the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2022) **111,** 1701–1716 doi: 10.1111/tpj.15918

Rearrangement with the *nkd2* promoter contributed to allelic diversity of the *r1* gene in maize (*Zea mays*)

Hao Wu[1,†], Guosheng Li[2], Junpeng Zhan[2,‡], Shanshan Zhang[2], Brandon D. Beall[1,3], Ramin Yadegari[2] and Philip W. Becraft[1,3,*]

[1]*Genetics, Development and Cell Biology Department, Iowa State University, Ames, Iowa, USA,*
[2]*School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA and,*
[3]*Agronomy Department, Iowa State University, Ames, Iowa 50011, USA*

Received 6 May 2022; revised 13 June 2022; accepted 15 July 2022; published online 25 July 2022.
*For correspondence (e-mail becraft@iastate.edu).
[†]Present address: School of Integrative Plant Science, Cornell University, Ithaca, New York, 14853, USA
[‡]Present address: Donald Danforth Plant Science Center, St. Louis, Missouri, 63132, USA

SUMMARY

The maize *red1* (*r1*) locus regulates anthocyanin accumulation and is a classic model for allelic diversity; changes in regulatory regions are responsible for most of the variation in gene expression patterns. Here, an intrachromosomal rearrangement between the distal upstream region of *r1* and the region of *naked endosperm 2* (*nkd2*) upstream to the third exon generated a *nkd2* null allele lacking the first three exons, and the *R1-st* (*stippled*) allele with a novel *r1* 5′ promoter region homologous to 5′ regions from *nkd2-B73*. *R1-sc:124* (an *R1-st* derivative) shows increased and earlier expression than a standard *R1-g* allele, as well as ectopic expression in the starchy endosperm compartment. Laser capture microdissection and RNA sequencing indicated that ectopic *R1-sc:124* expression impacted expression of genes associated with RNA modification. The expression of *R1-sc:124* resembled *nkd2-W22* expression, suggesting that *nkd2* regulatory sequences may influence the expression of *R1-sc:124*. The *r1-sc:m3* allele is derived from *R1-sc:124* by an insertion of a *Ds6* transposon in intron 4. This insertion blocks anthocyanin regulation by causing mis-splicing that eliminates exon 5 from the mRNA. This allele serves as an important launch site for *Ac/Ds* mutagenesis studies, and two *Ds6* insertions believed to be associated with *nkd2* mutant alleles were actually located in the *r1* 5′ region. Among annotated genomes of teosinte and maize varieties, the *nkd2* and *r1* loci showed conserved overall gene structures, similar to the B73 reference genome, suggesting that the *nkd2-r1* rearrangement may be a recent event.

Keywords: aleurone, anthocyanin, inversion, kernel, *Zea mays*.

INTRODUCTION

The maize *red1* (*r1*) gene is a classic system for studying allelic variation. It encodes a helix–loop–helix transcription factor that, in combination with C1 (COLORED ALEURONE1), regulates spatial and temporal anthocyanin pigmentation (Ludwig & Wessler, 1990). Anthocyanin provides a convenient marker that has allowed the identification of numerous alleles, including structural variants, of *r1*. The products of *r1* gene family members are functional in most or all plant tissues, suggesting that allelic variation in pigmentation patterns is due to variation in gene regulatory sequences and expression (Goff et al., 1990; Ludwig et al., 1989). A summary of the homozygous phenotypes of *r1* alleles relevant to this study is presented in Table 1.

The *R1* gene has also been instrumental in the study and utilization of transposable elements. In particular, the *r1-sc::m3* allele (usually denoted *r1-m3*) is a useful marker for *Activator* (*Ac*) activity due to the presence of a non-autonomous *Dissociation* (*Ds*), specifically a *Ds6*, insertion, which renders the *r1* gene inactive. In the absence of *Ac* activity, homozygous *r1-sc::m3* kernels show a stable absence of anthocyanin pigmentation. The presence of an active *Ac* element elsewhere in the genome can catalyze the transposition of *Ds* elements, causing reversion of *R1* function, which is visible as sectors of anthocyanin pigmentation, allowing selection of *Ds* transposition events that can be screened in mutagenesis studies (Dellaporta & Moreno, 1994). The *Ac/Ds* project was a large-scale transposon mutagenesis effort that utilized several selectable

mutant genes, including *r1-sc::m3*, as donor sites for *Ds* transpositions (Ahern et al., 2009; Vollbrecht et al., 2010).

Chromosome rearrangements can result in changes in gene copy number, structure or expression and rearrangements involving regulatory regions may directly alter the location, timing or level of gene expression. Indeed, there is structural variation at the *r1* locus, some of which results from chromosomal rearrangements (Walker et al., 1995), which impacts the pattern of *R1* gene expression and, hence, anthocyanin pigmentation. Structural variants range from a simple transcription unit such as *R1-nj* (Dellaporta et al., 1988), to complex alleles containing multiple transcription units encoding R1 proteins (Eggleston et al., 1995, Panavas et al., 1999, Walker & Panavas, 2001). The well-studied standard *R-r* allele (*R-r:std*) pigments both seed and plant, and these components are genetically separable by mutation or recombination (Dooner & Kermicle, 1971; Stadler, 1946; Stadler & Emmerling, 1956). The *R-r* allele is a gene complex consisting of four components: the *P* (plant) component is associated with pigmentation of vegetative tissues, *q* is a truncated inactive region, and two duplicated *S* (seed) components, *S1* and *S2*, confer pigmentation to the aleurone (AL). *S1* and *S2* are arranged in opposite orientations separated by a 381-bp region (called the σ region) that apparently acts as a bidirectional promoter (Walker et al., 1995). The distance between *P* and *q* components is approximately 190 kbp and the distance between *q* and *S* is approximately 20 kbp (Walker et al., 1997). Remnants of *Dopia*-like elements suggest transposon activity induced rearrangements during the genesis of this allele (Walker et al., 1995).

The *R1-stippled* (*R1-st*) and *R1-marbled* (*R1-mb*) alleles are also each multi-gene complexes. *R1-st* contains four components, the *R1-sc* (*self-colored*) gene with a transposable element (*I-R*, inhibitor of *R1*, which may be responsible for the stippled kernel phenotype) inserted in the seventh exon, followed by three tandem duplications of *R1-nc* (near-colorless) (Eggleston et al., 1995). The *R1-marbled* (*R1-mb*) complex contains three genes, *Scm* (similar sequence as *R1-sc*) with a transposable element, *Shooter* (*Sho*), which is associated with the marbled kernel phenotype, and two tandem duplicated *Lcm* genes (Panavas et al., 1999). *R1-st* and *R1-mb* each have paramutagenic properties that can induce a meiotically heritable suppression of standard *R1* alleles when combined as heterozygotes (Brink, 1973; Giacopelli & Hollick, 2015; Panavas et al., 1999). Duplicated components of *R1* complexes have similar coding sequences but the promoter regions are variable and may be associated with temporal or tissue-specific expression patterns (Li et al., 2001).

While it is documented that gene duplications and rearrangements have contributed to much of the allelic diversity at the *r1* locus, and that variation among gene promoters is responsible for much of the variation in expression, the origins of *r1* promoters remain largely unknown. Here we show that the upstream regulatory region (i.e., promoter) of *R1-st* shares sequence homology with the region upstream to the third exon of the *naked endosperm2* (*nkd2*) gene. The *nkd2* gene encodes an indeterminate domain transcription factor that, together with its duplicated partner gene, *nkd1*, regulates multiple processes in endosperm development (Gontarek et al., 2016; Yi et al., 2015). The *nkd2* gene is linked to *r1*, approximately 1.1 Mb away in the B73 reference genome (Jiao et al., 2017). The *r1-sc:m3* allele was derived by a *Ds6* insertion into *R1-sc:124* (Alleman & Kermicle, 1993; Kermicle, 1980) and *R1-sc:124* is in turn a simplex allele derived from the complex *R1-st* allele (Ashman, 1960; McWhirter & Brink, 1962). Sequence analysis suggests that the 5′ promoter regions of the *R1-st* allele originated from localized sequence rearrangements that translocated *nkd2* promoter sequences to the *r1* locus, and that *r1* alleles derived from or related to *R1-st*, including *r1-sc::m3*, *R1-sc:124* and *R1-mb*, share this same rearrangement. This event eliminated *nkd2* function and likely altered the expression of *r1*.

Table 1 Summary of *r1* alleles used in this study

R1 allele	Phenotype of homozygous	Reference
R1-r (*R1-r:standard*)	Anthocyanin-pigmented aleurone; anthocyanin-pigmented plant tissues	Stadler, 1946; Stadler & Emmerling, 1956; Dooner & Kermicle, 1971
R1-g	Anthocyanin-pigmented aleurone; green plant tissues	Walker et al., 1997
R1-Navajo (*R1-nj*)	A patch of anthocyanin-pigmented aleurone on the crown; anthocyanin-pigmented plant tissues	Dellaporta et al., 1988
R1-sc:124	Strongly anthocyanin-pigmented aleurone	Ashman, 1960, McWhirter & Brink, 1962
R1-stippled (*R1-st*)	Stippled endosperm (sporadic anthocyanin-pigmented aleurone)	Eggleston et al., 1995
R1-marbled (*R1-mb*)	Marbled anthocyanin-pigmented aleurone	Panavas et al., 1999
r1-sc:m3	Unpigmented aleurone (yellow kernels); no anthocyanin-pigmented plant tissues	Kermicle et al., 1989; Alleman & Kermicle, 1993; Conrad & Brutnell, 2005; Vollbrecht et al., 2010

RESULTS

W22 variant lacked *nkd2* expression

We identified a *nkd1* mutant allele, *nkd1-Ds*, and surprisingly, *nkd2* transcript was not detected in this mutant line (Figure S1), which was counter to previously published results where the *nkd2* transcript level showed a compensatory rise in *nkd1* mutants (Yi et al., 2015). The *nkd1-Ds* allele was identified in the *Ac/Ds* project, described above (Ahern et al., 2009; Vollbrecht et al., 2010). To explore the basis of this anomalous result, we examined *nkd2* expression in the W22 inbred line that served as a genetic background for this project. Both yellow- and purple-kerneled W22 variants were used in this project. The purple carried the *R1-g* allele, while this particular yellow variant harbored the *r1-sc:m3* allele, which was stable due to a lack of *Ac* transposase activity (Alleman & Kermicle, 1993; Conrad & Brutnell, 2005; Kermicle et al., 1989; Vollbrecht et al., 2010). Total RNA was isolated from endosperms of *R1-g* and *r1-sc:m3* W22 variants that were 12 and 24 days after pollination (DAP), and semi-quantitative reverse transcription–polymerase chain reaction (semi-qRT-PCR) was performed. Expression of *nkd2* was detected in the purple variants but not in *r1-sc:m3* (Figure 1a). PCR amplification from genomic DNA using the same *nkd2* primer set indicates that the lack of product was not due to sequence polymorphism, and semi-qRT-PCR of GAPDH showed all RNA samples were intact.

Lack of *nkd2* expression is associated with disruption of the *nkd2* locus

To explore the basis for the lack of *nkd2* expression, PCR primers were designed to amplify genomic DNA of the *nkd2* locus, based on the W22 reference genome (Zm-W22-REFERENCE-NRGENE-2.0) (Springer et al., 2018). Sets of primers covering the entire transcribed region as well as 731 bases of 5′ promoter region all produced the expected amplification products in the W22 *R1-g* line (Figure 1b,c and S2). However, in W22 *r1-sc:m3*, two of the primer pairs produced different size products, whereas three of the primer pairs did not amplify any product. To explore unknown regions of *nkd2* in the W22 *r1-sc:m3* variant, thermal asymmetric interlaced (TAIL)-PCR was performed (Figure S2). We were able to extend the 3′ fragment (SeqN2F8/R8) in the 5′ direction, and extend the 5′ fragment (SN2F2/R2) in both the 3′ and 5′ directions.

Combining PCR amplicons and TAIL products we were able to assemble two contigs manually corresponding to the 5′ and 3′ regions of the *nkd2* locus. Each contig also contained insertions and/or deletions in the *r1-sc:m3* variant relative to the normal W22 reference genome (Springer et al., 2018). The 5′ contig was, in total, 2833 bp and

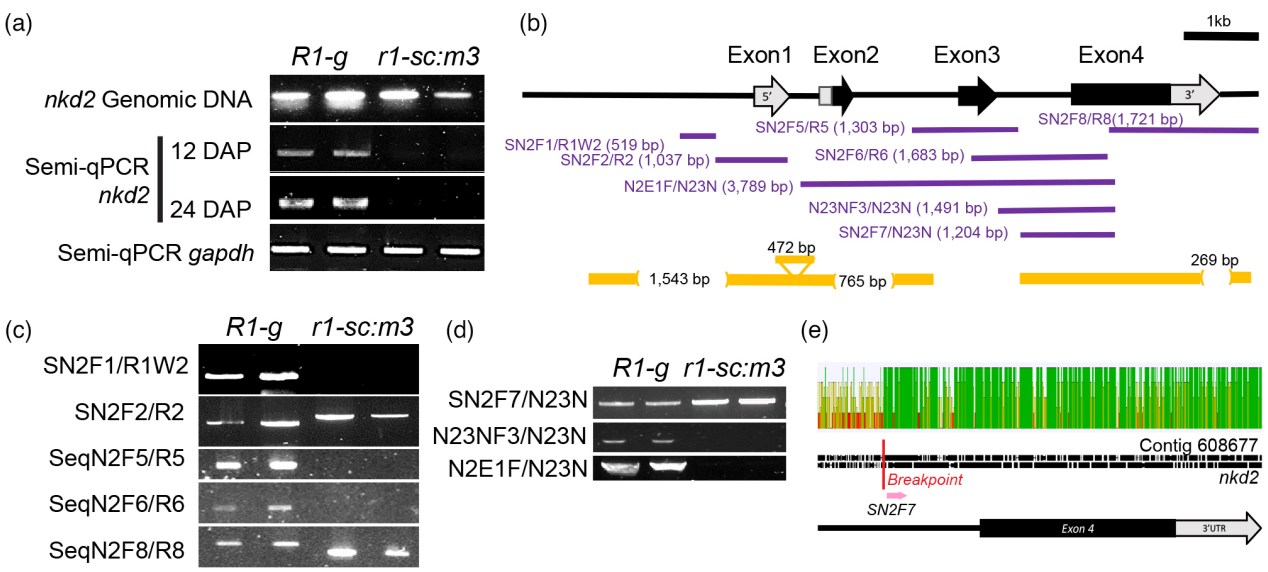

Figure 1. Sequence polymorphisms at the *nkd2* locus between purple (*R1-g*) and yellow (*r1-sc:m3*) W22 variants.
(a) No *nkd2* transcript is detectable in *r1-sc:m3* endosperm using primers that amplify a product from *R1-g* RNA or genomic DNA from either line. *GAPDH* control RT-PCR demonstrated intact RNA in each sample. DAP, days after pollination.
(b) Gene structure of *nkd2* in B73 RefGen_v4 and in W22 *R1-g*. Arrows represent intron, gray represents untranslated regions and black coding regions. Purple lines represent PCR products mentioned in (c,d). Orange lines represent sequenced regions of W22 *r1-sc:m3* mapped to the reference *nkd2* locus. Parentheses and triangle represent gaps and an insertion, respectively.
(c) Agarose gel showing different amplification patterns of 5 *nkd2* fragments in the two W22 variants.
(d) PCR tests of *nkd2* gene integrity suggest that sequences from the 5′ region are not contiguous with 3′ in the *r1-sc:m3* W22 variant.
(e) Nanopore long read assembly and alignments of contig #608677 to B73 *nkd2*. Color-coded histograms indicate nucleotide conservation, with green bar representing highly conserved regions and red/yellow bar representing diverse regions.

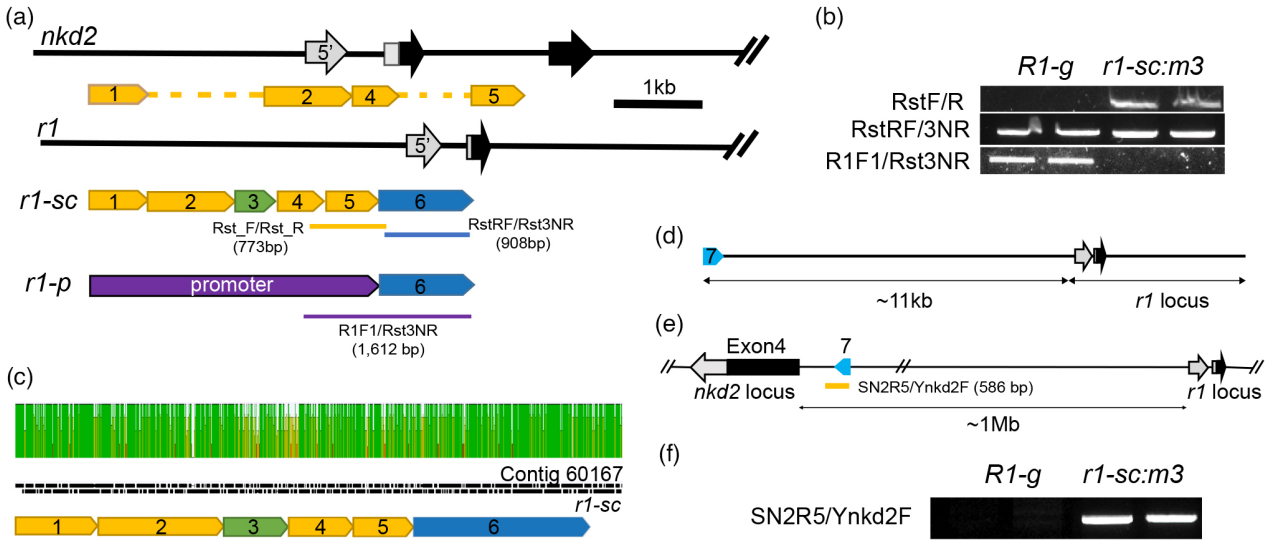

Figure 2. Chromosomal sequences rearranged between the *nkd2* and *r1* loci.
(a) Architecture of the upstream region of *nkd2-B73*, *R1-sc* and *R1-p*. The *R1-sc* upstream region showed six blocks based on homology with corresponding regions at *nkd2* and *R1-p*. Yellow blocks are found in both *nkd2* and *r1* alleles. Yellow, purple, and blue lines with primer names represent PCR amplicons to test the location, existence or integrity of corresponding regions.
(b) Agarose gel showing amplification products from the 5′ regions of different *r1* alleles.
(c) Alignment of Nanopore long read contig #60167 with *R1-sc* promoter and 5′ UTR. Color-coded histograms indicate nucleotide conservation, with green bars representing identity.
(d) The 5′ portion of contig #608677 (Figure 1e), that did not match *nkd2*, showed homology to sequences located about 11 kb upstream of *r1* in B73. This is designated block 7.
(e) Inferred chromosome structure in *R1-sc*.
(f) Agarose gel showing products of PCR with primers Ynkd2F and SeqN2R5 (shown in e) confirming the proximity of block 7 to the *nkd2* locus.

contained deletions of 1543 bp from upstream of the 5′ UTR and 765 bp from the middle of exon 2 to the middle of intron 2, as well as an insertion of 472 bp in intron 1 (Figure 1b,c). The 3′ contig spanned 1422 bp and included a deletion of 269 bases corresponding to the 3′ untranscribed region (Figure 1b,c).

The 5′ contig ended in the middle of intron 2 while the 3′ contig began in the middle of intron 3. Whereas PCR could readily amplify DNA between the contigs in W22 *R1-g*, this region could not be amplified from the *r1-sc:m3* variant (Figure 1d), suggesting these regions might not be contiguous. To explore further the structure of the *nkd2* gene in this variant, nanopore long-read sequencing and *de novo* assembly was performed. One nanopore contig (#608677) contained the 3′ region of *nkd2* from the middle of intron 3 to downstream of the 3′ UTR; however, the 5′ region of this contig did not match known genomic sequences associated with the *nkd2* locus (Figure 1e). Sequences associated with the 5′ portion of the *nkd2* gene were contained in a different contig (#60167; Figure 2); however, the 3′ end of this contig also did not match known *nkd2* sequence.

Sequences from the *nkd2* 5′ region are associated with the *r1-sc:m3* gene promoter region

Analysis of the *nkd2* locus suggested that the 5′ contig containing putative promoter sequences might not be contiguous with the 3′ contig containing exon 4, and furthermore, that each of these contigs contained sequences not associated with the normal *nkd2* gene. To explore the nature of each contig, the sequences were subjected to BLASTn performed against *Zea mays* (taxonomy ID: 4577) nucleotide sequences curated at NCBI. Surprisingly, the 5′ contig showed the strongest match (maximum score 2303; query coverage 97%) with promoter sequences of *R1-st* (accession number AF380388.1). The GenBank sequence is derived from the *sc* component of the complex *R1-st* allele (Li et al., 2001). *R1-sc:124* is a simplex allele, derived from *R1-st*, conferring strongly pigmented AL and scutellum (Ashman, 1960; Kermicle, 1980; McWhirter & Brink, 1962) and the *r1-sc:m3* allele was derived from *R1-sc:124* by insertion of a *Ds6* transposable element (Alleman & Kermicle, 1993; Kermicle, 1980). As such, we hypothesized that the 5′ contig was in fact associated with the *r1* locus rather than with *nkd2*. In the B73_v4 genome assembly, the *nkd2* and *r1* loci are in divergent orientations and separated by approximately 1.1 Mb. Figure 2a represents the blocks of sequence homology, numbered 1, 2, 4, and 5, as they are arranged in the *nkd2* gene compared with *R1-sc*. These blocks are variably dispersed in the *nkd2* location but directly adjacent to one another at *R1-sc*. The exception is blocks 2 and 4, which are adjacent at *nkd2* but separated by an insertion labeled "3" in *R1-sc*. This insertion of 471 bp has characteristics of non-long terminal repeat

retrotransposons and has been designated *Bnot* (Li et al., 2001). Region 6 was contained in the published *R1-st* promoter sequence but bore no homology to *nkd2*. It includes *r1* exon 1, intron 1 and part of exon 2, and matches sequences in the *R1 r-P* allele (accession number AF380390, Li et al., 2001).

To test the hypothesis that these sequences are in fact associated with *r1*, several PCR experiments were conducted (Figure 2b) based on the published W22 genome sequence (Springer et al., 2018). Primer pair RstF/R spans from block 4 (putatively derived from *nkd2*) to block 6 (affiliated with *r1*). In a W22 line carrying the standard *R1-g* allele, no amplification product was produced as expected for primer sites located at distant loci. However, in W22 *r1-sc:m3*, this primer pair produced a product indicating that in this genetic background these primer sites are contained in a contiguous region. The R1F1/Rst3NR primer pair spans from the *r1-P* promoter to block 6. As expected, this primer pair produced an amplification product in W22 *R1-g*; however, no product was produced from W22 *r1-sc:m3*. All PCR products were confirmed by sequencing.

To confirm further that *nkd2* 5′ sequences were contiguous with the *r1-sc:m3* allele, the nanopore long-read genomic sequences derived from W22 *r1-sc:m3* were aligned with the published *R1-st* promoter sequence. As shown in Figure 2c the long-read contig #60617 aligns throughout the entirety of the *R1-st* sequence, including both upstream regions as well as transcribed regions.

The association of *nkd2* sequences with the *r1* locus suggested that a chromosome rearrangement may have occurred. Consistent with this, when sequences from the 5′ end of long-read contig 608677 (Figure 1e), which did not match sequences from *nkd2-B73*, were used in a BLAST search of the maize genome, a match was detected to sequences located about 11 kb 5′ of the *r1* gene in the B73_v4 genome assembly (Figure 2d). Thus, sequences associated with the *r1* locus in B73 are associated with the *nkd2* in the *r1-sc:m3* line (block 7 in Figure 2e), and it appears that a chromosome rearrangement resulted in an exchange of DNA fragments between the *r1* and *nkd2* loci.

Re-evaluation of *nkd2-Ds* alleles

We previously reported the isolation of two *Ds*-containing *nkd2* mutants, *nkd2-Ds0766* and *-Ds0297*, identified from the *Ac/Ds* project (Yi et al., 2015). In light of the current findings, we reanalyzed these alleles and found the *Ds* insertions were located in rearranged *nkd2* sequences that were actually associated with the *r1* locus in this background (Figure 3a,b). Analysis of the original *Ds6* insertion site from the *R1-sc:m3* locus in each line showed that the *Ds6* element was no longer present but a footprint of the target site duplication remained (Figure 3c). Thus each of these events appears to represent a localized transposition that reinserted into the 5′ region of the donor locus. PCR using primers that spanned the breakpoint of the rearranged *nkd2* locus confirmed that both these lines contained the disrupted *nkd2* gene (Figure 3b).

R1 transcript levels are elevated in *R1-sc* alleles

As reported, the *Ds0766* line with a disrupted *nkd2* gene does not show *nkd2* expression (Yi et al., 2015). This line contained kernels pigmented with anthocyanin indicating that *R1* gene was functional (Figure 4a). Sequencing of RNA (RNA-seq) data generated from the 16 DAP

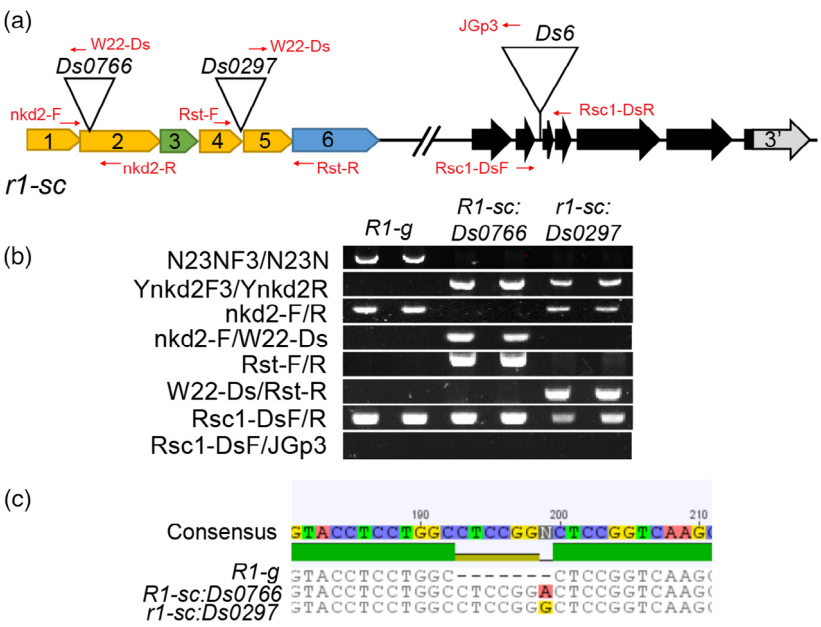

Figure 3. *Ds* insertions at *r1* locus.
(a) Architecture of *r1* locus and *Ds6* insertions in *r1-sc:m3* and two alleles originally ascribed as *nkd2* mutants, *R1-Ds0766* and *r1- Ds0297*. Red arrows show PCR primers used for detecting Ds elements. Hatch marks denote chromosomal regions that were omitted from the figure for the sake of space.
(b) Agarose gel shows PCR amplification products that confirm architecture and insertion sites shown in (a), compared with *R1-g*. The primer pairs tested the integrity of *nkd2* (N23NF3/N23N), the rearrangement of *nkd2* and *r1* (Ynkd2F3/Ynkd2R), and the *Ds6* insertions in *r1-Ds0766* (nkd2-F/R and nkd2-F/W22-Ds), *r1-Ds0297* (Rst-F/R and W22-Ds/Rst-R), and *r1-sc:m3* (Rsc1-DsF/R and Rsc1-DsF/JGp3).
(c) Sequences of intron 4 from several *r1* alleles at the site of the *Ds6* insertion in *r1-sc:m3*. *R1-sc:Ds0766*, and *r1-sc:Ds0297* contain putative Ds footprints suggesting transposition from this donor site contributed to the promoter insertions.

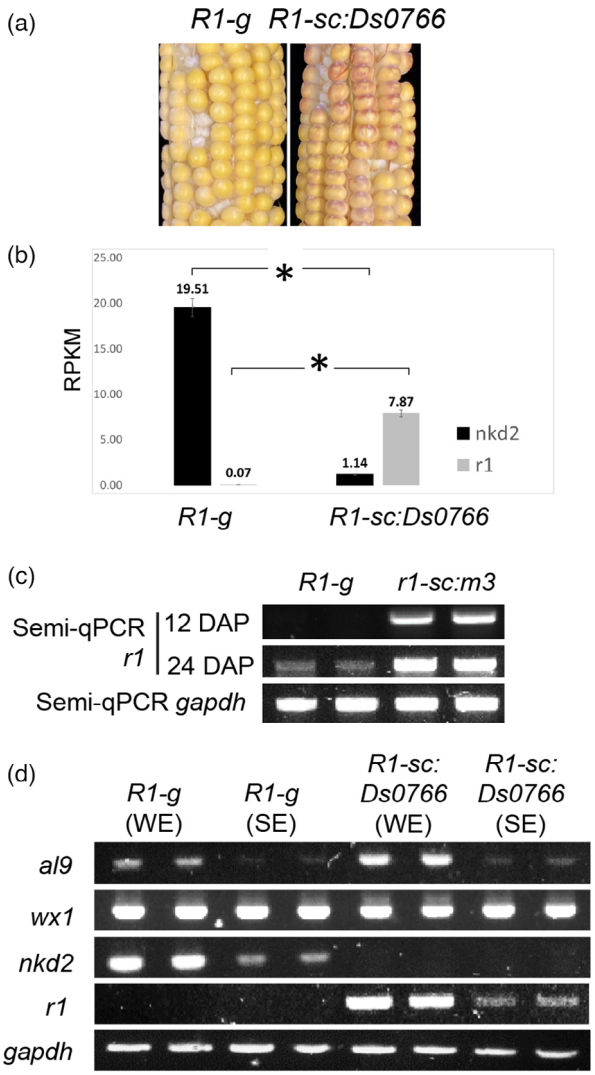

Figure 4. Expression of *r1* in lines with *R1-sc:124* derivative alleles.
(a) 16 days after pollination (DAP) cobs of purple-kernel W22 *R1-g* versus *R1-sc:Ds0766* at. The *R1-sc* allele confers early anthocyanin accumulation.
(b) Expression of *nkd2* and *r1* in W22 *R1-g* and *R1-sc:Ds0766* endosperms at 16 DAP. The numbers at the top of bars represent RPKM values. Error bars represent standard error of RPKM, and asterisk marks the significance at $P < 0.0001$ by Student's *t*-test.
(c) Semi-qRT-PCR testing the expression of *r1* in *R1-g* and *r1-sc:m3* at 12 and 24 DAP with *GAPDH* as internal control. (d) Semi-qRT-PCR testing expression of *R1-g* and *R1-sc:Ds0766*, compared with *nkd2*, in starchy endosperm (SE) compartment versus whole endosperm (WE) at 16 DAP. Marker genes included *al9* (AL), *ss-I* (SE), and *GAPDH* (constitutive control).

endosperm of the *Ds0766* line showed that the mean *r1* transcript levels were 114-fold higher than with a standard *R1-g* allele in a congenic W22 background (Figure 4b). Consistent with higher expression levels, kernels of the *R1-sc:Ds0766* line began accumulating pigment earlier than *R1-g* kernels grown under the same field conditions (Figure 4a).

Somewhat surprisingly, *r1* transcript levels were also elevated in the *r1-sc:m3* allele, even when using PCR primers 3′ of the *Ds6* insertion (Figure 4c), which raised the question of why this allele was non-functional. The insertion of *Ds6* in intron 4 between exons 4 and 5 suggested the possibility of disrupted RNA splicing and sequencing of cDNA fragments that spanned from exons 4 to 6 showed that indeed exon 5 was spliced out of the *r1-sc:m3* transcript (Figure 5a–cand S3). This mis-splicing causes an in-frame deletion of five codons encoding S-A-S-I-Q (Figure 5c). Protein sequence alignment of R1 homologs from a variety of grass species shows that four of these amino acids are invariant and the first S is conserved (Figure 5d). These five amino acids reside within the domain involved in heterodimerization with the C1 protein required for R1 activity in regulating anthocyanin biosynthesis (Figure S3) (Sainz et al., 1994). The *R1-sc:Ds0766* transcript showed normal splicing indicating that this represents a revertant of the *r1-sc:m3* allele (Figure 5b,c). In addition, *r1* transcript is ectopically expressed in starchy endosperm (SE) of *R1-sc:Ds0766*, and the pattern is similar to *nkd2* in SE of *R1-g* (Figure 4d).

R1-sc gene promoter is derived from *nkd2* by chromosome rearrangement

Our data support the notion that a chromosome rearrangement distinguishes *R1-B73*, as annotated in the B73_v.4 genome assembly, from *R1-st* and derivative *R1-sc* alleles (Jiao et al., 2017; Li et al., 2001). Given the syntenic relationship between *nkd2* and *nkd1*, it seemed more likely that the *R1-B73* state was ancestral because the *nkd2* gene is disrupted and non-functional in the *R1-sc* state. Several other *r1* alleles were examined; similar to *R1-g* (purple AL, green plant), the *R1-nj* (purple AL on crown of kernel, purple plant tissues) allele showed gene structures at the *r1* and *nkd2* loci that match B73 RefGen_v4 (Figure 6a). A similar pattern of PCR amplification to the *r1-sc:m3* allele was observed in an *R1-mb* and two *R1-st* accessions except that products from the promoter region of *r1* were slightly smaller. Sequencing of these products showed that they lacked the *Bnot* element (region 3) similar to *R1-sc:n5992* (Li et al., 2001) but that they otherwise matched the *R1-sc* allele (Figure S4). These alleles did not show evidence of a footprint making it difficult to determine whether the *Bnot* element inserted after occurrence of the rearrangement to produce the *R1-sc124* configuration or there was a perfect excision.

The *nkd2-r1* region was also examined among some annotated *Z. mays* genome sequences, including B73, W22, CML247, F7, EP1, PH207 and the teosinte *Z. mays* ssp. *mexicana*. Multiple alignment of *nkd2* genomic sequences revealed general conservation in the overall gene structure as well as sequence conservation throughout the gene (Figure 6b). In addition, BLAST of regions 4 and 5 (Figure 2a) against other grass species showed that they match well with genes encoding the

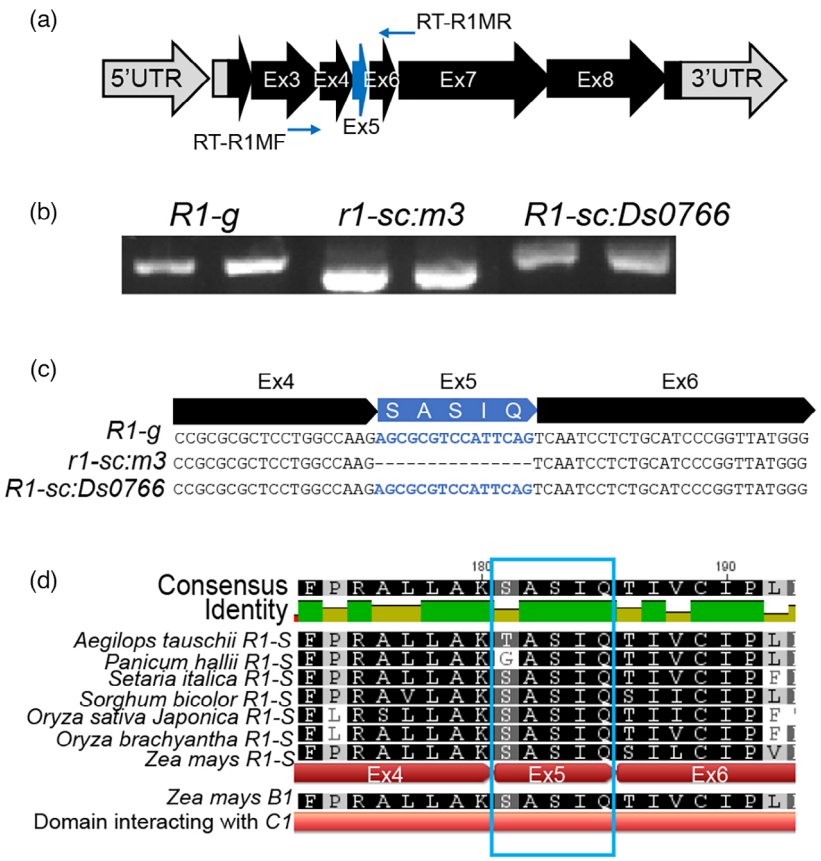

Figure 5. *Ds6* causes mis-splicing of the *r1-sc:m3* transcript.
(a) Architecture of *r1* canonical transcript. Blue arrow represents exon 5 (Ex5), which is mis-spliced out in *r1-sc:m3* transcripts. Thin blue arrows represent a primer pair to amplify the region flanking exon 5. UTR, untranslated region.
(b) Agarose gel showing size differences between *R1-g*, *r1-sc:m3*, and *R1-sc:Ds0766* PCR amplicons from the region containing exon 5 (primer pair RT-R1MF/RT-R1MR).
(c) Sequence alignment between *R1-g*, *r1-sc:m3*, and *R1-sc:Ds0766* at the Ex5-flanking region. Exon 5 is missing from *r1-sc:m3* transcripts.
(d) Amino acid sequence conservation at the exon 5 region among R1 homologous proteins from selected grass species and the maize syntelog B1.

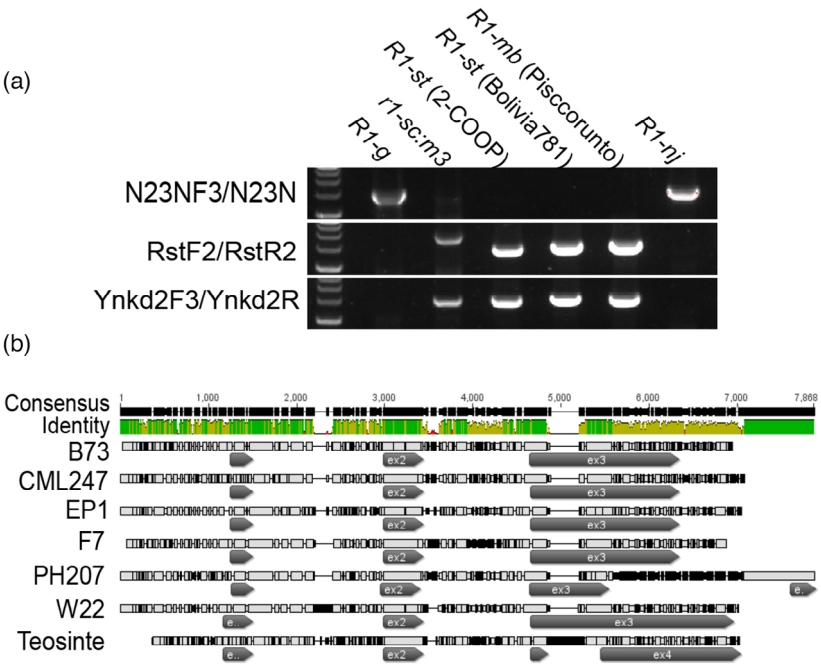

Figure 6. Integrity of the *nkd2* gene in select *r1* alleles and *Zea mays* genomes.
(a) PCR test for the presence of an intact *nkd2* gene or the *nkd2-r1* rearrangement in *r1* alleles *R1-g*, *r1-sc:m3*, *R1-st* (*2-COOP*), *R1-st* (*Bolivia781*), and *R1-mb* (*Pisccorunto*, and *R1-nj*. All the *R1-st* or *-mb* alleles contained the rearrangement but not *R1-g* nor *R1-nj*.
(b) Alignment of the *nkd2* locus among teosinte and maize lines with publicly assembly genome assemblies. All lines appeared to contain an intact *nkd2* locus with all the exons.

indeterminate domain transcription factors (Table S2), consistent with the expectation that an intact *nkd2* is the ancestral state.

Multiple alignment of *r1* genomic sequences also showed conservation in overall gene structure (Figure S4). Coding regions were well conserved while the highest

level of variability was seen in the large second intron. High levels of variability and repetitive sequences in the 5′ region made analysis difficult and generally uninformative. However, the sequence corresponding to region 7 produced a hit in each of the assembled genomes listed in Table 1. In each case, this fragment was located close to the *r1* locus, within 7–25 kb, rather than near the *nkd2* locus, again supporting the hypothesis that the ancestral state of this region is more likely to be the B73, rather than the *R1-sc:124*, configuration.

Ectopic expression of *R1-sc* appears to regulate gene expression in SE

The ectopic expression of *R1-sc* in the SE (Figure 5) raises the question of whether this transcription factor ectopically alters downstream gene expression in the SE, which may in turn alter metabolic or developmental processes. To investigate this question, laser capture microdissection (LCM) RNA-seq was performed on AL and SE compartments of three near-congenic genotypes: *R1-g* (wild-type [WT] *R1*, WT *Nkd2*), *R1-sc* (ectopic *R1*, null *nkd2*), and *r1-m3* (mutant *r1*, null *nkd2*). *R1* was exclusively expressed in the AL of *R1-g*, but in both AL and SE of *R1-sc* and *r1-m3*. In the AL, *R1* transcript levels were significantly higher for *R1-sc* and *r1-m3* than for *R1-g* (Figure 7a). Transcript levels for *nkd2* were significantly higher in *R1-g* than *R1-sc* or *r1-m3* in both the AL and SE (Figure 7b). These data are consistent with the semi-qRT-PCR tests of *r1* and *nkd2* between *R1-g* and *r1-m3* or *R1-sc:Ds0766* (Figures 1a and 4c,d). To verify the LCM RNA-seq data, we performed qRT-PCR analysis of select genes *Al9* (AL marker), *Ereb167* (SE marker), *Nkd2*, and *R1*. *Al9* and *Ereb167* showed

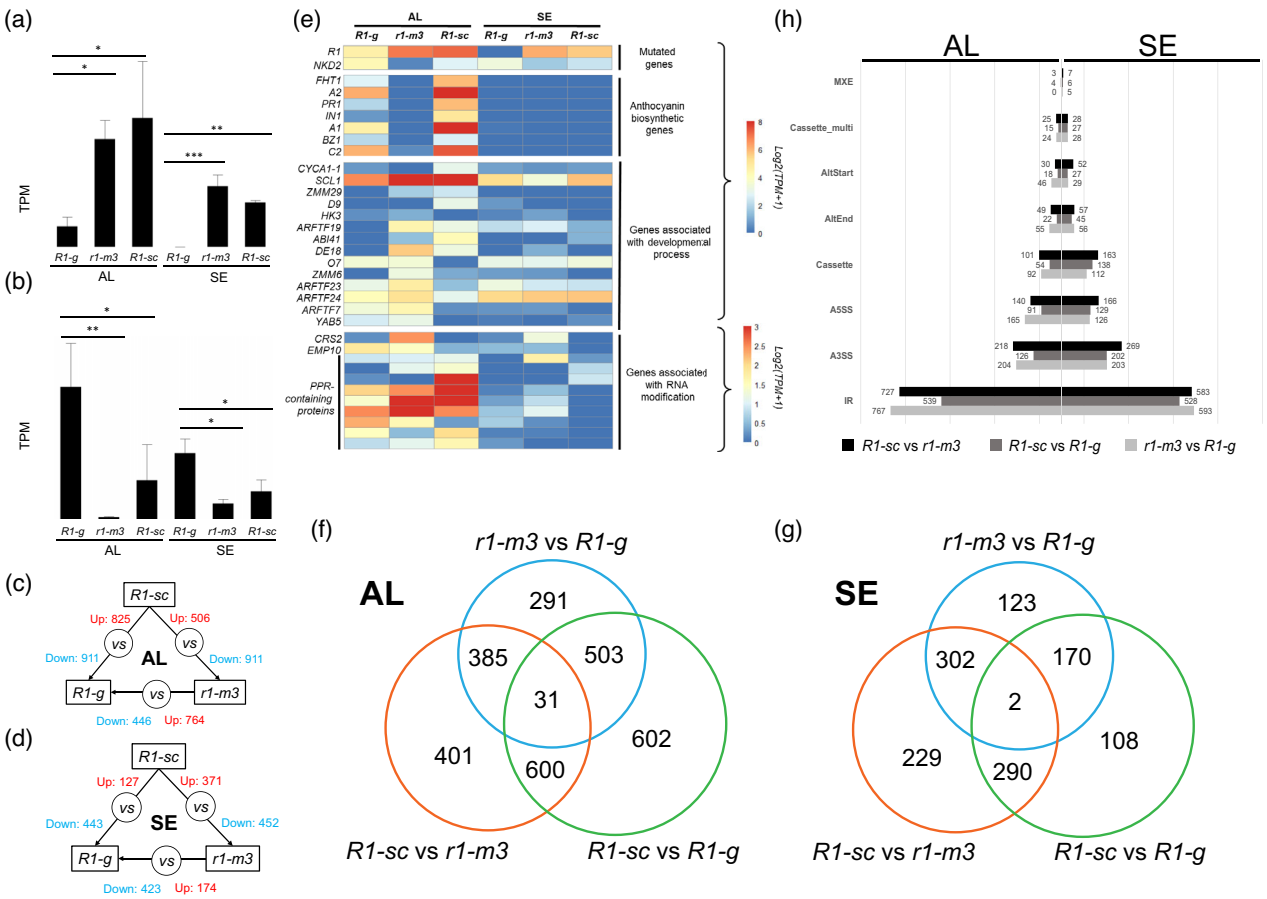

Figure 7. Laser capture microdissection RNA-sequencing of *R1-g*, *r1-m3*, and *R1-sc* aleurone (AL) and starchy endosperm (SE).
(a,b) Expression of (a) *R1* and (b) *NKD2* genes in the AL and SE among the three genotypes. Normalized expressed values are measured by transcript per million (TPM). Pairwise *t*-test was used to compare means of three biological replicates. Significance levels were marked by single (<0.05), double (<0.01) or triple (<0.001) asterisks and error bars represent standard errors.
(c,d) Differentially expressed genes (DEGs) of pair-wise comparisons among the three genotypes in AL (c) and SE (d).
(e) Expression heatmap of selected DEGs by functions among the three genotypes in AL and SE. Overlap of DEGs among the three pairwise comparisons in (f) AL and (g) SE.
(h) Number of significant (false discovery rate <0.05) differential splicing events by pairwise genotype comparison. A3SS, alternative 3′ splice site; A5SS, alternative 5′ splice site; AltEnd, alternative end exon; AltStart, alternative start exon; Cassette, skipped exon; Cassette_multi, multiple adjacent cassette exons; IR, intron retention; MXE, mutually exclusive exons.

preferential expression in the captured AL and SE, respectively, indicating we obtained relatively pure AL and SE tissues via LCM. In addition, *Nkd2* and *R1* showed trends consistent with the transcript per million data of LCM RNA-seq, which confirmed the differential expression pattern of *Nkd2* and *R1* between AL and SE among the three genotypes (Figure S5).

We detected differentially expressed genes (DEGs) and identified the corresponding annotated gene functions in the three pairwise comparisons of the genotypes for AL and SE (Table 3 and Figure 7c,d). We used these comparisons further to address the question of whether ectopically expressed *R1* is functional in SE. R1 has been shown to promote the expression of anthocyanin biosynthetic genes in the AL. In fact, some of the anthocyanin pathway genes show elevated expression in AL of *R1-sc* compared with *R1-g*, consistent with the increased anthocyanin accumulation in *R1-sc* (Figure S6), and confirming that elevated *R1* transcript levels translate to elevated R1 function. However, the expression of the same anthocyanin genes was not detected in *r1-m3*, confirming the requirement for exon 5 for promoting their expression. In the SE, no anthocyanin gene expression was detected for any of the *R1* genotypes, including *R1-sc*, indicating that the ectopic *R1* expression is not sufficient to activate these genes in the SE.

Comparison of gene expression profiles suggested that R1 may regulate expression of other genes in addition to those of the anthocyanin pathway. These comparisons are confounded by the *nkd2* mutation, as well as potential interactions between *nkd2* and *r1*. Nonetheless, examining the overlap in DEGs allowed the parsing of genes that are likely to be regulated by *R1* (Figure 7f,g and Table 3). For example, *R1-sc* and *r1-m3* both contain the same *nkd2* allele; thus in the SE, the 823 DEGs in this comparison are most likely due to differences between ectopic expression of the *R1-sc* and the mis-spliced *r1-m3* alleles. The 570 DEGs in the comparison of *R1-sc* versus *R1-g* in SE are likely due to (i) ectopic *R1-sc* expression versus no *R1* expression, and (ii) expression of mutant versus WT alleles of *Nkd2*. Of these, 290 genes occur among the DEGs in both comparisons and therefore represent high-confidence targets regulated (directly or indirectly) by ectopic *R1* expression from the *R1-sc* allele.

R1-sc alters RNA splicing in the SE

R1 and *Nkd2* expression variation may influence differential expression of downstream genes involved in diverse biological processes and molecular functions (Figure 7e and Table 4). In the SE, the 290 genes that were differentially expressed in the two aforementioned genotypic comparisons were enriched for functions associated with RNA modification, suggesting that ectopic *R1* expression could alter RNA processing. To test whether RNA splicing is altered, differential splicing events were identified using CASH v2.2.1 (Comprehensive Alternative Splicing Hunting) (Wu et al., 2018). Among the three pairwise genotypic comparisons, we identified 1152 significant differentially spliced events in *r1-m3* versus *R1-g*, 1102 in *R1-sc* versus *R1-g*, and 1325 in *R1-sc* versus *r1-m3* (Figure 7h). We also tested for differential RNA splicing in the AL, and identified 1353 significant events in *r1-m3* versus *R1-g*, 869 in *R1-sc* versus *R1-g*, and 1293 in *R1-sc* versus *r1-m3* (Figure 7h). By far the most common type of differential splicing in all tissues and genotypic comparisons, was intron retention, involving over 52% of all events (Figure 7h).

In addition, several genes encoding pentatricopeptide repeat (PPR) proteins were differentially expressed (Figure 7e) suggesting that RNA splicing in mitochondria or plastids may be altered (Dai et al., 2018; Dai et al., 2020; Qi et al., 2017; Zhu et al., 2019; Zoschke et al., 2016). This is supported by Figure S7 showing variation of certain exons among several organellar transcripts. In all, these observations support that the ectopic expression of *R1* transcript in the SE may alter RNA splicing processes, consistent with the enriched Gene Ontology (GO) term of overlapping DEGs associated with *R1-sc*.

DISCUSSION

The *r1* gene has long been a model for genetic studies due to its easily scorable phenotypes and extraordinary allelic diversity. The *r1* locus has been claimed to exhibit more known diversity in expression than any other in maize (Coe et al., 1988). This reflects, at least partly, on the ease in recognizing different expression patterns and the attention this locus has attracted. Transposable elements, epigenetic modifications, structural variation, copy number variation, unequal crossing over, and gene conversion, as well as many undefined factors, all contribute to the allelic diversity and variation in expression patterns observed at the *r1* locus (Kermicle et al., 1995; Li et al., 2001; Robbins et al., 1991; Walker et al., 1995; Walker et al., 1997; Walker & Panavas, 2001).

The simplex *R1-sc:124* allele was derived from the complex *R1-st* allele by unequal crossing over and is noteworthy for its intense level of anthocyanin pigmentation in the kernel AL (Ashman, 1960; Kermicle, 1980; McWhirter & Brink, 1962). The *r1-sc:m3* allele was subsequently derived by insertion of a *Ds6* transposon and this allele has been important as a marker for *Ac* transposon studies and for *Ac* or *Ds* transposon mutagenesis (Ahern et al., 2009; Dellaporta & Moreno, 1994). Here we found that a chromosome rearrangement preceded, or perhaps was responsible for, formation of *R1-st*.

As summarized in Figure 8, the intrachromosomal rearrangement involves the *nkd2* and *r1* loci, where *nkd2* upstream elements were moved to the *r1* promoter position, while a distal 5′ region of *r1* was moved to the *nkd2* locus. Analysis of diverse maize lines, including teosinte,

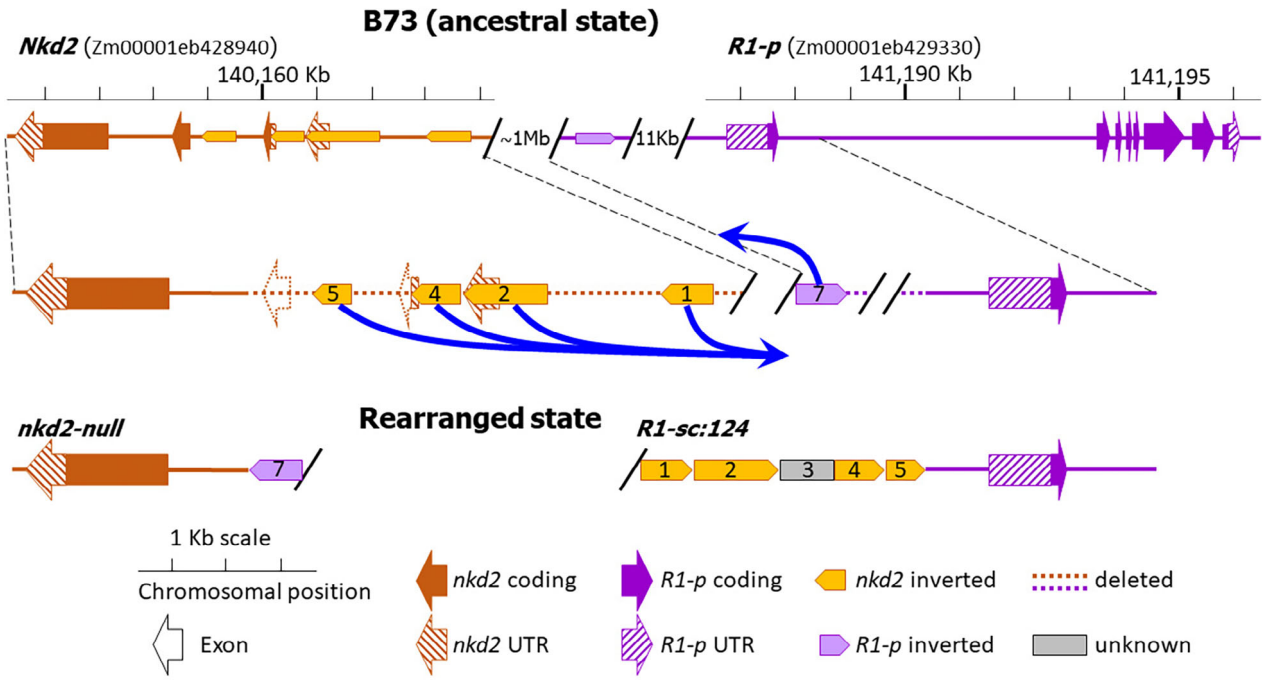

Figure 8. Summary of the rearrangement between *r1* and *nkd2* loci. Orange or gold colors represent regions originally associated with *nkd2* whereas purple or lavender represents regions from the *r1* locus. Dashed lines represent regions that were deleted in the rearrangement. Hatch marks indicate regions that were omitted from the figure for the sake of space. UTR, untranslated region.

showed that the overall structure of the *nkd2-r1* region is consistent with B73 (Figure 6b and Table 2). This suggests the ancestral state contains *nkd2* and *r1* genes in a head-to-head orientation on chromosome 10L (about 1.1 Mbp apart in B73) and it is therefore likely that an inversion occurred. In addition, part of the *nkd2* gene body, including exons 1–3, was deleted, as was most of the original *r1* promoter. Whether all these alterations occurred as part of a single rearrangement or due to sequential events is unknown.

The rearrangement generated a *nkd2* null allele with exon 4 located 3′ to sequences found about 11 kb 5′ to the *r1* transcription start site in B73 (Figure 2d,e). No *nkd2* transcript was detected by either RT-PCR nor RNA-seq (Figures 1a and 4b) suggesting that this region is not sufficient for promoter activity, at least in endosperm. There is no visible phenotypic change in the endosperm associated with the loss of *nkd2* function, most likely due to functional redundancy with its paralog, *nkd1* (Gontarek et al., 2016; Yi et al., 2015).

We had previously reported that two *nkd2* null mutants were caused by insertion of *Ds* elements (Yi et al., 2015). Here we found that the *Ds* elements originated from the *r1-sc:m3* allele but that the presumptive *nkd2* sequences into which the insertions occurred were actually located in the 5′ region of the *r1* gene. The overall conclusion, that each of these lines harbored a null *nkd2* allele that confirmed the gene identity, remains valid but the original report was erroneous in assigning the *Ds* insertions as the causal agents.

Table 2 Relationships between *nkd2* and *r1* loci in various *Zea mays* genomes

Genome	*nkd2* locus	*r1* locus	Distance between *nkd2* and *r1* (Mbp)	Distance between region 7 and *r1* (kbp)
B73	Chr 10	Chr 10	1	11
W22	Chr 10	Scaffold 282	Unknown	6.7*
EP1	Chr 10	Chr 10	1.4	25
F7	Chr 10	Chr 10	0.99	21
CML247	Chr 10	Chr 10	0.81	8
PH207	Chr 10	Chr 10	0.91	7
Z. mays ssp. *mexicana*	Chr 10	Chr 10	0.286	11

*Approximate distance between the breakpoint to 5′ of the *q* component of *r1* complex in W22.

Table 3 Summary of differentially expressed genes in AL and SE of pairwise genotype comparisons

Structure	Genotype	*R1* and *Nkd2* expression	Versus genotype	*R1* and *Nkd2* expression	No. genes upregulated*	No. genes downregulated*
AL	*R1-sc*	Elevated *R1*; *nkd2* null	*R1-g*	Normal *R1*; normal *Nkd2*	825	911
AL	*r1-m3*	*r1* loss-of-function; *nkd2* null	*R1-g*	Normal *R1*; normal *Nkd2*	764	446
AL	*R1-sc*	Elevated *R1*; *nkd2* null	*r1-m3*	*r1* loss-of-function; *nkd2* null	506	911
SE	*R1-sc*	Ectopic *R1*; *nkd2* null	*R1-g*	No *R1*; normal *Nkd2*	127	443
SE	*r1-m3*	*r1* loss-of-function; *nkd2* null	*R1-g*	No *R1*; normal *Nkd2*	174	423
SE	*R1-sc*	Ectopic *R1*; *nkd2* null	*r1-m3*	*r1* loss-of-function; *nkd2* null	371	452

AL, aleurone; SE, starchy endosperm.
*Up- or downregulated in the first genotype compared with the second genotype.

Table 4 Gene Ontology terms of differentially expressed genes at AL or SE

Genotype comparison	AL	SE
R1-sc vs. *R1-g*	Developmental process involved in reproduction (BP, 1.7E-3)[a] Regulation of cell morphogenesis (BP, 3.7E-3)	RNA modification (BP, 3.6E-2) Endonuclease activity (MF, 7.3E-3)
R1-sc vs. *r1-m3*	Developmental process involved in reproduction (BP, 1.8E-4) Multicellular organism development (BP, 1.6E-3)	DNA replication initiation (BP, 3.3E-4) Methylation (BP, 7.7E-3) Protein methyltransferase activity (MF, 2.5E-3)
r1-m3 vs. *R1-g*	Developmental process involved in reproduction (BP, 1.9E-3) Cell division (PP, 2.8E-2)	Nucleoside phosphate metabolic process (PP, 4.4E-3) Protein methyltransferase activity (MF, 3.4E-2)

AL, aleurone; BP, biological process; MF, molecular function; SE, starchy endosperm.
[a]Letters and numbers in the parentheses represent categories of Gene Ontology terms (BP and MF) and corresponding false discovery rate, respectively.

R1-sc:124 appears to be a hypermorphic allele that confers particularly high levels of anthocyanin pigmentation (McWhirter & Brink, 1962) and we found that the transcript accumulated earlier and to higher levels than a standard *R1-g* allele (Figure 4c). This expression was manifest phenotypically with earlier anthocyanin accumulation (Figure 4a). The *R1* gene complex contains variable copy numbers of the coding region and the expression properties of each component appear to be associated with its regulatory region (Li et al., 2001; Walker et al., 1995). The timing and levels of *R1-sc* transcript accumulation were reminiscent of *nkd2* transcript accumulation (Figure 4c) (Yi et al., 2015) making it tempting to speculate that *nkd2* regulatory sequences are responsible for the altered *R1-sc* expression. However, many other *R1-sc* alleles derived from the same *R1-st* progenitor confer much lower levels of anthocyanin than *R1-sc:124* (McWhirter & Brink, 1962), so the basis of the altered *R1* expression remains unclear.

In addition to high levels of expression, *R1-sc:124* also appeared to show ectopic expression in the SE (Figure 4d). Normal *R1* expression is generally assumed to be restricted to the AL, based on anthocyanin pigmentation and *in situ* mRNA localization (Procissi et al., 1997). The SE of *R1-sc:124* does not accumulate anthocyanin but that could be due to lack of expression of C1, which is required to complex with R1 to promote expression of anthocyanin biosynthesis genes (Goff et al., 1992; Grotewold et al., 2000). However, R1 can also homodimerize and bind G-box promoter elements independent of C1 (Kong et al., 2012). G-boxes are common promoter elements present on some, but not all, anthocyanin biosynthetic genes, and are known for other endosperm genes (Huang et al., 2016). Alternatively, promiscuous interactions with other MYB factors could allow regulation of gene expression. The Arabidopsis bHLH protein, GL3, is a homolog of R1 and can interact with alternative MYB proteins to regulate anthocyanin synthesis or other genes involved in trichome or root hair formation (Brkljacic & Grotewold, 2017). Furthermore, Arabidopsis bHLH proteins EGL3 and TT8, in the same homology subgroup with R1 and GL3, interact with MYB5 to regulate seed coat differentiation (Feller et al., 2011; Gonzalez et al., 2009). Thus, there are several feasible mechanisms by which ectopic *R1* expression could alter gene expression networks and metabolism in the SE without activating the anthocyanin pathway.

The *r1-sc:m3* allele contains a *Ds6* insertion located in intron 4, which completely eliminates the ability of the *r1* locus to promote anthocyanin synthesis (Alleman & Kermicle, 1993; Conrad & Brutnell, 2005). Surprisingly, transcript abundance remained elevated in the *r1-sc:m3* allele but

sequencing of the cDNA showed the mRNA was mis-spliced, eliminating exon 5 and resulting in an in-frame deletion of five amino acids (Figure 5c). The deletion occurs in the conserved region responsible for binding to C1 (Sainz et al., 1994). Because the interaction between R1 and C1 is necessary for binding the promoters and activating transcription of several anthocyanin biosynthetic genes (Goff et al., 1992, Grotewold et al., 2000), disrupted interaction between the R1 and C1 proteins may explain the lack of anthocyanin pigmentation.

LCM RNA-seq was used to analyze gene expression in AL and SE to address the question of whether ectopic *R1-sc* expression could indeed alter gene expression in the SE. Three genotypes were available in a near-congenic W22 background and substantial gene expression differences were apparent among all of them in both AL and SE (Table 3). These comparisons were complicated by variation at both the *nkd2* and *r1* genes at this locus and, overall, the results are most consistent with both genes (and potentially their interactions) contributing to the differences and assuming that the mis-spliced *r1-m3* gene product, missing exon 5, is partially functional in controlling genes other than the anthocyanin pathway. This hypothesis bears further study but focusing on overlapping sets of DEGs allowed us to identify genes putatively regulated by R1. In total, 290 SE genes were differentially expressed in comparisons of both *R1-sc* versus *R1-g* and *R1-sc* versus *r1-m3*, and thus represent high confidence genes that are regulated by ectopic *R1-sc* expression. Whether ectopic R1 functions via any of the aforementioned mechanisms requires further study. This group of putatively R1-regulated genes was enriched for the GO term "RNA modification," which led us to examine RNA splicing in our data.

Both *R1-sc* and *r1-m3* showed significant differential RNA splicing compared with WT *R1-g* and this was true in both the SE and the AL. Furthermore, *R1-sc* and *r1-m3* showed differential RNA splicing. The most common splice variant was intron retention in all compartment and genotype comparisons. The mechanism by which R1 regulates RNA splicing needs further study, but interestingly, a MYB transcription factor, CDC5, regulates RNA splicing and processing in yeast and Arabidopsis (Burns et al., 1999; Hirayama & Shinozaki, 1996; Zhang et al., 2013). A maize *CDC5* ortholog, *ZmMYB2/ZM1* (Zm00001d042287), is expressed in both AL and SE (Supplemental Data S1), and *ZmMYB2/ZM1* is homologous to *C1*. This raises the intriguing possibility that ZmMYB2/ZM1 could heterodimerize with R1 in place of C1, and thereby regulate RNA modification (Franken et al., 1994).

In addition, several PPR genes were differentially expressed among these samples. PPR proteins are associated with RNA modification in mitochondria or plastids (Dai et al., 2018; Dai et al., 2020; Qi et al., 2017; Zhu et al., 2019; Zoschke et al., 2016) suggesting that ectopic expression of R1 may affect organellar RNA modification processes in the SE. Indeed, several organellar transcripts appeared differentially spliced among samples (Figure S7).

An interesting feature of the *R1-st* and *R1-mb* alleles is they are both paramutagenic; that both contain the rearrangement prompts speculation that the *nkd2* promoter sequences or some other element of the rearrangement is responsible for this property, but evidence is to the contrary. Using recombination to replace the *R1-sc* promoter of the *R1-mb* complex with the non-rearranged *R1-nj* promoter did not eliminate paramutagenicity. Rather, for both the *R1-st* and *R1-mb* alleles, the level of paramutagenicity was correlated with the *r1* gene copy number (Kermicle et al., 1995; Panavas et al., 1999).

EXPERIMENTAL PROCEDURES

Plant materials

The *r1-sc:m3* and *R1-g* alleles were obtained from the *Ac/Ds* project (Ahern et al., 2009). *R1-st*:124, *R1-nj*, *R1-st* (2-COOP), *R1-st* (Bolivia781), and *R1-mb* (Pisccorunto) were obtained from the Maize Genetics Cooperation Stock Center, stock IDs X19A, 127B, X233S, X20U, and X13L, respectively. All plant materials were in a W22 inbred background. Plant materials were grown in a glasshouse or the field at the Curtiss Research Farm, Iowa State University, Ames, IA, USA during summers of 2018–2020.

RNA extraction and semi-qRT-PCR

Endosperms were dissected from developing kernels for RNA extraction and expression analysis. Total RNA was extracted as described (Li *et al.* 2014) except that GeneJET RNA Cleanup and Concentration Micro Kit (Thermo Fisher Scientific, United States) was used to clean up DNaseI-treated RNA. The cDNA was synthesized using SuperScript III First-Strand Synthesis SuperMix (Thermo Fisher Scientific) following the manufacturer's instructions. Semi-qPCR for *nkd2* and *r1* transcripts was performed using primers listed in Table S1. *GAPDH* was used as the internal standard (Lin et al., 2014). The semi-qRT-PCR was performed using GoTaq® Master Mixes (Promega Corporation, United States) following the manufacturer's protocol. The products were visualized by 1% agarose gel electrophoresis.

Genomic DNA analysis

Genomic DNA was extracted from leaves using a CTAB-based extraction method (Abdel-Latif & Osman, 2017) with the following modification: 5 ml CTAB extraction buffer (1% CTAB, 0.7 M NaCl, 10 mM Tris–HCl, 50 mM EDTA, 1% PVP [360 000 molecular weight]) was added to each sample (0.5 g ground leaf tissue in liquid nitrogen). Next, 12.5 μl 2-mercaptoethanol, 20 μl proteinase K (20 mg ml^{-1} stock), and 10 μl RNaseA (100 mg ml^{-1} stock) were added to the CTAB buffer-sample mix, followed by rotating incubation at 65°C for 1 h. An equal volume of chloroform/isoamyl alcohol (24:1) was added and mixed followed by centrifugation at 5000 *g* at 15°C for 20 min. The upper aqueous phase was mixed with an equal volume of isopropanol and centrifuged, followed by washing, air drying, and resuspension of the DNA pellet. The DNA was quantified by Qubit® 2.0 Fluorometer.

PCRs were performed to test the existence, integrity, or polymorphism of corresponding regions using GoTaq® Master Mix (Promega Corporation) and primers listed in Table S1. The thermal cycler conditions were 95°C, 2 min, followed by 30 cycles of 95°C, 30 sec; (Tm-2) °C, 30 sec; 72°C, 1 min/1 kb, with a final extension 72°C, 5 min. The Tm values were calculated by OligoCalc (Kibbe, 2007). The products were visualized by 1% agarose gel, and were Sanger sequenced at the DNA Facility in Iowa State University.

TAIL-PCR

The concept and cycle parameters of TAIL-PCR were adapted from previously published methods (Liu et al., 1995; Yi et al., 2009). Genomic DNA was digested by *Mse*I (AT sticky end) or *Msp*I (CG sticky end), and ligated with an AT-end adaptor (annealing of oligo adap-T1 and adap-T3) or an CG-end adaptor (annealing of oligo adap-T2 and adap-T3), respectively. In the first round PCR, adaptor-specific primer adap-B/P, and *nkd2* gene-specific primers with M13 sequence overhangs (NKD2-1, NKD2-2, and NKD2-3) were used to amplify specific products with the adaptor and M13 sequence. In the second round PCR, primer adap-B/P, and M13bt (biotinylated primer specific to M13) were used to amplify biotinylated products, which were then enriched by Dynabeads™ M-280 Streptavidin (Thermo Fisher Scientific) following the manufacturer's instructions. In the third round PCR, adap-B/P, and nested primers (N21N, N22N, and N23N, respectively) were used. The products were ligated to pGEM®-T Vector (Promega Corporation) and sequenced. All primer sequences are listed in Table S1. PCR products were manually assembled with the Align/Assemble tool of Geneious Pro 5.6.7 (https://www.geneious.com).

Nanopore long-read sequencing

The genomic DNA of W22 *r1-sc:m3* was gently extracted as described above, producing average fragment lengths of 27 078 as determined by an AATI Fragment Analyzer. Library preparation and long-read sequencing were performed by the DNA Facility at Iowa State University with one flowcell of an Oxford Nanopore GridIONx5 sequencer. In total, 1 090 196 reads were base-called via GUPPY 2.1.3 software (Wick et al., 2019) and 769 681 (70.6%) passed quality control. Mean read quality score was 8.9 and N50 was 6621. Reads were *de novo* assembled by Geneious Pro 5.6.7 (https://www.geneious.com). Relevant contigs (#60167 and #608677) were further confirmed by BLASTn against sequences of the *R1-sc* promoter, the *R1-B73* upstream region (approximately 11 kbp upstream of transcription start site in B73.v4 reference genome) and the *nkd2* gene. The sequencing data are available as SRA project accession number PRJNA835624.

Comparative sequence analyses

Manually assembled sequences and long-read contigs were used to search the NCBI nucleotide collection (nr/nt) of *Z. mays* (taxonomy ID: 4577) using BLAST. BLASTn was used to align the *nkd2*, *r1*, and region7 sequences to the assembled genomes of maize lines B73, W22, F7, PH207, CML247, EP1, and *Z. mays* ssp. *mexicana* at MaizeGDB (https://www.maizegdb.org/) using an e-value cutoff of 1e-100. Multiple sequence alignments were performed by Geneious Pro 5.6.7 (https://www.geneious.com).

RNA-seq and data analysis from 16 DAP endosperm

RNA was extracted as described above from 16 DAP whole endosperm of *R1-sc*:DS0766 and *R1-g* in a W22 background. Total RNA per sample (100–500 ng), with four biological replicates, was used to construct multiplexed sequencing libraries using the Illumina (Madison, WI, USA) TruSeq Stranded mRNA Library Preparation Kit with poly(A) selection. The resulting libraries were sequenced using the HiSeq 125 Cycle Paired-End Sequencing v4 protocol of the Illumina HiSequation 2500 platform at the High-Throughput Genomics and Bioinformatic Analysis Shared Resource at Huntsman Cancer Institute (HCI) at the University of Utah. Raw reads were quality checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and mapped to the maize reference genome (B73 RefGen_v4) using TOPHAT v2.1.1 as described (Trapnell et al., 2009). Reads mapped to each gene were counted using featureCounts and RPKM values were calculated by CUFFLINKS v2.2.1 (Liao et al., 2014; Trapnell et al., 2010). The RNA-seq data are available at GEO accession number GSE202269.

LCM RNA-seq and data analysis

Developing kernels (19 DAP) of *R1-g*, *R1-sc*, and *r1-m3* lines were harvested (three cobs each type as biological replicates) and fixed in ice-cold Farmer's fixative (ethanol/glacial acetic acid, 3:1) and stored in cold fixative overnight. Fixed kernels were then processed following Zhang et al. (2018). Laser capture of AL and SE, RNA-isolation, cDNA synthesis, amplification, library construction, and sequencing were performed as described previously (Zhan et al., 2015). The RNA quality and quantity were measured using Agilent (Santa Clara, CA, USA) RNA 6000 Pico Kit on a Bioanalyzer 2100. Four kernels from each sample were used for dissection of AL and SE. In all cases, approximately 5 ng of captured RNA was used for cDNA amplification using the previously described amplification procedures (Zhan et al., 2015). The construction of the cDNA paired-end libraries and their quality checking were carried out at the High-Throughput Genomics and Bioinformatic Analysis Shared Resource at HCI at the University of Utah, using a NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) with approximately 400 ng of amplified cDNA. Each bio-replicated set of cDNA samples was multiplexed and run on a single lane of a NovaSeq 6000 system (Illumina) at HCI using the NovaSeq S4 Reagent Kit v1.5 (Illumina).

The RNA-seq data were validated via real time qPCR. Three biological replicates of amplified cDNAs were tested for the expression of marker genes using qPCR. The gene-specific primer pair sequences of selected genes are listed in Supporting Table S1. One nanogram of cDNA was used for each qPCR assay. The amount of cDNA template for a specific gene was considered negligible when the C_t value was ≥36.

The raw reads of 18 samples (3 genotypes × 2 compartments × 3 biological replicates) were trimmed by Trimmomatic (v0.39) (Bolger et al., 2014) followed by quality control by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The trimmed reads were mapped to the maize reference genome (Zm-B73-REFERENCE-GRAMENE-4.0) (Jiao et al., 2017) via HISAT2 (v2.2.1) with parameters specific for paired-end reads (Kim et al., 2015; Kim et al., 2019), and 84%–88% of reads were successfully mapped. The output SAM files were converted to BAM files and were sorted and indexed by SAMTOOLS (v1.14). For differential expression analysis, mapped transcripts in sorted BAM files were assembled according to annotation file Zea_mays.B73_RefGen_v4.50.gtf (https://www.maizegdb.org/download) and normalized RPKM and transcript per million values were calculated with STRINGTIE (v2.1.6) (Pertea et al., 2016). Read counts were calculated by HTSEQ (v0.11.2) and DEGs were called via DESEQ2 (v3.14) using an empirical Bayes shrinkage approach with the criteria log2 fold-change >1 or < −1 and false discovery rate <0.05 between pairwise comparisons at AL or SE (Anders et al., 2015;

Love et al., 2014). GO term enrichment analysis was performed at AgriGOv2 (http://systemsbiology.cau.edu.cn/agriGOv2/) with false discovery rate <0.05 (Tian et al., 2017). The number of differential splicing events was identified by CASH (v2.2.1) based on sorted and indexed BAM files (Wu et al., 2018), and differential exon expression was visualized on IGV (Integrative Genomics Viewer) (v2.8.2) (Robinson et al., 2011; Thorvaldsdottir et al., 2013) based on bigwig files generated by DEEPTOOLS (v3.5.0) (Ramirez et al., 2016). The LCM RNA-seq data are available at GEO (accession number GSE200905).

ACKNOWLEDGEMENTS

The authors thank Jerry Kermicle and Bill Eggleston for helpful discussions about the origins and relationships among *r1* alleles. Tom Peterson and Sharu Paul Sharma shared unpublished data. This research was supported by the U.S. National Science Foundation award IOS-1444568 to R.Y. and P.W.B., and Hatch project IOW03649 to P.W.B. Open access funding provided by the Iowa State University Library.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest with this research.

DATA AVAILABILITY STATEMENT

All sequence data generated in this study have been archived at the National Center for Bioinformatic Information. Nanopore sequencing data are available as SRA BioProject accession PRJNA835624. The RNA-seq data are available at GEO accession GSE202269. The LCM RNA-seq data are available at GEO, accession GSE200905.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. *nkd2* expression in 16 DAP endosperm between W22 WT and *nkd1-Ds*. Error bars represent standard error of RPKM, and asterisk marks the significance at $P < 0.0001$ by Student's *t*-test.

Figure S2. Gene models and PCR products described in the text. (A) *nkd2* locus. (B) *r1* locus. (C) *nkd2-r1* rearrangement. Hatch marks denote chromosomal regions that were omitted from the figure for the sake of space.

Figure S3. *Ds6* insertion in the *r1-sc:m3* allele. (A) Architecture of *r1-sc:m3*. Red arrows represent primers used to test the *Ds6* insertion. (B) Tests of *Ds6* insertions in *R1-g*, *r1-sc:m3*, and *R1-sc:Ds0766*. The *Ds6* insertion in intron 4 was lost in *R1-sc:Ds0766*. (C) Amino acid alignment between selected grass species and maize B1 at the exon 5 region. The figure shows that exon 5 encodes highly conserved amino acids located within the C1 interaction domain.

Figure S4. Alignment of the *r1* gene among maize lines. (A) Alignment of sequences involved in the *nkd2-r1* rearrangement. In W22 carrying *R1-g* (WT) the sequence is located at the *nkd2* locus. In the other *r1* alleles shown, the sequence is located at the *r1* locus. Alleles of *r1* include *R1-st ref* (Li et al., 2001), *R1-st* (2-COOP), *R1-st* (Bolivia781), and *R1-mb* (Pisccorunto). Only *R1-st ref* contains the *Bnot* putative retroelement (region 3) with accompanying target site duplication (bold lettering). (B) The overall structure of the *r1* locus is conserved.

Figure S5. RT-qPCR results of selected genes between AL and SE. (A) *AL-9*, (B) *EREB167*, (C) *R1*, and (D) *NKD2*. The bars represent the average and standard error of the three biological replicates of LCM RNAs used in RNA-seq analysis.

Figure S6. Ear images of *r1-sc:m3* (A) 12 DAP, (B) 16 DAP, and (C) 19 (DAP), and segregating ear images of *R1-sc:124*XW22 (*R1-g*) (D) 12 DAP, (E) 16 DAP, and (F) 19 DAP. The ears were collected from 2020 greenhouse (A,B,D,E) and 2020 summer field (C,F).

Figure S7. Effect of ectopic expression of *R1* in SE on RNA processing variation in mitochondria and plastids.

Table S1. Primer sequences

Table S2. BLAST results of region 4 and 5 against other plant species

Table S3. Summary of *r1* standard promoter BLAST against B73 genome

Table S4. Summary of *r1* standard promoter BLAST against W22 genome

Table S5. Accession numbers of R1 DNA and protein sequences

Data S1. Expression data and differentially expressed genes from laser-capture microdissection and RNA sequencing

REFERENCES

Abdel-Latif, A. & Osman, G. (2017) Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. *Plant Methods*, **13**, 1.

Ahern, K.R., Deewatthanawong, P., Schares, J., Muszynski, M., Weeks, R., Vollbrecht, E. *et al.* (2009) Regional mutagenesis using dissociation in maize. *Methods*, **49**, 248–254.

Alleman, M. & Kermicle, J.L. (1993) Somatic variegation and germinal mutability reflect the position of transposable element dissociation within the maize R gene. *Genetics*, **135**, 189–203.

Anders, S., Pyl, P.T. & Huber, W. (2015) HTSeq–a python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

Ashman, R.B. (1960) Stippled aleurone in maize. *Genetics*, **45**, 19–34.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Brink, R.A. (1973) Paramutation. *Annual Review of Genetics*, **7**, 129–152.

Brkljacic, J. & Grotewold, E. (2017) Combinatorial control of plant gene expression. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, **1860**, 31–40.

Burns, C.G., Ohi, R., Krainer, A.R. & Gould, K.L. (1999) Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 13789–13794.

Coe, E.H., Neuffer, M.G. & Hoisington, D.A. (1988) The Genetics of Corn. In: Sprague, G.F. & Dudley, J.W. (Eds.) *Corn and Corn Improvement*. Madison, WI: American Society of Agronomy, Inc., Crop Society of America, Inc., & Soil Science Society of America, Inc., pp. 81–258.

Conrad, L.J. & Brutnell, T.P. (2005) Ac-immobilized, a stable source of activator transposase that mediates sporophytic and gametophytic excision of dissociation elements in maize. *Genetics*, **171**, 1999–2012.

Dai, D., Luan, S., Chen, X., Wang, Q., Feng, Y., Zhu, C. *et al.* (2018) Maize Dek37 encodes a P-type PPR protein that affects cis-splicing of mitochondrial nad2 intron 1 and seed development. *Genetics*, **208**, 1069–1082.

Dai, D., Jin, L., Huo, Z., Yan, S., Ma, Z., Qi, W. *et al.* (2020) Maize pentatricopeptide repeat protein DEK53 is required for mitochondrial RNA editing at multiple sites and seed development. *Journal of Experimental Botany*, **71**, 6246–6261.

Dellaporta, S.L. & Moreno, M.A. (1994) Gene Tagging with Ac/Ds Elements in Maize. In: Freeling, M. & Walbot, V. (Eds.) *The Maize Handbook*. New York, NY: Springer New York, pp. 249–233.

Dellaporta, S.L., Greenblatt, I., Kermicle, J.L., Hicks, J.B. & Wessler, S.R. (1988) Molecular Cloning of the Maize R-nj Allele by Transposon Tagging with Ac. In: Gustafson, J.P. & Appels, R. (Eds.) *Chromosome Structure and Function*. Boston: Springer, pp. 263–282.

Dooner, H.K. & Kermicle, J.L. (1971) Structure of the R r tandem duplication in maize. *Genetics*, **67**, 437–454.

Eggleston, W.B., Alleman, M. & Kermicle, J.L. (1995) Molecular organization and germinal instability of R-stippled maize. *Genetics*, **141**, 347–360.

Feller, A., Machemer, K., Braun, E.L. & Grotewold, E. (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal*, **66**, 94–116.

Franken, P., Schrell, S., Peterson, P.A., Saedler, H. & Wienand, U. (1994) Molecular analysis of protein domain function encoded by the myb-homologous maize genes C1, Zm 1 and Zm 38. *The Plant Journal*, **6**, 21–30.

Giacopelli, B.J. & Hollick, J.B. (2015) Trans-homolog interactions facilitating paramutation in maize. *Plant Physiology*, **168**, 1226–1236.

Goff, S.A., Klein, T.M., Roth, B.A., Fromm, M.E., Cone, K.C., Radicella, J.P. *et al.* (1990) Transactivation of anthocyanin biosynthetic genes following transfer of B regulatory genes into maize tissues. *The EMBO Journal*, **9**, 2517–2522.

Goff, S.A., Cone, K.C. & Chandler, V.L. (1992) Functional analysis of the transcriptional activator encoded by the maize B gene: evidence for a direct functional interaction between two classes of regulatory proteins. *Genes and Development*, **6**, 864–875.

Gontarek, B.C., Neelakandan, A.K., Wu, H. & Becraft, P.W. (2016) NKD transcription factors are central regulators of maize endosperm development. *Plant Cell*, **28**, 2916–2936.

Gonzalez, A., Mendenhall, J., Huo, Y. & Lloyd, A. (2009) TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. *Developmental Biology*, **325**, 412–421.

Grotewold, E., Sainz, M.B., Tagliani, L., Hernandez, J.M., Bowen, B. & Chandler, V.L. (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 13579–13584.

Hirayama, T. & Shinozaki, K. (1996) A cdc5+ homolog of a higher plant, Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 13371–13376.

Huang, H., Xie, S., Xiao, Q., Wei, B., Zheng, L., Wang, Y. *et al.* (2016) Sucrose and ABA regulate starch biosynthesis in maize through a novel transcription factor, ZmEREB156. *Scientific Reports*, **6**, 27590. https://doi.org/10.1038/srep27590

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M.C., Wang, B. *et al.* (2017) Improved maize reference genome with single-molecule technologies. *Nature*, **546**, 524–527.

Kermicle, J.L. (1980) Probing the component structure of a maize gene with transposable elements. *Science*, **208**, 1457–1459.

Kermicle, J.L., Alleman, M. & Dellaporta, S.L. (1989) Sequential mutagenesis of a maize gene, using the transposable element dissociation. *In Genome*, **31**, 712–716.

Kermicle, J.L., Eggleston, W.B. & Alleman, M. (1995) Organization of paramutagenicity in R-stippled maize. *Genetics*, **141**, 361–372.

Kibbe, W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, **35**, W43–W46.

Kim, D., Langmead, B. & Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**, 357–360.

Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**, 907–915.

Kong, Q., Pattanaik, S., Feller, A., Werkman, J.R., Chai, C., Wang, Y. *et al.* (2012) Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E2091–E2097.

Li, G., Wang, D., Yang, R., Logan, K., Chen, H., Zhang, S., et al. (2014) Temporal patterns of gene expression in developing maize endosperm identified through transcriptome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 7582–7587.

Li, Y., Bernot, J.P., Illingworth, C., Lison, W., Bernot, K.M., Eggleston, W.B. *et al.* (2001) Gene conversion within regulatory sequences generates maize r alleles with altered gene expression. *Genetics*, **159**, 1727–1740.

Liao, Y., Smyth, G.K. & Shi, W. (2014) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

Lin, Y., Zhang, C., Lan, H., Gao, S., Liu, H., Liu, J. *et al.* (2014) Validation of potential reference genes for qPCR in maize across abiotic stresses, hormone treatments, and tissue types. *PLoS One*, **9**, e95445.

Liu, Y.G., Mitsukawa, N., Oosumi, T. & Whittier, R.F. (1995) Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *The Plant journal: For Cell and Molecular Biology*, **8**, 457–463.

Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Ludwig, S.R. & Wessler, S.R. (1990) Maize R gene family: tissue-specific helix-loop-helix proteins. *Cell*, **62**, 849–851.

Ludwig, S.R., Habera, L.F., Dellaporta, S.L. & Wessler, S.R. (1989) Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 7092–7096.

McWhirter, K.S. & Brink, R.A. (1962) Continuous variation in level of paramutation at the R locus in maize. *Genetics*, **47**, 1053–1074.

Panavas, T., Weir, J. & Walker, E.L. (1999) The structure and paramutagenicity of the R-marbled haplotype of Zea mays. *Genetics*, **153**, 979–991.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. & Salzberg, S.L. (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and ballgown. *Nature Protocols*, **11**, 1650–1667.

Procissi, A., Dolfini, S., Ronchi, A. & Tonelli, C. (1997) Light-dependent spatial and temporal expression of pigment regulatory genes in developing maize seeds. *Plant Cell*, **9**, 1547–1557.

Qi, W., Tian, Z., Lu, L., Chen, X., Chen, X., Zhang, W. *et al.* (2017) Editing of mitochondrial transcripts nad3 and cox2 by Dek10 is essential for mitochondrial function and maize plant development. *Genetics*, **205**, 1489–1501.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S. *et al.* (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, **44**, W160–W165.

Robbins, T.P., Walker, E.L., Kermicle, J.L., Alleman, M. & Dellaporta, S.L. (1991) Meiotic instability of the R-r complex arising from displaced intragenic exchange and intrachromosomal rearrangement. *Genetics*, **129**, 271–283.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. *et al.* (2011) Integrative genomics viewer. *Nature Biotechnology*, **29**, 24–26.

Sainz, M.B., Goff, S.A., Krahn, J.M. & Chandler, V.L. (1994) Transcriptional Regulation of the Maize Anthocyanin Pathway. In: Coruzzi, G. & Puigdomenech, P. (Eds.) *Plant Molecular Biology*. Berlin Heidelberg: Springer-Verlag, pp. 381–390.

Springer, N.M., Anderson, S.N., Andorf, C.M., Ahern, K.R., Bai, F., Barad, O. *et al.* (2018) The maize w22 genome provides a foundation for functional genomics and transposon biology. *Nature Genetics*, **50**, 1282–1288.

Stadler, L.J. (1946) Spontaneous mutation at the R locus in maize. I. the aleurone-color and plant-color effects. *Genetics*, **31**, 377–394.

Stadler, L.J. & Emmerling, M.H. (1956) Relation of unequal crossing over to the interdependence of R elements (P) and (S). *Genetics*, **41**, 124–137.

Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, **14**, 178–192.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z. *et al.* (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, **45**, W122–W129.

Trapnell, C., Pachter, L. & Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-seq. *Bioinformatics*, **25**, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J. *et al.* (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, **28**, 511–515.

Vollbrecht, E., Duvick, J., Schares, J.P., Ahern, K.R., Deewatthanawong, P., Xu, L. *et al.* (2010) Genome-wide distribution of transposed dissociation elements in maize. *Plant Cell*, **22**, 1667–1685.

Walker, E.L. & Panavas, T. (2001) Structural features and methylation patterns associated with paramutation at the r1 locus of Zea mays. *Genetics*, **159**, 1201–1215.

Walker, E.L., Robbins, T.P., Bureau, T.E., Kermicle, J. & Dellaporta, S.L. (1995) Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. *The EMBO Journal*, **14**, 2350–2363.

Walker, E.L., Eggleston, W.B., Demopulos, D., Kermicle, J. & Dellaporta, S.L. (1997) Insertions of a novel class of transposable elements with a strong target site preference at the r locus of maize. *Genetics*, **146**, 681–693.

Wick, R.R., Judd, L.M. & Holt, K.E. (2019) Performance of neural network basecalling tools for Oxford nanopore sequencing. *Genome Biology*, **20**, 129.

Wu, W., Zong, J., Wei, N., Cheng, J., Zhou, X., Cheng, Y. *et al.* (2018) CASH: a constructing comprehensive splice site method for detecting alternative splicing events. *Briefings in Bioinformatics*, **19**, 905–917.

Yi, G., Luth, D., Goodman, T.D., Lawrence, C.J. & Becraft, P.W. (2009) High-throughput linkage analysis of mutator insertion sites in maize. *The Plant journal: For Cell and Molecular Biology*, **58**, 883–892.

Yi, G., Neelakandan, A.K., Gontarek, B.C., Vollbrecht, E. & Becraft, P.W. (2015) The naked endosperm genes encode duplicate INDETERMINATE domain transcription factors required for maize endosperm cell patterning and differentiation. *Plant Physiology*, **167**, 443–456.

Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N.M., Arakaki, A.M. *et al.* (2015) RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell*, **27**, 513–531.

Zhang, S., Xie, M., Ren, G. & Yu, B. (2013) CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 17588–17593.

Zhang, S., Thakare, D. & Yadegari, R. (2018) Laser-capture microdissection of maize kernel compartments for RNA-seq-based expression analysis. *Methods in Molecular Biology*, **1676**, 153–163.

Zhu, C., Jin, G., Fang, P., Zhang, Y., Feng, X., Tang, Y. *et al.* (2019) Maize pentatricopeptide repeat protein DEK41 affects cis-splicing of mitochondrial nad4 intron 3 and is required for normal seed development. *Journal of Experimental Botany*, **70**, 3795–3808.

Zoschke, R., Watkins, K.P., Miranda, R.G. & Barkan, A. (2016) The PPR-SMR protein PPR53 enhances the stability and translation of specific chloroplast RNAs in maize. *The Plant Journal*, **85**, 594–606.