.

Iron deficiency occurs in plants when iron is unavailable, either by a lack of iron or a lack of iron in the ferrous (Fe²⁺) form. IDC is a problem for soybeans in the upper Midwest where fields may have alkaline, calcareous soils. Although iron is usually abundant in soil, the plants' ability to uptake ferrous iron is hindered by various soil properties, such as high moisture content, high pH and an abundance of soluble salts (Hansen *et al.* 2003). Iron stress decreases chlorophyll production and photosynthetic rates, leading to yellow leaves with green veins (interveinal chlorosis, Spiller & Terry 1980; Terry 1980). Other symptoms of IDC include stunting and yield reduction.

In the last several years, microarray analyses have been used to identify soybean genes differentially expressed during iron stress and iron stress recovery (O'Rourke *et al.* 2007, 2009). O'Rourke *et al.* (2009) compared gene expression in leaves in response to iron stress between two near-isogenic lines (NILs, Clark and IsoClark) differing in iron efficiency. While the iron inefficient line IsoClark had very little response to iron stress, the iron efficient line Clark had significant differential expression of genes involved in iron acquisition, defense, stress and DNA replication. Of specific interest was a probe corresponding to *Replication Protein A subunit 3* (*GmRPA3c*, GmaAffx.36066.1.S1_at, Glyma20g24590), one

Correspondence: M. A. Graham. E-mail: michelle.graham@ars.usda.gov

[†]These authors have contributed equally to the manuscript.

of the most significantly differentially expressed probes between the two NILs when grown under iron deficient conditions. In addition, Glyma20g24590 mapped to an IDC QTL on soybean chromosome 20 (LG I) identified in two different segregating populations (Lin, Cianzio & Shoemaker *et al.* 1997; Lin *et al.* 1998).

RPA is the eukaryotic single-stranded DNA binding protein and is essential for maintaining genome integrity during DNA replication, repair of DNA lesions and cell-cycle checkpoint activation (Hass, Lam & Wold 2011). It is a heterotrimeric protein made of three subunits: RPA1 (70 kDa), RPA2 (32 kDa) and RPA3 (14 kDa) (Wold 1997). RPA1 acts as the ssDNA-binding subunit (Pfuetzner *et al.* 1997). RPA2 regulates RPA activity during the cell cycle and stress responses (Din *et al.* 1990; Binz, Sheehan & Wold 2004). The role of RPA3, however, is unclear, although studies have shown it may regulate the formation of the heterotrimeric protein complex (Daniely & Borowiec 2000; Kim *et al.* 2005). *RPA* gene expression is closely tied to replication, with high expression in proliferating tissues and low expression in mature tissues (Ishibashi *et al.* 2001; Chang *et al.* 2009). *RPA* gene expression is up-regulated after DNA damage from chemical mutagens (Takashi *et al.* 2009). The study by O'Rourke *et al.* (2009) was the first connection made between *RPA* gene expression and abiotic stress in plants.

In this study, we identified the homologs of all *Arabidopsis thaliana* *RPA* genes in the genome sequence of soybean (*G. max*) and examined their expression in leaves during iron stress in two NILs differing in iron efficiency. Of the 18 *RPA* homologs identified, eight were differentially expressed in response to iron stress in both NILs. However, while *RPA* gene expression decreased in iron stress conditions in the iron efficient line, expression increased in the iron inefficient line. We used virus induced gene silencing (VIGS, Zhang *et al.* 2010) of *GmRPA3* in the iron inefficient soybean line IsoClark to mimic the gene expression patterns observed in Clark. Silencing resulted in reduced IDC symptoms. RNA-Seq analysis of three biological replicates of VIGS-treated plants suggests the iron efficient soybean line Clark employs novel responses, including the repression of a suite of DNA replication genes, to survive iron deficiency conditions. A combination of genetic, molecular and bioinformatic approaches identify *GmRPA3* as a component of the IDC tolerance response in soybean.

MATERIALS AND METHODS

Germplasm

Soybean (*G. max* (L.) Merr.) lines Clark (PI 548533), IsoClark (PI 547430) and T203 (PI 54619) were used to study the role of RPA in iron homeostasis. Clark is iron efficient, while IsoClark and T203 are iron inefficient. IsoClark is a NIL of Clark, containing an introgressed region from parent T203 that is hypothesized to cause iron inefficiency (Severin *et al.* 2010b).

Homolog identification

Nine *RPA* homologs have been identified in *A. thaliana* (*RPA1*: At2g06510, At5g08020, At5g45400, At5g61000,

Table 1. Location of RPA subunits within the soybean genome

| RPA subunit | Glyma1 identifier | Genomic location |
|-----------------|----------------------|-----------------------------|
| <i>GmRPA1Aa</i> | <i>Glyma15g19090</i> | Gm15:16,104,809..16,107,391 |
| <i>GmRPA1Ab</i> | <i>Glyma09g07850</i> | Gm09:6,838,917..6,841,192 |
| <i>GmRPA1Ac</i> | <i>Glyma20g19560</i> | Gm20:27,433,375..27,433,575 |
| <i>GmRPA1Ad</i> | <i>Glyma14g12480</i> | Gm14:11,281,090..11,283,684 |
| <i>GmRPA1Ae</i> | <i>Glyma06g32800</i> | Gm06:33,675,732..33,676,304 |
| <i>GmRPA1Ba</i> | <i>Glyma17g08660</i> | Gm17:6,394,788..6,398,085 |
| <i>GmRPA1Bb</i> | <i>Glyma05g00370</i> | Gm05:160,592..166,261 |
| <i>GmRPA1Ca</i> | <i>Glyma09g34670</i> | Gm09:41,001,622..41,004,660 |
| <i>GmRPA1Cb</i> | <i>Glyma04g34970</i> | Gm04:41,271,806..41,273,410 |
| <i>GmRPA2a</i> | <i>Glyma08g18770</i> | Gm08:14,121,827..14,122,549 |
| <i>GmRPA2b</i> | <i>Glyma14g17270</i> | Gm14:18,982,337..18,986,301 |
| <i>GmRPA2c</i> | <i>Glyma17g29730</i> | Gm17:32,520,158..32,524,286 |
| <i>GmRPA2d</i> | <i>Glyma17g07020</i> | Gm17:5,099,119..5,103,243 |
| <i>GmRPA2e</i> | <i>Glyma13g00960</i> | Gm13:684,930..691,735 |
| <i>GmRPA3a</i> | <i>Glyma19g38100</i> | Gm19:45,105,359..45,105,808 |
| <i>GmRPA3b</i> | <i>Glyma03g35460</i> | Gm03:42,650,331..42,652,668 |
| <i>GmRPA3c</i> | <i>Glyma20g24590</i> | Gm20:34,220,298..34,220,889 |
| <i>GmRPA3d</i> | <i>Glyma10g42530</i> | Gm10:49,451,777..49,453,797 |

Hidden Markov Models (HMMs) were developed from *Arabidopsis thaliana* *RPA* subunits (Shultz *et al.* 2007) and HMMER (Durbin *et al.* 1998) was used to search all predicted coding sequences in the soybean genome (Glyma1, Schmutz *et al.* 2010) translated in all six reading frames. Glyma1 identifiers were queried against the SoyBase genome browser (soybase.org/gb2/gbrowse/gmax1.01/) to identify potential homeologous *RPA* subunits. Homeology was confirmed by BLAST (Altschul *et al.* 1997) analyses. Homeologous pairs are boxed and shaded grey.

At4g19130, *RPA2*: At2g24490, At3g02920 and *RPA3*: At2g24490, At3g02920) (Shultz *et al.* 2007). *Arabidopsis* *RPA* protein sequences were aligned using ClustalW (Thompson, Higgins & Gibson 1994) and HMMER (Durbin *et al.* 1998) was used to build a hidden Markov model (HMM) for each *RPA* subunit. The HMM was then searched against all predicted coding sequences in the soybean genome (Glyma version 1.0, Schmutz *et al.* 2010, <http://www.phytozome.net>), which were translated into all six reading frames. Any soybean gene above the default e-value cut-off (E-value = -1) was considered in our analysis. This comparison identified nine homologs of the *RPA1* gene, five homologs of the *RPA2* gene and four homologs of the *RPA3* gene in soybean (Table 1).

For our study, each homolog was given a name corresponding to the *Arabidopsis* homolog for which it had the greatest homology (Table 1). *GmRPA1A*, *GmRPA1B*, *GmRPA1C*, *GmRPA2* and *GmRPA3* names correspond to genes matching *AtRPA1A* (*AtRPA70a*), *AtRPA1B* (*AtRPA70b*), *AtRPA1C* (*AtRPA70c*), *AtRPA2* and *AtRPA3*, respectively. The SoyBase Genome Browser (<http://soybase.org/gb2/gbrowse/gmax1.01/>) was used to identify potential homeologous *RPA* subunits. To confirm homeology, the 100 000 bases surrounding each *RPA* homolog was divided into 2000 base intervals. BLASTN (Altschul *et al.* 1997, $E < 10E^{-30}$) was used to compare the intervals to the whole genome assembly (Schmutz *et al.* 2010) and confirm homeology.

Phylogenetic analyses

Amino acid sequences for RPA homologs in the species *A. thaliana*, *Oryza sativa*, *Medicago truncatula*, *Ricinus communis* and *G. max* were assessed for sequence conservation. Peptide sequences were obtained for RPA homologs already identified in *A. thaliana* and *O. sativa* (Shultz *et al.* 2007). BLASTP (Altschul *et al.* 1997) was used to compare the Arabidopsis RPA homologs to predicted proteins in the *M. truncatula* and *R. communis* genomes (<http://www.phytozome.net>). Proteins with >50% amino acid similarity were used in phylogenetic analyses. Protein sequences were aligned with Pileup in the Accelrys GCG software (Accelrys Inc., San Diego, CA, USA). The sequence alignment was visually inspected and trimmed to eliminate gaps and nonconserved regions. Sequence alignments for all three RPA subunits were visualized with Multiple Align Show (http://www.bioinformatics.org/SMS/multi_align.html, Supporting Information Figs S1, S2 & S3).

The unrooted phylogenetic tree for the RPA1, RPA2 and RPA3 subunits was created in MEGA5 (Tamura *et al.* 2011) from trimmed amino acid sequence alignments (Fig. 1). The evolutionary history was inferred using the neighbour-joining method (Saitou & Nei 1987). A percentage of replicate trees in which the associated taxa clustered together was calculated from 500 bootstrap replicates (Felsenstein 1985). Evolutionary distances were computed using the p-distance method (Nei & Kumar 2000) and are in the units of the number of amino acid differences per site.

Expression analyses of RPA subunits in soybean

Plant growth conditions

Clark, IsoClark and T203 seed were germinated for 5 to 7 d in a growth chamber at 27° C until unifoliate had emerged but were not fully expanded. Seedlings were removed from germination paper and placed in iron sufficient hydroponic conditions (100 µM Fe(NO₃)₃•9H₂O) in the greenhouse. When the first trifoliolate was fully expanded (13 d after placing in hydroponics), plant roots from each bucket were rinsed six times in fresh double distilled water, each for 15 s minimum and returned to a new hydroponic bucket. Six buckets were returned to iron sufficient conditions (100 µM Fe(NO₃)₃•9H₂O), while the other six were placed in iron insufficient conditions (50 µM Fe(NO₃)₃•9H₂O). Nutrient solutions were based on growth conditions described in Chaney *et al.* (1992), with volumes adjusted for 10 L buckets.

For testing qPCR primers (Supporting Information Table S1), tissues were pooled from six Clark and six IsoClark plants grown in iron insufficient conditions. For the qPCR analyses, tissue from two Clark and two IsoClark plants was pooled from each bucket at the time points 1 h, 6 h and 24 h after introduction into new iron conditions, for a total of six biological replicates at each time point and iron condition. First trifoliate was harvested and flash frozen in liquid nitrogen and stored at -80 °C for later RNA extraction.

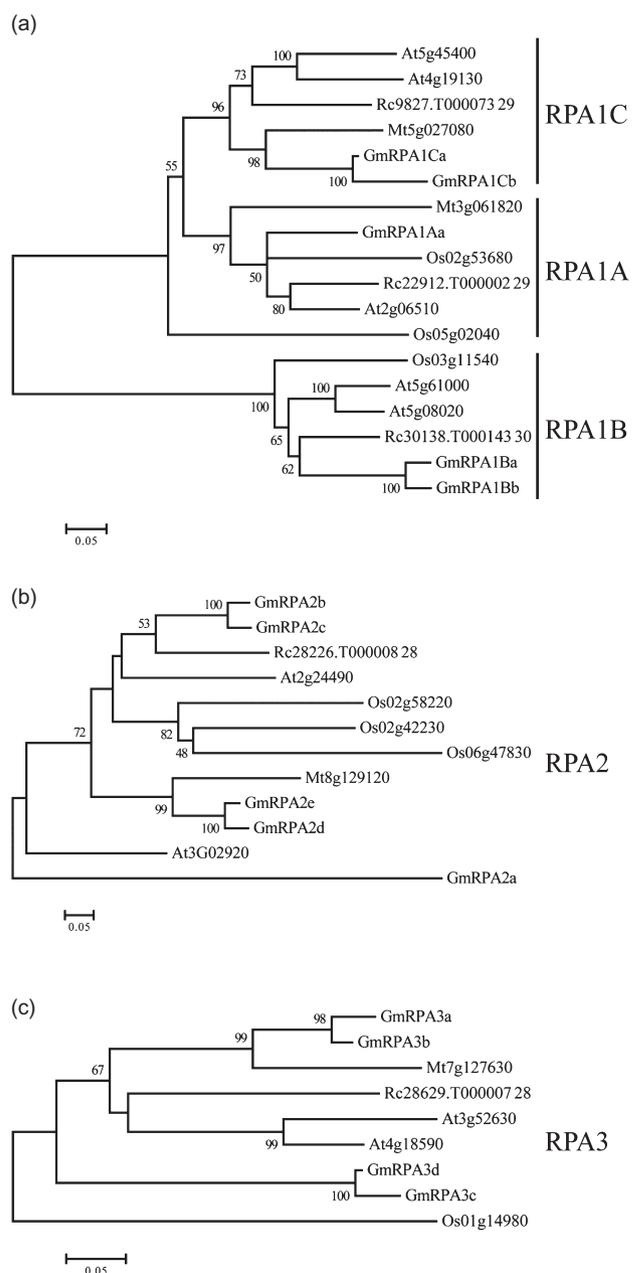


Figure 1. Phylogenetic analyses of the RPA subunits in *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Medicago truncatula* (Mt), *Ricinus communis* (Rc), and *Glycine max* (Gm). (a–c) Phylogenetic analysis of the three RPA1 subunits (a), RPA2 (b) and RPA3 (c). The amino acid alignments shown in Supporting Information Figs S1, S2 and S3 were used to infer evolutionary history using the Neighbor-Joining method (Saitou & Nei 1987) in the MEGA5 software package (Tamura *et al.* 2011). Phylogenetic trees represent the consensus tree from 500 bootstrap replicates. Branches supported by greater than 50% of the bootstrap replicates are indicated (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei & Kumar 2000) and are in the units of the number of amino acid differences per site.

RNA isolation and quality assessment

Flash frozen leaf tissue was ground in liquid nitrogen and RNA was extracted using a Qiagen® RNeasy® Plant Mini Kit (Qiagen®, Germantown, MD, USA). The manufacturer's recommended protocol was used with the following specifications or changes: ~200 mg tissue was lysed with RLT buffer, tubes were incubated at 56 °C for 2 min with 800 rpm shaking to aid in tissue disruption and columns were incubated at room temperature for 10 min during elution. RNA was then DNased with an Ambion® TURBO DNA-free™ kit (Ambion®, Austin, TX, USA) to remove all DNA. After isolation, RNA was assessed for quality using a Thermo Fisher Scientific® NanoDrop™ ND-1000 Spectrophotometer (Thermo Fisher Scientific®, Waltham, MA, USA). RNA was considered to be of good quality for qPCR if the 260/280 ratio was greater than 2.0 and the 260/230 ratio was above 1.7. RNA was also analyzed for quality using an Agilent® 2100 Bioanalyzer™ (Agilent®, Santa Clara, CA, USA). RNA was considered to be of good quality if the RNA was not degraded or was only marginally degraded.

qPCR primer design and quality assessment

qPCR primers were designed for all RPA homologs using the program Primer 3 (Rozen & Skaletsky 2000). Primers were designed using the Primer 3 defaults, specifying an amplicon size (125–175 bp). Primers were designed based on coding sequences of RPA homologs (<http://www.phytozome.net>, Table 1 and Supporting Information Table S1). RPA coding sequences were compared to each other using BLASTN (Altschul *et al.* 1997, $E < 10E^{-30}$) and only unique sequences were used in primer design in order to distinguish between homeologs located in duplicated genomic regions. Primers were tested on Clark and IsoClark total RNA harvested from an iron-insufficient bucket at 14 days post iron stress. mRNA was amplified using the Brilliant® II SYBR® Green QRT-PCR Master Mix Kit from Agilent Technologies following the manufacturer's recommendations. cDNA synthesis was carried out at 60 °C, initial denaturing time was 10 min and a total of 45 cycles were carried out with an additional extension time of 15 s at 72 °C. Reactions were run on polyacrylamide gels. Primers were used in subsequent qPCR reactions if the amplification product yielded a single band and few primer-dimers were found. Primers that had no amplification or had multiple bands were not used for later qPCR studies and were subsequently redesigned and retested. If redesigned primers did not amplify, the gene of interest was removed from the study. Seven of the original 18 RPA homologs were found to not amplify, coinciding with documented low expression (Severin *et al.* 2010a) or predicted pseudogenes.

RPA expression analyses

Prior to qPCR analyses, two 96 well plates were organized for amplification by a specific RPA homolog primer pair and the reference gene *CYP2* (*cyclophilin 2*). *CYP2* was chosen as a reference gene based upon previous qPCR reference gene studies in plants as well as an in-house study (Phillips *et al.* 2009; Wang *et al.* 2011). Each plate contained either Clark or

IsoClark genotype at all three time points and both iron conditions. Three biological replicates were chosen at random for qPCR analysis in order to maintain all time points and iron conditions on one plate, allowing for direct comparison of expression level (Rieu & Powers 2009).

RNA isolated from Clark and IsoClark at 1, 6 and 24 h post stress (hps) was amplified using the Invitrogen™ SuperScript™ III Platinum® SYBR® Green One-Step qRT-PCR Kit (Invitrogen™, San Diego, CA, USA). Reactions were carried out according to the SYBR® Green protocol with the following specifications: total starting RNA was 100 ng and reactions had a final volume of 25 µL instead of 50 µL. RNA was diluted to 9.52 ng µL⁻¹ for greater pipetting accuracy. All experiments included a standard curve of 600, 400, 100, 50 and 10 ng concentration as well as No Reverse-Transcriptase (NRT) and No Template Control (NTC) wells for each primer. NRT wells replaced Superscript III with Invitrogen® Platinum® *Taq* DNA polymerase at the same volume. NTC wells replaced RNA with double distilled, nuclease-free H₂O at the same volume. qPCRs were carried out on a Stratagene Mx3000P™ Real-Time PCR System. After amplification, a dissociation reaction was performed for later analysis of reaction quality. Amplification conditions were from the Invitrogen™ SuperScript™ III Platinum® SYBR® Green One-Step qRT-PCR Kit with the following modifications: cDNA synthesis was carried out at 60 °C, initial denaturing time was 10 min and a total of 45 cycles were carried out with an additional extension time of 15 s at 72 °C. A measurement of fluorescence was taken after each cycle. The default SYBR® Green dissociation reaction conditions were used from the Stratagene Mx3000P™ Real-Time PCR software.

Gene expression data analyses

Amplicons were considered of good quality for further data analysis if the NRT and NTC cycle thresholds (Cts) were greater than five cycles away from the lowest data point and the dissociation curve showed only one peak per reaction well. RNA quantities were determined for each well by aligning to the standard curve for that primer set. A normalized value for RNA quantity was calculated as a ratio of gene-of-interest RNA quantity to reference gene RNA quantity for each sample. Averages of the normalized data were then calculated over technical replicates. Relative expression is a ratio of normalized values in insufficient conditions over normalized values in sufficient conditions at each time point. This ratio is then log base two transformed. Log-transformed data was analysed for standard deviation and standard error. Differences in relative quantity were analysed with analysis of variance (ANOVA; Chambers, Freeny & Heiberger 1992) and then Tukey's Honestly Significant Difference test (Yandell 1997) for pairwise comparisons, with a significance cut-off of 0.05 (Fig. 2, Supporting information Tables S2 & S3).

Identification of transcription factor binding sites in the promoters of RPA homologs

Clover (Frith *et al.* 2004) was used in conjunction with the TRANSFAC transcription factor database (version 7.0,

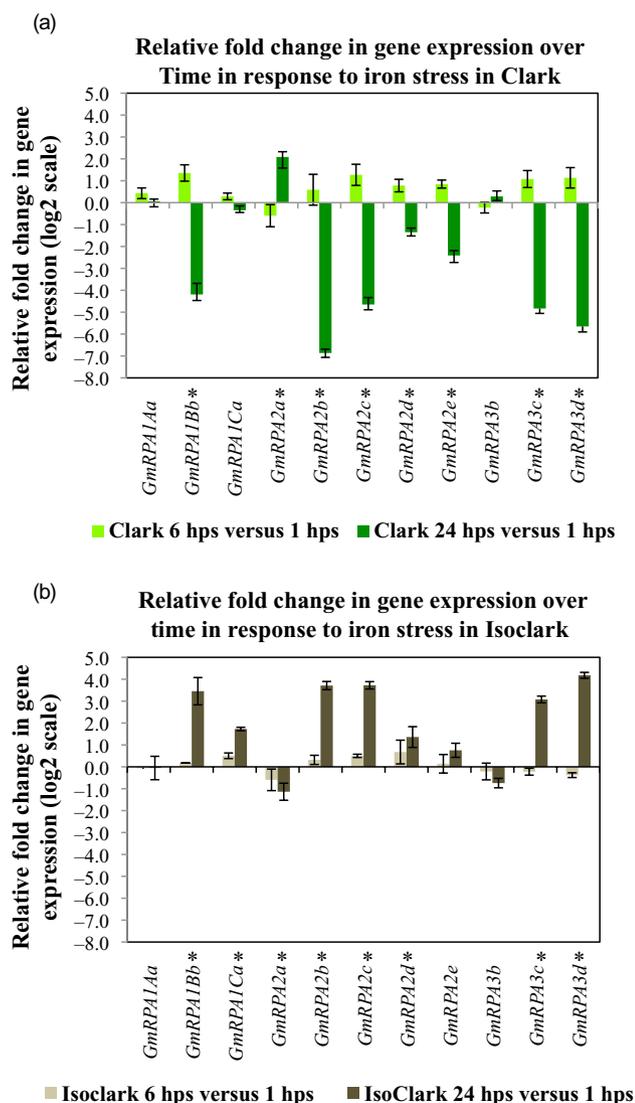


Figure 2. RPA homologs are differentially expressed in response to iron stress. Relative fold change in gene expression over time in response to iron stress in the NILs Clark (a) and IsoClark (b). Gene expression was determined via qPCR. Relative gene expression was determined as a ratio relative to iron sufficient conditions at a given time point, averaged over three biological replicates, then \log_2 transformed. Values above zero would indicate greater expression in iron insufficient conditions, while values below zero indicate lesser expression in iron insufficient conditions. Relative fold change over time was determined comparing the 6 and 24 hps ratios to the 1 hps ratio. The standard error for the fold change was determined using the following equation $SE = \sqrt{SD1 + SD2}/N$, where SD1 and SD2 are the standard deviations of the two ratios being compared and N is the total number of plants sampled (24). Fold change ratios significantly ($P < 0.05$) different from 0 (time point 1hps) were identified using Tukey's Honestly Significant Difference test (Yandell 1997) and are indicated by an asterisk.

Matys *et al.* 2006) to identify transcription factor binding sites overrepresented within the promoters of the 10 differentially expressed RPA homologs when compared to all promoters in the soybean genome (65 197 promoters,

Supporting Information Table S4). The promoter sequences were limited to the 1000 base pairs upstream of the start ATG. Clover analysis was performed with default settings and a t-value cut-off of $t < 0.05$. Our analysis was limited to the 126 transcription factors identified from plants. The same approach was used to identify transcription factor binding sites overrepresented in the promoters of differentially expressed genes identified by O'Rourke *et al.* (2009).

Identification of putative soybean replication and repair orthologs

To find the orthologs of *Arabidopsis* replication and repair genes in soybean, we took advantage of the work by Shultz *et al.* (2007) and Singh *et al.* (2010). The *Arabidopsis* protein sequences were compared to all predicted proteins in the soybean genome (Schmutz *et al.* 2010) using BLASTP ($E < 10^{-4}$, Altschul *et al.* 1997). To account for genome duplication in soybean yet remain stringent, each *Arabidopsis* protein was allowed to hit two soybean proteins. BLASTP ($E < 10^{-4}$, Altschul *et al.* 1997) was then used to compare to soybean proteins back to all *Arabidopsis* proteins. Soybean proteins were only considered putative orthologs if they identified the original *Arabidopsis* query sequence. Note that 23 genes were in common between the publication by Shultz *et al.* (2007) and Singh *et al.* (2010).

VIGS of GmRPA3c and GmRPA3d

Vector construction

A 302 base pair segment of *Glyma20g24590* (GmRPA3c) was amplified with forward (5'-ATGCGGATCCTCCTTC TGTATTTGTAATGCTCAG-3') and reverse (5'-ATGCGGATCCGAAACAGACCCTTAAATTCACCA-3') primers containing *Bam*HI sites using Invitrogen™ Platinum® Taq DNA Polymerase High Fidelity and Clark cDNA as template. The resulting amplicon was cloned into the *Bam*HI cloning site in pBPMV-IA-V1 RNA2 (Zhang *et al.* 2010). Vectors with RPA3 target sequence in the sense (GmRPA3S) and antisense (GmRPA3AS) direction were confirmed via sequencing. Empty vector, GmRPA3S, or GmRPA3AS constructs were bombarded into Clark and IsoClark seedlings (~10 days old). After 42 days, tissue was flash frozen with liquid nitrogen and stored at -80 °C. Infection was confirmed with a DAS-ELISA PathoScreen® BPMV Kit from Agdia® (Agdia®, Elkhart, IN, USA). All samples were tested in duplicate. Stored tissue was used for rub inoculation in subsequent experiments.

To confirm silencing of *GmRPA3c* and *GmRPA3d*, we developed gene-specific primers from the 3'UTR of each gene. The *GmRPA3c*UTR primers (5'-TATACTTGCACC TGTTTACATG-3' and TAAGCCAAACTCAACCTAAC AT-3') amplified a 129 bp product. The *GmRPA3d*UTR primers (5'-CCAGCAATTAGATGGGGTTT-3' and 5'-GC TAACATCAGAGATAATGGAACA-3') amplified a 158 bp product. RNA isolation and qPCR were performed as described above.

Silencing of *GmRPA3c* and *GmRPA3d* during iron stress

Forty-eight Isoclark seed, 12 Clark seed and six T203 seed were germinated on paper for 5 to 7 d at 27 °C before moving into hydroponic buckets. Eight Isoclark, two Clark and one T203 seedlings were placed in 10 L buckets with nutrient solutions that were either iron sufficient (100 µM Fe(NO₃)₃•9H₂O; three buckets) or iron insufficient (50 µM Fe(NO₃)₃•9H₂O; three buckets). Nutrient solutions were based on growth conditions described in Chaney *et al.* (1992), with solutions adjusted for 10 L buckets. Fully expanded unifoliates of six Isoclark seedlings in each treatment (two Isoclark seedlings in each bucket) were rub-inoculated with one of four treatments: Mock (buffer control), vector, *GmRPA3S*, or *GmRPA3AS*. Clark (iron efficient) and T203 (iron inefficient) served as bucket controls to ensure proper conditions for developing an IDC phenotype.

Vegetative growth, IDC score and chlorophyll content were assessed at 21 days post inoculation on the expanding trifoliolate (Fig. 3). Vegetative growth was measured as height in centimetres. IDC score was rated on a scale of 1 to 5, with 1 being green and healthy leaves and 5 being yellow and necrotic leaves. Chlorophyll content was assessed with a Minolta SPAD-520 Chlorophyll Meter. Phenotypic differences were analysed with a Student's t-test (Ramsey & Schafer 2002) with a significance cut-off of 0.05. Equal variance was assumed among data sets.

Scored leaf tissue was harvested at 21 days post inoculation, flash frozen with liquid nitrogen and stored at -80 °C. Stored tissue was used for RNA extraction and confirmation of BPMV infection by DAS-ELISA.

Silencing of *GmRPA3c* and *GmRPA3d* in soil grown plants

Six Clark and six Isoclark seeds were germinated on paper for 5 to 7 d at 27 °C before transplanting into pots (two plants per pot) with sterile soil. Fully expanded unifoliates were dusted with carborundum and rubbed with virus-infected tissue ground in phosphate buffer. Two Clark and two Isoclark seedlings were rub-inoculated with each treatment: vector, *GmRPA3S* or *GmRPA3AS*. The experiment was repeated and scores were averaged across experiments. Vegetative growth, measured as height in centimetres, was taken at four time points: 21, 30, 36 and 42 days post inoculation (Fig. 4). Plants were grown in a growth chamber at 20 °C with 16 h light (light intensity: 860 µmoles m⁻² s⁻¹). Internode lengths were measured in centimetres between each node beginning at the unifoliolate node.

RNA-Seq of VIGS-treated plants

RNA was isolated from empty vector and *RPA3AS* silenced plants grown in sufficient and deficient conditions 21 days after VIGS treatment using the protocols outlined above. For each treatment, two plants were used per biological replicate and a total of three biological replicates per treatment were collected. RNA samples were sent to the Iowa State

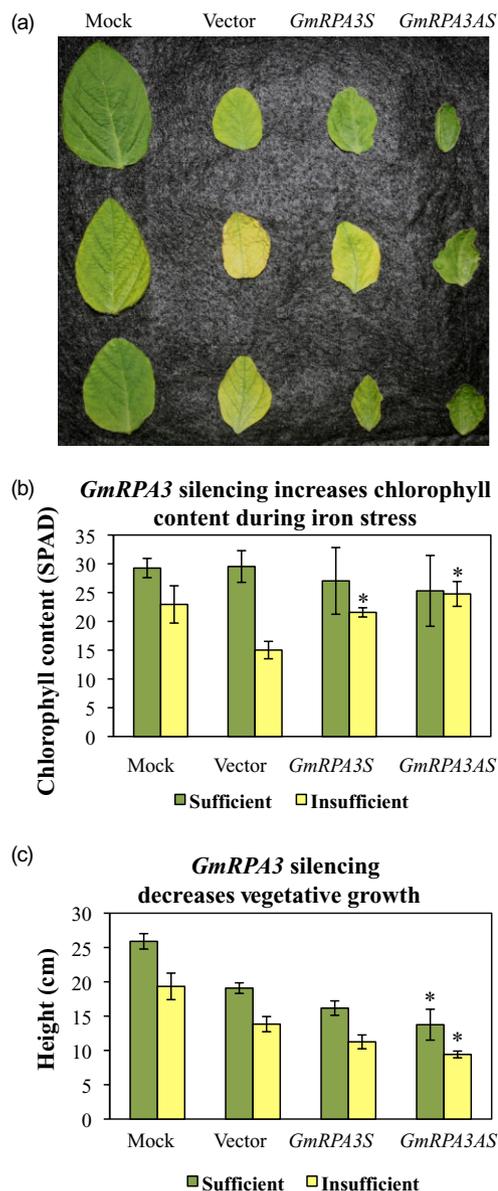


Figure 3. Silencing of *GmRPA3* in Isoclark reduces IDC symptoms during iron stress. (a) Expanding trifoliolate leaves of plants inoculated with *GmRPA3AS* constructs had an improved IDC visual score when compared to plants inoculated with vector alone when grown under iron insufficient conditions. Each row represents plants from an individual bucket. BPMV infection produces mild chlorotic symptoms, thus, the mock-infected leaves remain greener under iron stress. However, plants inoculated with the *GmRPA3AS* do not show the yellowing associated with the BPMV inoculation. (b) Silencing *GmRPA3* increases chlorophyll content under iron stress. Third trifoliolate leaves of plants inoculated with *GmRPA3AS* and *GmRPA3S* constructs significantly ($P < 0.01$, indicated by an asterisk) improved SPAD readings when compared to plants inoculated with vector alone when grown under iron insufficient conditions. Each data point is the average of six plants \pm standard error. (c) Silencing *RPA3* decreases vegetative growth under iron stress. Plants inoculated with *GmRPA3AS* were significantly shorter ($\sim 30\%$, $P < 0.05$, indicated by asterisk) on average than vector only inoculated plants 21 dpi in both iron conditions. Each data point is the average of six plants \pm standard error.

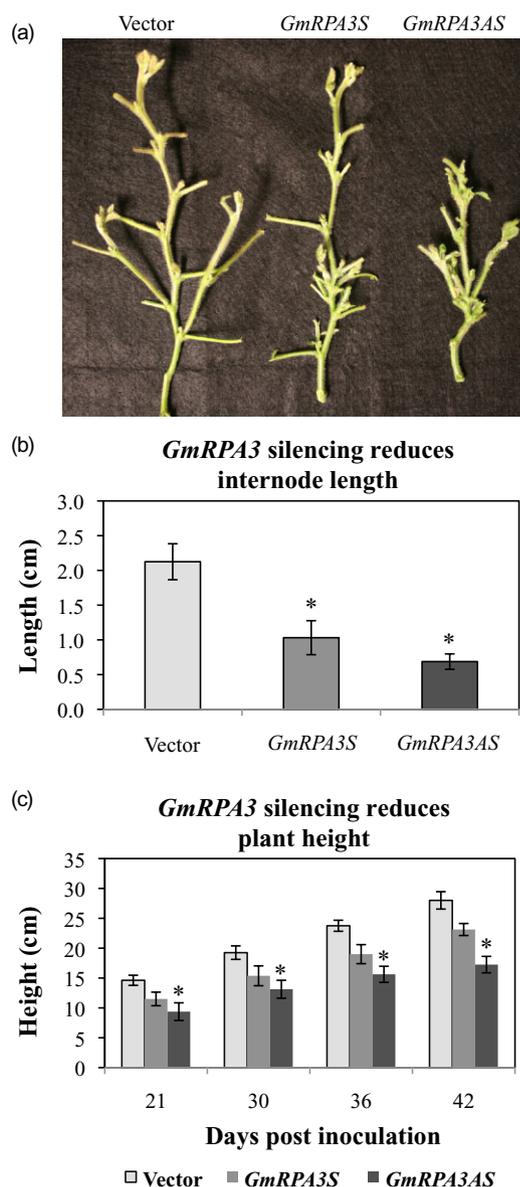


Figure 4. Silencing of *GmRPA3* decreases internode length and plant height in plants grown in soil (iron sufficient conditions). (a) *GmRPA3* silencing alters internode length and plant height. (b) Comparison of internode lengths of plants inoculated with empty vector, *GmRPA3AS* and *GmRPA3S*. Internode lengths were measured between the second and fourth trifoliolate nodes at 42 dpi. *GmRPA3AS* and *GmRPA3S* inoculation significantly reduced internode length ($P < 0.001$) when compared to vector only inoculated plants. (c) Comparison of plant heights from plants inoculated with empty vector, *GmRPA3AS* and *GmRPA3S*. Plant heights were measured 21, 30, 36 and 42 days after inoculation. Plants inoculated with *GmRPA3AS* were significantly shorter than vector only inoculated plants at all time points ($P < 0.01$, indicated by asterisk).

University DNA Facility for library preparation and paired-end sequencing using the Illumina® HiSeq 2000 platform (San Diego, CA, USA). To minimize variation caused by run differences, all twelve samples were indexed and run on a

single lane. Libraries were prepared from 0.4 μg of total RNA and indexed using the Illumina® TruSeq™ RNA Sample Prep Kit and TruSeq™ adaptors.

The resulting 100 base pair reads from all 12 sequenced samples were analysed with the programs Scythe (UC Davis Bioinformatics Core, <https://github.com/ucdavis-bioinformatics>) and Sickle (UC Davis Bioinformatics Core, <https://github.com/ucdavis-bioinformatics>, parameters $-q$ 20 and $-l$ 50) to remove adaptors and low quality sequences. The program DeconSeq (Schneider & Edwards 2011) was used to remove reads corresponding to the VIGS vector (GenBank Accessions GQ996949 and GQ996952, 20% coverage and 80% nucleotide identity). TopHat (version 2.0.3, Trapnell, Pachter & Salzberg 2009) was used to align paired reads to the Williams 82 reference genome sequence (version Gmax_109, Schmutz *et al.* 2010). All default settings were used except the distance between mate pairs ($-r$) was set at 150 and the maximum intron length ($-I$) was set at 10 000 base pairs. The 12 resulting mapping files (bam) were imported into the statistical program R (R Development Core Team 2006) using the Bioconductor package Rsamtools (Morgan & Pages 2010). The Bioconductor package rtracklayer (Lawrence, Gentleman & Carey 2009) was used to import the gene feature file corresponding to Gmax_109 (Schmutz *et al.* 2010). The package GenomicRanges (Aboyoun, Pages & Lawrence 2013) was used to count reads for genes and output a matrix containing gene counts for each sample. Prior to statistical analysis, counts assigned to *GmRPA3c* and *GmRPA3d* were removed, as they could be of viral origin. In the quality control stage of the analysis, one sample corresponding to empty vector grown under iron deficient conditions was removed because diagnostic graphics of this sample relative to the other 11 samples indicated that these gene counts were substantially different. Analyses conducted with and without this sample confirmed that it affected the downstream dispersion estimation and analysis. The remaining 11 samples were analysed together using Bioconductor package edgeR (Robinson & Smyth 2007, 2008; Robinson, McCarthy & Smyth 2010; McCarthy, Chen & Smyth 2012) by fitting a negative binomial generalized log-linear model to the read counts for each gene and then performing genewise likelihood ratio tests. Log-2 fold change, P -value and False discovery rate (FDR) are reported (Supporting Information Table S6). The design matrix used in the linear model is a 2×2 experimental design, where two factors (VIGS vector, Iron conditions) are involved and each factor has two levels. To estimate trended dispersion, which depends on overall gene expression, we used a Cox-Reid approximate conditional inference and applied an empirical Bayes method to shrink the tagwise dispersion towards the trended dispersion, as suggested in the edgeR documentation. Differential gene expression in response to differences in the VIGS vectors (Empty vector versus *GmRPA3AS*) was tested using the full linear model (FDR < 0.1). To identify genes whose iron responsiveness was significantly different (FDR < 0.1) between empty vector and *GmRPA3AS* constructs, contrasts were specified to identify genes differentially expressed in response to iron in either empty vector or *RPA3AS* samples.

This list was then reduced by selecting genes where the FDR < 0.1 for one construct but FDR > 0.1 for the other (Supporting Information Tables S7 & S8). Interaction plots produced using the graphics packages ggplot2 (CRAN, Wickham 2009) and ggbio (Bioconductor, Yin, Cook & Lawrence 2012) were used to visualize and confirm all differentially expressed genes.

Annotation of differentially expressed genes

Differentially expressed genes were annotated using the SoyBase Genome Annotation report page (<http://soybase.org/genomeannotation>). In brief, the longest predicted protein sequences of all genes in the soybean genome were compared to the UniRef100 database (version 11/26/2012, Apweiler *et al.* 2004) protein database using BLASTP (E < 10⁻¹⁰, Altschul *et al.* 1997). Custom perl scripts were used to identify the most informative BLASTP hit (not putative, hypothetical or predicted). Proteins were also compared to predicted proteins from the *Arabidopsis* (*A. thaliana*) genome (version 10; The *Arabidopsis* Information Resource [<http://www.arabidopsis.org>]) using BLASTP (E < 10⁻¹⁰, Altschul *et al.* 1997). The *Arabidopsis* Information Resource gene ontology annotations (Berardini *et al.* 2004) for the top *Arabidopsis* hit were then assigned to the corresponding soybean gene. Custom perl scripts were used to pull out annotation information for differentially expressed genes of interest.

The Ontologizer 2.0 software (Bauer *et al.* 2008) was used to identify gene ontology terms overrepresented among differentially expressed genes relative to the soybean genome. Gene ontology information from *Arabidopsis* (described above) was used to create a gene associate file for soybean. Ontologizer 2.0 was run using the Parent-Child-Union calculation method with the Westfall-Young-Single-Step multiple testing correction with 1000 resamplings (Fig. 5, Supporting Information Tables S9 & S10). Ontologizer 2.0 was also used to identify overrepresented GO terms in the Ren *et al.* (2012) RNA-Seq data (Table 2). In this case, overrepresentation analysis was conducted relative to all genes in the *Arabidopsis* genome (TAIR version 10).

RESULTS

Identification of RPA homologs in soybean

The protein sequences of all RPA subunits of *A. thaliana* were used to develop hidden Markov models (HMMs, Durbin *et al.* 1998) for each of the three RPA subunits. The HMMs were screened against six-frame translations of all predicted soybean genes. We identified 18 RPA subunits in soybean; five homologs of *RPA1A*, two homologs of *RPA1B*, two homologs of *RPA1C*, five homologs of *RPA2* and four homologs of *RPA3* (Table 1, Fig. 1). Given soybean's duplicated genome (Schmutz *et al.* 2010), the SoyBase Genome Browser (<http://soybase.org/gb2/gbrowse/gmax1.01/>) was used to identify homeologs within the identified *RPA* subunits (Table 1). Of the five *GmRPA1A* homologs, only two could clearly be identified as homeologs (*GmRPA1Aa* and *GmRPA1Ab*). The two homologs of *GmRPA1B* represent a

single homeologous pair (*GmRPA1Ba* and *GmRPA1Bb*). No homeolog could be identified for *GmRPA2a*. The four homologs of *GmRPA3* corresponded to two homeologous pairs. Four of the 18 *RPA* subunits identified (*GmRPA1Ab*, *GmRPA1Ac*, *GmRPA1Ad* and *GmRPA1Ae*) likely represent pseudogenes since they are truncated at either the 5' or 3' end of the gene.

To determine which soybean homologs were most closely related to the characterized genes in rice and *Arabidopsis*, we developed multiple sequence alignments from the *RPA* subunits of several plant species (Supporting Information Figs S1, S2 & S3). Priority was given to plant species with nearly complete genome sequences (<http://www.phytozome.net>). In addition to rice and *Arabidopsis*, we included two additional representatives of the subclass Fabidae: *R. communis* and *M. truncatula*, a related legume (Fig. 1). The predicted pseudogenes, *GmRPA1Ab*, *GmRPA1Ac*, *GmRPA1Ad* and *GmRPA1Ae* (Table 1) were not included in the analyses.

Arabidopsis and rice have five and three homologs, respectively, of *RPA1*, the largest *RPA* subunit involved in binding ssDNA (Shultz *et al.* 2007). The *RPA1* homologs can be divided further into three classes: *RPA1A*, *RPA1B* and *RPA1C* (Sakaguchi *et al.* 2009). *RPA1A* is represented by a single gene in both *Arabidopsis* and rice (*At2g06510* and *Os02g53680*, respectively), but corresponds to one homeologous pair in soybean (*GmRPA1Aa* and *GmRPA1Ab*, Table 1 & Fig. 1A). *RPA1B* has two homologs in *Arabidopsis* (*At5g08020* and *At5g61000*) and a single homolog in rice (*Os03g11540*). Again, these sequences correspond to a single homeologous pair in soybean (*GmRPA1Ba* and *GmRPA1Bb*). *RPA1C* corresponds to two genes in *Arabidopsis* (*At4g19130* and *At5g45400*), one gene in rice (*Os05g02040*) and two genes in soybean (*GmRPA1Ca* and *GmRPA1Cb*). The phylogenetic relationship of *GmRPA1Ca* and *GmRPA1Cb* could not be determined.

RPA2 is the second largest *RPA* subunit and regulates *RPA* during the cell cycle and during genotoxic stress (Binz *et al.* 2004). *Arabidopsis* and rice have two and three homologs of *RPA2*, respectively (Fig. 1B). The soybean homologs of *RPA2* are made up of two homeologous pairs (*GmRPA2b* and *GmRPA2c*, *GmRPA2d* and *GmRPA2e*) and a fifth homolog (*GmRPA2a*, Table 1). Of the five soybean *RPA2* homologs, the predicted protein corresponding to *GmRPA2a* is smaller than the other subunits and highly divergent.

RPA3 is the least studied *RPA* subunit and its role in DNA replication and repair remains largely unknown. In *Arabidopsis* and rice, the subunit has little representation, with two and one homologs, respectively (Fig. 1C). In soybean, we identified two homeologous pairs (Table 1). One pair (*GmRPA3a* and *GmRPA3b*) corresponded most closely with the two homologs in *Arabidopsis*, while the other pair (*GmRPA3c* and *GmRPA3d*) corresponded most closely to the *RPA3* subunit in rice (*Os01g14980*, Fig. 1C).

RPA homologs are expressed during soybean development

To examine the expression of *RPA* homologs during development, we took advantage of the soybean RNA-Seq atlases

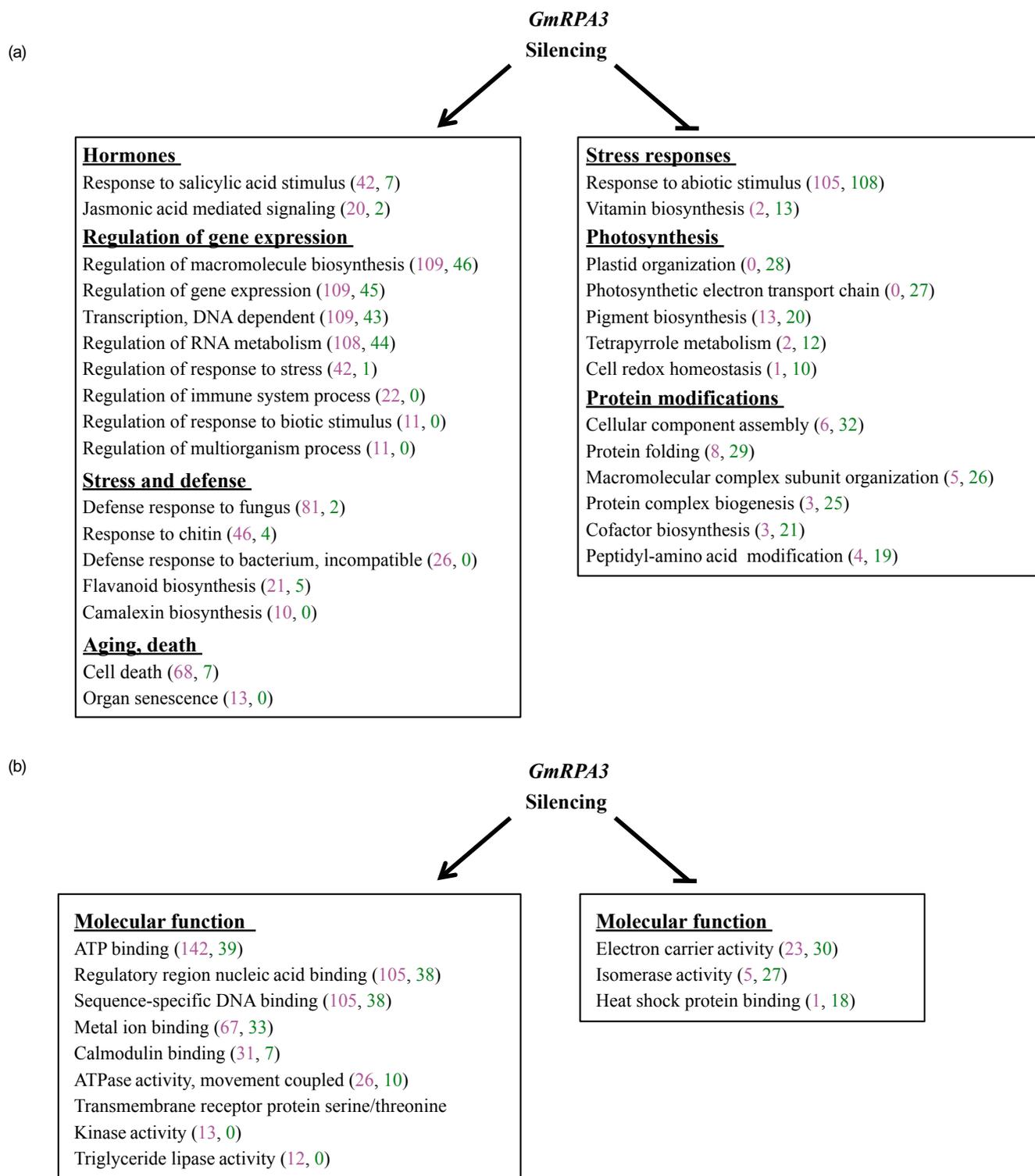


Figure 5. Gene ontology biological process and molecular function terms significantly (corrected $P < 0.05$) overrepresented in *GmRPA3* silenced plants. The analysis was conducted separately for induced genes (arrows) and repressed genes (vertical bars). The Ontologizer 2.0 software (Bauer *et al.* 2008) was used to find overrepresented GO categories among differentially expressed genes relative to the soybean genome. The Parent-Child-Union option was used to correct for relationships between GO terms and the Westfall-Young Single Step multiple testing correction was applied. Only GO terms corresponding to 10 or more genes are shown. GO terms overrepresented among genes induced by *GmRPA3* silencing are indicated by an arrow, while GO terms overrepresented among genes repressed by *GmRPA3* silencing are indicated by a bar. The number of genes induced (magenta) or repressed (green) within each GO term is indicated in parentheses. For a full list of GO terms identified prior to parent-child term correction see Supporting Information Tables S9 and S10.

Table 2. GO terms significantly overrepresented among genes induced by *GmRPA3* silencing and TOR inhibition (Ren et al. 2012)

| GO ID | GO description | Soybean: GO categories induced by <i>GmRPA3</i> silencing | | | | Arabidopsis: GO categories induced by TOR inhibition | | | |
|------------|--|---|-------------------------------------|------------------|----------------------------|--|-------------------------------------|------------------|----------------------------|
| | | Genome: Genes/Term | Differentially Expressed Genes/Term | Adjusted P-value | % Differentially Expressed | Genome: Genes/Term | Differentially Expressed Genes/Term | Adjusted P-value | % Differentially Expressed |
| GO:0044700 | single organism signalling | 3720 | 175 | 0 | 4.7 | 2166 | 112 | 0 | 5.17 |
| GO:0023052 | signalling | 3729 | 176 | 0 | 4.72 | 2167 | 112 | 0 | 5.17 |
| GO:0007165 | signal transduction | 3697 | 175 | 0 | 4.73 | 2110 | 112 | 0 | 5.31 |
| GO:0007154 | cell communication | 4261 | 191 | 0 | 4.48 | 2543 | 122 | 0 | 4.8 |
| GO:0052317 | camalexin metabolic process | 35 | 10 | 0 | 28.57 | 10 | 4 | 0.02 | 40 |
| GO:0052314 | phytoalexin metabolic process | 40 | 10 | 0 | 25 | 11 | 4 | 0.04 | 36.36 |
| GO:0050832 | defense response to fungus | 604 | 81 | 0 | 13.41 | 342 | 48 | 0 | 14.04 |
| GO:0010200 | response to chitin | 401 | 46 | 0 | 11.47 | 420 | 78 | 0 | 18.57 |
| GO:0002376 | immune system process | 1176 | 114 | 0 | 9.69 | 971 | 101 | 0 | 10.4 |
| GO:0006955 | immune response | 1075 | 103 | 0 | 9.58 | 859 | 81 | 0 | 9.43 |
| GO:0006952 | defense response | 2735 | 223 | 0 | 8.15 | 1631 | 129 | 0 | 7.91 |
| GO:0009607 | response to biotic stimulus | 2311 | 165 | 0 | 7.14 | 1442 | 118 | 0 | 8.18 |
| GO:0051707 | response to other organism | 2228 | 157 | 0 | 7.05 | 1404 | 116 | 0 | 8.26 |
| GO:0016265 | death | 1098 | 68 | 0 | 6.19 | 497 | 54 | 0 | 10.87 |
| GO:0008219 | cell death | 1098 | 68 | 0 | 6.19 | 497 | 54 | 0 | 10.87 |
| GO:0006950 | response to stress | 7197 | 313 | 0 | 4.35 | 4081 | 198 | 0 | 4.85 |
| GO:0051716 | cellular response to stimulus | 5074 | 205 | 0 | 4.04 | 3282 | 139 | 0 | 4.24 |
| GO:0050896 | response to stimulus | 13916 | 456 | 0 | 3.28 | 6830 | 271 | 0 | 3.97 |
| GO:0009867 | jasmonic acid mediated signalling pathway | 197 | 20 | 0.01 | 10.15 | 281 | 39 | 0.02 | 13.88 |
| GO:0009751 | response to salicylic acid stimulus | 466 | 42 | 0 | 9.01 | 470 | 63 | 0 | 13.4 |
| GO:0009753 | response to jasmonic acid stimulus | 582 | 44 | 0 | 7.56 | 471 | 58 | 0 | 12.31 |
| GO:0010243 | response to organic nitrogen | 450 | 50 | 0 | 11.11 | 435 | 78 | 0 | 17.93 |
| GO:1901698 | response to oxygen compound | 603 | 52 | 0 | 8.62 | 596 | 78 | 0 | 13.09 |
| GO:1901700 | response to nitrogen-containing compound | 3510 | 152 | 0 | 4.33 | 2350 | 158 | 0 | 6.72 |
| GO:0043900 | regulation of multi-organism process | 74 | 11 | 0.01 | 14.86 | 119 | 20 | 0 | 16.81 |
| GO:0080134 | regulation of response to stress | 319 | 42 | 0 | 13.17 | 554 | 66 | 0 | 11.91 |
| GO:0002682 | regulation of immune system process | 179 | 22 | 0 | 12.29 | 418 | 57 | 0 | 13.64 |
| GO:0048583 | regulation of response to stimulus | 866 | 52 | 0 | 6 | 770 | 67 | 0 | 8.7 |
| GO:0050794 | regulation of cellular process | 9723 | 287 | 0 | 2.95 | 5187 | 160 | 0 | 3.08 |
| GO:0050789 | regulation of biological process | 11019 | 308 | 0 | 2.8 | 6075 | 183 | 0 | 3.01 |
| GO:0065007 | biological regulation | 11647 | 320 | 0 | 2.75 | 6520 | 197 | 0 | 3.02 |
| GO:0000975 | regulatory region DNA binding | 4397 | 105 | 0 | 2.39 | 1687 | 58 | 0 | 3.44 |
| GO:0051252 | regulation of RNA metabolic process | 5144 | 108 | 0 | 2.1 | 2617 | 67 | 0 | 2.56 |
| GO:0010556 | regulation of macromolecule biosynthetic process | 5290 | 109 | 0.02 | 2.06 | 2721 | 67 | 0 | 2.46 |
| GO:0010468 | regulation of gene expression | 5445 | 109 | 0 | 2 | 2895 | 68 | 0 | 2.35 |
| GO:0001067 | regulatory region nucleic acid binding | 4397 | 105 | 0 | 2.39 | 1687 | 58 | 0 | 3.44 |
| GO:0043565 | sequence-specific DNA binding | 4468 | 105 | 0 | 2.35 | 1717 | 58 | 0 | 3.38 |
| GO:0006351 | transcription, DNA-dependent | 5351 | 109 | 0 | 2.04 | 2837 | 68 | 0 | 2.4 |
| GO:0032774 | RNA biosynthetic process | 5388 | 109 | 0 | 2.02 | 2852 | 68 | 0 | 2.38 |
| GO:0003677 | DNA binding | 5993 | 115 | 0 | 1.92 | 2392 | 61 | 0 | 2.55 |
| GO:0051704 | multi-organism process | 2906 | 169 | 0 | 5.82 | 1783 | 122 | 0 | 6.84 |
| GO:0007568 | aging | 383 | 17 | 0 | 4.44 | 149 | 11 | 0.04 | 7.38 |
| GO:0044711 | single-organism biosynthetic process | 2129 | 75 | 0 | 3.52 | 1822 | 86 | 0 | 4.72 |
| GO:0006464 | cellular protein modification process | 6221 | 197 | 0 | 3.17 | 2766 | 47 | 0 | 1.7 |
| GO:0003824 | catalytic activity | 23032 | 539 | 0 | 2.34 | 8936 | 221 | 0 | 2.47 |
| GO:0019438 | aromatic compound biosynthetic process | 7121 | 161 | 0 | 2.26 | 4152 | 124 | 0 | 2.99 |
| GO:1901362 | organic cyclic compound biosynthetic process | 7377 | 164 | 0 | 2.22 | 4391 | 127 | 0 | 2.89 |
| GO:0009987 | cellular process | 31067 | 671 | 0 | 2.16 | 14943 | 330 | 0 | 2.21 |
| GO:0008152 | metabolic process | 32423 | 679 | 0 | 2.09 | 15139 | 336 | 0 | 2.22 |

Overrepresented GO terms were identified using the Ontologizer 2.0 (Bauer et al. 2008) software using the Parent-Child-Union and Westfall-Young-Step Corrections and 1000 replicates. For genes induced by *GmRPA3* silencing, comparisons were made relative to all genes in the soybean genome (Glyma version 1). For all genes induced by TOR inhibition (Ren et al. 2012), comparisons were made to all genes in the Arabidopsis genome (TAIR version 10). All identified GO terms are listed.

described by Severin *et al.* (2010a, <http://soybase.org/soyseq/>) and Libault *et al.* (2010a,b; <http://soykb.org/>). The atlases were complementary, as the Severin *et al.* (2010a) atlas focused largely on above-ground tissues of the plant while the Libault *et al.* (2010a,b) atlases focused on below-ground tissues. Of the 18 RPA homologs, no expression was detected for the four predicted pseudogenes (*GmRPA1Ab*, *GmRPA1Ac*, *GmRPA1Ad* and *GmRPA1Ae*) or *GmRPA1Cb* in any atlas (Supportive Information Fig. S4). Of the remaining 13 RPA homologs, expression of 12 was shared across atlases. In general, if RPA expression was detected, it was expressed in all tissues sampled, usually at low levels (< 20 Reads/Kb/Million (RPKM)). *GmRPA3c* was the most highly expressed RPA homolog in the Severin *et al.* (2010a) atlas, with greatest expression found in young leaves (>40 RPKM). *GmRPA1Bb* was the most highly expressed RPA homolog in the Libault *et al.* (2010a,b) atlases, with greatest expression in the root tip (>120 RPKM). High expression of *GmRPA3c* and *GmRPA1Bb* in young leaves and root tips suggests RPA is involved in soybean developmental processes, especially in rapidly growing tissues.

RPA homologs are differentially expressed in response to iron stress

GmRPA3c (*Glyma20g24590*) was previously identified as differentially expressed in leaves after 14 days of iron stress (O'Rourke *et al.* 2009). However, iron homeostasis-related genes, such as *FRO* and *IRT*, can change expression levels in as little as 6 h after iron insufficiency (Buckhout, Yang & Schmidt 2009). To determine if and when other RPA subunits respond to iron stress, we used quantitative reverse transcription PCR (qPCR) to measure the expression of all RPA subunits in leaves of the NILs Clark and Isoclark. We compared expression in first trifoliolate leaf tissue after 1, 6 and 24 h of iron insufficiency (50 μ M Fe(NO₃)₃•9H₂O) to the same tissue and time points grown in iron sufficient conditions.

Primers were designed to amplify individual RPA subunits and differentiate between homeologous genes (Supporting Information Table S1). The predicted transcript sequences of the homeologous pairs shared between 93 and 97% nucleotide identity. Primers were tested on RNA from Clark and Isoclark first trifoliolate leaf tissue grown under 14 d of iron stress and were used for qPCR analysis if they produced a single amplicon of the predicted size. Of the 18 RPA subunits tested, 11 met this criterion and were used for further gene expression analyses (Supporting Information Table S1). Of the seven RPA subunits that failed to amplify, four (*GmRPA1Ab*, *GmRPA1Ac*, *GmRPA1Ad* and *GmRPA1Ae*) represented likely pseudogenes. The remaining three (*GmRPA1Ba*, *GmRPA1Cb* and *GmRPA3a*) were high-confidence gene predictions, but were not expressed under test conditions. Expression of *GmRPA1Cb* was also not detected in any of the gene atlases (Libault *et al.* 2010a,b; Severin *et al.* 2010a). One homolog of RPA1A (*GmRPA1Aa*), one homolog of RPA1B (*GmRPA1Bb*), one homolog of RPA1C (*GmRPA1Ca*), five homologs of RPA2 (*GmRPA2a*, *GmRPA2b*, *GmRPA2c*, *GmRPA2d* and

GmRPA2e) and three homologs of RPA3 (*GmRPA3b*, *GmRPA3c* and *GmRPA3d*) were analyzed by qPCR (Supporting Information Table S1, Fig. 2A & B).

The relative gene expression ratio was determined by dividing the RNA quantity in iron insufficient conditions to the RNA quantity at the same time point in iron sufficient conditions, averaged over three biological replicates, then log transformed (base 2) to create a normal distribution of data. ANOVA (Chambers, Freeny & Heiberger 1992) was performed on normalized gene expression values in iron sufficient conditions at 1, 6 and 24 hps to determine if RPA gene expression was stable over the time points in iron sufficient conditions. Normalized gene expression values in iron sufficient conditions were also log transformed (base 2) to ensure a normal distribution. Gene expression was considered stable if the ANOVA analysis was insignificant, indicating no change in expression among time points.

In the iron inefficient genotype (Isoclark) grown under iron sufficient conditions, none of the 11 RPA genes tested changed expression significantly over time (data not shown, $P > 0.05$). Similarly, nine of the 11 RPA genes had no significant changes in expression in the iron efficient line Clark (data not shown, $P > 0.05$). *GmRPA1Bb* and *GmRPA2c* were significantly ($P < 0.05$) differentially expressed, suggesting their expression was not stable across time points in iron sufficient conditions. However, pairwise comparisons between time points found a significant change in expression only between 6 hps and 24 hps for *GmRPA1Bb* and *GmRPA2c*. This result suggests change in gene expression under iron sufficient conditions is for a short period around 6 hps before returning to their baseline levels. Therefore, when gene expression values are compared across time points in iron insufficient conditions, any changes in relative gene expression are due to the iron stress response and not the developmental time points. To determine how relative expression changes over time, the relative expression at 6 and 24 hps was compared to relative expression at 1 hps. Values above zero indicate greater expression in iron insufficient conditions, while values below zero indicate lesser expression in iron insufficient conditions. Statistical significance data can be found in Supporting Information Tables S2 and S3.

Over the course of this study, we identified a total of nine RPA subunits that were differentially expressed in response to iron stress (Fig. 2A & B). In general, RPA expression decreased in leaves of the iron efficient line Clark under iron stress (Fig. 2A). Of the three *GmRPA1* genes, *GmRPA1Bb* showed decreased expression at 24 hps while the expression levels of *GmRPA1Aa* and *GmRPA1Ca* remained unchanged under iron stress. Four of the five *GmRPA2* homologs exhibited decreased expression levels, while the fifth, *GmRPA2a*, showed increased expression at 24 hps. Two of the three *GmRPA3* genes showed decreased expression at 24 hps, while *GmRPA3b* remained unchanged. In all cases, the change in expression from 1 to 6 hps is insignificant (Supporting Information Table S2), suggesting changes in gene expression occur between 6 hps and 24 hps in Clark.

RPA subunits exhibit the opposite expression patterns in the iron inefficient line Isoclark, with expression largely

increasing during iron stress (Fig. 2B). In IsoClark, *GmRPA1Bb* and *GmRPA1Ca* showed increased expression at 24 hps, while *GmRPA1Aa* again remained unchanged. Increased expression in IsoClark is also observed for the *GmRPA2* genes, except for *GmRPA2e*, which remained unchanged. For subunit RPA3, the same genes that exhibit decreased expression in Clark 24 hps were up regulated at 24 hps in IsoClark (*GmRPA3c* and *GmRPA3d*). *GmRPA3b* showed a slight decrease in expression in IsoClark. In all cases, the change in expression from 1 to 6 hps is insignificant (Supporting Information Table S3), suggesting response to iron occurs between 6 hps and 24 hps in IsoClark as well.

In addition to examining the changes in gene expression in response to iron stress, we also studied the relationship between gene expression and phylogenetic distribution. Expression analyses revealed RPA homologs are differentially expressed in response to iron stress (Fig. 2). In several cases, we could see clear differences in gene expression between homeologous genes. For example, *GmRPA1Ba* was not expressed in our initial study, but its homeolog *GmRPA1Bb* was significantly differentially expressed in response to iron stress in Clark and IsoClark. In some cases, both homeologs were expressed similarly in response to iron stress. For example, *GmRPA2b* and *GmRPA2c* were both differentially expressed in response to iron stress in both Clark and IsoClark.

To determine which transcription factors regulate RPA expression, we examined the promoters of the nine significantly differentially expressed RPAs using Clover (Cis-Element OVER representation) software (Frith *et al.* 2004) and the TRANSFAC transcription factor database (Matys *et al.* 2006). Six plant transcription factor-binding sites were significantly ($P < 0.05$) overrepresented among the promoters of the RPA homologs when compared to all promoters in the soybean genome (Supporting Information Table S4). These included ERF2 (M01057, $P = 0$), ABF1 (M00401, $P < 0.001$), E2F (M01114, $P < 0.001$), OSBZ8 (M00654, $P < 0.003$), TRAB1 (M00507, $P < 0.026$) and KNOX3 (M00819, $P < 0.02$). Their frequency and positions within promoters are listed in Supporting Information Table S4. The ABF1, ERF2, OSBZ8 and TRAB1 transcription factors are associated with abscisic acid signalling in response to abiotic stress (Mukherjee *et al.* 2006; Agarwal *et al.* 2010; Mizoi, Shinozaki & Yamaguchi-Shinozaki 2012). KNOX3 is involved in ethylene signalling (Osnato *et al.* 2010). E2F is involved in the control of cell cycle and DNA replication (Mariconti *et al.* 2002; Blanchet *et al.* 2011).

Genes involved in DNA replication are repressed in response to iron stress in the iron efficient line Clark

Since RPA has roles in both DNA replication and DNA repair, we examined the differentially expressed genes in Clark and IsoClark (O'Rourke *et al.* 2009) for differential expression of replication or repair associated genes (Supporting information Table S5). Fifty-eight DNA replication proteins have been characterized in *Arabidopsis* (Shultz *et al.*

2007). To find putative orthologous proteins in soybean, we used best match reciprocal BLASTP ($E < 10^{-4}$, Altschul *et al.* 1997) to compare the *Arabidopsis* replication proteins against all predicted proteins in the soybean genome. Given soybean's duplicated genome (Schmutz *et al.* 2010), each *Arabidopsis* protein was limited to two potential soybean matches. Of the 58 *Arabidopsis* proteins, 47 had putative orthologs in soybean. In total, we identified 78 soybean genes orthologous to known *Arabidopsis* DNA replication proteins (data not shown). Of these, 44 genes were represented on the Soybean Affymetrix gene chip (<http://soybase.org/AffyChip/>). In Clark, 14 (32%) of these genes exhibited significantly decreased expression in response to iron stress (fold change < -2.981 , O'Rourke *et al.* 2009), while none were differentially expressed in IsoClark. The same approach was used to identify putative soybean orthologs of *Arabidopsis* DNA repair and recombination genes (Singh *et al.* 2010). Of the 229 *Arabidopsis* genes, 175 had orthologs in soybean, corresponding to 281 soybean genes. Of these, 170 genes were represented on the Soybean Affymetrix gene chip (<http://soybase.org/AffyChip/>). Only three were differentially expressed in Clark, while none were differentially expressed in IsoClark (data not shown). In the O'Rourke *et al.* experiment, Clark and IsoClark seedlings were transferred to a hydroponics system containing iron sufficient or deficient media following germination. Two weeks later, when plants were approximately 21 days old, leaf tissue was collected for microarray analysis, giving insight into later stages of the iron deficiency response. To examine the early stages of iron stress, we took advantage of the work of Peiffer *et al.* (2012, Supporting information Table S5). In this case, germinated Clark seedlings were placed in hydroponics containing iron sufficient media for 14 days. Plant roots were rinsed and plants were returned to hydroponics with either sufficient or deficient iron. Twenty-four hours later, when plants were approximately 21 days old, leaf tissue was collected for RNA-seq. Of the 78 soybean sequences orthologous to known *Arabidopsis* DNA replication proteins, 47 (60%) showed significant ($Q < 0.05$) decreased expression in response to iron stress (fold change < -9.83). Of the 281 soybean genes associated with DNA repair, 43 (15%) were significantly ($Q < 0.05$) differentially expressed in response to iron stress. Thirty-six genes showed decreased expression (fold change < -8.06) while seven genes exhibited increased expression (fold change > 7.33). Our analysis of RPA homolog expression followed the same methods used by Peiffer *et al.* (2012). By combining our work and that of O'Rourke *et al.* (2009) and Peiffer *et al.* (2012), we have demonstrated that differential expression of RPA homologs in response to iron stress occurs between 6 and 24 h after iron stress and is maintained for at least 2 weeks, suggesting a prominent role in the iron deficiency response.

To identify the transcription factors regulating the expression of all iron responsive genes in Clark, we used Clover (Frith *et al.* 2004) and the TRANSFAC transcription factor database (version 7.0, Matys *et al.* 2006) to examine the promoters of the 610 unique genes differentially expressed after 14 d of iron stress (O'Rourke *et al.* 2009). The same E2F

transcription factor-binding site (M01114) over-represented in the promoters of RPA subunits was significantly overrepresented ($P=0$) and was identified at least once in the promoters of 262 differentially expressed genes. The E2F transcription factor is involved in replication and cell cycle control (Mariconti *et al.* 2002; Blanchet *et al.* 2011). The same approach was used to examine the 250 probes (198 unique genes) differentially expressed 14 days after iron stress in the iron inefficient line Isoclark (O'Rourke *et al.* 2009). The E2F transcription factor-binding site (M01114) was not significantly overrepresented. Our results, using different time points in the iron stress response and different gene expression platforms, suggest that repression of DNA replication in the leaves is a striking component of the iron stress response unique to the iron efficient Clark.

Silencing of *GmRPA3* in Isoclark reduces IDC symptoms during iron stress

Our analyses to this point demonstrate that RPA genes are differentially expressed in response to iron stress and show opposing patterns of expression in response to iron deficiency in the two NILs. Further, a large suite of DNA replication genes is similarly affected. Given that *GmRPA3c* had the greatest level of differential expression, mapped with an IDC QTL associated with chlorophyll content in two other populations and was one of the most differentially expressed genes identified by O'Rourke *et al.* (2009), we decided to test the function of *GmRPA3c* by VIGS coupled with RNA-Seq. *GmRPA3c* is repressed in the iron efficient line Clark. We hypothesized silencing *GmRPA3c* expression in the iron inefficient line Isoclark, would mirror decreased expression found in Clark and would improve IDC symptoms. Given that *GmRPA3c* and *GmRPA3d* are homeologs (Table 1) differentially expressed in response to iron (Fig. 2) and share more than 97% nucleotide identity, both genes were simultaneously targeted for VIGS using the *Bean pod mottle virus* (BPMV) vector described by Zhang *et al.* (2010). The BPMV vector has been used successfully to identify genes responsible for resistance (Meyer *et al.* 2009; Liu *et al.* 2012), defense (Liu *et al.* 2011; Pandey *et al.* 2011; Zhang *et al.* 2012) and growth and development (Zhang *et al.* 2010; Liu *et al.* 2011).

Previously, sense and antisense constructs have been shown to have different silencing activity, with greatest silencing observed in antisense constructs (Zhang *et al.* 2010). We developed two VIGS constructs targeting both *GmRPA3c* and *GmRPA3d*: *GmRPA3S* has a portion of *GmRPA3c* in the sense orientation, while it was inserted in the antisense orientation in *GmRPA3AS*. In addition to constructs developed to silence *GmRPA3c* and *GmRPA3d*, we included an empty vector VIGS construct (Empty Vector) as a control for viral symptoms in our experiments.

The optimal time for assessing VIGS phenotypes in soybean leaves is 14 to 21 d after virus inoculation (Meyer *et al.* 2009; Zhang *et al.* 2010; Pandey *et al.* 2011). Plants are inoculated 10 to 14 d after germination, when unifoliates have emerged. Inoculated plants are phenotyped 2 to 3

weeks later to allow the virus to spread systemically through the plant. For our VIGS experiment, 7-day-old germinated seedlings were transferred to hydroponics with either sufficient or insufficient iron, following the approach used by O'Rourke *et al.* (2009). After 7 d in hydroponics, when plants were approximately 14 days old, plants were inoculated with *GmRPA3AS*, *GmRPA3S*, empty vector, or mock VIGS treatment. After inoculation, plants were maintained in the same growth conditions to allow the VIGS vectors to spread systemically. Plants were phenotyped and tissue for RNA-seq was collected 21 d later. Phenotyping was performed using visual score and chlorophyll content. Plants were grown hydroponically as described by Chaney *et al.* (1992). Each hydroponic bucket included 11 plants: two untreated Clark and one untreated T203 (IDC symptom controls) and two Isoclark plants for each treatment (Mock, Vector only, *GmRPA3S* and *GmRPA3AS*). For iron sufficient and deficient treatments, the experiment was repeated in triplicate, yielding six biological replicates for each treatment.

Comparisons of mock and vector only plants (Fig. 3) demonstrated the vector alone increased IDC symptom severity during iron deficiency. Therefore, all plants treated with RPA VIGS constructs were compared to vector only plants, not mock treated plants. As hypothesized, 21 d post inoculation (dpi), roughly 28 d after iron stress, plants inoculated with antisense constructs (*GmRPA3AS*) exhibited reduced chlorotic symptoms compared to those inoculated with the BPMV virus alone (Fig. 3A & B). In plants inoculated with *GmRPA3AS*, IDC visual scores improved by 1 point, from an average of 3 (interveinal chlorosis) in plants inoculated with empty vector to an average of 2 (slight yellowing) (data not shown). Soil and plant analyzer development (SPAD) readings demonstrated significantly ($P < 0.01$) greater chlorophyll content in *GmRPA3AS* and *GmRPA3S* inoculated plants grown under iron insufficient conditions compared to vector only inoculated plants (Fig. 3B). *GmRPA3AS* plants were ~30% shorter on average than plants inoculated with empty vector (Fig. 3C, $P < 0.01$) when grown under iron insufficient conditions. In iron sufficient conditions, height was affected but chlorophyll content was not significantly different among any of the four treatments (Fig. 3B & C). All plants had virus symptoms and BPMV infection was confirmed with a double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) (data not shown).

To determine the affect of VIGS on *GmRPA3c* and *GmRPA3d* expression, we developed gene-specific primers from the 3' UTR (untranslated region) of each gene. Primers were used for qPCR of RNA isolated from VIGS treated plants. *GmRPA3c* silencing was confirmed in *GmRPA3AS* and *GmRPA3S* plants at an average level of 12.0 and 7.3-fold less expression relative to empty vector plants, respectively ($P < 0.01$, Fig. S5). Similarly, *GmRPA3d* silencing in *GmRPA3AS* and *GmRPA3S* plants resulted in an average of 3.3 and 2.0-fold less expression compared to empty vector plants ($P < 0.01$, Fig. S5). While silencing of *GmRPA3d* may contribute to the VIGS phenotype, these results confirm that the VIGS phenotype we observed is largely due to silencing of *GmRPA3c*.

To determine whether the stunted phenotype was a result of the VIGS treatment or the hydroponics system, we repeated the experiment growing plants in soil. *GmRPA3*-silencing in soil revealed a shortened internode phenotype in addition to the vegetative stunting phenotype (Fig. 4A–C). Two plants of each genotype (Clark and Isoclark) were inoculated via rub inoculation for the treatments empty vector, *GmRPA3S* and *GmRPA3AS*, for a total of four plants per treatment. All plants had viral symptoms and BPMV infection was confirmed with a DAS-ELISA (data not shown). When compared to vector only inoculated plants, silencing with *GmRPA3AS* decreased internode length by nearly 70% on average at the third and fourth trifoliolate stage (Fig. 4A & B, $P < 0.01$). We also measured plant height at 21, 30, 36 and 42 dpi to track the effect of silencing on growth over time. *GmRPA3AS* plants were ~25% shorter on average than empty vector plants at all time points (Fig. 4C, $P < 0.01$). The stunting and internode length phenotypes confirm a role for *GmRPA3* in plant growth and development.

RNA-Seq of VIGS-treated plants reveals massive transcriptional reprogramming in response to *GmRPA3* silencing

To understand how silencing of *GmRPA3* promotes IDC tolerance, we performed RNA-Seq on VIGS treated plants to identify differentially expressed genes. Leaf RNA from three biological replicates of empty vector and *GmRPA3AS* plants grown in iron sufficient or deficient conditions, and 21 d after VIGS treatment, was used for library preparation and paired end sequencing using the Illumina HiSeq2000 platform. Following the bioinformatic pipeline detailed in the materials and methods, a total of 157 338 354 reads (corresponding to all 12 samples) were mapped to the soybean genome. The Illumina reads generated by this study were deposited in the National Center for Biotechnology Short Read Archive (NCBI SRA Bioproject accession PRJNA190191). As part of the quality control process, one sample corresponding to empty vector grown under iron deficient conditions was removed from further analysis because diagnostic analyses indicated it was an outlier that biased downstream analyses.

To identify genes differentially expressed in response to *GmRPA3* silencing, regardless of iron conditions, we used edgeR (Robinson *et al.* 2010) to conduct two single factor experiments comparing empty vector and *RPA3AS* treated plants in iron sufficient or deficient conditions. A total of 2076 differentially expressed genes (RPKM ≥ 1 , fold change ≥ 1.5) were identified using a false discovery rate FDR < 0.05 (Supporting Information Table S6). Of these, 935 were repressed by *GmRPA3AS* while 1140 were induced.

The Ontologizer 2.0 software (Bauer *et al.* 2008) was used to identify gene ontology (GO) terms significantly overrepresented among the differentially induced and repressed genes relative to the soybean genome (Fig. 5, Supporting Information Tables S9 & S10). The software corrects for multiple testing, sampling and parent-child relationships between terms. Biological process GO terms significantly overrepresented (Corrected $P < 0.05$) among genes induced

by *GmRPA3* silencing included GO terms associated with salicylic and jasmonic acid signalling, regulation of gene expression, stress, defense, immunity, cell death and organ senescence (Fig. 5A, Supporting Information Table S9). Overrepresented molecular function terms included terms associated with signalling, transcription factor activity, transport and metal ion binding (Fig. 5B, Supporting Information Table S9). In contrast, genes repressed by *GmRPA3* silencing were overrepresented with GO terms associated with photosynthesis and protein modification (Fig. 5A, Supporting Information Table S10).

These results suggest that inhibition of DNA replication induces massive transcriptional reprogramming. To determine how *GmRPA3* silencing could affect such diverse genes, we used the SoyDB transcription factor database (<http://casp.rnet.missouri.edu/soydb>; Wang *et al.* 2010) to identify all differentially expressed transcription factors. Of the 5683 transcription factors present in SoyDB, 173 were differentially expressed in response to *GmRPA3* silencing, representing 34 different transcription factor families (Fig. 6, Supporting Information Table S11). Fifty-one transcription factors were repressed with fold change levels ranging from -1.65 to -8 . In contrast, 122 transcription factors were induced with fold changes ranging from 1.77 to 490. Several orthologs of transcription factors with known function in *Arabidopsis* were identified. For example, Glyma10g33060 and Glyma20g34570 are orthologs of ethylene responsive factor 1 (ERF1) and are induced 37- and 89-fold, respectively in response to *GmRPA3* silencing. ERF1 integrates jasmonate and ethylene signalling during plant defense (Lorenzo *et al.* 2003). Similarly, Glyma10g32410 and Glyma20g35180, orthologs of MYB15 are induced 10- and fivefold, respectively. MYB15 overexpression confers drought tolerance (Ding *et al.* 2009). Glyma19g26400 is an ortholog of WRKY75, a modulator of phosphate acquisition (Devaiah, Karthikeyan & Raghothama 2007) and is induced 33-fold by *GmRPA3* silencing. Glyma17g06290 is an ortholog of CGA1, which regulates chlorophyll biosynthesis in *Arabidopsis* (Hudson *et al.* 2011). The expression of Glym17g06290 and Glyma17g34670 (a CGA1 homolog) are repressed two- and fivefold respectively, by *GmRPA3* silencing, suggesting they regulate the downstream photosynthetic genes identified above.

In addition to identifying genes that respond to silencing regardless of iron availability, we also identified genes that either lost or gained iron responsiveness as a result of *GmRPA3* silencing. We used edgeR to identify genes responsive to iron conditions in either an empty vector or *GmRPA3* silenced background and then used contrast statements to identify differences in iron responsiveness between *GmRPA3* and empty vector treated plants. We identified 74 genes that were iron responsive in empty vector treated plants (FDR < 0.1) but were less responsive in *GmRPA3* silenced plants (FDR > 0.1 , Supporting Information Table S7). Similarly, we identified 71 genes that were largely nonresponsive to iron status in empty vector treated plants (FDR > 0.1) but became iron responsive in *GmRPA3* silenced plants (FDR < 0.1 , Supporting Information

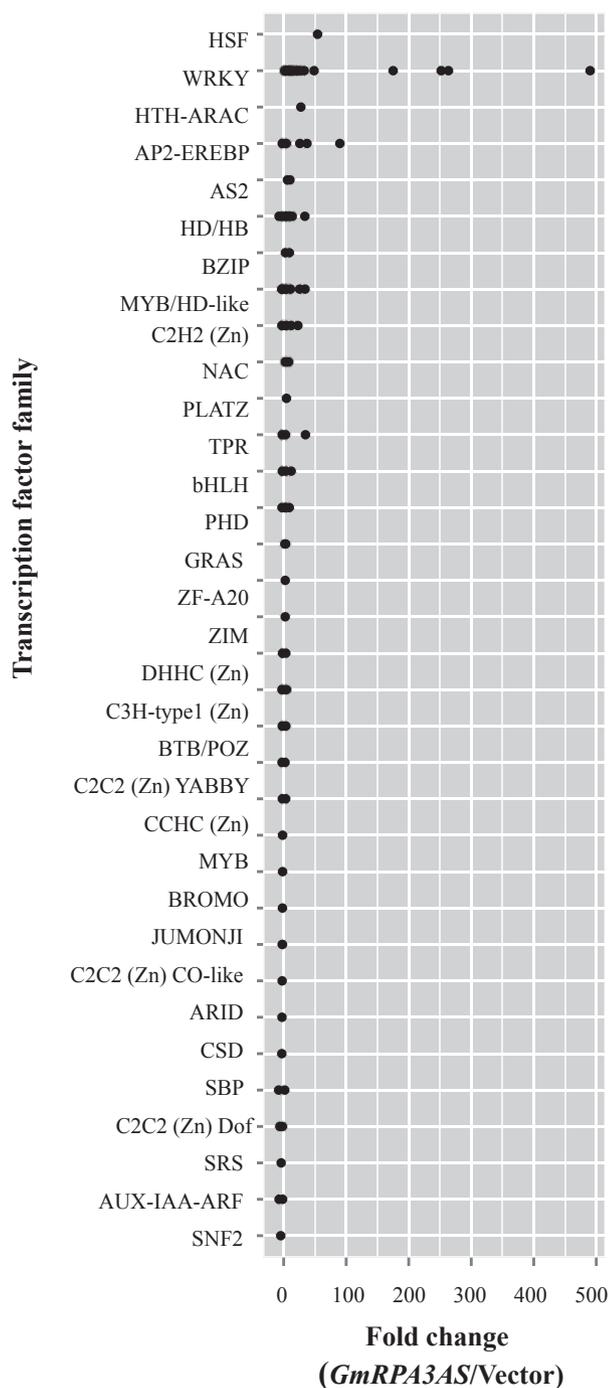


Figure 6. Expression patterns of transcription factors differentially expressed between *GmRPA3* silenced and empty vector treated plants. Transcription factor families refer to the SoyDB transcription factor database (Wang *et al.* 2010). Multiple differentially expressed transcription factors per family may be plotted. For additional data, see Supporting Information Table S11.

Table S8). We used the Ontologizer 2.0 software (Bauer *et al.* 2008) to identify significantly overrepresented GO terms (Corrected $P < 0.05$) within the 145 genes whose iron responsiveness was changed by *GmRPA3* silencing. As anticipated, identified GO terms included ferric iron binding

(GO:0008199). Additionally, GO terms related to single organism transport (GO:0044765), generation of precursor metabolites and energy (GO:0006091), cellular homeostasis (GO:0019725) and photosynthesis (GO:0015979) were also significantly overrepresented.

DISCUSSION

Previous research suggested a link existed between IDC tolerance and control of cell cycle during iron limiting conditions. O'Rourke *et al.* (2009) found DNA replication and repair genes were overrepresented among differentially expressed genes during the response to iron stress. Specifically, *GmRPA3c* was one of the most differentially expressed genes between two NILs differing in iron efficiency during iron stress. Further, the location of *GmRPA3c* corresponded to an IDC QTL in two other soybean populations (O'Rourke *et al.* 2009).

RPA has been studied most extensively in humans and yeast, but recently orthologs of RPA have been discovered in plants such as rice and *Arabidopsis* (Ishibashi *et al.* 2001; Shultz *et al.* 2007). Knockout and RNAi studies of the RPA subunits in *Arabidopsis* and yeast have provided insight into the functions of each subunit. A T-DNA insertion mutant of *AtRPA1A* in *Arabidopsis* was lethal, as was an RNAi line, suggesting an essential role in DNA replication (Ishibashi *et al.* 2005). However, a separate study found that *AtRPA1A* knockouts were viable, with only an increased sensitivity to DNA mutagens and increased telomere length (Takashi *et al.* 2009). A T-DNA insertion mutant of *AtRPA1B* and an *AtRPA1B* RNAi line were viable and showed greater sensitivity to DNA damaging agents, suggesting *AtRPA1B* is involved in DNA repair (Ishibashi *et al.* 2005). Knock out of *AtRPA1C* is lethal, again suggesting an essential role in DNA replication (Ishibashi, Kimura & Sakaguchi 2006). Functional studies have tied RPA2 to both DNA replication and repair. Three independent studies found that T-DNA insertion mutants of the *AtRPA2* gene resulted in stunted plants, earlier flowering and an increased sensitivity to the DNA damaging agent MMS (Elmayan, Proux & Vaucheret 2005; Kapoor *et al.* 2005; Xia *et al.* 2006). While mutants are available for *AtRPA3* homologs (<http://www.arabidopsis.org>), no phenotypic data is available. No additional functional studies of RPA3 have been reported in plants, although a knockout of RPA3 in yeast was lethal (Brill & Stillman 1991).

Studies of RPA homologs are further complicated by the interactions of different subunits in multiple complexes. Rice has three homologs of RPA1 (*OsRPA1A*, *OsRPA1B* and *OsRPA1C*), three homologs of RPA2 (*OsRPA2-1*, *OsRPA2-2* and *OsRPA2-3*) and one RPA3 subunit (*OsRPA3*) (Ishibashi *et al.* 2006). While two complexes have been shown to localize to the nucleus, a third localizes to the chloroplast (Ishibashi *et al.* 2006). With 18 homologs of RPA subunits in soybean, subspecialization may occur more readily, potentially resulting in more complexes. By combining phylogenetic and expression analyses, we can visualize how duplicated genes have changed in either function or expression. The potential for subfunctionalization and functional redundancy in

soybean may provide novel opportunities to tease apart the roles of different RPA subunits without the risk of lethality.

Inhibition of the cell cycle in response to stress and nutrient deficiency has been well studied in yeast and mammalian systems (Sengupta, Peterson & Sabatini 2010; Smeekens *et al.* 2010; Maddocks & Vousden 2011). Nutrient deficiency activates an AMP-activated protein kinase (SnRK1 in plants) that inhibits the mammalian target of rapamycin (mTOR). While SnRK1 activity in *Arabidopsis* is induced by sucrose, heavy carbon demand and virus infection, it is repressed by trehalose-6-phosphate and glucose-6-phosphate (Hey, Byrne & Halford 2010). When TOR is active, it promotes growth, development and biosynthesis. Inhibition of TOR induces cellular recycling and represses cell replication. In *Arabidopsis* and other plant systems TOR, RAPTOR (Regulatory Associated Protein Of TOR), and LST8 (Target of rapamycin complex subunit LST8) form a complex (TORC1, Moreau *et al.* 2010; Dobrenel *et al.* 2011) that phosphorylates S6K (ribosomal subunit S6 Kinase), which regulates DNA replication and cell cycle progression by associating with the E2F transcription factor (Henriques *et al.* 2010).

In mammals and yeast, rapamycin binds to the FK506 binding protein 12, which then binds to mTOR inhibiting its activity. However, the plant FK506 binding protein is resistant to rapamycin, making it difficult to identify genes in the TOR pathway, as silencing of TOR itself can be lethal. Recently, Ren *et al.* (2012) transformed the yeast *FK506 binding protein 12* into *Arabidopsis* making it sensitive to rapamycin and allowing inhibition of TOR. TOR inhibited plants exhibited smaller leaves, shorter hypocotyls, greater chlorophyll content, slower phase change, delayed flowering and senescence and increased life span relative to wild type. Similarly, *GmRPA3* silenced plants had smaller leaves, decreased internode lengths, greater chlorophyll content and did not senesce under iron limiting conditions. TOR inhibited plants resembled nutrient starved plants and were unable to respond to increased nutrition. Similarly, *GmRPA3* silenced plants failed to respond to increased nutrient available in soil and remained stunted. Further, RNA-Seq comparing gene expression between TOR inhibited and wild type plants and *GmRPA3* silenced and empty vector plants, reveal extensive and overlapping transcriptome reprogramming.

The phenotypic similarities between our *GmRPA3* silenced plants and TOR inhibited plants suggest TOR could be involved in regulating the expression of *GmRPA3*, likely via the E2F transcription factor. Of the 510 *Arabidopsis* genes induced by TOR inhibition, only 203 had at least one soybean ortholog (BLASTP, $E < 10^{-10}$). Of these, 26 (12.8%) were induced by *GmRPA3* silencing. Similarly, of the 405 *Arabidopsis* genes repressed by TOR inhibition, only 174 had at least one soybean ortholog. Of these, 17 (9.7%) were repressed by *GmRPA3* silencing. However, this analysis ignored approximately 60% of TOR regulated genes for which no soybean orthologs could be identified. Therefore, we decided to use the Ontologizer 2.0 software to identify gene ontology terms significantly overrepresented among the TOR regulated genes and compare the results to those obtained for genes differentially expressed in response to

GmRPA3 silencing (Fig. 5, Supporting Information Tables S9 & S10). The same approach described previously was used except all predicted genes in the *Arabidopsis* genome were used as the reference for the TOR regulated genes (Ren *et al.* 2012). We identified 127 GO terms that were significantly overrepresented (corrected $P < 0.05$) among genes induced by TOR inhibition. Of these, 50 were also over represented in genes induced by *GmRPA3* silencing (Table 2). Overlapping gene ontology categories include four GO terms related to signalling, 14 GO terms related to stress, defense and immunity, three GO terms related to jasmonic acid and salicylic acid signalling, three GO terms related to responses to oxygen and nitrogen, 11 GO terms related to regulation, five GO terms related to transcription and one GO term related to aging. To confirm these patterns in greater detail, we used the program MAPMAN (Thimm *et al.* 2004) to identify genes related to signalling and defense. MAPMAN analysis found that members of the DUF26, LRRXI and LRRVII receptor-like kinases (RLKs) were significantly ($P < 0.05$) overrepresented among *GmRPA3* silencing induced genes. When we examine the expression of the 173 differentially expressed RLKs, we can see that greater than 85% of RLKs are induced by *GmRPA3* silencing (Fig. 7). Similarly, when we examine GO:0050832 (defense response to fungus), we see that 98% of differentially expressed genes were induced in response to *GmRPA3* silencing (Fig. 7). Interestingly, Bao, Yang & Hua (2013), recently found that perturbation of the cell cycle triggers plant immunity by activating resistance genes.

We also compared GO terms repressed by TOR inhibition and *GmRPA3* silencing. While 27 GO terms were overrepresented in the TOR data set, only a single GO term (GO:005514, oxidation reduction process) was shared between genes repressed by TOR inhibition or *GmRPA3* silencing. However, Ren *et al.* (2012) associated TOR inhibition with repression of photosynthesis and decreased protein synthesis. In the GO analysis of genes repressed by *GmRPA3* silencing the GO terms photosynthesis (GO:0015979), photosynthetic electron transport chain (GO:0009767), plastid organization (GO:0009657), photosynthesis light reaction (GO:0019684), chlorophyll binding (GO:0016168) were significantly overrepresented among *GmRPA3* silencing repressed genes (Fig. 5, Supporting Information Tables S9 & S10). Similarly, genes associated with protein folding (GO:0006457) and protein complex biogenesis (GO:0070271) were also significantly overrepresented among genes repressed by *GmRPA3* silencing. Using the MAPMAN software (Thimm *et al.* 2004), we identified differentially expressed genes associated with photosynthetic electron transport, protein synthesis and protein degradation. All 40 differentially expressed genes associated with photosynthetic electron transport were repressed in response to *GmRPA3* silencing (Fig. 7). While the expression of genes associated with protein synthesis decreased in response to *GmRPA3* silencing, genes associated with protein degradation had increased expression in response to *GmRPA3* silencing. The net effect of these processes results in decreased protein levels.

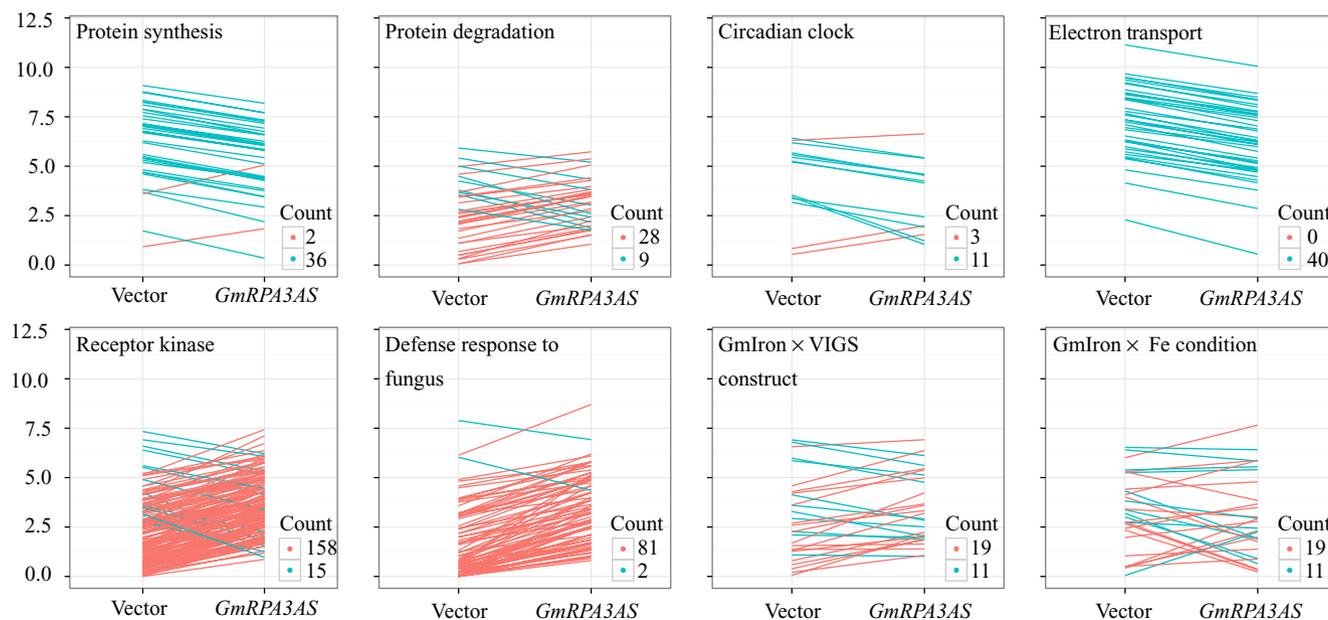


Figure 7. Silencing of *GmRPA3* alters the expression of genes involved protein synthesis and degradation, circadian clock, defense, photosynthesis and iron uptake and mobilization. To demonstrate that *GmRPA3* silencing mirrors TOR inhibition, we examined gene expression changes in plant pathways that are differentially expressed in response to TOR inhibition. We used the program MAPMAN (Thimm *et al.* 2004) to identify receptor kinases and genes associated with protein synthesis and degradation, photosynthesis and defense. Genes associated with the circadian clock were identified based on homology to circadian clock genes in Arabidopsis (TAIR version 10). Genes associated with iron regulation and uptake was identified based on homology to Arabidopsis iron genes (Kobayashi & Nishizawa 2012). For each panel, the y-axis is the \log_2 (normalized counts per million + 1). For the first six panels, genes induced in response to *GmRPA3* silencing are coloured pink while repressed genes are coloured blue. To study the effect of *GmRPA3c* silencing on iron regulation and uptake genes, we examined these genes looking at the both the VIGS construct effect (GmIron × VIGS construct) and iron availability effect (GmIron × Fe condition). If homologs of known iron genes were differentially expressed in response *GmRPA3c* but were not affected by iron availability, we would expect the lines plotted in GmIron × Fe condition would match lines plotted in GmIron × VIGS constructs. Differences in these two plots indicate that silencing of *GmRPA3c* can affect iron responsiveness. For additional data, refer to Supporting Information Tables S6, S7 and S8.

TOR inhibition also altered growth phase and delayed development and maturity. While no circadian clock genes were identified by Ren *et al.* (2012), our analyses of *GmRPA3* silenced plants identified a number of differentially expressed circadian clock associated genes including homologs of *Cryptochrome 3* (*CRY3*, Glyma04g07870), response regulators *ARR4* (Glyma04g29250) and *GLK1* (Glyma06g443330) and pseudo response regulator *PRR5* (Glyma16g02050), *Early flowering 3* (*ELF3*, Glyma07g01600 and Glyma08g21110), *Timeless* (*ATIM*, Glyma10g37220), *CGA1* (Glyma17g06290 and Glyma17g34670) and *Timing of CAB expression 1* (*TOC1*, Glyma17g11040). The expression of most of these genes was repressed (Fig. 7). In *Arabidopsis*, *ELF3* is required for photoperiodic flowering and normal circadian regulation (Hicks, Albertson & Wagner 2001). Similarly, *TOC1* mutants have altered circadian rhythms and flowering time (Somers, Webb & Kay 1998). In *Drosophila*, TOR signalling affects the timing of nuclear accumulation of TIMELESS (Zheng & Sehgal 2010), suggesting TOR is also involved in regulating components of the circadian clock.

With these results in mind, we developed a model (Fig. 8) to explain the role of GmRPA3 and DNA replication in the Clark iron deficiency response. TOR and SnRK1 act antagonistically to regulate cellular homeostasis and growth

(Robaglia, Thomas & Meyer 2012). While TOR is activated by nutrient availability, SnRK1 is activated by nutrient stress. Under nutrient limiting conditions, SnRK1 is activated to slow growth and induce stress and defense responses. In mammalian systems, the SnRK1 ortholog, AMPK, inactivates RAPTOR by phosphorylation (Gwinn *et al.* 2008). A similar system is likely in plants. Inactivation of RAPTOR inhibits TOR, S6K1 and eventually E2F activity (Henriques *et al.* 2010). A lack of E2F available to bind to promoters of replication genes, including *GmRPA3*, results in growth inhibition. We hypothesize that in the iron efficient soybean line Clark, iron deficiency activates SnRK1, which results in inhibition of RAPTOR, TOR, S6K, E2F and eventually DNA replication, which reduces growth until iron becomes available. Silencing *GmRPA3* in the iron inefficient line IsoClark bypasses the RAPTOR, TOR, S6K and E2F phosphorylation relay, but still inhibits DNA replication and induces a TOR-like response. Since E2F is one of several downstream TOR targets, we would expect genes differentially expressed in response to *GmRPA3* silencing to be a subset of those differentially expressed in response to TOR inhibition. Since this response is activated only in the iron efficient line Clark, it suggests Clark recognizes iron limitation and activates nutrient starvation responses.

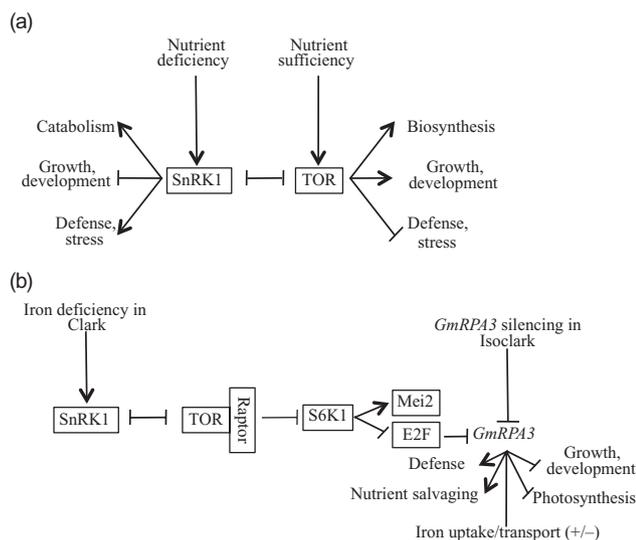


Figure 8. Model for the regulation of the SnRK1/TOR pathway in response to iron stress in Clark. (a) The SnRK1/TOR pathway is regulated by nutrient availability. Nutrient sufficiency regulates TOR, promoting biosynthesis and growth. In contrast, nutrient deficiency regulates SnRK1 promoting nutrient recycling, defense gene expression and inhibition of growth and development. (b) In mammalian systems AMPK, the ortholog of SnRK1, can phosphorylate RAPTOR, inactivating the TOR complex (Smeekens *et al.* 2010). It is assumed a similar mechanism operates in plants. In this experiment, we silenced *GmRPA3c* in the iron inefficient soybean NIL IsoClark to mirror expression found in the iron efficient line Clark. This bypassed SnRK1 activation and TOR deactivation, and resembled the results found for TOR inactivation (Ren *et al.* 2012). This suggests that iron deficiency in Clark activates SnRK1-mediated signalling.

The similar gene expression patterns and phenotypes of *GmRPA3*-silenced plants and TOR inhibited *Arabidopsis* plants suggest the SnRK1/TOR pathway promotes plant survival during iron deficiency conditions. Therefore, we examined the expression of iron regulation, uptake and translocation genes (reviewed by Kobayashi & Nishizawa 2012) in the *Arabidopsis* TOR inhibited plants (Ren *et al.* 2012) and *GmRPA3*-silenced plants. In TOR inhibited *Arabidopsis* plants grown in nutrient sufficient conditions, the iron regulation genes *BHLH100* and *BHLH38* were induced relative to wild type, as were the iron uptake genes *NAS4*, *YSL1*, *YSL7*, *AHA4*, *OPT3* and *NRAMP4*. However, the iron uptake genes *IRT1*, *FRO2*, *FRO6*, *FRO7* and *NAS2* were repressed. In *GmRPA3*-silenced plants we observed induction of *OPT3* (Glyma15g06510 and Glyma19g32400), *AHA* (Glyma09g06170 and Glyma09g06250), *YSL1* (Glyma10g36100, Glyma20g16600 and Glyma20g35980) homologs and repression of homologs of *VIT* (Glyma08g05230), *IREG* (Glyma10g28760), *ILR3* (Glyma12g34300 and Glyma1506860) and *FER4* (Glyma18g02800), regardless of iron availability (Supporting Information Table S6, Fig. 7). Other genes lost or gained iron responsiveness in *GmRPA3* silenced plants (Supporting Information Tables S7 & S8, Fig. 7). For example Glyma03g10790, a homolog of *IRT1*, was induced by iron stress, but only in *GmRPA3*-silenced plants. An additional twelve

iron-related genes were either significantly more or less iron responsive in *GmRPA3* silenced plants than empty vector plants. These included homologs of *FER1* (Glyma02g43040, Glyma03g06420, Glyma07g19060 and Glyma14g06160), *OPT3* (Glyma07g39780 and Glyma17g01000), *FRO2* (Glyma18g47060), *VITI* (Glyma16g28340) and *YSL3* (Glyma06g13820).

Vert, Briat & Curie (2003) proposed two models for understanding how iron deficiency is sensed in the leaf and transmitted to the root. In the promotive model, iron deficiency in the shoot induces the expression of iron acquisition and uptake genes. In the repressive model, iron sufficient conditions in the shoot repress iron acquisition and uptake genes in the root. Ren *et al.* (2012) found that inhibition of TOR activity repressed growth of primary and lateral roots and also root hairs. They hypothesized that TOR acts as a nutrient sensor by regulating the development of the root system and related functions. We began our analysis with an interest in understanding how the regulation of DNA replication and specifically *GmRPA3* in the leaves is related to iron deficiency in soybean. Our model predicts that SnRK1 recognizes iron deficiency in Clark and inhibits the TOR signalling pathway and *GmRPA3*. Our analysis and the Ren *et al.* (2012) data, both found that the expression of iron uptake and transport genes was altered by *GmRPA3* silencing and TOR inhibition. Taken together, these data suggest that the TOR/SnRK1 pathway and by association, *GmRPA3*, are involved in signaling iron deficiency from the shoot to the root. Future experiments will tease apart the pathway controlling *GmRPA3* expression.

What is the benefit of slowing growth in the face of nutrient deficiency? The relationship between nutrient availability, growth and yield is clear when surveying the available soybean germplasm. In iron sufficient environments, soybean growers prefer IDC susceptible lines because they offer higher yields than IDC resistant lines. IDC resistant lines, such as Clark, only yield higher when usable iron is in short supply. Research has also demonstrated that even after recovery from IDC there is a long-term effect on yield (Froehlich & Fehr 1981). Our study demonstrates that *RPA* plays a vital role in the iron stress response of soybean. Reducing *RPA3* gene expression in the iron inefficient line IsoClark resulted in stunted plants with improved iron deficiency symptoms, confirming the importance of *RPA3* in the iron efficient soybean lines response to iron deficiency. Controlling *RPA* gene expression, likely via the SnRK1/TOR pathway, allows iron efficient soybeans to regulate, even delay, energetically expensive processes such as reproduction until adequate resources are available. The inhibition of plant growth in response to iron deficiency, and possibly other stresses, may explain the reduced yield observed in IDC resistant lines. Understanding this response will lead to the development of better soybeans and other crops for the future.

ACKNOWLEDGMENTS

We would like to thank David Grant and Carroll Vance for their helpful discussions and insights. The authors are

grateful for the financial support from the United Soybean Board, North Central Soybean Research Program, the Iowa Soybean Association, the NSF Plant Genome Research Program (award number 0820642) and the USDA-Agricultural Research Service.

REFERENCES

- Abouyou P, Pages H. & Lawrence M. (2013) GenomicRanges: representation and manipulation of genomic intervals (R package version 1.4.8). Software. Available from: <http://www.bioconductor.org/packages/release/bioc/html/GenomicRanges.html>.
- Agarwal P, Agarwal P.K., Joshi A.J., Sopory S.K. & Reddy M.K. (2010) Over expression of *PgDREB2A* transcription factor enhances abiotic stress tolerance and activates downstream stress-responsive genes. *Molecular Biology Reports* **37**, 1125–1135.
- Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. & Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.
- Apweiler R., Bairoch A., Wu C.H., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R. & Magrane M. (2004) UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **32**, D115–D119.
- Bao Z., Yang H. & Hua J. (2013) Perturbation of cell cycle regulation triggers immune response via activation of disease resistance genes. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2407–2412.
- Bauer S., Grossmann S., Vingron M. & Robinson P.N. (2008) Ontologizer 2.0 – a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* **24**, 1650–1651.
- Berardini T.Z., Mundodi S., Reiser L., et al. (2004) Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiology* **135**, 745–755.
- Binz S.K., Sheehan A.M. & Wold M.S. (2004) Replication protein A phosphorylation and the cellular response to DNA damage. *DNA Repair* **3**, 1015–1024.
- Blanchet E., Annicotte J.S., Lagarrigue S., et al. (2011) E2F transcription factor-1 regulates oxidative metabolism. *Nature Cell Biology* **13**, 1146–1152.
- Brill S.J. & Stillman B. (1991) Replication factor-A from *Saccharomyces cerevisiae* is encoded by three essential genes coordinately expressed at S phase. *Genes & Development* **5**, 1589–1600.
- Buckhout T.J., Yang T.J.W. & Schmidt W. (2009) Early iron-deficiency-induced transcriptional changes in *Arabidopsis* roots as revealed by microarray analyses. *BMC Genomics* **10**, 147.
- Chambers J.M., Freeny A. & Heiberger R.M. (1992) Analysis of variance; designed experiments. In *Statistical Models in S* (eds J.M. Chambers & T.J. Hastie), pp. 145–193. Wadsworth & Brooks/Cole, Pacific Grove, CA.
- Chaney R.L., Coulombe B.A., Bell P.F. & Angle J.S. (1992) Detailed method to screen dicot cultivars for resistance to Fe-chlorosis using FeDTPA and bicarbonate in nutrient solutions. *Journal of Plant Nutrition* **15**, 2063–2083.
- Chang Y., Gong L., Yuan W., Li X., Chen G., Li X., Zhang Q. & Wu C. (2009) Replication protein A (*RPA1a*) is required for meiotic and somatic DNA repair but is dispensable for DNA replication and homologous recombination in rice. *Plant Physiology* **151**, 2162–2173.
- Daniely Y. & Borowiec J.A. (2000) Formation of a complex between nucleolin and replication protein A after cell stress prevents initiation of DNA replication. *Journal of Cell Biology* **149**, 799–810.
- Devaiah B.N., Karthikeyan A.S. & Raghothama K.G. (2007) WRKY75 transcription factor is a modulator of phosphate acquisition and root development in *Arabidopsis*. *Plant Physiology* **143**, 1789–1801.
- Din S., Brill S.J., Fairman M.P. & Stillman B. (1990) Cell-cycle-regulated phosphorylation of DNA replication factor A from human and yeast cells. *Genes & Development* **4**, 968–977.
- Ding Z., Li S., An X., Liu X., Qin H. & Wang D. (2009) Transgenic expression of *MYB15* confers enhanced sensitivity to abscisic acid and improved drought tolerance in *Arabidopsis thaliana*. *Journal of Genetics and Genomics* **36**, 17–29.
- Dobrenel T., Marchive C., Sormani R., Moreau M., Mozzo M., Motané M., Menand B., Robaglia C. & Meyer C. (2011) Regulation of plant growth and metabolism by the TOR kinase. *Biochemical Society Transactions* **39**, 447–481.
- Durbin R., Eddy S., Krogh A. & Mitchison G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge, UK.
- Elmayan T., Proux F. & Vaucheret H. (2005) *Arabidopsis RPA2*: a genetic link among transcriptional gene silencing, DNA repair and DNA replication. *Current Biology* **15**, 1919–1925.
- Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- Frith M.C., Fu Y., Yu L., Chen J., Hansen U. & Weng Z. (2004) Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Research* **32**, 1372–1381.
- Froehlich D.M. & Fehr W.R. (1981) Agronomic performance of soybeans with differing levels of iron-deficiency chlorosis on calcareous soil. *Crop Science* **21**, 438–441.
- Guerinot M.L. & Yi Y. (1994) Iron: nutritious, noxious and not readily available. *Plant Physiology* **104**, 815–820.
- Gwinn D.M., Shackelford D.B., Egan D.F., Mihaylova M.M., Mery A., Vasquez D., Turk B.E. & Shaw R.J. (2008) AMPK phosphorylation of Raptor mediates a metabolic checkpoint. *Molecular Cell* **30**, 214–226.
- Hansen N.C., Schmitt M.A., Anderson J.E. & Stroock J.S. (2003) Iron deficiency of soybean in the upper Midwest and associated soil properties. *Agronomy Journal* **95**, 1595–1601.
- Hansen N.C., Jolley V.D., Naeve S.L. & Goos R.J. (2004) Iron deficiency of soybean in the north central U.S. and associated soil properties. *Soil Science & Plant Nutrition* **50**, 983–987.
- Hass C.S., Lam K. & Wold M.S. (2011) Repair-specific functions of replication protein A. *Journal Biological Chemistry* **287**, 3908–3918.
- Henriques R., Magyar Z., Monardes A., Khan S., Zaleski C., Orellana J., Szabados L., de la Torre C., Koncz C. & Bögre L. (2010) *Arabidopsis* S6 kinase mutants display chromosome instability and altered RBR1-E2FB pathway activity. *EMBO Journal* **29**, 2979–2993.
- Hey S.J., Byrne E. & Halford N.G. (2010) The interface between metabolic and stress signaling. *Annals of Botany-London* **105**, 197–203.
- Hicks K.A., Albertson T.M. & Wagner D.R. (2001) *EARLY FLOWERING3* encodes a novel protein that regulate circadian clock function and flowering in *Arabidopsis*. *The Plant Cell* **13**, 1281–1292.
- Hudson D., Guevara D., Yaish M.W., Hannam C., Long N., Clarke J.D., Bi Y. & Rothstein S.J. (2011) GNC and CGA1 modulate chlorophyll biosynthesis and glutamate synthase (GLU1/Fd-GOGAT) expression in *Arabidopsis*. *PLoS ONE* **6**, e26765.
- Ishibashi T., Kimura S., Furukawa T., Hatanaka M., Hashimoto J. & Sakaguchi K. (2001) Two types of replication protein A 70 kDa subunit in rice, *Oryza sativa*: molecular cloning, characterization and cellular & tissue distribution. *Gene* **272**, 335–343.
- Ishibashi T., Koga A., Yamamoto T., Uchiyama Y., Mori Y., Hashimoto J., Kimura S. & Sakaguchi K. (2005) Two types of replication protein A in seed plants. *FEBS Journal* **272**, 3270–3281.
- Ishibashi T., Kimura S. & Sakaguchi K. (2006) A higher plant has three different types of RPA heterotrimeric complex. *Journal Biochemistry* **139**, 99–104.
- Kapoor A., Agarwal M., Andreucci A., Zheng X., Gong Z., Hasegawa P.M., Bressan R.A. & Zhu J.K. (2005) Mutations in a conserved replication protein suppress transcriptional gene silencing in a DNA-methylation-independent manner in *Arabidopsis*. *Current Biology* **15**, 1912–1918.
- Kim K., Dimitrova D.D., Carta K.M., Saxena A., Daras M. & Borowiec J.A. (2005) Novel checkpoint response to genotoxic stress mediated by nucleolin-replication protein A complex formation. *Molecular and Cellular Biology* **25**, 2463–2474.
- Kobayashi T. & Nishizawa N.K. (2012) Iron uptake, translocation, and regulation in higher plants. *Annual Review of Plant Biology* **63**, 131–152.
- Lawrence M., Gentleman R. & Carey V. (2009) rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842.
- Libault M., Farmer A., Brechenmacher L., et al. (2010a) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiology* **152**, 541–552.
- Libault M., Farmer A., Joshi T., Takahashi K., Langley R.J., Franklin L.D., He J., Xu D., May G. & Stacey G. (2010b) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *The Plant Journal* **63**, 86–99.
- Lin S., Cianzio S. & Shoemaker R. (1997) Mapping genetic loci for iron deficiency chlorosis in soybean. *Molecular Breeding* **3**, 219–229.

- Lin S.F., Baumer J.S., Ivers D., Cianzio S.R. & Shoemaker R.C. (1998) Field and nutrient solutions tests measure similar mechanisms controlling iron deficiency chlorosis in soybean. *Crop Science* **38**, 254–259.
- Liu J., Horstman H., Braun E., et al. (2011) Soybean homologs of MPK4 negatively regulate defense responses and positively regulate growth and development. *Plant Physiology* **157**, 1363–1378.
- Liu S., Kandath P.K., Warren S.D., et al. (2012) A soybean cyst nematode gene points to a new mechanism of plant resistance to pathogens. *Nature* **492**, 256–260.
- Lorenzo O., Piqueras R., Sánchez-Serrano J.J. & Solano R. (2003) ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *The Plant Cell* **15**, 165–178.
- McCarthy D.J., Chen Y. & Smyth G.K. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**, 4288–4297.
- McLean E., Cogswell M., Egli I., Wojdyla D. & de Benoist B. (2008) Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993–2005. *Public Health Nutrition* **12**, 444–454.
- Maddocks O.D.K. & Vousden K.H. (2011) Metabolic regulation by p53. *Journal of Molecular Medicine* **89**, 237–245.
- Mariconti L., Pellegrini B., Cantoni R., Stevens R., Bergounioux C., Cella R. & Albani D. (2002) The E2F family of transcription factors from *Arabidopsis thaliana*. Novel and conserved components of the retinoblastoma/E2F pathway in plants. *Journal of Biological Chemistry* **277**, 9911–9919.
- Matys V., Kel-Margoulis O.V., Fricke E., et al. (2006) TRANSFAC and its module TRANSCompel@: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research* **34**, D108–D110.
- Mayer J.E., Pfeiffer W.H. & Beyer P. (2008) Biofortified crops to alleviate micronutrient malnutrition. *Current Opinion in Plant Biology* **11**, 166–170.
- Meyer J.D.F., Silva D.C.G., Yang C., et al. (2009) Identification and analyses of candidate genes for *Rpp4*-mediated resistance to Asian soybean rust in soybean. *Plant Physiology* **150**, 295–307.
- Mizoi J., Shinozaki K. & Yamaguchi-Shinozaki K. (2012) AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta* **1819**, 86–96.
- Moreau M., Sormani R., Menand B., Veit B., Robaglia C. & Meyer C. (2010) The TOR complex and signaling pathway in plants. *Enzymes* **27**, 285–302.
- Morgan M. & Pages H. (2010) Rsamtools: binary alignment (BAM), variant call (BCF) or tabix file import (R package version 1.4.3). Software. Available from: <http://bioconductor.org/packages/release/bioc/html/Rsamtools.html>.
- Mukherjee K., Choudhury A.R., Gupta B., Gupta S. & Sengupta D.N. (2006) An ABRE-binding factor, *OSBZ8*, is highly expressed in salt tolerant cultivars than in salt sensitive cultivars of indica rice. *BMC Plant Biology* **6**, 18.
- Nei M. & Kumar S. (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- O'Rourke J.A., Graham M.A., Vodin L., Gonzalez D.O., Cianzio S.R. & Shoemaker R.C. (2007) Recovering from iron deficiency chlorosis in near-isogenic lines of soybeans: a microarray study. *Plant Physiology Biochemistry* **45**, 287–292.
- O'Rourke J.A., Nelson R.T., Grant D., Schmutz J., Grimwood J., Cannon S., Vance C.P., Graham M.A. & Shoemaker R.C. (2009) Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genomics* **10**, 376.
- Osnato M., Stile M.R., Wang Y., et al. (2010) Cross talk between the KNOX and ethylene pathways is mediated by intron-binding transcription factors in barley. *Plant Physiology* **154**, 1616–1632.
- Pandey A.K., Yang C., Zhang C., Graham M.A., Horstman H.D., Lee Y., Zabolina O.A., Hill J.H., Pedley K.F. & Whitham S.A. (2011) Functional analysis of the Asian soybean rust resistance pathway mediated by *Rpp2*. *Molecular Plant-Microbe Interactions* **24**, 194–206.
- Pfeiffer G.A., King K.E., Severin A.J., May G.D., Cianzio S.R., Lin S.F., Lauter N.C. & Shoemaker R.C. (2012) Identification of candidate genes underlying an iron efficiency QTL in soybean. *Plant Physiology* **158**, 1745–1754.
- Pfuetzner R.A., Bochkarev A., Frappier L. & Edwards A.M. (1997) Replication protein A: characterization and crystallization of the DNA binding domain. *Journal of Biological Chemistry* **272**, 430–434.
- Phillips M.A., D'Auria J.C., Luck K. & Gershenzon J. (2009) Evaluation of candidate reference genes for real-time quantitative PCR of plant samples using purified cDNA as template. *Plant Molecular Biology Reporter* **27**, 407–416.
- R Development Core Team (2006) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, ISBN 3-900051-07-0.
- Ramsey F.L. & Schafer D.W. (2002) *The Statistical Sleuth. A Course in Methods of Data Analysis* second edn, pp. 68–73. Duxbury Press, Pacific Grove, CA.
- Ren M., Venglat P., Qiu S., et al. (2012) Target of rapamycin signaling regulates metabolism, growth and life span in *Arabidopsis*. *The Plant Cell* **24**, 4850–4874.
- Rieu I. & Powers S.J. (2009) Real-time quantitative RT-PCR: design, calculations and statistics. *The Plant Cell* **21**, 1031–1033.
- Robaglia C., Thomas M. & Meyer C. (2012) Sensing nutrient and energy status by SnRK1 and TOR kinases. *Current Opinion in Plant Biology* **15**, 301–307.
- Robinson M.D. & Smyth G.K. (2007) Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881–2887.
- Robinson M.D. & Smyth G.K. (2008) Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, 321–332.
- Robinson M.D., McCarthy D.J. & Smyth G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.
- Rozen S. & Skaletsky H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds S. Krawetz & S. Misener), pp. 365–386. Humana Press, Totowa, NJ.
- Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.
- Sakaguchi K., Ishibashi T., Uchiyama Y. & Iwabata K. (2009) The multi-replication protein A (RPA) system – a new perspective. *FEBS Journal* **276**, 943–963.
- Schneider R. & Edwards R. (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE* **6**, e17288.
- Schmutz J., Cannon S.B., Schlueter J., et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183.
- Sengupta A., Peterson T.R. & Sabatini D.M. (2010) Regulation of the mTOR complex 1 pathway by nutrients, growth factors, and stress. *Molecular Cell* **40**, 310–322.
- Severin A.J., Woody J.L., Bolon Y.T., et al. (2010a) RNA-seq atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biology* **10**, 160.
- Severin A.J., Pfeiffer G.A., Xu W.W., et al. (2010b) An integrative approach to genomic introgression mapping. *Plant Physiology* **154**, 3–12.
- Shultz R.W., Tatineni V.M., Hanley-Bowdoin L. & Thompson W.F. (2007) Genome-wide analysis of the core DNA replication machinery in the higher plants *Arabidopsis* and rice. *Plant Physiology* **144**, 1697–1714.
- Singh S.K., Roy S., Choudhury S.R. & Sengupta D.N. (2010) DNA repair and recombination in higher plants: insights from comparative genomics of *Arabidopsis* and rice. *BMC Genomics* **11**, 443.
- Smekens S., Ma J., Hanson J. & Roland F. (2010) Sugar signals and molecular networks controlling plant growth. *Current Opinion in Plant Biology* **13**, 273–278.
- Somers D.E., Webb A.A. & Kay S.A. (1998) The short-period mutant, *toc1-1*, alters circadian clock regulation of multiple outputs throughout development in *Arabidopsis thaliana*. *Development* **125**, 484–494.
- Spiller S. & Terry N. (1980) Limiting factors in photosynthesis: II. Iron stress diminishes photochemical capacity by reducing the number of photosynthetic units. *Plant Physiology* **65**, 121–125.
- Takashi Y., Kobayashi Y., Tanaka K. & Tamura K. (2009) *Arabidopsis* replication protein A 70a is required for DNA damage response and telomere length homeostasis. *Plant Cell and Physiology* **50**, 1965–1976.
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739.
- Terry N. (1980) Limiting factors in photosynthesis: I. Use of iron stress to control photochemical capacity *in vivo*. *Plant Physiology* **65**, 114–120.
- Thimm O., Blaessing O., Gibon Y., Nagel A., Meyer S., Krüger P., Selbig J., Müller L.A., Rhee S.Y. & Stitt M. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal* **37**, 914–939.
- Thompson J.D., Higgins D.G. & Gibson T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.
- Trapnell C., Pachter L. & Salzberg S. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111.

- Vert G.A., Briat J. & Curie C. (2003) Dual regulation of the Arabidopsis high-affinity root iron uptake system by local and long-distance signals. *Plant Physiology* **132**, 796–804.
- Wang A., Libault M., Joshi T., Valliyodan B., Nguyen H.T., Xu D., Stacey G. & Cheng J. (2010) SoyDB: a knowledge database of soybean transcription factors. *BMC Plant Biology* **10**, 14.
- Wang Y., Yu K., Poysa V., Shi C. & Zhou Y. (2011) Selection of reference genes for normalization of qRT-PCR analysis of differentially expressed genes in soybean exposed to cadmium. *Molecular Biology Reporter* **39**, 1585–1594.
- Wickham H. (2009) Ggplot2: elegant graphics for data analysis. Use R! Series Springer New York, Inc.
- Winterbourn C.C. (1995) Toxicity of iron and hydrogen peroxide: the Fenton reaction. *Toxicology Letters* **83**, 969–974.
- Wold M.S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annual Review of Biochemistry* **66**, 61–92.
- Xia R., Wang J., Liu C., et al. (2006) ROR1/RPA2A, a putative replication protein A2, functions in epigenetic gene silencing and in regulation of meristem development in Arabidopsis. *The Plant Cell* **18**, 85–103.
- Yandell B.S. (1997) *Practical Data Analysis for Designed Experiments*, pp. 88–104. Chapman and Hall, Boca Raton, FL.
- Yin T., Cook D. & Lawrence M. (2012) ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biology* **13**, R77.
- Zhang C., Bradshaw J.D., Whitham S.A. & Hill J.H. (2010) The development of an efficient multipurpose bean pod mottle virus viral vector set for foreign gene expression and RNA silencing. *Plant Physiology* **153**, 52–65.
- Zhang C., Grosic S., Whitham S.A. & Hill J.H. (2012) The requirement of multiple defense genes in soybean Rsv1-mediated extreme resistance to Soybean mosaic virus. *Molecular Plant-Microbe Interactions* **25**, 1307–1313.
- Zheng X. & Sehgal A. (2010) AKT and TOR signaling set the pace of the circadian pacemaker. *Current Biology* **20**, 1203–1208.

Received 25 February 2013; received in revised form 9 May 2013; accepted for publication 28 May 2013

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Alignment of RPA1 subunits from *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Medicago truncatula* (Mt), *Ricinus communis* (Rc) and *Glycine max* (Gm) used for phylogenetic analysis. Alignment is based on the amino acid sequence of each predicted protein. Predicted pseudogenes from soybean were excluded from the analysis. Sequences were aligned using Pileup in GCG (Accelrys Inc., San Diego, CA, USA). Extraneous sequence was trimmed to include the most conserved regions.

Figure S2. Alignment of RPA2 subunits from *Arabidopsis thaliana*, *Oryza sativa*, *Medicago truncatula*, *Ricinus communis* and *Glycine max* used for phylogenetic analysis. Alignment is based on the amino acid sequence of each predicted protein. Sequences were aligned using Pileup in GCG (Accelrys Inc., San Diego, CA, USA). Extraneous sequence was trimmed to include the most conserved regions.

Figure S3. Alignment of RPA3 subunits *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Medicago truncatula* (Mt), *Ricinus communis* (Rc), and *Glycine max* (Gm) used for phylogenetic analysis. Alignment is based on the amino acid sequence of each predicted protein. Sequences were aligned using Pileup in GCG (Accelrys Inc., San Diego, CA, USA). Extraneous sequence was trimmed to include the most conserved regions.

Figure S4. RPA homologs are broadly expressed during soybean development. To examine the expression patterns of RPA homologs, we took advantage of the soybean RNA-Seq atlases described by Severin *et al.* (2010a, <http://soybase.org/soyseq/>) and Libault *et al.* (2010a, 2010b, <http://soykb.org/>). The Glyma identifiers of all eighteen soybean RPA homologs were used as queries; however, only 13 RPAs were expressed. A. Expression of RPA homologs in the Severin *et al.* (2010a) atlas, focused largely on above ground tissues. B. Expression of RPA homologs in the Libault *et al.* (2010a,b) atlases focused largely on below ground tissues. Genes are coded the same colour in each panel.

Figure S5. VIGS constructs *GmRPA3S* and *GmRPA3AS* reduce expression *GmRPA3c* and *GmRPA3d* relative to vector only controls. To confirm silencing of *GmRPA3c* and *GmRPA3d*, gene-specific primers specific to the 3' UTR of each gene were developed for qPCR of VIGS plants 21 days after VIGS treatment. RNA was pooled from three biological replicates of vector, *GmRPA3AS* and *GmRPA3S* treated Isoclark plants grown in iron sufficient and deficient conditions, 21 days after treatment. Replicates from different iron conditions were combined to give six replicates per vector. Silencing vector *GmRPA3AS* and *GmRPA3S* reduced *GmRPA3c* expression by an average of 12.0 and 7.3 fold, respectively. In contrast, *GmRPA3d* expression was reduced 3.3 and 2.0 fold, respectively. Each data point is the average of six replicates \pm standard error. Statistically significant reduction in expression relative to vector controls is indicated by an asterisk ($P < 0.05$).

Table S1. Primer sequences for *GmRPA* homologs. RT-PCR primers were designed for all RPA homologs using the program Primer 3 (Rozen & Skaletsky 2000). Primers were designed using the Primer 3 defaults, specifying an amplicon size (125–175 bp). Primers were designed based on coding sequences of RPA homologs (<http://www.phytozome.net>, Table 1). RPA coding sequences were compared to each other using BLASTN (Altschul *et al.* 1997, $E < 10E-30$), and only unique sequences were used in primer design in order to distinguish between homeologs located in duplicated genomic regions. Primers were tested on Clark and Isoclark total RNA harvested from an iron-insufficient bucket at 14 days post iron stress.

Table S2. Relative gene expression values of *GmRPA* homologs over three time points in iron efficient line Clark. Relative gene expression was determined by qPCR of the gene of interest in iron sufficient or deficient conditions. Relative gene expression values are presented as a ratio to the value at the same time point in iron sufficient conditions, averaged over three biological replicates, then log₂ transformed. Values above zero indicate greater expression in iron insufficient conditions, while values below zero indicate lesser expression in iron insufficient conditions. Log-transformed data was analysed for standard deviation and standard error. Differences in relative quantity were analysed with ANOVA (Chambers *et al.* 1992) and then Tukey's Honestly Significant Difference test (Yandell 1997) for pairwise comparisons, with a significance cut-off of 0.05. For fold change comparisons relative to 1 hps, standard error

was calculated using the equation $SEFC = \text{SQRT} \left(\frac{SD1 + SD2}{N} \right)$ (SEFC = Standard error of fold change comparison relative to time point 1, SD1 is the standard deviation of time point 1, SD2 is the standard deviation of time point 2 and N is the total number of samples compared).

Table S3. Relative gene expression values of *GmRPA* homologs over three time points in iron inefficient line Iso-clark. Relative gene expression was determined by qPCR of the gene of interest in iron sufficient or deficient conditions. Relative gene expression values are presented as a ratio to the value at the same time point in iron sufficient conditions, averaged over three biological replicates, then log transformed (base 2). Values above zero indicate greater expression in iron insufficient conditions, while values below zero indicate lesser expression in iron insufficient conditions. Log-transformed data were analysed for standard deviation and standard error. Differences in relative quantity were analysed with ANOVA (Chambers *et al.* 1992) and then Tukey's Honestly Significant Difference test (Yandell 1997) for pairwise comparisons, with a significance cut-off of 0.05. For fold change comparisons relative to 1 hps, standard error was calculated using the equation $SEFC = \text{SQRT} \left(\frac{SD1 + SD2}{N} \right)$ (SEFC = Standard error of fold change comparison relative to time point 1, SD1 is the standard deviation of time point 1, SD2 is the standard deviation of time point 2 and N is the total number of samples compared).

Table S4. Identification of significantly overrepresented transcription factor binding sites in the promoters of differentially expressed *RPA* homologs. Clover (Frith *et al.* 2004) was used in conjunction with the TRANSFAC transcription factor database (Matys *et al.* 2006) to identify transcription factor binding sites overrepresented in the promoters of the differentially expressed *RPA* homologs when compared to all promoters in the soybean genome. Analysis was limited to the plant transcription factors present in TRANSFAC. Promoter sequences were defined as 1000 base pairs upstream of the transcription start site.

Table S5. Genes involved in DNA replication are repressed in response to iron stress in the iron efficient line Clark. To determine the role of RPA during iron deficiency, we examined the expression of soybean homologs of known Arabidopsis replication and repair proteins under iron deficient conditions. Best reciprocal BLASTP was used to identify soybean proteins with homology to known DNA replication (Shultz *et al.* 2007) or repair (Singh *et al.* 2010) proteins. Identified soybean proteins were then queried against the O'Rourke *et al.* (2009) and Peiffer *et al.* (2012) data sets.

Table S6. Annotation of genes significantly (FDR < 0.05) differentially expressed between *GmRPA3c* silenced and empty vector treated plants.

Table S7. Genes significantly differentially expressed in response to iron deficiency in empty vector treated plants (FDR < 0.1), but not in *GmRPA3* silenced plants (FDR > 0.1).

Table S8. Genes significantly differentially expressed in response to iron deficiency in *GmRPA3* silenced plants (FDR < 0.1), but not in empty vector treated plants (FDR > 0.1).

Table S9. GO terms significantly overrepresented among genes induced by *GmRPA3* silencing. Overrepresented GO terms were identified using the Ontologizer 2.0 (Bauer *et al.* 2008) software using the Parent-Child-Union and Westfall-Young-Single-Step Corrections and 1000 replicates. All identified GO terms are indicated.

Table S10. GO terms significantly overrepresented among genes repressed by *GmRPA3* silencing. Overrepresented GO terms were identified using the Ontologizer 2.0 software (Bauer *et al.* 2008) using the Parent-Child-Union and Westfall-Young-Single-Step Corrections and 1000 replicates. All identified GO terms are indicated.

Table S11. Annotation of transcription factors significantly (FDR < 0.05) differentially expressed between *GmRPA3c* silenced and empty vector treated plants.