Mutations in *Argonaute5* Illuminate Epistatic Interactions of the *K1* and *I* Loci Leading to Saddle Seed Color Patterns in *Glycine max*

Young B. Cho, Sarah I. Jones, and Lila O. Vodkin

Plant Cell. Advance Publication Mar. 28, 2017; doi: 10.1105/tpc.17.00162

Corresponding author: Lila O. Vodkin, l-vodkin@illinois.edu.

Review timeline:

TPC2016-00834-RA	Submission received:	Nov. 5, 2016
	1st Decision:	Dec. 13, 2016 *manuscript declined*
TPC2017-00162-RA	Submission received:	Feb. 24, 2017
	1st Decision:	Mar. 16, 2917 *acceptance pending, sent to science editor*
	Final acceptance:	Mar. 28, 2017
	Advance publication:	Mar. 28, 2017

REPORT: (The report shows the major requests for revision and author responses. Minor comments for revision and miscellaneous correspondence are not included. The original format may not be reflected in this compilation, but the reviewer comments and author responses are not edited, except to correct minor typographical or spelling errors that could be a source of ambiguity.)

TPC2016-00834-RA 1st Editorial decision – *declined* Dec. 13, 2016

During the post-review consultation session, we agreed that if you could address the major point raised by the reviewers by new experiments, we would welcome a resubmission. This would be treated as a new submission, but we would attempt to use at least some of the same reviewers. Nevertheless, reviewers will be asked to assess as a new manuscript (i.e. are the claims fully supported by the data; do the results presented move the field forward), and not only whether previous reviewer comments have been addressed.

Overall, there were a number of positive comments about your work, particularly from Reviewer #2, but also Reviewer #1, who captured the strength of your work with their comment that "this is an interesting study that demonstrates how a recessive mutation can epistatically suppress a dominant locus (siRNA producer) and which is associated with phenotypic variation."

The major issue mentioned by both Reviewer 1 and Reviewer 3 is the lack of evidence demonstrating causality of the mutation in *AGO5b*. Both reviewers mention the ideal experiment, transgenic complementation, but other possibilities are proposed (a second allele, or high-resolution genetic data) that could better address this weakness. This journal does have the following text on our instructions for authors page:

"Generally, characterization of multiple, independent alleles is necessary to establish that a mutation (such as a T-DNA insertion or chemically induced mutation) is responsible for an observed phenotype, as opposed to an undetected, linked mutation. Complementation tests via transformation can be valuable, although their interpretation may be limited by the possibility of redundancy, and thus statements of proof should be limited to cases in which multiple independent alleles have been characterized." In other words, the concern of both reviewers is entirely justified and codified in our instructions. But, if you can directly address this concern with new data (and we fully recognize the challenges of some of the possible experiments in soybean), we would be quite interested to see your manuscript again. In addition, Reviewer #3 requests the characterization of the i^k allele to better understand why it is a modifier of the *K1* phenotype, which seems like a reasonable request as it would make your story more complete.

There was also a concern expressed by Reviewer 1 and Reviewer 3 and shared by our board members that the manuscript is overly long, particularly with respect to the number of figures, but with careful editing and consolidation of figures, this could be remedied without much difficulty.

[Reviewer comments shown below along with author responses.]

TPC2017-00162-RA Submission received Feb. 24, 2017

Reviewer comments on previously declined manuscript and **author responses:**

RESPONSE: On December 6, 2016, we received a decision letter for TPC2016-00834-RA for our manuscript entitled "A Mutation in an Argonaute Protein Explains the Epistatic Interaction of the *K* and *I* Loci which Control Chimeric Seed Color Patterns in *Glycine max*". A resubmission was encouraged if we could address the major point raised by reviewers with new data on additional alleles or mutations in order to establish causality of the *AGO5b* mutation as the *k1* gene that specifies the chimeric black saddle seed color phenotype. Thank you for this opportunity.

We are now re-submitting the manuscript under the slightly revised title: "Mutations in an Argonaute Protein Illuminate the Epistatic Interaction of the *K1* and *I* Loci which Control Chimeric Seed Color Patterns in *Glycine max*". In the new data presented in Figure 5, Supplemental Figure 4, and Supplemental Data File 5, we show the results of sequencing amplicons of the entire 6.2 kb *AGO5b* gene by next-generation sequencing for several independent spontaneous mutations that lead to the black saddle phenotype (*i*^{i}*, k1* genotype) as compared to their parent lines that have yellow seed coats (*i*^{i}*, K1* genotype). In addition to the Clark 18a black saddle mutation described in the previous manuscript, and which is a 129 bp deletion leading to premature protein termination of the AGO5b protein, we here show that a spontaneous mutation in the Lincoln cultivar background is a frameshift due to a small deletion, and another one in the Calland cultivar is a small deletion that includes the initiation codon of the *AGO5b* gene. In addition, we sequenced two Kurakake black saddle Japanese varieties. Kurakake is a parent source of the *k1* allele that was used to establish *K1* as a genetic locus in 1929. In total, there were predicted protein lesions in all six of the black saddle mutations or varieties and no lesions in the four *K1* parent or standard lines that were sequenced.

We believe that we have satisfied the criteria of the journal that "statements of proof should be limited to cases in which multiple independent alleles have been characterized" especially since four of the cases we show are clearly derived from independent, spontaneous mutations found in different genetic backgrounds at different locations and different times.

Reviewer #1:

1. The figure layout of this study needs to be heavily modified in a way that combines figures and takes advantage of panels (A, B, etc). Figure 1 and 2 should be combined. Figure 3, 4 and 5 should be combined and Figure 5 should be represented as a bar plot (or a ratio). Figure 7, 8, 9, 10 and 11 should be combined. The manuscript and discussion should also be condensed.

> **RESPONSE: As suggested by Reviewer 1, we have consolidated figures (and eliminated some) so that the present manuscript has only 7 figures (and two of these are from new data), whereas the original manuscript had 12 figures.**

2. It has not been proven that AGO5 "explains" the epistatic interaction as described in the title. It is by far the most logical explanation. However, in the absence of a second mutant allele or a rescue of the black to a yellow pigment using a wild-type AGO5 the authors must refrain from claiming causality. Please change the title accordingly and ensure that no mention of causality is described in the main text (e.g. line 408). I don't believe this diminishes the quality of this study or the contribution to the field by making these clarifications.

> **RESPONSE: As discussed above, we now have data showing lesions in the AGO5b protein with four cases derived from independent, spontaneous mutations found in different genetic backgrounds at different locations and different times which establishes it as the *K1* locus. However, we still changed the title from "explains" to a more general verb "illuminates".**

3. The siRNAs are presented as a single size class, but it is well established that siRNAs function as discrete size classes. The data presented do not distinguish 21 from 24 for example. Is there a reason? It seems knowing if a specific size class is dominant would help strengthen the link between AGO5 and kl given the known products of AGO5 in Arabidopsis.

RESPONSE: In previous articles (Tuteja et al., 2009 and Cho et al., 2013), we have discussed the size classes of the *CHS* siRNAs and this is summarized briefly on page 14, "The primary *CHS* siRNAs were lower in abundance but composed of a higher proportion of 22-nt small RNAs, whereas the more abundant secondary *CHS* siRNAs were primarily 21 nt".

Reviewer #2:

The gel in Fig 8A has a light ~500-bp band in the C18a k1 genotype. Why is that there? Is it Ago5a?

RESPONSE: That is a reasonable assumption since they are about the same size and AGO5a does not have a deletion. The band had variable appearances depending on the experiment (see the new Figure 7).

Supp. Fig. 2 the small RNA blot does not have a loading control.

RESPONSE: We have eliminated the blot shown in the previous Supplemental Figure 2 as it was superfluous, given the large amount of small RNA sequencing in the main manuscript.

Reviewer #3:

However, when looking at the pedigree of these lines (shown in Supplemental Figure 1), these lines appear in fact extremely distant and far from being isogenic. Moreover, the nature of the i^k locus (carried by line Clark 8) is totally uncharacterized … The i^k locus needs to be characterized.

RESPONSE: The Clark 8 (i^k *K1)* line was created by 6 generations of backcrossing with Clark as the recurrent parent which is standard for near isoline creation to homogenize to the recurrent parent background for all of these and other near isogenic lines released by the USDA into the germplasm collection. This description was in the legend but it was unclear in the boxes of the pedigree chart and we have added words to this effect in the chart to clarify this.

As to characterizing the i^k allele, we certainly agree and are working on this, along with descriptions of genomic resequencing and copy number variation by digital PCR in all of the other allelic types of the *I* locus including *I*, i^k and *i*. Genomic resequencing of the *I* alleles is not trivial because of the highly repeated nature of the *CHS* genes (over 6 at the well characterized i^i allele). Given the fact that the present manuscript is already long, inclusion of these data is not suggested and will also dilute from the focus on the *K1* locus.

The authors need to explain why they compare Clark 18a (i^i, k1) with Clark 8 (i^k, *K1*) instead of Clark (i^i, *K1*). In fact, given the phenotypes shown in Figure 1, wouldn't it be more straightforward to compare (*I*, *K1*) with (*I*, *k1*)?

RESPONSE: For genomic re-sequencing, we did compare the Clark (i^i, *K1*, yellow) to the Clark 18a (i^i, *k1*, black saddle) which is a spontaneous mutant found in Clark. However, if the question is why didn't we compare the RNA-Seq data of Clark (i^i, *K1*, fully yellow seed coats) to the black saddle chimeric seed coats of Clark 18a (i^i, *k1*), we have the following statement on page 7: "We compared the transcriptome data of the pigmented region of Clark 8 (*K1* allele, black saddle) to the pigmented region of Clark 18a (*k1* allele, black saddle) in order to minimize variation due to position on the seed coat while searching for genetic differences specific to the *K1* locus. Likewise the non-pigmented regions were compared between the two varieties." The expression of the flavonoid pathway genes or its products influence the downstream expression of a number of other genes (not unlike a transcription factor does) and comparing RNA-Seq from a fully yellow seed coat to a chimeric one (even one derived as a spontaneous mutation) could confound the interpretation. In addition, we wanted to show the variation in the small RNAs in the different chimeric tissues and to have both the small RNA and RNA-Seq derived from the same positional samples.

Segregation analysis should be performed on the F_2 of a cross between *k1* and *K1* lines to determine if the *k1* phenotype always segregates with the mutation in *AGO5b* (a very large number of F_2 plants need to be analyzed). This should be done with *k1* and *K1* lines that carry the same i^i or *I* allele and not using Clark 18a (i^i, *k1*) and Clark 8 (i^k, *K1*).

RESPONSE: Segregation from a cross, even with large numbers of F_2 plants analyzed, would still always be just a correlation because of the possibility of extremely close linkage. We do show in the new Figure 7, that the 129 bp

deletion of Clark 18a is found in isolines created to demonstrate the epistatic interaction of the *k1* allele with the dominant *I* allele.

Better than a correlation, transformation of *k1* by a copy of the wild-type *AGO5b* gene would definitely address whether *AGO5b* is *K1*."

RESPONSE: This has been addressed in response to Reviewer 1 above with finding multiple lesions in the *AGO5b* gene in independent lines.

The soybean genome has 20 *AGO* genes, including members of the *AGO1* family, which are usually involved in mRNA cleavage guided by 21-nt siRNA, and which are much highly expressed than *AGO5* members in seed coats. If the mutation in *AGO5b* is responsible for the *k1* phenotype, the fact that part of the seed still undergoes *CHS* silencing may be due to the action of these other *AGOs*. This seems a more reasonable hypothesis than saying that *AGO5* affects the spatial distribution of *CHS* siRNAs.

RESPONSE: We had discussed this in the manuscript that the AGO5b is not exerting the effect directly on the secondary *CHS* siRNAs that accumulate later but is likely exerting its effect earlier in development. This is summarized on page 18 "More likely, the need for functional AGO5 in mediating patterning may be manifested only very early in development of the seed coat either directly in the function of the low abundance primary *CHS* siRNAs or indirectly by altering pattern development through the action of other miRNAs or siRNAs."

We feel we have adequately addressed the concerns of the reviewers and reviewing editors in the revised manuscript. Finally, summarizing the manuscript:

Finally, summarizing the manuscript: First in Figures 1-2, we determined that the naturally occurring silencing system of the *I* (inhibitor) locus exhibits spatially regulated, pattern-specific *CHS* siRNA and chimeric pigment patterns on seed coats carrying the dominant I^i and I^k alleles. Then with Figures 3-4, by combining RNA-Seq, knowledge of map position, and structural variation from whole genome resequencing of a mutant line, we determined the molecular identity of the independent *K1* locus which modifies the spatial regulation of *CHS* siRNA production within seed coats. In Figures 5-6, we used amplicon sequencing of additional spontaneous mutations to definitively show that the *K1* locus encodes AGO5b, a member of the Argonaute family of proteins of which there are 20 in soybean (Figure 7). Thus, the function of AGO5b appears to be integral to the spatial distribution of the *CHS* siRNAs, thus explaining how the *k1* allele reverses the phenotype of the seed coat regions from yellow to pigmented, even in the presence of the normally dominant I^i or *I* alleles. The results help unravel the molecular basis for the genetic interactions of these two classical loci.

We feel these findings are of broad interest as there are relatively few known loci in non-model systems that shed light on the interaction of small RNA pathways controlling plant traits. Our delineation of a naturally occurring AGO5 mutant should also open up additional research on the function of this critical protein in the regulation of small RNA processes and control of plant traits.

TPC2017-00162-RA 1st Editorial decision – *acceptance pending* Mar. 16, 2017

We are pleased to inform you that your paper entitled "Mutations in an Argonaute Protein Illuminate Epistatic Interactions of the *K1* and *I* Loci Leading to Chimeric Seed Color Patterns in *Glycine max*" has been accepted for publication in *The Plant Cell*, pending a final minor editorial review by journal staff. At this stage, your manuscript will be evaluated by a Science Editor with respect to scientific content presentation, compliance with journal policies, and presentation for a broad readership.

---------- Reviewer comments:

Reviewer #1:

This revised manuscript takes into account the comments and suggestions of the reviewers and is improved as a result. The manuscript has been shortened and several figures consolidated making it more readable. The identification of four independent mutants supports the conclusion that *Ago5b* encodes the *K1* locus. I have no further comments.

Reviewer #2:

The authors have satisfactorily addressed the concerns I raised in my original review. The addition of the additional mutation that are spontaneous in nature provide strong evidence. This evidence could be strengthened with allelic complementation experiments, but the fact that so many spontaneous mutations have repeatedly occur is strong evidence. The condensed nature of the revised version is also appreciated.

Final acceptance from Science Editor **Mar. 28, 2017**