Aberrant splicing in maize *rough endosperm3* reveals a conserved role for U12 splicing in eukaryotic multicellular development

Christine M. Gault\(^{a,b,1,2}\), Federico Martin\(^{a,b,1,3}\), Wenbin Mei\(^{c}\), Fang Bai\(^{b}\), Joseph B. Black\(^{b}\), W. Brad Barbazuk\(^{a,c,d}\), and A. Mark Settles\(^{a,b,d,4}\)

\(^{a}\)Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL 32611; \(^{b}\)Horticultural Sciences Department, University of Florida, Gainesville, FL 32611; \(^{c}\)Department of Biology, University of Florida, Gainesville, FL 32611; and \(^{d}\)Genetics Institute, University of Florida, Gainesville, FL 32611

Edited by Julia Bailey-Serres, University of California, Riverside, CA, and approved January 30, 2017 (received for review September 29, 2016)

RNA splicing of U12-type introns functions in human cell differentiation, but it is not known whether this class of introns has a similar role in plants. The maize ROUGH ENDOSPERM3 (RGH3) protein is orthologous to the human splicing factor, ZRSR2. ZRSR2 mutations are associated with myelodysplastic syndrome (MDS) and cause U12 splicing defects. Maize *rgh3* mutants have aberrant endosome cell differentiation and proliferation. We found that most U12-type introns are retained or misspliced in *rgh3*. Genes affected in *rgh3* and *ZRSR2* mutants identify cell cycle and protein glycosylation as common pathways disrupted. Transcripts with retained U12-type introns can be found in polysomes, suggesting that splicing efficiency can alter protein isoforms. The *rgh3* mutant protein disrupts colocalization with a known ZRSR2-interacting protein, U2AF2. These results indicate conserved function for RGH3/ZRSR2 in U12 splicing and a deeply conserved role for the minor spliceosome to promote cell differentiation from stem cells to terminal fates.

**Results**

**Missplicing of U12-Type Introns in rgh3.** We determined the genome-wide effect of *rgh3* on mRNA splicing with RNA-seq and isoform expression analysis from homozygous *rgh3* and WT sibling root and shoot tissues. Cufflinks (20) predicted 46 genes had altered isoform use in *rgh3* (Table S1). Localized differences in read coverage depth between *rgh3* and WT libraries were used to design semiquantitative RT-PCR assays for nine of the genes identified as differentially spliced (Fig. L4 and Fig. S2). Sequencing of the amplified products revealed that intron retention occurs more frequently in *rgh3* mutants at seven of these loci (Fig. IB and Figs. S3 and S4). Some *rgh3* amplifications exhibited slow-migrating bands due to heteroduplex formation of differently sized RT-PCR products (Fig. S5). Six of the validated splicing differences have U12-type introns as defined by the ERISdb plant splice site database (21). The intron retained in

![Significance](https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616173114/-/DCSupplemental)

The last eukaryotic common ancestor had two spliceosomes. The major spliceosome acts on nearly all introns, whereas the minor spliceosome removes rare, U12-type introns. Based on in vitro RNA-splicing assays, the RGH3/ZRSR2 RNA-splicing factor has functions in both spliceosomes. Here, we show that the maize *rgh3* mutant allele primarily disrupts U12 splicing, similar to human ZRSR2 mutants, indicating a conserved in vivo function in the minor spliceosome. These mutant alleles block cell differentiation leading to overaccumulation of stem cells in endosperm and blood, respectively. We found extensive conservation between maize and human U12-type intron-containing genes, demonstrating that a common genetic architecture controls at least a subset of cell differentiation pathways in both plants and animals.

**Author contributions:** C.M.G., F.M., W.B.B., and A.M.S. designed research; C.M.G., F.M., F.B., J.B.B., and A.M.S. performed research; C.M.G., F.M., W.B.B., and A.M.S. analyzed data; and C.M.G., F.M., W.B.B., and A.M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Ferely available online through the PNAS open access option.

Data deposition: The RNA-seq data files reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE57466).

\(^{1}\)C.M.G. and F.M. contributed equally to this work.

\(^{2}\)Present address: Institute of Genomic Diversity, Cornell University, Ithaca, NY 14853.

\(^{3}\)Present address: Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80511.

\(^{4}\)To whom correspondence should be addressed. Email: settles@ufl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616173114/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1616173114

*PNAS* | Published online February 27, 2017 | E2195–E2204
U12-type intron-containing genes account for <2% of the B73 filtered gene set. With 20% of Cufflinks predictions involving U12-type introns, we hypothesized that U12-type introns are misspliced on a global scale in rgh3. To allow for novel transcript isoforms to be detected, we analyzed intron read counts normalized by gene expression for all introns in the filtered gene set of the maize genome (Fig. 1C). Of the 446 nonredundant U12-type introns in ERISdb, 340 were expressed in the seedling libraries. Significantly more RNA-seq reads mapped to 240 U12-type introns in rgh3 libraries vs. WT libraries (Fig. 1D and Dataset S1). This represents 71% of the maize U12-type introns tested. Only three U12-type introns showed significantly fewer reads in rgh3 libraries, which is below the expected false-positive rate for the number of t tests completed. By contrast, 4% of the remaining 113,345 nonredundant introns within expressed genes had differences in the number of normalized reads mapped in rgh3 vs. WT (Fig. 1D and Dataset S2). Globally, fewer U2-type introns showed significant differences than the expected number of false-positive t tests, indicating that rgh3 specifically affects genes with U12-type introns.

To quantify the extent of intron splicing defects, we calculated percent spliced out (PSO) for individual introns using exon-exon junction and intron reads (Fig. 1E). Based on Fisher’s exact tests of read counts, splicing defects were detected for 77% of U12-type introns (Dataset S1). The median difference between WT and rgh3 PSO (ΔPSO) was 62%, indicating extensive retention of U12-type introns in rgh3 (Fig. 1F). More than 13% of all other introns also showed statistically significant increased retention in rgh3, but the median ΔPSO for these introns was only 4%, indicating minor impacts on U2 splicing (Dataset S2).

An additional nine genes showing significant differences in U12-type intron read depth and PSO were randomly selected for RT-PCR (Fig. 2A–C and Fig. S7). All of these genes had splicing defects in rgh3 with three patterns of altered splicing: intron retention (Fig. 2D and Fig. S8A–D), missplicing of the U12-type intron concomitant with retention of a downstream U2-type intron (Fig. 2E and Fig. S8E), and activation of cryptic, U2-type, 5′- and 3′-splice sites at the U12-type intron (Fig. 2F and Fig. S8E and F).

There were no differences in splice site consensus sequences between the U12-type introns that are misspliced in rgh3 and those with no significant difference in rgh3 mutants (Fig. S9). The range of splicing defects found in rgh3 is also observed in human bone marrow samples from MDS patients with ZRSR2 mutations (18). We conclude that rgh3 mutants, like human ZRSR2 mutants, are impaired in minor spliceosome function.

**Predicted Functions of rgh3 Misspliced Genes.** Pfam domains in the 230 maize genes with increased intron read depth in rgh3 showed no significant enrichment or deenrichment compared with all 308 expressed U12-type intron-containing genes in the RNA-seq experiment (Dataset S3). These analyses suggest most biological processes that are dependent upon U12 splicing are affected in rgh3. Many U12-type introns are found within genes involved in DNA replication, DNA repair, transcription, RNA processing, and translation (22, 23). *Arabidopsis* U12-type introns were used for prior cross-kingdom comparisons, but only 58% of maize U12-type intron-containing genes have an *Arabidopsis* ortholog that also contains a U12-type intron (21).

Pfam domains were analyzed at a global level to identify U12-dependent biological processes affected in rgh3 (Fig. 3A and Dataset S3). More than one-half of the domains with annotated roles in translation, endomembrane dynamics, and unknown functions were enriched in the 230 *rgh3* misspliced genes relative to all genes tested for intron read depth differences. When all U12-type intron-containing genes are compared genome-wide, more than one-half of domains with roles in cell cycle, RNA processing,
and protein folding/degradation also are enriched (Fig. 3 A). These analyses support conserved functions for maize U12-type intron-containing genes with U12 spliceosome targets in other species. Endosperm cells in rgh3 show aberrant differentiation of embryo surrounding region (ESR) and basal endosperm transfer layer (BETL) cells (19), which is analogous to blood cell differentiation defects in MDS patients (18). BETL and ESR differentiation occurs early in seed development. To examine developmental expression of rgh3-affected genes, we reanalyzed a transcriptome profile of maize seed development (24). Based on reported clusters of expression profiles, the 230 misspliced genes in rgh3 were overrepresented in early endosperm but not in early embryo development (Fig. 3 C and D). Approximately 78% of misspliced genes have a peak of expression between 6 and 12 d after pollination (Fig. 3 E). These data are consistent with U12-type intron retention in rgh3. (E) GRMZM2G083620 had adjacent U12- and U2-type intron retention. (F) GRMZM2G587327 activated cryptic U2-type splice sites in rgh3.

To explore mechanistic similarities between U12-dependent cell differentiation pathways in maize and human, we identified human homologs of maize genes with U12-type introns. BLASTP searches of human RefSeq proteins identified 233 maize U12-type intron-containing genes with a human homolog (Dataset S4). The 233 maize genes correspond to 154 human genes due to differences in gene redundancy in the maize and human genomes. We found two biological processes that account for 35% of the U12-type intron-containing maize genes with human homologs. First, 50 maize U12-type intron-containing genes have roles in cell cycle. Of these cell cycle genes, 29 are differentially spliced in rgh3, which correspond to 24 human homologs (Fig. 3 B). Second, 33 maize genes with U12-type introns have predicted roles in protein glycosylation including the following: synthesis and transport UDP-xylose, synthesis of dolichyl-diphosphooligosaccharide, secretion of glycosylated protein complexes, and the unfolded protein response pathway (Table S2). Twenty-five of the protein glycosylation genes are misspliced in rgh3 and correspond to 10 human homologs.

To find genes with altered splicing in both rgh3 and human ZRSR2 mutants, we identified the current gene symbols for human U12-intron–containing genes in U12DB (Dataset S5) (25). This revealed 36 human genes that are homologs of 57 maize genes with U12-type introns in both species. Maize/human homology enriches for U12 splicing defects in both rgh3 and ZRSR2 mutants (Table S3). Fifty of these maize genes were tested in the RNA-seq analysis, and 96% (48 of 50) had evidence of splicing defects in rgh3, as determined by intron read depth or junction read tests. Of the 758 human genes in U12DB, only 216 (28%) genes had significant splicing defects reported in ZRSR2 mutants, but 47% (17 of 36) of human U12-type intron-containing genes with maize homologs have splicing defects (18).
Mutations in three of these genes, GPN2, MAEA, and TAPT1, have documented roles in animal and plant development. The *Arabidopsis* homolog of GPN2 encodes the QQT1 protein and is required for embryos to complete periclinal divisions to establish epidermal and internal cell layers (26). The mouse ortholog of MAEA is required for final differentiation of erythrocytes (27). The vertebrate TAPT1 gene is required for skeletal patterning and normal function of the primary cilium (28). The *Arabidopsis* TAPT1 ortholog, *POD1*, is required for apical–basal patterning of the early embryo and endomembrane protein sorting (29).

**Fig. 3.** U12 splicing defects in *rgh3*-affected genes involved in cell differentiation and growth. (A) Heat map of Pfam domains found in U12-type intron-containing genes. Red domains are enriched in genes with *rgh3* U12 splicing defects relative to all maize genes tested for splicing defects. Blue indicates additional domains enriched in the domain relative either to genes tested for splicing defects or to all maize genes, respectively. (B) Cell cycle schematic showing human homolog gene symbols for maize U12-type intron-containing genes. Bold indicates U12 splicing defects in *rgh3*. Asterisks indicate splicing defects in human *ZRSR2* mutants. (C–E) Endosperm expression of *rgh3* misspliced genes (24). Cluster analysis of all genes (open triangles) and *rgh3* misspliced genes (blue squares) is plotted for embryo (C) and endosperm (D). E/L, early/late developmental expression. Const., constitutive expression. (F) Example endosperm expression profiles for maize/human homologs with U12-type introns. (G and H) Scatterplots of expression levels for misspliced genes with human homologs that also have U12-type introns (orange circles) and all other misspliced genes (blue squares). Dotted black line shows 1:1 ratio of WT to *rgh3* expression. Dotted gray lines show a twofold ratio change.

E2198 | www.pnas.org/cgi/doi/10.1073/pnas.1616173114
Gault et al.
Our analysis also revealed divergent loss of U12-type introns for genes with conserved biological functions (Datasets S4 and S5). The protein glycosylation enzymes, ALG12 and PG1B, have U12-type introns in both maize and human. U12-type introns are also found in other protein glycosylation genes including the following: ALG3, ALG6, ALG8, PIGN, and PGIP in human as well as the maize homologs of DOST, MPDU1, MGAT1, RFT1, and PIGM. The DNA origin of replication complex subunit, ORC3, in humans contains a U12-type intron, whereas the maize ORC4 homolog contains a U12-type intron. For ubiquitin-specific peptidases, U12-type introns are found in human USP7, USP10, and USP14 as well as in maize homologs of USP36 and USP42. Similarly, molybdenum cofactor (Moco) biosynthesis genes contain U12-type introns in maize. In humans, two Moco cofactor-requiring enzymes, aldehyde oxidase (AOX1) and xanthine dehydrogenase (XDH), contain U12-type introns. Loss of Moco in maize does not affect endosperm development but does cause seedling lethality (30, 31). There are at least 25 additional examples where different members of protein complexes or biochemical pathways have U12-type introns in human and maize including TRAPP and adaptor related protein complexes, Rab, importins, DNA polymerase, activating signal cointegrator-1, TFIIA, RNA polymerase III, the exosome, ribosomal proteins, and the autophagy pathway. These homologies suggest a common genetic architecture of minor spliceosome targets in human and maize where splicing defects in U12-type introns disrupt conserved biological processes and result in stem cell-like phenotypes in ZRSR2 mutant MDS cells and rgh3 endosperm cells.

Conservation of U12-Type Intron Positions in Maize and Human. Divergence of gene sets with U12-type introns in a given species is almost exclusively due to loss of U12-type introns or mutation toward a U2-type intron (22, 32). Protein alignments of the human and maize isoforms were used to identify U12-type intron positions for maize/human homologs (Fig. S10 and Dataset S4). Approximately one-half of the maize/human homologs (19 of 36 human genes and 26 of 57 maize genes) have at least one U12-type intron in a conserved position within the protein coding sequence (Fig. S11 and Table S3). The U12-type introns are randomly distributed within the coding sequences of both maize and human homologs (Fig. 3F). The intron positions fit the normal distribution (Shapiro-Wilk P > 0.05) with very little skew (human = −0.04; maize = −0.03) but have a low kurtosis (human = −0.53; maize = −0.91). The near-uniform distribution suggests a diversity of coding sequence impacts with most defective transcripts expected to trigger nonsense-mediated decay (NMD). However, rgh3 splicing defects do not appear to significantly alter transcript abundance. For the rgh3 misspliced genes, 86 and 79% of shoot and root transcripts showed a nonsense-mediated decay allele has a loss-of-function all disrupt the dynamic colocalization of U2AF1, RGH3 and U2AF2 had YFP signal in nuclear speckles, whereas U2AF1 did not show overlap with U2AF2 (Fig. 5D and Fig. S14 E and F).

Missplicing of E2F3 was identified as a candidate for mediating human myeloid malignancies observed in ZRSR2 mutant cells (18). U12-type intron retention in the maize E2F3 homologs truncate the C-terminal transactivation domain similar to the endogenous GRMZM2G041701_T02 isoform (Fig. 4F). The predicted E2F proteins from the intron retention transcripts contain all domains necessary for cell proliferation and the endocycle in plants (33). By contrast, the splicing defects observed in human E2F3 are primarily expected to be NMD targets, except for a transcript that skips exons 3–4, which flank the U12-type intron (18). If translated, the human exon skip transcript would produce a nonfunctional E2F3 protein lacking the DP-dimerization and transactivation domains. This difference in U12-type intron position may contribute to the contrasting cell proliferation phenotypes with rgh3 being proliferative in culture and ZRSR2 mutants arresting the cell cycle.

**Mutant RGH3 Proteins Disrupt Localization with U2AF2.** It is surprising that rgh3 disrupts a larger proportion of U12-type introns than ZRSR2 mutations in MDS patients. The rgh3 mutant is predicted to encode a weak allele, whereas most ZRSR2 mutations are loss-of-function alleles (14–16, 19). The rgh3-umu1 allele has a Mutator transposon insertion that partially splices from the transcript to delete 12 aa and insert 47 aa coded by the transposon sequence in the N-terminal acidic domain of the RGH3 protein (Fig. S4). The normal RGH3 allele produces multiple splice variants. Only the Rgh3a isoform encodes a full-length ortholog of ZRSR2 (19). Both the mutant and splice isoforms coding for protein truncations are likely to be expressed as protein in maize, because RGH3 antibodies cross-react with multiple proteins that migrate similarly to in vitro transcribed/translated cDNA clones of the alternatively spliced variants (Fig. S13). We investigated subcellular localization of RGH3 protein variants to gain additional insight into the nature of the rgh3 mutant allele.

Human ZRSR2 interacts with the major spliceosome subunit, U2AF2, through a U2AF homology motif (UHM) that is related to RNA recognition motifs but mediates protein–protein interactions (13, 34). These interactions are likely conserved in maize (19). Colocalization is observed when maize RGH3a and U2AF2 are transiently expressed in Nicotiana benthamiana as GFP and RFP fusions (Fig. S14A). U2AF2 is the large subunit of U2 auxiliary factor (U2AF), and maize U2AF2 also colocalizes with the small subunit of U2AF2, U2AF1 (Fig. S14B). These data are consistent with maize RGH3a and U2AF2 acting in the same subnuclear compartment. The mutant, RGH3a-umu1a, protein has aberrant subnuclear localization with low levels of diffuse signal in the nucleoplasm (Fig. 5B) instead of localizing to splicesomal speckles as seen for RGH3a (Fig. S14A). When RGH3a-umu1a is coexpressed with U2AF2, the proteins typically localize to different subnuclear compartments. Some cells showed partial overlap of RFP and GFP fusions, indicating reduced colocalization (Fig. S14C). An in-frame deletion of the UHM domain (RGH3aΔUHM) showed intermittent colocalization like RGH3a-umu1a (Fig. 5C and Fig. S14D). Splice isoforms coding for protein truncations of the R domain (RGH3a) or the UHM and RS domain (RGH3β, RGH3γ) did not show overlap with U2AF2 (Fig. 5D and Fig. S14 E and F).

Bimolecular fluorescent complementation (BiFC) assays of U2AF2 with RGH3a, U2AF1, RGH3a-umu1a, and RGH3aΔUHM all resulted in YFP signal in the nucleus (Fig. 5E–H and Fig. S15). Similar to colocalization experiments, U2AF2 and RGH3a showed YFP signal in nuclear speckles and the nucleolus, whereas U2AF1 and U2AF2 had YFP signal in nuclear speckles. By contrast, BiFC signals from U2AF2 with RGH3a-umu1a or RGH3aΔUHM appear aggregated in larger subnuclear foci. Reconstitution of YFP in BiFC assays is irreversible with transient interactions able to give stable YFP signal (35). These data show that U2AF1, U2AF2, and RGH3 colocalize as predicted from human protein–protein interaction studies. Mutations or truncations affecting the acidic, UHM, or RS domain all disrupt the dynamic colocalization of
Fig. 4. A subset of rgh3 misspliced transcripts may be translated. (A–H) RT-PCR analysis using RNA extracted from total, nuclei, and polysome fractions of normal and rgh3 samples. U12-type intron regions were amplified from the following: (A) GRMZM2G052515, (B) GRMZM2G097568, (C) GRMZM2G033430, (D) GRMZM2G177026, and (E) GRMZM2G093716. The larger two mRNA variants of Rsp31B (F) and mir156 premiRNA (G) are expected to be excluded from polysomes, whereas actin (H) serves as loading control. All transcript diagrams are drawn to the same scale as in A with U12-type introns indicated. Arrows show forward and reverse primers used for RT-PCR. (I) Transcript and protein diagrams for maize E2F3 genes and their human E2F3a homolog. Retention of the maize U12-type introns introduce premature termination codons (red octagon) that would have minimal effect on E2F protein domains. Retention and cryptic splicing of the U12-type intron in human are predicted NMD targets, whereas skipping exons 4–5 is predicted to produce a nonfunctional protein.
RGH3 with U2AF2 equivalently. Combined with the RNA-seq results, these localization experiments suggest that rgh3-umu1 is more likely a strong allele and indicate an important role for the acidic domain of ZRSR2/RGH3 in U12 splicing.

**Discussion**

**RGH3/ZRSR2 Are U12 Splicing Factors in Vivo.** Our data provide an independent genetic analysis of ZRSR2 function in a distantly related species from humans. RGH3/ZRSR2 has a conserved role in splicing U12-type intron-containing genes. ZRSR2 and rgh3 mutants exhibit U12-type intron retention, activation of cryptic 5′- and 3′-splice sites within U12-type introns, and retention of U2-type introns that are adjacent to misspliced U12-type introns.

A near-exclusive in vivo function for ZRSR2/RGH3 in U12 splicing contradicts biochemical experiments that conclusively show ZRSR2 copurifies and is required in both spliceosomes (12, 13). It is possible that copurification of ZRSR2 and U2AF in human cell extracts is due to independent binding of common pre-mRNA species. However, ZRSR2 interacts with U2AF2 in yeast two-hybrid assays (13). Yeast lacks U12-type introns, suggesting direct protein–protein interactions are more likely between these spliceosome subunits. In maize, colocalization of RGH3 with U2AF2 appears to be an indicator of RGH3 function in the minor spliceosome. Potentially, interactions between the major and minor spliceosomes promote efficient splicing of either class of intron.

**Minor Splicing as a Regulatory Process.** Minor splicing factors are at low abundance, and U12 splicing can be a rate-limiting step to produce protein-coding mRNA (36). For example, U6atac levels in HeLa cells affect splicing efficiency as well as expression level of genes with U12-type introns, and the minor spliceosome was proposed to regulate cellular responses to stimuli (4). Under this model, minor spliceosome activity determines the balance of coding mRNA vs. NMD or translation into alternative protein isoforms (4). We found little evidence of splicing defects resulting in expression level changes of maize U12-type intron-containing genes. Similarly, *Drosophila* U6atac mutants have little impact on U12-type intron-containing genes. Similarly, *Drosophila* U6atac mutants have little impact on U12-type intron-containing gene expression levels (37). In maize, reduced U12 splicing generally leads to predicted NMD targets being retained in the nucleus. A smaller subset of U12-type intron retention transcripts in rgh3 are associated with polysomes and are likely to be translated into alternative protein isoforms.

Although different molecular mechanisms seem to act downstream of U12 splicing efficiency in maize and humans, there are a substantial number of homologous genes with U12-type introns. We found 233 maize U12-type intron-containing genes with easily identified human homologs representing 154 human genes. Nearly 25% of the maize–human homolog pairs contain U12-type introns in both species (57 of 233 for maize genes or 36 of 154 human homologs). There are also overlapping functions among non-homologous genes subject to U12 splicing. These nonhomologous
overlapping genes are subunits of the same protein complexes, members of conserved gene families, or members of the same metabolic pathways. These overlaps suggest selection for specific biological processes, such as cell cycle and protein glycosylation, to be dependent upon U12 splicing efficiency and support the idea that the minor spliceosome could be regulatory.

Minor Splicing Is Required for Cell Differentiation. ZRSR2 and rgh3 mutants both disrupt cell differentiation programs. MDS patients with ZRSR2 mutations accumulate myeloid blast precursors (18). Maize rgh3 endosperm retains proliferative capacity and shows cell fate switching to aleurone in the basal endosperm transfer cell layer and the embryo-surrounding region (19). Mutations in other minor spliceosome factors lead to developmental defects in many species. In humans, Taybi–Linder syndrome is caused by mutations in ZRSR2 with the test with the Normal and mutations accumulate myeloid blast precursors (18).

Gault et al. Arabidopsis in cell differentiation becomes and the embryo-surrounding region becomes clear (18, 19). Extending to a more general interpretation of U12 splicing mutant phenotypes, the data suggest a role for U12 splicing to promote differentiation of a subset of cell types in both plants and animals. It is unlikely for U12 splicing to be needed in all differentiation processes, because the minor spliceosome has been lost in multiple eukaryotic lineages (2).

Conservation of Minor Splicing During Evolution. The losses of minor splicing during evolution raise the possibility that the minor spliceosome evolved in a convergent manner to function in plant and animal cell differentiation. However, there appears to be more selective pressure to maintain U12 splicing in multicellular eu-

karyotes. Although the minor spliceosome is missing in some lineages with multicellular development, a larger fraction of unicellular eukaryotic genomes have lost U12 splicing (2, 40). The few unicellular species that retain the minor spliceosome, such as Acanthamoeba castellani, tend to have amoeboid cellular organization (2, 40). The A. castellani genome sequence revealed a high frequency of horizontal gene transfer, including potential eukaryote-to-eukaryote gene transfer that would confound deep evolutionary comparisons (41).

Volvocine green algae illustrate convergent evolutionary innovation in multicellular development. Volvocines do not have a minor spliceosome and independently evolved multicellular species relative to higher plants (2, 42). Recent genome comparisons within this clade found that cell cycle genes are expanded in multicellular species with mutation of retinoblastoma as the primary change needed for unicellular species to become multicellular (41). In addition, genes encoding glycosylated extracellular matrix proteins are expanded in Volvox carteri, which has two distinct cell types. Both processes are highly represented in maize U12-type intron-containing genes. However, neither retinoblastoma nor the volvocine-type extracellular matrix proteins have U12-type introns in any species (21, 25).

Importantly, the U12-type introns affected in rgh3 and ZRSR2 mutants were maintained since the divergence of plants and animals (32). The number of U12-type introns in the last eukaryotic common ancestor is not known, but all lineages with a minor spliceosome have a small fraction of genes with U12-type introns. Assuming random loss of introns, the number of homologous genes that still have U12-type introns in human and maize is large, especially with one-half of the homologs having divergent U12-type intron positions. These observations argue for selective pressure to maintain a subset of genes with U12-type introns to carry out essential roles in cell differentiation.

Materials and Methods

RNA-Seq. Normal and rgh3 kernels from W22 inbred background were sown in soil (29 °C/20 °C, 12 h/12 h, day/night). Individual seedlings were harvested when about 5 cm high. Normal seedlings were genotyped to select homozygous WT individuals. Total mRNA was extracted from three biological replicates of rgh3 and WT roots and shoots using TRIzol Reagent (Life Technologies) with RNase-free glycerogen (fermentas) as an RNA carrier. Nonstrand-specific TrueSeq (illumina) CDNA libraries were built from 2 μg of RNA input with a 200-bp median insert length. Twelve libraries were pooled using the KAPA Library Quantification Kit (Kapa Biosystems) and sequenced on the HiSeq 2000 platform with 100-bp paired-end reads.

Read Mapping. Bases with a quality score <20 were trimmed with the FASTQQA utility in the FASTX toolkit (hannonlab.cshl.edu/fastx_toolkit/). Low-quality reads were discarded if >20% of bases had a quality score <20. Reads from repetitive elements in the MIPS database (43) were removed using Bowtie (44), version 0.12.7. Single-nucleotide polymorphisms (SNPs) between W22 and B73 RefGen_v2 genome were identified by mapping WT libraries to the B73 RefGen_v2 genome (45) (Dataset S6). Mosaic 2, version 2.13.33, was run with a hash size of 14, up to 100 hash positions per seed, and allowing 2% mismatch (46). Duplicate reads were removed using Picard, version 1.54 (broadinstitute.github.io/picard/). W22 SNPs were identified with FreeBayes, version 0.9.5, assuming 2% pairwise differences with B73 and requiring a 99% confidence statistic that an individual variant exists (47). FreeBayes parameters included at least 10 mapped reads with >40% of reads supporting the variant and at least one read having a minimum quality score of 20.

GatLib version 2012-11, with all SNP calls being reported, was used to identify individual intron SNPs. Transcript isoforms were assembled using Cufflinks, version 1.3.0 (20). Each library was assembled independently with a minimum intron size of 10 bp and an overlap radius of 10 bp. Isoform expression analysis was restricted to the ZmB73_5b filtered gene set. Differential splicing was tested for roots and shoots independently using Cuffdiff with the bias detection and correction algorithm and upper-quartile normalization.

Tests for Altered Splicing of Individual Introns. Nonredundant introns were identified from the ZmB73_5b annotation. HTSeq, version 0.6.1p1 (49), tallied intron counts, which were normalized by use of the number of transcript per million mapped reads (FPKM) of the gene model in each library. Differences in intron reads were determined by a two-tailed t test with the Welch modification for degrees of freedom with no correction for multiple testing. Genes that were not expressed in every RNA-seq library were excluded from the analysis. Genomic coordinates for U12-type introns were identified by BLASTN searches against the B73 RefGen_v2 genome using sequences from the ERISdb database (21, 45).

PSO was calculated from exon–exon junction reads and reads mapping to shortest nonredundant intron for each unique 5′- and 3′-splice site (50). Read counts were calculated with the intersectBed command from Bedtools, version 2.24.0 (51). Fisher’s exact test was calculated for each intron with the summed reads across all WT and all mutant libraries and a false-discovery rate of 0.05 using the Benjamini-Hochberg method.

RT-PCR. RNA from 12 d after pollination (DAP) whole kernels was isolated by freezing rgh3 and normal kernels in liquid nitrogen with extraction using TRIZol Reagent followed by RNeasy RNA cleanup with on-column DNase I digestion (Qiagen). Embryo, starchy endosperm, and basal endosperm transfer cell-enriched tissue was hand dissected from 12 DAP whole kernels at –18 °C (52). RNA from these tissues was extracted with the Arcturus PicoPure RNA Isolation Kit (Applied Biosystems) with on-column DNase I digestion. Endosperm cultures were established as described (19). RNA was extracted from endosperm culture, seedling roots, and seedling shoots using the RNeasy Plant
Gene-specific primers were designed to distinguish multiple splice variants from individual genes using one primer pair. For genes predicted by Cufflinks to have splicing differences in rgh3, annotated alternative transcripts along with raw read coverage depth, visualized by the Integrated Genome Browser (53), were used to target specific regions for RT-PCR. Amplification conditions consisted of the following: 1 min denaturation at 95 °C; 30–31 cycles of 95 °C for 1 min, 55 °C for 1 min, 72 °C for 30 s, and a final extension step of 72 °C for 5 min. RT-PCR products were cloned into the pCR4-TOPO vector for Sanger sequencing.

Conserved Protein Domain Enrichment Analysis. All predicted protein isoforms in the maize filtered gene set were queried with HHMMER 3.0 against the Pfam database using an inclusion threshold of 0.01 (54, 55). Hypergeometric P values were calculated for three samples and populations (Dataset S3): (i) rgh3 misspliced genes relative to all expressed U12-type intron-containing genes, (ii) rgh3 misspliced genes relative to all expressed maize genes, and (iii) all U12-type intron-containing genes relative to all genes in the filtered gene set. A gene was considered expressed if at least one intron was tested for normalized intron read depth differences between rgh3 and WT samples. Pfam domains were manually categorized into biological functions based on the annotation of each domain.

Human Homology Analysis. The B73 RefGen_v3 maize protein isoforms for the 408 genes in ERISdb with U12-type introns were downloaded from MaizeGDB (21, 56). Each protein isoform was queried against the human RefSeq protein database using BLASTP. Human homologs were defined as the best match with an identity cutoff of 50. Dataset S5, National Center for Biotechnology Information (NCBI) Gene database unique identifiers were retrieved for each human RefSeq accession along with the corresponding HUGO Gene Nomenclature Committee (HGNC) human gene symbols (57). Human gene symbols containing U12-type introns were identified using the “intron FASTA” query at U12DB to download the complete set of U12-type introns (25). ENSEMBL gene identifiers were parsed from the FASTA file, and identifiers that were still current were converted to NCBI Gene database unique identifiers and human gene symbols using DAVID (58, 59). Predicted protein sequences from retired ENSEMBL gene identifiers were used to query the human RefSeq protein database using BLASTP. Intron sequences that did not have an ENSEMBL identifier were queried against the NCBI human genomic plus transcript database using BLASTN. The complete list of U12DB intron identifiers with current NCBI Gene database unique identifiers and human gene symbols is given in Dataset S5. The human gene symbols from maize–human homologs were used to cross-reference the human gene symbols identified from U12DB to determine maize and human genes that both contained a U12-type intron (Table S3).

To determine U12-type intron positions relative to maize protein sequences, the 5′-exon at ERISdb was queried with BLASTX against the maize protein isoforms from the B73 RefGen_v3 annotation. Human protein isoforms were downloaded from the Consensus CDS protein set at the NCBI CCDS database (60). The 5′-exon sequence given at U12DB was used for pattern match searches of the consensus transcripts. The relative amino acid position of the U12-type intron in the maize and human homologs were determined using Clustal Omega (61). All maize protein isoforms for each U12-type intron were aligned along with protein isoforms that were truncated at the last complete codon in the exon 5′ of the U12-type intron. The U12-type intron position was scored as identical if the C-terminal residues of the maize and human truncated isoforms were at the same position in the alignment. Normalized U12-type intron positions were determined based on the amino acid residues for each protein isoform and then averaged by gene. When a protein had more than one U12-type intron, normalized positions were calculated independently for each intron. U12-type intron retention models were constructed for E2F gene family members using the first annotated transcript and protein isoform for each gene.

Isolation of Polysomes. Normal and rgh3 seedlings were collected at 3- to 5-cm stage. For each biological replicate, seedlings from two self-pollinated families were combined, flash frozen, and pulverized in liquid nitrogen. Total, nuclear, and polysomal RNA fractions were extracted from 2.5–3 mL of ground tissue using the conventional isolation of plant polysomes protocol as described (62). Initial pellets to clarify the tissue extract were saved as the nuclear fraction, and 500 µL of the clarified supernatant was saved as the total RNA fraction. RNA from all fractions was extracted with Qiagen RNeasy kits and treated with Ambion TURBO DNA-free kit (Thermo Fisher Scientific) following the manufacturer’s instructions. RT-PCR was performed from 1 µg of RNA using M-MLV reverse transcriptase (Promega) to synthesize cDNA as described (63). PCR cycles were optimized for each gene.

Subcellular Localization. The RGH3-GFP and U2AF2 fluorescent protein fusion constructs were previously described (19). Rgh3ΔUHM was cloned from Rgh3Δ by amplifying an N-terminal fragment from the start codon to the end of the first zinc finger using primers Rgh3ΔUHM-1 and Rgh3ΔUHM-2 (Table S4). A C-terminal fragment was amplified from the second zinc finger to the C-terminal stop codon using primers Rgh3ΔUHM-3 and Rgh3ΔUHM-4. The fragments were ligated by overlap extension PCR (64). U2AF2 and U2AF1 were amplified from leaf cDNA of normal 14-d-old seedlings. The RT-PCR products were subcloned into pENTR vector (Invitrogen) and then cloned into pB7-WGF2, pB7-RWGV, or pB7-FWGV (65, 66) as described (19). BiFC constructs were created by transferring coding sequences from pENTR vector clones to pSAT4-DEST or pSATS-DEST vectors (67, 68) by LR clonal reactions following manufacturer’s instructions (Invitrogen). The recombinant transcription cassette were digested with †CelI or †Scel and ligated into the pEPP-RC52 binary vector (67). Transient expression in Nicotiana benthamiana and microscopy analyses were completed as described with some modifications (19). Binary vectors were transformed into Agrobacterium tumefaciens strain GV3101. Protein expression was visualized 24–48 h after transient transformation. YFP was excited at 514 nm and detected with an emission band of 525–565 nm.

Protein Analyses. Rabbit polyclonal N-terminal RGH3 antibodies was raised against the synthetic peptide: SAEQLVDKVAQETPINFGE (BioSynthesis). Total protein from seed tissue was extracted as described (69) except that the fresh tissue extracts were cleared with a single filtration step. For the in vitro transcription/translation, ORFs from described Rgh3 variants (19) were amplified with Phusion polymerase (Finnzymes) using primers containing an N-terminal 6×His tag and restriction sites for Sgfp (Promega) and Pme1 (New England Bio Labs). PCR products were digested and ligated into pF3AWG vectors (Promega) and expressed using the TNT SP6 High-Yield Wheat Germ Protein Expression System (Promega) following the manufacturer’s instructions.

ACKNOWLEDGMENTS. We thank Yuqing Xiong and Byung-Ho Kang for technical assistance in dissecting endosperm tissue. This work was supported by National Science Foundation (NSF) Graduate Research Fellowship DGE-0802270 (to C.M.G.), NSF Awards MCB-1412218 and IOS-1031416, the University of Florida Genetics Institute Center for RNA Research, and the Vasil–Monsanto Endowment.

1. Braunschweig U, Gueroussov S, Flock AM, Graveley BR, Blencowe BJ (2013) Dynamic integration of splicing within gene regulatory pathways. Cell 152(2):1252–1269.
2. Dávila López M, Rosenblad MA, Samuelsson T (2008) Computational screen for splicing events in the Arabidopsis thaliana pre-mRNA. Plant Cell 20(12):3951–3962.
3. He H, et al. (2011) Mutations in U4atac snRNA, a component of the minor spliceosome RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal proteins. Genes Dev 25(6):599–604.
4. Kim WY, et al. (2010) The Arabidopsis pre-mRNA splicing component U4atac snRNA. Mol Cell 38(2):240–249.
5. Kim JW, et al. (2006) Human U2AF1 protein interacts with U2AF65 and SR proteins in splicing of pre-mRNA. Nature 446:397–400.
6. Markmiller S, et al. (2014) Minor class splicing shapes the zebrafish transcriptome during development. Proc Natl Acad Sci USA 111(8):3062–3067.
7. Tronchère H, Wang J, Fu XD (1997) A protein related to splicing factor U2AF35 that interacts with U2AF proteins in splicing of pre-mRNA. Proc Natl Acad Sci USA 94(11):5915–5920.
8. Vasil et al. (2017) PNAS | Published online February 27, 2017 | E2203
