Aberrant RNA splicing due to genetic incompatibilities in sunflower hybrids

Chris C. R. Smith, Loren H. Rieseberg, Brent S. Hulke, and Nolan C. Kane

1Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, Colorado 80309
2E-mail: chriscs@colorado.edu
3Department of Botany, University of British Columbia, Vancouver, BC VCR 2A5, Canada
4Edward T. Schafer Agricultural Research Center, USDA-ARS, Fargo, North Dakota 58102

Received February 5, 2021
Accepted September 1, 2021

Genome-scale studies have revealed divergent mRNA splicing patterns between closely related species or populations. However, it is unclear whether splicing differentiation is a simple byproduct of population divergence, or whether it also acts as a mechanism for reproductive isolation. We examined mRNA splicing in wild × domesticated sunflower hybrids and observed 45 novel splice forms that were not found in the wild or domesticated parents, in addition to 16 high-expression parental splice forms that were absent in one or more hybrids. We identify loci associated with variation in the levels of these splice forms, finding that many aberrant transcripts were regulated by multiple alleles with nonadditive interactions. We identified particular spliceosome components that were associated with 21 aberrant isoforms, more than half of which were located in or near regulatory QTL. These incompatibilities often resulted in alteration in the protein-coding regions of the novel transcripts in the form of frameshifts and truncations. By associating the splice variation in these genes with size and growth rate measurements, we found that the cumulative expression of all aberrant transcripts was correlated with a significant reduction in growth rate. Our results lead us to propose a model where divergent splicing regulatory loci could act as incompatibility loci that contribute to the evolution of reproductive isolation.

KEY WORDS: Alternative splicing, genetic incompatibilities, hybridization, quantitative trait loci, RNA-seq, speciation.

Gene expression in multicellular organisms involves alternative splicing, which uses the information encoded in a single gene to create an assortment of mRNA transcripts. The removal of introns and joining of exons is accomplished via the spliceosome—a complex consisting of hundreds of protein and RNA components (Will and Lührmann 2011)—and additional regulatory factors, for example, serine/arginine-rich (SR) proteins. Genetic regulation of splicing is highly complex, partly due to the sheer number of genes involved, but also due to its intricate epistatic networks (Munding et al. 2010; Julien et al. 2016; Pacini and Koziol 2018). It was previously thought that these regulatory networks were tightly conserved (Sorek and Ast. 2003; Ast 2004; Sugnet et al. 2004). However, genome-scale studies have shown that closely related populations often have divergent splicing patterns; that is, pre-mRNA transcripts are spliced differently between genotypes (Thatcher et al. 2014; Smith et al. 2018; Zhang and Xiao 2018; Huang et al. 2019; Khokhar et al. 2019; Lin et al. 2020). This raises the question: is splicing differentiation a simple byproduct of population divergence, or is it also a mechanism of reproductive isolation?

Theory predicts that genetic incompatibilities in hybrids may lead to reproductive barriers between species, and empirical research has shown that sterility and other issues in hybrids are sometimes caused by genetic incompatibilities (Coyne and Orr 2004). Such incompatibilities might in some cases cause misregulation of RNA splicing. The simplest prediction for splicing in hybrids is that an individual will inherit regulatory alleles from each parent and thus can create a set of transcripts resembling that of either or both parents (Fig. 1). This outcome is expected if splicing patterns segregate as populations diverge and do not play a role in reducing gene flow between taxa. On the other hand, it is possible that regulatory alleles that were benign in one parent genetic background may interact with alleles in the opposite parent genome (Bateson 1909; Dobzhansky 1934; Muller 1942),
causing a malfunction in splicing (Fig. 1). Specifically, the hybrid may produce an isoform not found in either parent due to unsuccessful splice site recognition or an incorrect splice site (Scascitelli et al. 2010). Similarly, the hybrid could fail to produce an important isoform that was spliced by both parents. It is reasonable to expect that an accumulation of splicing disorders would lead to a reduction in hybrid fitness.

Other links have been found between alternative splicing and species differentiation. Terai et al. (2003) showed that the putative pigmentation gene *hagoromo* has increased splicing diversity in cichlid species, in particular ones that have recently radiated, and different species exhibited fixed splicing differences. Likewise, the well known Agouti gene produced multiple isoforms in the skin of deer mice and one isoform was associated with light-sand camouflage (Mallarino et al. 2017). Of particular relevance, Scascitelli et al. (2010) found peculiar splice isoforms in poplar hybrids that were not present in either parent genotype; however, these observations were limited to relatively few isoforms and the genetic regulatory mechanisms have not been identified.

Here, we investigate splicing irregularities and genetic incompatibilities in the common sunflower, *Helianthus annuus*. Although wild and domesticated common sunflower are the same species, domesticated sunflowers vary in distinct clade in phylogenetic analyses; *F*<sub>ST</sub> between domesticated and wild *H. annuus* genotypes ranges from 0.32 to 0.46 (Park and Burke 2020). Plant domestication frequently includes the evolution of partial reproductive isolation between the domesticate and its wild progenitor (Dempewolf et al. 2012). Reproductive barriers between domesticated and wild sunflowers include a loss of self-incompatibility (Burke et al. 2002), making domesticated genotypes less likely to outcross, and a moderate reduction in hybrid fitness (Snow et al. 1998). In the current study, we surveyed the transcriptomes of 100 recombinant inbred lines (RILs) derived from a wild × domesticated cross. This survey revealed several splicing abnormalities in the hybrids and we subsequently mapped the genetic basis of the abnormal splice isoforms. Our findings help us understand how genetic incompatibilities contribute to the speciation process and how alternative splice forms evolve.

**Materials and Methods**

**PLANT MATERIAL AND RNA SEQUENCING**

The plants and raw sequencing data used in the current study were previously described by Smith et al. (2018). The inbred line HA 89 is commonly used for research and breeding (PI 599773; http://www.ars-grin.gov). The second parent line, Ann1238, was derived from material collected at Cedar Point Biological Station, Keith County, Nebraska. Each parental accession was represented by three seedlings. One hundred RILs were grown in greenhouse conditions to address stochastic variation in expression: Ann1238 pollen was donated in a cross with HA 89, and a single F1 was self-pollinated to produce F2s, which were each self-pollinated for six generations. Above-ground tissue was frozen in liquid
nucleotide, and total RNA was extracted following standard protocols. Non-normalized Illumina RNA-sequencing (RNA-seq) libraries were sequenced on a HiSeq2000 system. These sequencing data were downloaded from the Sequence Read Archive Bio-Project PRJNA417714.

IDENTIFICATION OF ALTERNATIVE SPLICING

The following data and analyses are presented for the first time. Custom scripts were used except when software is explicitly named. Our code is available at https://github.com/chriscrsmith/SunflowerAberrantSplicing. First, we built a de novo transcriptome assembly including all RILs and parental samples using the program Trinity (version 2.8.1) with default settings. De novo mode was used to avoid biased transcript assembly due to one parent genotype being more closely related to the reference. The Trinity pipeline infers alternative splice forms, which were the focus in all subsequent analyses.

After completion of the transcriptome assembly, the following steps were used to verify the assembled transcripts. All transcripts were aligned to the HA 412HO version 2.0 reference genome (Staton and Rieseberg 2019; Todesco et al. 2020), a domesticated sunflower genotype, using BLAST (McGinnis and Madden 2004). Blast hits with percent identity below 95% were ignored. We required the entire transcript to have aligned unambiguously to the same genomic region, but with up to 10% of the transcript length missing from the ends. Additionally, the genomic alignment was required to be at least 50% single copy—we delineated single copy sites by aligning whole genome sequencing data to the reference and visualizing the distribution of read depths per site (see Supporting Information). If isoforms that were assigned the same gene ID by Trinity aligned to different genomic locations, then we considered them to represent different genes.

Next, we used the following procedure to deal with erroneously inferred alternative splice forms identified by Trinity that were actually alternative alleles. We created a multiple-sequence alignment of all isoforms from a gene using MUSCLE (version 3.8.31), or, if only two isoforms, a pairwise sequence alignment using the EMBOSS (version 6.6.0.0) needle program. For each pair of transcripts in the alignment: if at least one insertion/deletion or substitution larger than our minimum exon size cutoff—25 bp—was present, we considered the transcripts to be alternative splice forms. If any shorter runs of differences, for example, a single nucleotide polymorphism (SNP), were present, we considered the transcripts to be alternative alleles of the same splice form. If both a >25bp run and a shorter run were present, we deemed the relationship ambiguous. After identifying alternative splicing events, we used the Trinity utility align_and_estimate_abundance.pl with default settings (bowtie2 aligner) to quantify the abundance of individual transcripts using the transcripts per million (TPM) measurement.

IDENTIFICATION OF SPLICING ABNORMALITIES IN HYBRIDS

Next, we applied the following filters to identify isoforms that were spliced exclusively in hybrids and absent in the parents, hereafter “novel” isoforms. (1) The novel isoform was expressed at >1 TPM in at least one hybrid. Because these isoforms were expressed at a substantial level, they are presumably distinct from stochastic splicing errors or noise. (2) All parent samples had zero reads aligning to the novel isoform. (3) There was at least one alternative splice form from the same gene as the novel isoform (>1 TPM) in every parent. The latter filter was used to ensure that the examined genes were expressed at a substantial level in the parents.

Separately, we identified isoforms that were expressed at a substantial level in the parents, but were absent in one or more hybrids. The following filters were used to identify missing isoforms: (1) the isoform was absent in one or more hybrids, but (2) the hybrid must have >1 TPM total expression from the alternative isoform(s) of the same gene. And, (3) each isoform from the corresponding gene must be expressed at >1 TPM in all parent samples. We refer to the latter cases as “missing isoformsm” and we use “aberrant isoforms” to describe both novel isoforms and missing isoforms together.

TESTING FOR SPLICING DIFFERENTIATION

To test for divergent splicing composition between wild and domesticated sunflower genotypes, we first applied an isometric log ratio transformation to the proportions of each isoform. Then we applied a t-test, or a MANOVA in cases with more than two parent isoforms.

GENE ANNOTATION

Transcripts were aligned to the Araport11 table from The Arabidopsis Information Resource (Berardini et al. 2015) using BLASTX (McGinnis and Madden 2004) with e-value < 10^-20 cutoff. The best hit was retained for each transcript.

OPEN READING FRAMES

The program ORFfinder (National Center for Biotechnology Institute website; www.ncbi.nlm.nih.gov/orffinder/) with default settings was used to identify open reading frames (ORFs) for each aberrant isoform and all parent isoforms corresponding to the same gene.

GROWTH RATE EXPERIMENT AND ANALYSIS

HA 89 × Ann1238 RIL seedlings, and inbred parental lines, were grown in a common environment in greenhouses at the
University of British Columbia in the spring of 2011, with plants watered and fertilized daily. Plant height, leaf lengths and widths, and other measurements were taken at 4 and 6 weeks after germination. At 6 weeks, watering was stopped entirely, with daily observations checking for wilting and death of each plant. Soil moisture content was measured with a HydroSense CS620 soil moisture system (Campbell Scientific) at 6 days after the initiation of the drought and on the day of first observed wilting for each plant. Plant measurements from this experiment are in Table S2.

The dimensionality of the 11 plant size measurements was reduced using a principal component analysis. We modeled novel isoform splicing using a cumulative novel splicing value for each RIL; this value was obtained by finding the proportion of each novel isoform spliced relative to the overall expression of the gene, rescaling the proportions to the range \([0, 1]\), and then averaging among isoforms. Likewise, a cumulative missing isoform value was obtained for each RIL. Last, we used the same procedure to calculate a cumulative aberrant splicing value, which averaged across all novel isoforms and missing isoforms together.

**SINGLE NUCLEOTIDE VARIANT IDENTIFICATION AND FILTERING**

SNPs were identified by aligning RNA-seq reads to the HA 412HO version 2.0 transcriptome (Renaut et al. 2013) with bwa mem version 0.7.15, with default settings. SNPs were called using SAMtools version 1.4.1 with the multiallelic caller option and otherwise default settings. We filtered SNPs that aligned to non-single-copy genomic regions (see Supporting Information) and marked genotypes with fewer than 10 read depth as missing data. After the initial filters, we kept 66,893 SNPs that showed fixed differences between wild and domesticated parents. Next, we obtained genetic map positions for each fixed SNP by aligning transcriptome contigs to the reference genome using BLAST (\(e\text{-value} \leq 10^{-20}\); \(ID \geq 90\%\)) and linearly interpolating cM positions from the HA 412HO version 2.0 genetic map (Staton and Rieseberg 2019; Todesco et al. 2020). Markers that aligned well were examined in the RILs, and genotypes with fewer than 10 reads were marked as missing. Markers were excluded if present in fewer than 35% of RILs, resulting in 13,874 filtered SNPs. Last, we applied a conservative imputation step to the filtered SNPs: if a marker with missing data occurred between two markers with the same genotype that were within 10cM, we assigned the same genotype to the missing marker. Otherwise, missing genotypes were left as missing. This step filled in 31% of missing data, and 1.4% of genotypes overall.

**SINGLE-QTL SCAN**

To identify potential regulatory loci involved in aberrant isoform splicing, we conducted a single-QTL scan using R/qtl with standard interval mapping and a nonparametric model. The mapped phenotype value was the proportion of aberrant isoform spliced relative to overall gene expression. The densest genetic map grid (step = 0) was used, and the genotyping error rate was set to zero (error.prob = 0) because the genotypes were already carefully filtered (above). We generated a null distribution of LOD scores for each aberrant isoform being mapped by permuting the phenotypic measurements relative to the genotypes: for 10,000 iterations, the splicing measurements were shuffled before applying the QTL scan, and the largest LOD score was recorded from each permutation; the 95th percentile of this distribution was used as the significance threshold (Broman and Sen 2009). We allowed a single QTL per chromosome and assigned the QTL peak as the marker position with the highest LOD score. We delineated the QTL region as the range of SNP markers with LOD scores within 25% of the peak. The same procedure was used to identify QTLs associated with isoforms missing in the hybrids.

For each identified QTL, we calculated the degree of dominance at the QTL peak. The degree of dominance is the ratio of the dominance effect to the additive effect, the former being the difference between the mean phenotype of the heterozygote group and the midpoint between the mean phenotypes of each homozygote group, and the latter being half the difference between the mean phenotypes of the two homozygous groups (Kenney-Hunt et al. 2006; Ishikawa 2009). A degree of dominance equal to zero means perfect codominance, \(0–1.5\) means one allele is dominant, and greater than 1.5 was interpreted as overdominance. Here, “overdominance” describes a phenotype that is more pronounced in heterozygotes compared to either homozygote (Young 1996; Rieseberg et al. 1999; Renaut et al. 2009; Stelkens and Seehausen 2009): either a greater proportion of novel isoforms or a greater proportion of missing isoforms. However, “overdominance” is frequently used to describe heterozygote advantage, implying a positive fitness effect. These two definitions are not interchangeable in our case; we will use the first definition.

**EPISTASIS SCAN**

Independent from the single-QTL scan, we next scanned the genome for pairs of loci where alleles from each parent interacted to produce the aberrant isoform. The proportion of aberrant isoform spliced relative to overall gene expression was used as the phenotype value, and if an individual had insufficient expression (<1 TPM) at the gene level they were assigned missing data. We limited the search to pairs of loci on different chromosomes that met the following criteria: (i) at least three individuals were represented that had nonmissing phenotype data and were homozygous for the HA 89 allele at both loci—notated AAbb, (ii) at least three individuals were represented that were homozygous for the Ann1238 allele at both loci—aabb, and (iii) at least three
individuals were represented for at least one mixed genotype—
aabb, AABB, AAbb, AaBb, AAaB, or AaBb (upper- and
lower-case alleles do not imply dominance or recessiveness,
here). Encoding the genotypes at each locus as allele dosages—
0, 1, or 2, we tested for an interaction effect between markers
using a multiple linear regression. The same procedure was used
to identify epistasis associated with isoforms missing in the hy-
brids.

We used a Bonferroni correction to obtain a signifi-
cance threshold for the epistasis scan: the total number of tests for all 45
novel isoforms, plus 16 missing isoforms, was $4.2 \times 10^6$, and the
significance threshold became $P < 1.19 \times 10^{-11}$. Last, we applied
the following post hoc filter to the aberrant isoform cases to avoid
outlier effects: if the genotype group with the most aberrant splic-
ning had fewer than three individuals represented, we counted an
otherwise significant test as nonsignificant. We allowed a single
QTL pair for each pair of chromosomes and assigned the QTL
peak as the marker position with the smallest $P$-value. We delin-
eated the QTL region as the range of SNP markers with $P$-values
within 25% of the peak in log-space: $P < P_{\text{min}}^{(1-0.25)}$. We used a
custom scan because the R/qtl program did not meet the specific
needs of our dataset (see Supporting Information); specifically,
false positives were encountered due to outlier individuals when
using R/qtl.

ASSOCIATIONS WITH INDIVIDUAL SPLICEOSOME
COMPONENTS

Sequences for Arabidopsis thaliana spliceosome protein and
RNA components were obtained from KEGG and arabidop-
sis.org. Spliceosome sequences were aligned to our sunflower
transcriptome using TBLASTN (McGinnis and Madden 2004),
and the best hit with $e$-value $< 10^{-20}$ was retained for each com-
ponent. For each identified spliceosome component, we used a
simple linear regression to test whether the expression level of
the spliceosome component was associated with splicing of each
aberrant isoform. These statistical tests modeled spliceosome ex-
pression as a continuous variable (TPM), although for the pur-
pose of visualization we plotted spliceosome expression as cate-
gories (“with” or “without”). The total number of tests was 5795
including both novel isoforms and missing isoforms. With a Bon-
ferroni correction, the significance threshold for this analysis was
$P < 8.63 \times 10^{-6}$.

Once associations were identified between spliceosome component expression and individual aberrant isoforms, we parti-
tioned the variance explained by each factor using a linear model.
Each model included the aberrant splicing proportion as the re-
sponse, and each spliceosome component and QTL associated
with the aberrant isoform as predictors. Interaction terms were
included for each combination of spliceosome component and
QTL. We partitioned the variance explained using eta-squared
and type three sum of squares.

Results

CHARACTERIZATION OF NOVEL SPICE FORMS IN
HYBRIDS

After initial filtering, 15,674 different mRNA isoforms were
retained in the parental lines, and 18,836 isoforms were ex-
pressed in the offspring. These numbers do not represent the
total transcript diversity, as stringent filters were used. We
identified 45 isoforms that were spliced in at least one hybrid at
a substantial level, were absent in the parents, and represented
alternative splice forms of a corresponding parent isoform
(Fig. 1; Table S1). All exons and retained intron sequences
in the novel transcripts were present in the reference genome;
this check supports that these transcripts represent alternative
splice forms instead of sequence mutations, for example, indels.
Additional splicing errors potentially exist, but these represented
the most convincing examples of novel splicing in the current
dataset.

The number of novel isoforms spliced per gene was some-
times greater than one: three genes each had two novel isoforms
(isoforms 4, 5; 9, 10; and 35, 36), and a single gene had three
novel isoforms (isoforms 37, 38, 39), for a total of 40 genes with
novel splicing. Most novel isoforms were relatively uncommon:
only 20% were spliced in more than half of the hybrids (Fig. 2).
The proportion of alternative splicing composition at the gene
level was usually low for each novel isoform (mean $= 0.11$; SD
$= 0.12$), although three novel isoforms were spliced at high levels
(>60%) in some hybrids. All examined hybrids spliced at least
five different novel isoforms, but some had up to 18 different
novel isoforms.

Genes with novel splicing patterns were distributed across
cross 17 chromosomes. The length of novel transcripts ranged from
368 to 4783 bp (median $= 1431$ bp), which was longer than the
most similar parent isoform in 95.6% of cases (Fig. 2). Genes
with novel splicing patterns had higher splicing complexity than
the genomic background ($P= 2.6 \times 10^{-7}; t= 6.2$), with up to 11
isoforms in the parents (median $= 4$). However, higher splicing
complexity was not caused by divergence: the majority of genes
with novel splicing had similar splicing patterns between wild
and domesticated parents, although 25% were differentiated
between parent taxa. This indicates that divergence in splicing of
a particular gene is not a prerequisite to produce novel isoforms.
Intron retention was the most common splicing type in the novel
isoforms, which is consistent with general alternative splicing
in plants (Wang et al. 2006; Chamala et al. 2015; Ullah et al. 2018).
CHARACTERIZATION OF MISSING ISOFORMS IN HYBRIDS

In addition to hybrid-only isoforms, we identified 16 isoforms that were expressed in the parents at a substantial level but missing in at least one hybrid. As before, additional splicing errors potentially exist, but these represented the most convincing examples of missing isoforms in the current dataset. Most genes with missing isoforms in the RILs had exactly two isoforms, but two genes each had three isoforms. In one of the two-isoform genes, both isoforms were missing in the RILs. Genes with missing isoforms were distributed across 10 different chromosomes and ranged in length from 205 to 1575 bp (median = 680.5). As with the novel isoforms, only a small number of hybrids failed to splice each of the missing isoforms (Fig. 2). In the missing isoforms, intron retention was the most common “basic” splice type (Fig. 2).

ORF ANALYSIS

In mRNA, an ORF is a stretch of codons that begins with an initiation codon and terminates with a stop codon (Min et al. 2005), or is bounded by stop codons (Claverie 1997; Sieber et al. 2018); these segments can be translated into protein. Thus, identifying ORFs, or lack thereof, is important for understanding the biological function of nucleotide sequences. Importantly, if the mRNA sequence is changed—for example, through nonsynonymous SNPs, indels, or splicing—the corresponding protein function can be affected, especially if the altered ORF has undergone a frameshift or truncation.

The longest ORF had been altered in most novel transcripts. Specifically, 31 novel isoform ORFs were different versions of the corresponding parent ORF, or different than all parent ORFs if the parents spliced multiple isoforms. Twenty-two were shorter than the most similar parent ORF, eight were longer, and one novel ORF was the same length as the parent version but differed in a six amino acid sequence on the 5′ end (Fig. 2). Twenty-five of the novel ORFs differed by >25 amino acids compared to the most similar parent ORF. In 13 novel splicing cases, an intact parent ORF was present in the novel isoform. Two novel isoforms contained SNPs that led to nonsynonymous amino acid changes, but did not cause a premature termination codon. In the remaining—and shortest—gene with novel splicing, no ORF of substantial length was detected in the parent isoform or the

Figure 2. Characterizations of the identified novel isoforms (top row) and missing isoforms (bottom row). Splice type categories are described for the aberrant isoform relative to the most similar parent isoform determined by sequence alignment. The splice type “multiple” refers to cases that involve more than one of: intron retention, exon skipping, or alternate splice site.
novel isoform. We did not attempt to quantify the effect of these frameshifts on protein function.

Among the isoforms that were missing in some hybrids, only one shared a longest ORF with the alternate isoform (Fig. 2). This indicates that the alternative splice forms within each gene probably have different biological functions. Twelve missing isoforms had shorter ORFs than the alternate splice form, and two had longer ORFs than the alternate splice form.

**FUNCTIONAL ANNOTATION OF ABERRANT TRANSCRIPTS**

Novel isoforms had sequence homology with—but were not identical to—35 Arabidopsis proteins (Table S1). The annotations spanned a diverse array of functions, including the critical roles of cell membrane transport, sucrose biosynthesis, the electron transport chain, glycolysis, ribosomal RNA, and DNA replication. No functional category appeared predominant. Missing isoforms were homologous with 11 Arabidopsis proteins, also with a diverse array of functional annotations (Table S1).

Most novel isoforms aligned to the same Arabidopsis splice variant as one of the corresponding parent isoforms. This result is consistent with the novel isoforms being previously characterized, although the protein database is not comprehensive. Only two novel isoforms aligned to different proteins than the parent isoform. In one of these cases, the novel isoform aligned to a different Arabidopsis splice variant than the parent isoform, but they had the same annotation: uncharacterized protein family UPF0016, a suspected calcium transporter (Demaegd et al. 2014). In the remaining case, the novel isoform aligned to a protein with a different accession than the parent isoform, but only a subtly different functional annotation. The parent isoform was annotated as an F-box/RNI-like superfamily protein, whereas the novel isoform was most similar to the RNI-like superfamily protein.

**ABERRANT SPLICING ASSOCIATED WITH STRESS RESPONSE**

We evaluated potential relationships between aberrant splicing and 15 measurements related to hybrid stress response. The measurements included: (i, ii) the number of days until wilting and the number of days until death after watering ceased; (iii, iv) soil water content at two timepoints; and (v–xv) a set of correlated measurements related to plant size including: height at three timepoints, dry and wet weight, longest leaf length and width at two time points, and most recent fully expanded leaf length and width.

The first principle component for plant size (88% of variance explained) was correlated with novel splicing using a simple linear regression ($P = 1.4 \times 10^{-2}$; Fig. 3). This analysis predicted the seedlings with the highest levels of novel splicing to be 47% smaller (dry weight) than the healthiest seedlings, although novel splicing only partly explained variation in plant size. Plant size was not correlated with the proportion of missing isoforms in the RILs ($P = 9.0 \times 10^{-2}$). However, after combining the missing isoforms and novel isoforms, plant size was correlated with aberrant splicing ($P = 8.0 \times 10^{-3}$; $R^2 = 0.07$) using the first principal component. No individual isoform splicing levels had significant correlations with plant size after correcting for multiple tests. Genome-wide heterozygosity was not correlated with plant size, and overall gene expression for the genes with aberrant splicing did not predict plant size.

**INTERACTIONS BETWEEN PARENT ALLELES UNDERLY ABERRANT SPLICING**

We first conducted a single-QTL scan to identify potential regulatory loci involved in aberrant isoform production. This scan found QTLs associated with 19 novel isoforms and 13 missing isoforms. Most QTLs were associated with a single aberrant splice form, but 10 QTLs each regulated two to five isoforms. The mean variance explained by each QTL was 34.7% (SD 25.3%). The genomic position for most QTLs (78.6%) was more than 5 cM away from the gene with aberrant splicing.

For each of the identified QTLs, we checked whether the heterozygous genotype was associated with aberrant splicing. Eleven novel isoforms and four missing isoforms had QTLs with overdominance effects: individuals that were heterozygous at the QTL had a higher proportion of the novel isoform—or a smaller proportion of the missing isoform—than either homozygous genotype group (Fig. 4; Table S1). This indicates that alleles from each parent interacted nonadditively to produce the aberrant splice type.

Figure 3. Stress response phenotypes versus aberrant splicing proportion. X-axes show the average proportion of novel isoform spliced, and the average missing isoform proportion.
epistasis \((1.8 \times 10^{-4} < P < 3.8 \times 10^{-2})\). There were no QTL-pairs where alleles from different parents combined additively to cause aberrant splicing.

Separate from the single-locus scan, we next scanned the genome for pairs of loci where the effect of one locus on aberrant splicing depended on the genotype at the other locus. Fourteen novel splice forms could be explained by interacting regulatory QTLs \((2.2 \times 10^{-20} < P < 1.7 \times 10^{-11}; 0.44 < R^2 < 0.92;\) Fig. 4; Table S1). There were no significant hits from the two-locus scan affecting missing isoforms. Epistatic QTLs were distributed across all chromosomes except Chr. 4, and we did not find interactions involving cpDNA. One epistatic QTL on Chr. 13 (\(\sim 72\) cM) was associated with six novel isoforms. Most epistatic QTLs were not near the gene being spliced, although 16.7% of interactions involved a QTL within 5 cM of the gene with novel splicing. Looking back to the set of QTLs identified in the single-locus scan, three single-QTLs had genomic positions that overlapped with an epistatic QTL from the two-locus scan affecting the same isoform. This suggests that some QTLs identified in the single-locus scan may have been detected due to interactions with other loci.

**INTERACTIONS INVOLVING THE SPLICEOSOME**

The spliceosome is a protein and snRNA complex that is responsible for the removal of introns and splicing together of exons during alternative splicing. With this in mind, we carefully inspected spliceosome expression levels to see whether individual protein or RNA components induced splicing abnormalities in sunflower hybrids. We found associations between individual spliceosome component expression levels and the splicing of 15 novel isoforms and six missing isoforms (Fig. 5; Table S1). An association suggests that the individual spliceosome component may regulate splicing of a particular transcript; however, we did not experimentally verify any statistical associations. The identified spliceosome genes did not overlap with the genomic positions of our focal isoforms. Although most isoforms had only one spliceosome component association, nine aberrant isoforms each had more than one association. Likewise, some spliceosome components were used by more than one aberrant isoform. For example, the expression of splicing factor PRP38 was negatively correlated with four aberrant splicing events.

In 36.4% of spliceosome component associations, the genomic location of the spliceosome gene overlapped with a QTL.
associated with the same gene; in these examples, it is likely that cis-regulatory elements inside the detected QTL region directly regulate the nearby spliceosome component. Another 21.2% of associated spliceosome components were located within 5 cM of the corresponding regulatory QTL. Among the remaining spliceosome associations, where the spliceosome component fell on a different chromosome than the corresponding QTLs, 61.5% had a statistical interaction with an associated QTL \((1.9 \times 10^{-6} < P < 1.5 \times 10^{-2})\), that is, the effect of the spliceosome expression level depended on the genotype at the QTL (Fig. 5; Table S1). Looking back to the parents, one of the spliceosome components that interacted with a regulatory QTL had differentiated gene expression between parents despite low statistical power \((P = 1.3 \times 10^{-3}; n = 6)\); this component was associated with three novel isoforms and interacted with two QTLs. This suggests that the parents have heritable variation in spliceosome expression that may contribute to genetic incompatibilities in the hybrids.

Although the effects of spliceosome associations and QTLs were interdependent in most cases, there were five spliceosome components that did not interact with QTLs and were not physically linked to QTLs. For these, we examined which type of regulatory factor—(a) the QTLs or (b) the associated spliceosome component—explained more variation in aberrant isoform splicing. In one case, the spliceosome component was more tightly associated with the missing isoform than the QTL was (17.4% variance explained vs. 4.5%); this pattern of variance propagation suggests that the QTL affects splicing less directly, perhaps through regulation of the spliceosome. In a second case, the spliceosome component was the most important individual factor in the model, but still explained less variance than the associated pair of epistatic QTLs (20.6% variance explained vs. 26.6%). The remaining three spliceosome components did not have significant effects after controlling for QTL effects. The latter three spliceosome components may play a smaller role in splicing of the focal isoforms. Alternatively, additional regulatory components may have been missed in our scans.

**Discussion**

**IRREGULAR SPLICING PATTERNS CAUSED BY GENETIC INCOMPATIBILITIES**

In our study there were 45 novel mRNA isoforms found in sunflower hybrids and 16 highly expressed parental isoforms that were missing in one or more hybrids. Although the final set of novel isoforms represent likely candidates for hybrid-only splicing, it is possible that these isoforms were expressed at a low, undetectable level in the parent lines. The converse is true for the missing isoforms. Increasing the number of biological replicates of a genotype would increase confidence that a transcript is missing in the parents or hybrids. Similarly, many wild and cultivated sunflower individuals will need to be sampled from each population to learn about the evolutionary novelty of these transcripts.

Aberrant splicing in individuals has been observed previously, and such errors are thought to be mitigated by nonsense-mediated decay (NMD; Brogna and Wen 2009; Shaul 2015). Furthermore, new transcripts are continually discovered and catalogued in public databases for model organisms (e.g., Mezlini et al. 2013). However, the only study to our knowledge that has reported novel isoforms resulting from hybridization is Scascitelli et al. (2010). Our study is one of the first to report missing isoforms resulting from hybrid incompatibilities.

It is not uncommon for abnormal splicing of a gene to be harmful: 10–15% of mutations underlying human diseases affect splicing (Nissim-Rafinia and Kerem 2002). Several genes with splicing abnormalities in sunflowers had critical functional annotations, for example, sucrose biosynthesis, glycolysis, ribosomal RNA, and DNA replication (Table S1), although experiments are needed to evaluate the precise function of these transcripts. If aberrant splice forms in hybrids disrupt the normal function of these genes, we would expect hybrid fitness to be affected to some extent.

Most novel isoforms had compromised ORFs. Frameshifts in the novel mRNAs are most likely costly for the organism, for example, by affecting protein stability or binding domains, or by causing more serious disorders. NMD may reduce some of these effects; however, plant transcripts are often insensitive to NMD (Kalyna et al. 2012; Göhring et al. 2014; Shaul 2015). Successful NMD in plants depends on—among other factors—delicate control of the NMD factor \(UPF3\) by its feedback loop and restriction of its transcription, which likely varies between cell types and conditions (Degtiar et al. 2015). Even when an aberrant transcript is ordinarily tagged for decay, some amount of transcript may escape decay depending on tissue, life stage, and environmental conditions. Successful NMD itself comes with a small energetic cost. Even if the longest ORF was unchanged, the stability or translation may have been affected in the altered transcript. Noncoding mRNA or nonfunctional proteins may have inherent regulatory function through limiting the production of functional proteins from the same gene (Keren et al. 2010); three novel isoforms in the current study were the predominant splice form in some hybrids, displacing the parent isoform to a large extent. Any of these alternate possibilities would likely cause the erroneous transcripts to exert some metabolic or other costs. This is borne out by our observation that, although no individual splicing error affected sunflower development substantially, the accumulation of many splicing errors was associated with a measurable reduction in growth rate in seedlings.

Similarly, most of the missing isoforms had different ORFs than the alternate splice form, indicating that each missing
isoform may have a unique function. The missing isoforms were expressed at a substantial level in both parent genotypes; therefore, the hybrids who fail to splice these isoforms could be lacking a biologically important transcript. However, this effect might be alleviated if the function of the missing isoforms is redundant with other genes.

The most common type of genetic incompatibility underlying the observed novel isoforms was epistatic interaction between parent alleles. The conservative Bonferroni correction was used in our epistasis scan; therefore, we expect there to be additional cases of epistasis that were missed by our analysis. Examples of epistasis included interactions between particular spliceosome components and other regulatory loci. This finding is consistent with the Dobzhansky-Muller (DM) model of hybrid incompatibility describing that deleterious epistatic interactions accumulate as lineages diverge (Bateson 1909; Dobzhansky 1934; Muller 1942; Orr 1995; Orr and Turelli 2001). It is possible that additional interactions were missed due to complex epistasis involving more than two loci, which we did not check, but is expected to be the more common form of hybrid incompatibility (Orr 1995; Presgraves 2010). Moreover, an incompatible allele may have interacted with the opposite genomic background in a diffuse manner involving many small interaction effects.

Most examples of DM incompatibilities from empirical research have been observed during the later stages of speciation and have large effects that individually cause substantial inviability or sterility in hybrids (Coyne and Orr 2004; Presgraves 2010). The heavy representation of large-effect DM incompatibilities in the literature may be partly due to the inherent muddiness in describing which loci contribute to speciation early on, before significant reproductive isolation has progressed. In fact, incompatibilities may not cause devastating effects on hybrid fitness (Orr 1997). The relatively small, quantitative fitness effects observed in the current study represent an important aspect of reproductive isolation that is usually not detectable. Another recent study, by Schumer et al. (2014), has shown that hundreds of small-effect incompatibilities can accumulate between lineages with incomplete reproductive isolation. Similarly, epistatic interactions with negative fitness effects can occur within species due to segregating variation (e.g., Corbett-Detig et al. 2013), called “variable reproductive isolation” by Cutter (2012). Therefore, reproductive isolation could arise through divergence in frequencies of many preexisting, latent incompatibilities (Corbett-Detig et al. 2013). These findings warrant additional research to explore whether many weak incompatibilities distributed throughout the genome can restrict gene flow between species (Orr 1997; Corbett-Detig et al. 2013; Schumer et al. 2014). Although speciation researchers often test for epistasis associated with a single hybrid inviability phenotype, for example, fecundity, aberrant splicing may provide numerous potential targets for detecting genetic incompatibilities. Also, expression QTLs tend to have a high signal to noise ratio, therefore we have improved statistical power to detect eQTLs and splicing QTLs. Through studying splicing abnormalities, we might gain precision for studying the accumulation of incompatibilities at various stages of the speciation continuum, in particular the early stages of divergence.

In genome-scale analyses that uncover tens or hundreds of negative epistatic effects in the form of statistical interactions between loci, it can be difficult to know how each genetic incompatibility manifests on a cellular level. We want to draw attention to aberrant RNA splicing as a possible mechanistic basis for hybrid inviability caused by DM, and other types of, hybrid incompatibilities. It is intuitive that incompatibilities might affect splicing, because alternative splicing ordinarily relies on epistatic gene networks (Munding et al. 2010; Julien et al. 2016; Pacini et al. 2018) involving hundreds of regulatory loci for the assembly and regulation of the spliceosome (Will and Lührmann 2011). An incompatibility involving any of these loci could disrupt the regulation of RNA splicing. In addition, an interaction that affects the spliceosome could theoretically disrupt splicing of many downstream loci and thereby reduce hybrid fitness. Our findings may align with this model because several of the detected QTLs in the current study overlapped with or had statistical interactions with individual spliceosome components, one of which had divergent gene expression in the parent taxa. Future research is warranted to investigate splicing malfunctions in other hybridizing taxa, particularly in wild populations or species.

**SPlicing EVOLUTION DURING SUNFLOWER DOMESTICATION**

The cultivated sunflowers’ transition to domestication began approximately 5000 years ago (Baute et al. 2015; Badouin et al. 2017). This evolutionary shift was accompanied by a reduction in genetic diversity and differentiation from wild *H. annuus* (Park and Burke 2020). Additionally, domestication was associated with changes in mRNA splicing patterns, particularly for genes involved in seed development (Smith et al. 2018). Thus, our theory for the role of alternative splicing during sunflower domestication is multilayered: first, divergence at splicing regulatory loci has the capacity to affect mRNA splicing for a large swath of downstream transcripts. Second, genetic incompatibilities that affect splicing may enhance reproductive isolation. In theory such incompatibilities could explain reduced fitness in wild-crop sunflowers hybrids; these hybrids have fewer branches, flower heads, and seeds than wild plants (Snow et al. 1998). Consistent with this idea, a sucrose biosynthesis gene was mis-spliced in one or more hybrid offspring in the current study, and aberrant splicing was correlated with a reduction in plant size. We noted that aberrant splicing at a particular locus did not require divergence in splicing between parent genotypes at the same time. Therefore, the hybrids who fail to splice these isoforms could be lacking a biologically important transcript. However, this effect might be alleviated if the function of the missing isoforms is redundant with other genes.

The most common type of genetic incompatibility underlying the observed novel isoforms was epistatic interaction between parent alleles. The conservative Bonferroni correction was used in our epistasis scan; therefore, we expect there to be additional cases of epistasis that were missed by our analysis. Examples of epistasis included interactions between particular spliceosome components and other regulatory loci. This finding is consistent with the Dobzhansky-Muller (DM) model of hybrid incompatibility describing that deleterious epistatic interactions accumulate as lineages diverge (Bateson 1909; Dobzhansky 1934; Muller 1942; Orr 1995; Orr and Turelli 2001). It is possible that additional interactions were missed due to complex epistasis involving more than two loci, which we did not check, but is expected to be the more common form of hybrid incompatibility (Orr 1995; Presgraves 2010). Moreover, an incompatible allele may have interacted with the opposite genomic background in a diffuse manner involving many small interaction effects.

Most examples of DM incompatibilities from empirical research have been observed during the later stages of speciation and have large effects that individually cause substantial inviability or sterility in hybrids (Coyne and Orr 2004; Presgraves 2010). The heavy representation of large-effect DM incompatibilities in the literature may be partly due to the inherent muddiness in describing which loci contribute to speciation early on, before significant reproductive isolation has progressed. In fact, incompatibilities may not cause devastating effects on hybrid fitness (Orr 1997). The relatively small, quantitative fitness effects observed in the current study represent an important aspect of reproductive isolation that is usually not detectable. Another recent study, by Schumer et al. (2014), has shown that hundreds of small-effect incompatibilities can accumulate between lineages with incomplete reproductive isolation. Similarly, epistatic interactions with negative fitness effects can occur within species due to segregating variation (e.g., Corbett-Detig et al. 2013), called “variable reproductive isolation” by Cutter (2012). Therefore, reproductive isolation could arise through divergence in frequencies of many preexisting, latent incompatibilities (Corbett-Detig et al. 2013). These findings warrant additional research to explore whether many weak incompatibilities distributed throughout the genome can restrict gene flow between species (Orr 1997; Corbett-Detig et al. 2013; Schumer et al. 2014). Although speciation researchers often test for epistasis associated with a single hybrid inviability phenotype, for example, fecundity, aberrant splicing may provide numerous potential targets for detecting genetic incompatibilities. Also, expression QTLs tend to have a high signal to noise ratio, therefore we have improved statistical power to detect eQTLs and splicing QTLs. Through studying splicing abnormalities, we might gain precision for studying the accumulation of incompatibilities at various stages of the speciation continuum, in particular the early stages of divergence.

In genome-scale analyses that uncover tens or hundreds of negative epistatic effects in the form of statistical interactions between loci, it can be difficult to know how each genetic incompatibility manifests on a cellular level. We want to draw attention to aberrant RNA splicing as a possible mechanistic basis for hybrid inviability caused by DM, and other types of, hybrid incompatibilities. It is intuitive that incompatibilities might affect splicing, because alternative splicing ordinarily relies on epistatic gene networks (Munding et al. 2010; Julien et al. 2016; Pacini et al. 2018) involving hundreds of regulatory loci for the assembly and regulation of the spliceosome (Will and Lührmann 2011). An incompatibility involving any of these loci could disrupt the regulation of RNA splicing. In addition, an interaction that affects the spliceosome could theoretically disrupt splicing of many downstream loci and thereby reduce hybrid fitness. Our findings may align with this model because several of the detected QTLs in the current study overlapped with or had statistical interactions with individual spliceosome components, one of which had divergent gene expression in the parent taxa. Future research is warranted to investigate splicing malfunctions in other hybridizing taxa, particularly in wild populations or species.

**SPlicing EVOLUTION DURING SUNFLOWER DOMESTICATION**

The cultivated sunflowers’ transition to domestication began approximately 5000 years ago (Baute et al. 2015; Badouin et al. 2017). This evolutionary shift was accompanied by a reduction in genetic diversity and differentiation from wild *H. annuus* (Park and Burke 2020). Additionally, domestication was associated with changes in mRNA splicing patterns, particularly for genes involved in seed development (Smith et al. 2018). Thus, our theory for the role of alternative splicing during sunflower domestication is multilayered: first, divergence at splicing regulatory loci has the capacity to affect mRNA splicing for a large swath of downstream transcripts. Second, genetic incompatibilities that affect splicing may enhance reproductive isolation. In theory such incompatibilities could explain reduced fitness in wild-crop sunflowers hybrids; these hybrids have fewer branches, flower heads, and seeds than wild plants (Snow et al. 1998). Consistent with this idea, a sucrose biosynthesis gene was mis-spliced in one or more hybrid offspring in the current study, and aberrant splicing was correlated with a reduction in plant size. We noted that aberrant splicing at a particular locus did not require divergence in splicing between parent genotypes at the same
locus. Third, as Scascitelli et al. (2010) proposed, novel isoforms in plant hybrids represent a form of transgressive segregation. Extreme phenotypes caused by interacting alleles from each parent could contribute to adaptive evolution (Rieseberg et al. 1999). Additional research is needed to investigate splicing patterns that differ quantitatively between parents and offspring.

AUTHOR CONTRIBUTIONS
Conceptualization, original draft preparation, and review and editing were contributed by CCRS, LHR, BSH, and NCK. Analysis was performed by CCRS and NCK.

ACKNOWLEDGMENTS
This work was funded by Genome Canada and Genome BC (LSARP2014-223SUN) and the National Science Foundation Plant Genome Program (DBI-0820451 and DBI-1444522). This work used both the BioFrontiers Computing Core at the University of Colorado at Boulder (UCB) and the RMACC Summit Supercomputer. The authors are also thankful to anonymous reviewers who provided feedback on the original manuscript draft.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA ARCHIVING
Sequence data reported in this article are available as part of the Sequence Read Archive: Bioproject no. PRJNA417714. All custom code is available at https://github.com/chrisrcsmith/SunflowerAberrantSplicing.

LITERATURE CITED

Ast, G. 2004. How did alternative splicing evolve? Nat. Rev. Genet. 5:773–782.
Badouin, H., J. Gouzy, C. J. Grassa, F. Murat, S. E. Staton, L. Cottret, C. Lelandais-Briè, G. L. Owens, S. Carrè, B. MayJonade, et al. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature 546:148–152.
Bateson, W. 1909. Heredity and variation in modern lights. Pp. 85–101 in A. C. Seward, ed. Darwin and modern science. Cambridge University Press, Cambridge, U.K.
Baute, G. J., N. C. Kane, C. J. Grassa, Z. Lai, and L. H. Rieseberg. 2015. Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. New PhytoL 206:830–838.
Berardini, T. Z., L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, and E. Huala. The arabidopsis information resource: making and mining the “gold standard” annotated reference plant genome. Genesis 53:474–485.
Brogna, S., and J. Wen. 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. Nat. Struct. Mol. Biol. 16:107–113.
Broman, K. W., and S. Sen. 2009. A guide to QTL mapping with R/qtl. Springer, New York.
Burke, J. M., S. Tang, S. J. Knapp, and L. H. Rieseberg. 2002. Genetic analysis of sunflower domestication. Genetics 161:1257–1267.
Chamala, S., G. Feng, C. Chavarro, and W. B. Barbazuk. 2015. Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. Front. Bioeng. Biotechnol. 3:33.
Claverie, J. M. 1997. Computational methods for the identification of genes in vertebrate genomic sequences. Hum. Mol. Genet. 6:1735–1744.
Corbett-Detig, R. B., J. Zhou, A. G. Clark, D. L. Harl, and J. F. Ayroles. 2013. Genetic incompatibilities are widespread within species. Nature 504:135–137.
Coyne, J. A., and H. A. Orr. 2004. Speciation Sinauer Associates. Sunderland, MA 276:281.
Cutter, A. D. 2012. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. Trends Ecol. Evol. 27:209–218.
Degtiar, E., A. Fridman, D. Gottlieb, K. Vexler, I. Berezin, R. Farhi, L. Golani, and O. Shaul. 2015. The feedback control of UPE3 is crucial for RNA surveillance in plants. Nucleic Acids Res. 43:4219–4235.
Demaeegd, D., A. S. Colinet, A. Deschamps, and P. Morosonne. 2014. Molecular evolution of a novel family of putative calcium transporters. PLoS ONE 9:e100851.
Dempewolf, H., K. A. Hodgins, S. E. Rummell, N. C. Ellstrand, and L. H. Rieseberg. 2012. Reproductive isolation during domestication. Plant Cell 24:2710–2717.
Dobzhansky, T. 1934. Studies on hybrid sterility. Z. Zellforsch. Mikrosk. Anat. 21:169–223.
Göhring, J., J. Jacak, and A. Barta. 2014. Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. Plant Cell 26:754–764.
Huang, Y., J. B. Lack, G. T. Hoppel, and J. E. Pool. 2019. Evolution of gene expression and splicing in parallel cold-adapted fly populations. bioRxiv https://doi.org/10.1101/795716.
Ishikawa, A. 2009. Mapping an overdominant quantitative trait locus for heterosis of body weight in mice. J. Hered. 100:501–504.
Julien, P., B. Mihana, P. Baesa-Centurion, J. Valcarcel, and B. Lehner. 2016. The complete local genotype–phenotype landscape for the alternative splicing of a human exon. Nat. Commun. 7:1–8.
Kalyna, M., C. G. Simpson, N. H. Syed, D. Lewandowska, Y. Marquez, B. Kusenda, J. Marshall, J. Fuller, L. Cardle, J. McNicol, et al. 2012. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res. 40:2454–2469.
Kenney-Hunt, J. P., T. T. Vaughn, L. S. Pletscher, A. Peripato, E. Routman, K. Cothran, D. Durand, E. Norgard, C. Perel, and J. M. Cheverud. 2006. Quantitative trait loci for body size components in mice. Mamm. Genome 17:526–537.
Keren, H., G. Lev-Maor, and G. Ast. 2010. Alternative splicing and evolution: diversification, exon definition and function. Nat. Rev. Genet. 11:345–355.
Khokhar, W., M. A. Hassan, A. S. Reddy, S. Chaudhary, I. Jabre, L. Byrne, and N. H. Syed. 2019. Genome-wide identification of splicing quantitative trait loci (sQTL) in diverse ecotypes of Arabidopsis thaliana. Front. Plant Sci. 10:1160.
Lin, A., J. Ma, F. Xu, W. Xu, H. Jiang, H. Zhang, C. Qu, L. Wei, and J. Li. 2020. Differences in alternative splicing between yellow and black-seeded rapeseed. Plants 9:977.
Mallarino, R. T., A. Linden, C. R. Limmen, and H. E. Hoekstra. 2017. The role of isoforms in the evolution of cryptic coloration in Peromyscus mice. Mol. Ecol. 26:245–258.
McGinnis, S., and T. L. Madden. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32(suppl_2):W20–W25.
Mezzini, A. M., E. J. Smith, M. Fiume, O. Buske, G. L. Savich, S. Shah, S. Aparicio, D. Y. Chiang, A. Goldenberg, and M. Brudno. 2013. iRecon: simultaneous isoform discovery and abundance estimation from RNA-seq data. Genome Res. 23:519–529.

ABERRANT SPlicing IN SUNFLOWER HYBRIDS
Min, X. J., G. Butler, R. Storms, and A. Tsang. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 33(suppl_2):W677–W680.

Muller, H. 1942. Isolating mechanisms, evolution, and temperature. Biol. Symp. 6:71–125.

Munding, E. M., A. H. Igel, L. Shiue, K. M. Dorighi, L. R. Treviño, and M. Ares. 2010. Integration of a splicing regulatory network within the meiotic gene expression program of Saccharomyces cerevisiae. Genes Dev. 24:2693–2704.

Nissim-Rafinia, M., and B. Kerem. 2002. Splicing regulation as a potential genetic modifier. Trends Genet. 18:123–127.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805–1813.

Orr, H. A. 1997. Haldane’s rule. Ann. Rev. Ecol. Syst. 28:195–218.

Orr, H. A. 1997. Haldane’s rule. Ann. Rev. Ecol. Syst. 28:195–218.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805–1813.

Orr, H. A. 1997. Haldane’s rule. Ann. Rev. Ecol. Syst. 28:195–218.

Pacini, C., and M. J. Koziol. 2018. Bioinformatics challenges and perspectives when studying the effect of epigenetic modifications on alternative splicing. Philos. Trans. R. Soc. B Biol. Sci. 373:20170073.

Park, B., and J. M. Burke. 2020. Phylogeography and the evolutionary history of sunflower (Helianthus annuus L.): wild diversity and the dynamics of domestication. Genes 11:266.

Presgraves, D. C. 2010. Darwin and the origin of interspecific genetic incompatibilities. Am. Nat. 176:S45–S60.

Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaptation and speciation. Heredity 83:363–372.

Renaut, S., C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai, N. C. Kane, J. E. Bowers, J. M. Burke, and L. H. Rieseberg. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat. Commun. 4:1–8.

Renaut, S., A. W. Nolte, and L. Bernatchez. 2009. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (Coregonus spp. Salomonidae). Mol. Biol. Evol. 26:925–936.

Scascitelli, M., M. Cognet, and K. L. Adams. 2010. An interspecific plant hybrid shows novel changes in parental splice forms of genes for splicing factors. Genetics 184:975–983.

Schumer, M., R. Cui, D. L. Powell, R. Dresner, G. G. Rosenthal, and P. Andolfatto. 2014. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. Elife 3:e02535.

Sieber, P., M. Platzer, and S. Schuster. 2018. The definition of open reading frames revisited. Trends Genet. 34:167–170.

Shaal, O. 2015. Unique aspects of plant nonsense-mediated mRNA decay. Trends Plant Sci. 20:767–779.

Smith, C. C., S. Tittes, J. P. Mendieta, E. Collier-Zans, R. HC, L. H. Rieseberg, and N. C. Kane. 2018. Genetics of alternative splicing evolution during sunflower domestication. Proc. Natl. Acad. Sci. USA 115:6768–6773.

Snow, A. A., P. Moran-Palma, L. H. Rieseberg, A. Wszelaki, and G. J. Seiler. 1998. Fecundity, phenology, and seed dormancy of F1 wild–crop hybrids in sunflower (Helianthus annuus, Asteraceae). Am. J. Bot. 85:794–801.

Sorek, R., and G. Ast. 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res. 13:1631–1637.

Staton, S. E., and L. H. Rieseberg. 2019. Sunflower genome database. Available via https://www.sunflowergenome.org/.

Sugnet, C. W., W. J. Kent, M. Ares Jr., and D. Haussler. 2004. Transcriptome and genome conservation of alternative splicing events in humans and mice. Pac. Symp. Biocomput. https://doi.org/10.1142/9789812704856_0007.

Stelkens, R., and O. Seehausen. 2009. Genetic distance between species predicts novel trait expression in their hybrids. Evolution 63:884–897.

Teraí, Y., N. Morikawa, K. Kawakami, and N. Okada. 2003. The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. Proc. Natl. Acad. Sci. USA 100:12798–12803.

Thatcher, S. R., W. Zhou, A. Leonard, B. B. Wang, M. Beatty, G. Zastrow-Hayes, X. Zhao, A. Baumgarten, and B. Li. 2014. Genome-wide analysis of alternative splicing in Zea mays: landscape and genetic regulation. Plant Cell 26:3472–3487.

Todesco, M., G. L. Owens, N. Bercovich, J. S. Légaré, S. Soudi, D. O. Burge, K. Huang, K. L. Ostevik, E. B. Drummond, I. Imrrovski, et al. 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature 584:602–607.

Ullah, F., M. Hamilton, A. S. Reddy, and A. Ben-Hur. 2018. Exploring the relationship between intron retention and chromatin accessibility in plants. BMC Genomics 19:1–11.

Wang, B. B., and V. Brendel. 2006. Genomewide comparative analysis of alternative splicing in plants. Proc. Natl. Acad. Sci. USA 103:7175–7180.

Will, C. L., and R. Lührmann. 2011. Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3:a003707.

Yan, C., R. Wan, R. Bai, G. Huang, and Y. Shi. 2016. Structure of a yeast activated spliceosome at 3.5 Å resolution. Science 353:904–911.

Young, N. D. 1996. QTL mapping and quantitative disease resistance in contrasting rice cultivars under drought stress and association of differential splicing genes with drought response QTLs. Euphytica 2018. 214:73.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

S1 Text. Identification of single-copy genomic regions

S2 Text. Epistasis scan using R/qtl

Supplementary information

Supplementary information