TITLE: Gene expression and alternative splicing contribute to adaptive divergence of ecotypes

Authors: Peter A Innes[1], April M. Goebl[1,2], Chris C.R. Smith[1,3], Kaylee Rosenberger[1], Nolan C. Kane[1]

Affiliations: [1]University of Colorado, Boulder, Department of Ecology and Evolutionary Biology. [2]Denver Botanic Gardens, Research and Conservation Department. [3]University of Oregon, Institute of Ecology and Evolution.

ABSTRACT

Regulation of gene expression is a critical link between genotype and phenotype explaining substantial heritable variation within species. However, we are only beginning to understand the ways that specific gene regulatory mechanisms contribute to adaptive divergence of populations. In plants, the post-transcriptional regulatory mechanism of alternative splicing (AS) plays an important role in both development and abiotic stress response, making it a compelling potential target of natural selection. AS allows organisms to generate multiple different transcripts/proteins from a single gene and thus may provide a source of evolutionary novelty. Here we examine whether variation in alternative splicing and gene expression levels might contribute to adaptation and incipient speciation of dune-adapted prairie sunflowers in Great Sand Dunes National Park, Colorado, USA. We conducted a common garden experiment to assess transcriptomic variation among ecotypes and analyzed differential expression, differential splicing, and gene coexpression. We show that individual genes are strongly differentiated for both transcript level and alternative isoform proportions, even when grown in a common environment, and that gene coexpression networks are disrupted between ecotypes. Furthermore, we examined how genome-wide patterns of sequence divergence correspond to divergence in transcript levels and isoform proportions and find evidence for a substantial contribution of *trans*-regulation. Genes that show significant differential expression and/or differential alternative splicing are enriched for functions related to seed development, abiotic stress response, nutrient acquisition, and photosynthesis. Together our results emphasize that alternative splicing has been an underappreciated mechanism providing source material for natural selection at micro-evolutionary time scales.

INTRODUCTION

Heritable variation in gene expression is recognized as a major source for adaptive evolution (e.g. Jones et al., 2012; Martin & Orgogozo, 2013; Signor & Nuzhdin, 2018; Whitehead & Crawford, 2006). However, research on this topic has primarily focused on gene expression level, i.e. variation in transcript abundance, while contributions of other gene regulatory processes to adaptation and population divergence remain understudied (Singh & Ahi, 2022; Verta & Jacobs, 2022).

Alternative splicing of pre-mRNA (AS) is one such post-transcriptional regulatory process, present in all eukaryotes, that produces multiple unique mRNA transcripts (i.e. isoforms) from a single gene. This occurs via the spliceosome, a dynamic ribonucleoprotein complex that catalyzes the retention (or exclusion) of different combinations of exons and introns in mature transcripts of a particular gene. Alternative isoforms can be translated and take on different functions, though AS often results in pre-mature stop codons, for example, a retained intron might disrupt the open reading frame of a gene. The resulting transcripts would either be “detained” in the nucleus for further processing at a later point, degraded in the cytoplasm via nonsense mediated decay (NMD), or translated into truncated proteins (Petrillo, 2023). Despite involving non-functional transcripts, nuclear detention and NMD have important roles in modulating transcript levels/turnover (Göhring et al., 2014; Kalyna et al., 2012), and truncated proteins can have other regulatory impacts (Filichkin & Mockler, 2012; Liu et al., 2013). All together, these outcomes make alternative splicing a core regulatory process that enhances transcriptome and proteome complexity and generates diverse phenotypes (Bush et al., 2017; Wright et al., 2022).

Alternative splicing, alongside changes to gene expression level, may facilitate evolutionary adaptation to new environments and subsequent population divergence (Singh & Ahi, 2022; Verta & Jacobs, 2022; Wright et al., 2022). Knowledge of how alternative splicing variation contributes to these processes is only beginning to emerge, though evidence is taxonomically broad, with examples from fish (Howes et al., 2017; Jacobs & Elmer, 2021; Singh et al., 2017), insects (Y. Huang et al., 2021), and mammals (Mallarino et al., 2017), among others. In plants, studies that examine differences in AS

between populations or genotypes of the same species have mainly been limited to crops, comparing different domesticated varieties or domesticated versus wild populations (Lin et al., 2020; Ner-Gaon et al., 2007; Smith et al., 2018, 2021; Thatcher et al., 2014; Vitulo et al., 2014; Zhang & Xiao, 2018) or within model organisms (Khokhar et al., 2019; Lutz et al., 2015; Wang et al., 2019). Therefore, although such studies have shown alternative splicing to be important, we are only beginning to understand how it could be involved in adaptation and divergence under natural selection rather than artificial selection, particularly in plants.

Alternative splicing contributes to multiple plant developmental processes, in particular seed maturation, seed dormancy/germination, seedling establishment, and transition to flowering (Posé et al., 2013; Sugliani et al., 2010; Szakonyi & Duque, 2018; Tognacca et al., 2022). AS is also known to underlie plastic responses of plants to environmental stressors including drought, heat, cold, and salt (Laloum et al., 2018; Tognacca et al., 2022). Notably, the function of AS in stress response appears to be more significant in plants than in animals (Martín et al., 2021).

Plastic changes in splicing or expression level in response to stress have been theorized to provide indirect contributions to adaptive divergence, for instance by supporting population persistence following colonization of a new environment and allowing genetic differences to subsequently accrue—the ‘plasticity-first’ hypothesis (Levis & Pfennig, 2016; Pavey et al., 2010). At the same time, when heritable variation in alternative splicing and expression exists, this directly allows for adaptive divergence. Variation in gene expression level and/or splicing between individuals/populations can be due to sequence variation within or nearby that gene (*cis* regulation, e.g. promoters, enhancers, splice sites) or variation in distantly located genes whose products diffuse to influence transcription/splicing (*trans* regulation, e.g. transcription and splicing factors) (Hill et al., 2021). Natural selection might act on *cis* or *trans* regulatory variants to promote canalization of gene expression levels/splicing patterns, such that a previously plastic phenotype associated with the new habitat becomes fixed, a process known as genetic assimilation (Ehrenreich & Pfennig, 2016). Standing variation at *cis* or *trans* regulatory loci could

also produce alternative isoforms with functions that, while previously neutral or slightly deleterious, become beneficial in the new environment (Wright et al., 2022).

To what extent does genetically based variation in alternative splicing and gene expression level contribute to local adaptation and population divergence? We investigated this question using common garden transcriptomic data from a pair of prairie sunflower (*Helianthus petiolaris fallax*) ecotypes originating from Great Sand Dunes National Park (GSD) in southern Colorado, USA. These well-studied 'dune' and 'non-dune' ecotypes represent an example of local adaptation, in this case to an extreme sand dune environment (Andrew et al., 2012, 2013; Andrew & Rieseberg, 2013). The dune and non-dune populations are in close proximity to each other, exchanging migrants and genes, but they are differentiated genetically and phenotypically due to selection, with low (but non-zero) survival when seeds are moved between habitats (Andrew et al., 2012; Andrew & Rieseberg, 2013; Goebl et al., 2022; Ostevik et al., 2016). Gene flow is high enough that genetic divergence between dune and non-dune plants is close to zero across much of the genome, though there are important regions of the genome with alleles strongly associated with the dune ecotype (Goebl et al., 2022; K. Huang et al., 2020). Several regions of elevated sequence divergence were recently shown to harbor large chromosomal inversions that vary in frequency across the landscape and are associated with divergent traits and environmental variables, including seed size, vegetation cover, and NO_3 nitrogen levels (K. Huang et al., 2020; Todesco et al., 2020). Within these inversions and other loci under strong divergent selection, analysis of functionally annotated expression and splicing variation could add insight into the molecular mechanisms of adaptive divergence between GSD sunflower ecotypes.

The GSD system thus represents an excellent opportunity to connect existing knowledge of natural and evolutionary history to patterns of variation in different forms of gene regulation. We sought to: (1) characterize genome-wide differences in alternative splicing and expression (transcript levels) between ecotypes, (2) gain a more holistic view of the transcriptomic changes underlying local adaptation by comparing patterns of gene coexpression, (3) explore how regulatory divergence relates

to genome-wide sequence divergence, and (4) determine putative functional roles of genes experiencing divergent regulation.

MATERIALS & METHODS

Study system

GSD is home to the tallest sand dunes in North America, which are marked by shifting sands, low nutrient availability, and intense exposure (Andrew et al., 2012). *Helianthus petiolaris* is among just a few plant species that live in the dunefield, and the divergence of dune and non-dune ecotypes is estimated to have occurred in the last 10,000 years (Andrew et al., 2013). One of the most striking phenotypic differences between the GSD ecotypes is the dune type has seeds that are more than twice as large as the non-dune type, on average. The latter has much higher fecundity, producing many small seeds in comparison, and these differences are maintained in a common garden (Ostevik et al., 2016; Todesco et al., 2020). Larger seeds have previously been shown to have higher emergence rates both on and off the dunes (Ostevik et al., 2016). Increased seed size is thus believed to be an adaptation that aids with seedling provisioning in the depleted sand dune environment, and/or may also help prevent deep burials of seeds by shifting sands (Ostevik et al., 2016). Beyond larger seed size, the dune ecotype displays comparatively thicker stems, reduced branching, and faster seedling development (Andrew et al., 2013; Ostevik et al., 2016).

Plant material, RNA extractions, and sequencing

We collected *H. petiolaris* seed from three sites in the dune habitat and three sites in the non-dune habitat, in and around the Great Sand Dunes National Park, Colorado, USA in 2017 (Fig 1A). Seeds were cold stratified and germinated on filter paper prior to planting 20 seedlings from each of the six sites. Germination and planting occurred from July 3–6, 2018. For one of the non-dune sites, only nine seedlings were planted due to limited germination. Seedlings were grown in an even mix of sand and potting soil and were kept in a greenhouse setting under natural light. Plants received an approximately

equal amount of water each day, to saturation. We randomly selected 12 plants of both ecotypes for RNA extraction and sequencing, with even sampling across sites (N = 4 per site). On July 30, we harvested approximately 100mg of tissue from the tops of the selected ~3.5 week-old seedlings, which included meristem, leaves, and upper stem tissue. Immediately before harvest we measured height (sand to meristem, cm) and number of leaves of all seedlings, including those that weren't sequenced. At this time we also removed one fully expanded leaf for measurement of leaf dry mass (g). After harvest, we measured total dry mass (above-ground tissue, g) for seedlings not chosen for RNA-sequencing. Plants of both ecotypes were harvested at the same time and developmental stage, with most plants at the 8-leaf seedling stage (Fig. S1). We immediately flash-froze the tissue in liquid nitrogen, stored it at -80°C, and extracted RNA the following day using a Qiagen Plant RNA Mini Kit. Library prep (KAPA mRNA HyperPrep Kit) was performed for each of the 24 samples by the CU Boulder Biofrontiers Sequencing Core, followed by sequencing on an Illumina NextSeq using a 75-cycle High Output v2 reagent kit. This produced 19–23 million reads (75bp single-end) per library. We chose to use seedlings for our RNA-sequencing experiment because this is a consequential life stage for GSD sunflowers, especially in the dunes (Goebl et al., 2022).

Filtering and read mapping

Adapter sequence and low-quality reads were trimmed using *fastp* v*0.23.1* (Chen et al., 2018). We aligned trimmed reads to the *Helianthus annuus* reference genome assembly Ha412HOv2 (Badouin et al., 2017; K. Huang et al., 2022) using STAR *v2.7.10a* in two-pass mode (Dobin et al., 2013).

Variant calling and F_{st} estimates

We identified SNPs using GATK *v4.2.5.0* (McKenna et al., 2010). We first processed sorted bam files from STAR using AddOrReplaceReadGroups, MarkDuplicates, and SplitNCigarReads. We then ran HaplotypeCaller in -gvcf mode and used CombineGVCFs and GenotypeGVCFs for genotyping. We selected only bi-allelic SNPs using SelectVariants and applied a generic set of hard filters (--window 35

157 --cluster 3 --filter “FS > 30.0” --filter “QD < 2.0”) using VariantFiltration. We then used vcftools *v0.1.15*
158 (Danecek et al., 2011) to filter for (1) phred quality score above 30, (2) minor allele frequency threshold
159 of 0.05, (3) minimum read depth of at least 5 per genotype, and (4) 0% missingness. Lastly, to avoid
160 spurious SNPs due to paralogous alignments, we counted heterozygotes per-site using vcftools --hardy
161 and filtered out sites with >60% heterozygosity.

162 We calculated genome-wide F_{st} (Weir & Cockerham, 1984) in 2Mb non-overlapping windows
163 between dune and non-dune ecotypes using vcftools, to visualize divergence across the genome and
164 compare to previous estimates. For correlations between expression/splicing divergence and F_{st} in *cis*,
165 we estimated F_{st} along sliding windows comprising five consecutive genes using the python library
166 scikit-allel *v1.3.3* (Miles et al., 2021). We chose this approach due to the limitations of using
167 transcriptomic data: F_{st} estimates from a single gene could be biased by allele-specific expression,
168 while windows based on a fixed length in bp would capture a varying number of genes and thus SNPs
169 per window. By setting a window length based on number of genes, we aimed to average out any
170 confounding effects of allele specific expression.

171 Lastly, we performed a principal components analysis of filtered SNPs using the R package
172 SNPRelate *v1.30.1* (Zheng et al., 2012). For the PCA we pruned SNPs on linkage disequilibrium with an
173 r^2 threshold of 0.2, a sliding window of 500kb, and a step size of 1 site, using PLINK *v2* (Chang et al.,
174 2015).

175

176 <u>Read counting and differential expression</u>

177 Reads mapping to each gene in the reference assembly annotations were counted using HTSeq
178 (Anders et al., 2015), which counts only uniquely mapped reads by default. We excluded genes with
179 total read counts less than 24 as a pre-filter and then used DESeq2 to analyze differential expression
180 (Love et al., 2014). We identified significant differentially expressed (DE) genes at FDR < 0.05 and $\log_2$
181 fold-change (LFC) > 0, and applied a shrinkage function to LFC values using lfcShrink() in DESeq2 in
182 order to better visualize and rank DE genes (Zhu et al., 2019).To assess overall divergence in gene

183 expression between ecotypes, we performed principal components analysis on regularized log-
184 transformed count data (n = 32,308 genes) using the R package vegan *v2.6* (Oksanen et al., 2020).
185

186 Differential splicing

187 We used two approaches to analyze differential splicing between ecotypes: (1) rMATS *v4.1.2* (Shen et
188 al., 2014), which identifies alternative splicing events using reference genome read alignments
189 produced by STAR, and (2) the approach from Smith et al (2018, 2021), which uses a custom pipeline
190 for analyzing *de novo* transcriptome assemblies. The latter approach complements the reference-
191 guided analysis by avoiding reference bias during transcript assembly and potentially characterizing
192 more complex or novel splicing events.

193 The rMATS program is capable of detecting five major types of splice events: skipped exon (SE),
194 intron retention (IR), mutually exclusive exons (MXE), alternative 3′ splice site (A3SS), and alternative 5′
195 splice site (A5SS). Conveniently, it does not require alternative isoforms to be annotated in the
196 reference genome, making it useful for non-model species. rMATS counts reads that align across splice
197 junctions and within exons to estimate the "percent spliced in" (PSI) value of each event, for each
198 individual. PSI ranges from 0 to 1 and represents the proportion of reads mapping to one of two
199 alternative isoforms (dubbed the "inclusion" and "skipping" isoform; importantly, while each event
200 comprises two alternative isoforms, rMATS can identify multiple splicing events per gene, which would
201 be expected if a gene has three or more isoforms). The degree of differential splicing for each event is
202 then calculated as the difference in PSI between ecotypes [$\Delta PSI = mean(PSI_{dune}) - mean(PSI_{non\text{-}dune})$].
203 ΔPSI ranges from 1 to -1, with the extremes representing fixed differences in isoform proportions
204 between ecotypes. By default, rMATS excludes splice events where one or both of the ecotypes have
205 zero reads to support the event and also removes events where neither ecotype has at least one read
206 for either the inclusion or skipping isoform. We opted to increase this threshold to require at least 12
207 reads in each case (up from at least 1 read) in order to increase statistical power by avoiding splice
208 events with low support. We implemented this more conservative filter within the rMATS source code

(rmats.py) because there is no option to adjust rMATS read filtering with the program's command line arguments. We also enabled the detection of novel splice sites using the rMATS option --novelSS, since the Ha412HOv2 reference annotations have only a single transcript annotated per gene and divergence between *H. annuus* and *H. petiolaris* may be significant enough to alter exact splice site positions. Significant differentially spliced (DS) events were called at the default ΔPSI threshold of .01% and an alpha level of .05 after FDR correction. After significance testing we removed events that had 40% or greater missingness (N=193 events). Lastly we performed a PCA of alternative splicing using the event PSI scores, with missing PSI values set to the average PSI for that event.

For the isoform-based differential splicing analysis, we first assembled a transcriptome for *H. petiolaris* using Trinity *v2.13.2* (Grabherr et al., 2011) using all 24 samples. We first removed redundant transcripts with CD-HIT-EST at a threshold of 99% similarity (Fu et al., 2012). Next we aligned the transcriptome to the Ha412HOv2 reference genome with BLASTN and only considered hits that had at least 85% identity and aligned for at least 75% of the transcript length. We calculated expression levels with RSEM (Li & Dewey, 2011) and imposed the following filters for low expressed genes/isoforms: (1) we removed Trinity 'genes' (and thus each of its corresponding isoforms) that were only expressed in one ecotype (minimum total read count per ecotype of 24, with at least 8 samples per ecotype having at least 3 reads) and (2) we removed isoforms with total read count across all samples less than 24. Subsequent steps to filter the transcriptome, retain high confidence alternative isoforms, and test for differential splicing were the same as described previously (Smith et al., 2021). Briefly, we required alternative isoforms to align to the same genomic region, otherwise we labeled the isoforms as separate genes. We performed pairwise or multiple sequence alignments of isoforms of each Trinity 'gene' using EMBOSS needle *v6.6.0.0* or MUSCLE *v5.1* (Edgar, 2021), respectively, to determine if the isoforms assembled by Trinity indeed represented alternative isoforms, or if they were more likely different alleles of the same isoform. In the latter case, we clustered alleles and summed their abundance estimates (transcripts per million, TPM). Next we converted isoform TPM values to proportions of overall gene TPM values and reduced the dimensionality of each gene's isoform composition matrix using an

isometric log ratio (ILR) transformation. Finally, we tested for significant splicing differentiation between ecotypes using t-tests—or MANOVA for genes with more than two isoforms—and used an FDR threshold of .05 to correct for multiple testing.

To assess the congruence between rMATS results and those from the 'Smith et al' *de novo* transcriptome pipeline, we aligned the longest isoform of each *de novo* Trinity gene, including those filtered out in the above steps, to Ha412HOv2 gene sequences using BLASTN and retained all hits with a bit score greater than 100. We generated the Ha412HOv2 gene fasta file using AGAT *v0.8.0 (Dainat, n.d.)* and included both intronic and exonic regions, since Trinity genes might harbor intron retention events. PCA was performed using the ILR transformations of the isoform proportions of genes with just two alternative isoforms. We did not include genes with more than two isoforms in the PCA because their isoform composition matrix remains multidimensional even after ILR transformation.

Tests of DE and DS overlap, spatial enrichment, and associations with F_{st}

We performed hypergeometric tests to determine the significance of overlap between DE and DS gene sets, using the R function dhyper(). The representation factor of the overlap was calculated as the observed divided by the expected number of overlap genes, where the expected overlap is the number of DS genes times the number of DE genes divided by the total number of genes expressed in our experiment.

We counted DE and DS genes within and outside of four major inversion regions: pet5.01, pet9.01, pet11.01, and pet17.01, using BEDTOOLS *v2.26.0* (Quinlan & Hall, 2010). We focused on these regions out of seven previously identified putative inversions (K. Huang et al., 2020) because they were by far the most divergent between ecotypes in our dataset (Fig. S2). While other inversions exist and are segregating, they are not strongly divergent between populations (K. Huang et al., 2020; Todesco et al., 2020). We performed Fisher's exact tests with the R function fisher.test() to determine whether these inversion regions as a whole were enriched for DE of DS genes.

We used linear models to assess the association between sequence divergence (see above for F_{st} methods) and expression/splicing divergence. For each gene, expression ($\log_2$ fold change) and splicing divergence (ΔPSI) were set to zero if deemed not significant.

Gene coexpression network analysis

We created signed, weighted gene coexpression networks for each sunflower ecotype with the R package WGCNA *v1.71* (Langfelder & Horvath, 2008). Here we used a slightly more stringent read count filter to reduce noise in the networks, keeping genes with mean of at least 3 reads per sample, and again we rlog-transformed the count data for input into WGCNA. Networks were built using the function blockwiseModules() with a soft-thresholding power of β=18 to achieve a good model fit for scale free topology. Modules (clusters of highly interconnected genes) were defined using hierarchical clustering and dynamic tree cut algorithm with the default minimum size of 30 genes. Similar modules were merged at a cut height of 0.25, corresponding to a correlation of 0.75. The remaining function parameters were set to the default values.

We compared gene coexpression networks between dune and non-dune ecotypes using the functions matchLabels() and modulePreservation() from WGCNA. The former function matches modules across networks using Fisher's exacts tests to judge overlap and assigns these matches the same color label. In other words, similar modules will be assigned the same color identifier in both networks, though the extent of their similarity may vary. The latter function more carefully assesses the similarity between matched modules, i.e. it calculates the extent to which the composition of a module in the "reference" (non-dune) network is preserved in the "test" (dune) network. The function returns pairwise statistics for all modules in the reference dataset compared to the test dataset, including the preservation "Zsummary," which is based on the Z statistic. A Zsummary score > 10 indicates strong preservation, and $2 <$ Zsummary < 10 indicates weak to moderate evidence of preservation.

Gene ontology analysis

Gene ontology annotations are lacking for the Ha412HOv2 genome, so we based our GO enrichment analyses on *Arabidopsis thaliana* GO annotations (ATH_GO_GOSLIM.txt) downloaded from arabidopsis.org (Berardini et al., 2004). We first matched Ha412HOv2 transcripts to Araport11 *A. thaliana* peptides (Cheng et al., 2017) using blastx with an e-value threshold of 1e-20, retaining only the top hit for each transcript. Next, we merged the list of *H. annuus–A. thaliana* homologs with the *A. thaliana* GO association file to annotate Ha412HOv2 genes with GO terms. We then used the python package GOATOOLS (Klopfenstein et al., 2018) to perform Fisher's exact tests of GO term enrichment, for DE genes (separately for dune up-regulated and non-dune up-regulated sets) and rMATS DS genes, plus the genes that were both DE and DS (DE∩DS). We used the complete set of expressed genes (total read count >= 24, N=32,308) and expressed multi-exonic genes (N=28,389) as the background gene lists, respectively, and an FDR threshold of .05 for significance. We clustered redundant GO terms from each enrichment test using GOMCL (Wang et al., 2020).

RESULTS

Variant calling and sequence divergence between ecotypes

The amount of uniquely mapped reads per library (i.e. per plant) ranged from 72% to 84% and averaged 80%. Un-mapped reads averaged 5.9%; the remaining reads mapped to multiple loci. Variant calling with GATK HaplotypeCaller produced 3,007,876 variable transcriptomic sites across the Ha412HOv2 reference genome, which includes 17 chromosomes plus unplaced contigs. These variants were filtered to 295,383 high quality bi-allelic SNPs. Subsequent pruning of SNPs that were in linkage disequilibrium resulted in 29,113 independent loci. PCA of LD-pruned SNPs showed distinct clustering of ecotypes along the first PC, which explained ~9.3% of the overall SNP variation among samples (Fig 1B).

Gene expression divergence between ecotypes

RNA-seq libraries from bulk seedling tissue produced detectable expression for 32,308 genes,

representing roughly 70% of the total genes annotated in the Ha412HOv2 genome. PCA based on the transcript levels of these genes clearly separated dune and non-dune ecotypes along the first principal component, which explained 18.14% of the total variation (Fig 1C). We found significant differential expression between dune and non-dune ecotypes for 5,103 genes (|Log2FC| > 0, FDR < .05), approximately 15% of genes expressed in our study (Extended Data S1). Of these, 2,480 were up-regulated in the dune environment, while 2,623 were up-regulated in the non-dune environment.

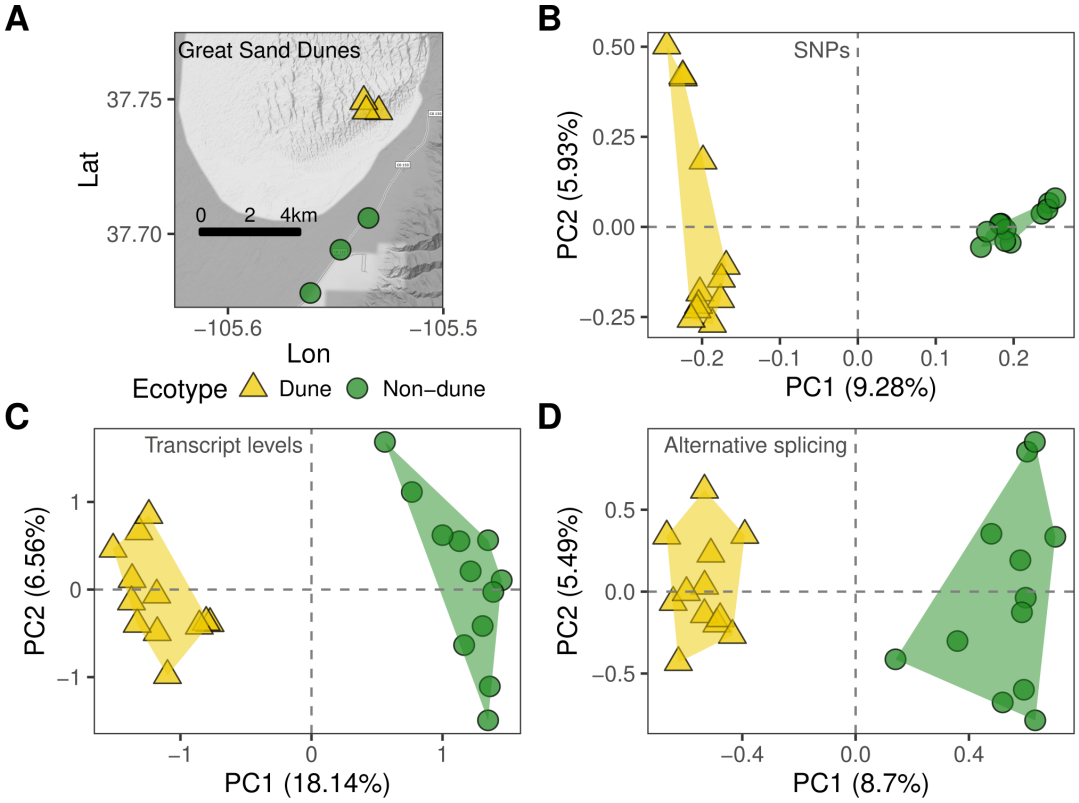

Figure 1. Prairie sunflower ecotypes cluster according to sequence and regulatory variation. A) Map of sampling locations for dune (yellow triangles) and non-dune ecotypes (green circles). Plants were grown from seed in a common garden and four plants from each of the six locations (n = 24; 12 dune, 12 non-dune) were randomly selected for the RNA sequencing experiment. **B–D)** Principal component analyses showing variation along the first and second principal components for **B)** 29,113 LD-pruned transcriptomic SNPs **C)** transcript levels of 32,308 genes and **D)** isoform proportions of 17,845 alternative splicing events identified with rMATS. Percent of total variation explained by each PC is given in parentheses.

Alternative splicing divergence between ecotypes

We detected 17,845 alternative splicing events across 6,551 genes using rMATS (Extended Data S2). This means approximately 23% of the 28,839 multi-exonic genes expressed in our study showed evidence of alternative splicing. Intron retention was the most common event type (9,946 RI events,

~55%), as expected for a plant species. PCA of isoform proportion values (percent spliced in, PSI) of all AS events produced separation between ecotypes along PC1, similar to what was found for SNPs and transcript levels, with the first axis explaining 8.7% of the total variation (Fig. 1D).

We identified 1,442 differential splicing (DS) events among 1038 unique genes (rMATS ΔPSI > .01%, FDR < .05), which represent around 16% of alternatively spliced genes. Intron retention remained the most prevalent event type among the significant DS events, though to a slightly reduced extent (669 significant RI events, ~46%). The dune ecotype tended to retain more introns (360 events with positive ΔPSI) compared to the non-dune ecotype (309 events with negative ΔPSI). Exon skipping was the least frequent AS event type overall but had a comparatively higher frequency in the set of DS events (7.9% vs 13.7%, Fig. S3).

Compared to rMATS we obtained similar results with the Smith et al (2018, 2021) *de novo* isoform-based analysis of alternative splicing. We identified 6,007 Trinity 'genes' with clear cases of alternative splicing, and 76% of these genes had strong BLAST hits to the 6,551 rMATS AS genes (Fig. S4A). Of the 6,007 Trinity AS genes, 1,299 were significantly differentially spliced (DS), representing a similar fraction compared to rMATS (1038 DS / 6,551 AS). The 1,299 Trinity DS genes had high confidence BLAST hits (bit score > 100) to approximately 14% of rMATS DS genes (289 of 1038, Extended Data S2). Considering only reciprocal best BLAST hits, the overlap between rMATS and Trinity DS genes was 53 genes, representing the highest confidence examples of differential splicing (Extended Data S2). PCA of isoform proportions for Trinity AS genes with just two isoforms (4,042 of 6,007) again showed distinct clustering of ecotypes along the first principal component (Fig. S4B), further supporting the results from rMATS.

<u>Gene coexpression network divergence between ecotypes</u>

We observed substantial differences in patterns of gene coexpression between ecotypes, which can provide insight into how the transcriptome is evolving as a whole. coexpression networks for dune and non-dune ecotypes differed in the number and size of clusters of co-expressed genes (i.e. modules),

and in general the composition of modules was not well preserved between ecotypes. The dune network had more modules compared to non-dune network (dune N = 239, non-dune N = 205), but the dune modules had significantly fewer genes on average (dune module mean size = 115, non-dune = 129, Figure 2A). Each network had some modules that were unique to that network (27 in nondune, and 61 in dune), while we identified comparable matches between networks for 179 modules. However, in terms of the exact composition of these modules, their overall preservation between ecotypes was low. Only 8 modules had a high level of preservation (Zsummary > 10); most non-dune modules had little to no preservation in the dune network (Fig. 2B, Fig. S5).

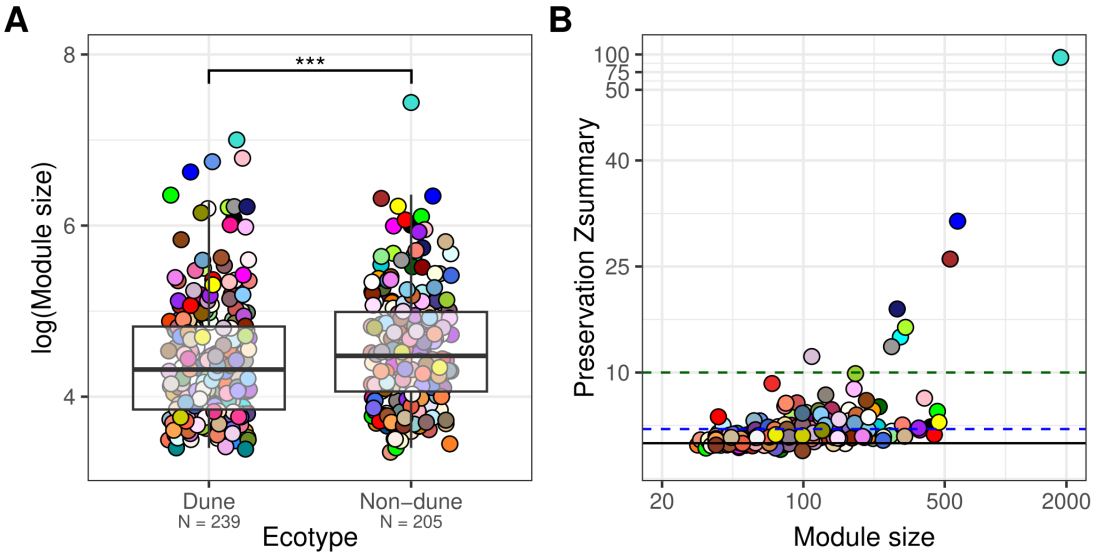

Figure 2. Gene coexpression networks are restructured between ecotypes. A) The dune ecotype has more but smaller coexpression modules compared to the non-dune ecotype (linear model, p << .001). **B)** Module preservation of non-dune modules in the dune network, versus module size (number of genes). Preservation Zsummary > 10 indicates strong preservation (green dashed line); 2 < Zsummary < 10 indicates weak to moderate evidence of preservation; very low preservation indicated under the blue dashed line. The x axis is on the log scale.

<u>Distribution of sequence, expression, and splicing divergence across the genome</u>

In general, divergence in expression and splicing tended to be greater for genes in regions of higher sequence divergence, but this trend explained only a small amount of the total variation in expression and splicing differences between ecotypes.

F_{st} estimates in 2Mb sliding windows revealed similar patterns of sequence divergence between ecotypes as reported previously (Andrew & Rieseberg, 2013; K. Huang et al., 2020; Todesco et al.,

2020). We found F_{st} peaks in regions known to harbor large chromosomal inversions that segregate strongly between ecotypes, though numerous peaks existed outside these inversion regions as well (Fig. 3A, Fig. S2). This closely matches findings based on WGS data (Todesco et al 2020); previous reduced representation approaches seemingly hid some of these smaller peaks of differentiation in non-inversion regions (Andrew & Rieseberg, 2013; K. Huang et al., 2020).

Divergence in transcript levels was widespread across the genome and not confined to individual regions, although some regions of high F_{st} also harbored numerous genes with high expression divergence, for example the inversion regions (Figure 3B). Splicing differentiation was likewise distributed across the genome, with a notable peak of significantly differentiated splicing events within or near the pet11.01 inversion (Fig. 3C). Genes inside the four inversion loci pet5.01, pet9.01, pet11.01, pet17.01 overall were more than twice as likely to be differentially regulated (DE or DS) compared to genes outside of these inversions (Fig. S6, Fisher's exact tests, $p << .001$). Correlations between F_{st} (estimated in windows surrounding each gene) and $\log_2$ fold-change or ΔPSI across the genome were positive and significant, though weak (R^2=.006 and R^2=.026, respectively; Fig. 3D,E).

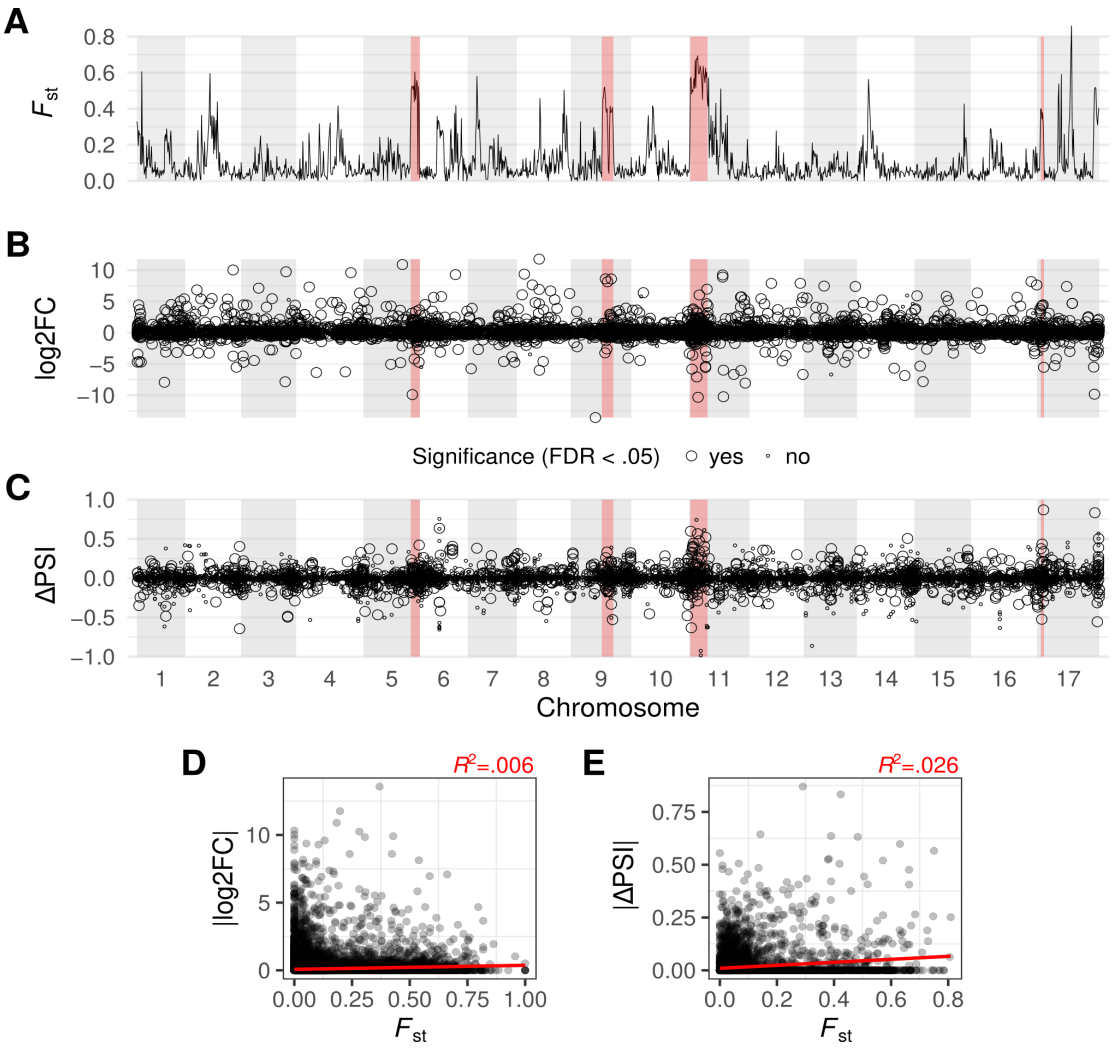

Figure 3. Both *cis* and *trans* regulation appear to contribute to expression and splicing divergence. Red shaded regions indicate positions of putative chromosomal inversions. **A)** Mean F_{st} values along 2Mbp non-overlapping sliding windows. Note putative inversion regions correspond with F_{st} peaks. **B)** Differentially expressed (DE) transcripts along the genome. Positive $\log_2$fold-change indicates up-regulation in the dune ecotype compared to the non-dune ecotype, while negative values indicate the reverse. **C)** Differential splicing (DS) along the genome. ΔPSI of 1 or -1 indicates a fixed difference in isoform proportions between dune and non-dune ecotypes. Circles represent alternative splicing events (a single gene can have multiple splice events if it has more than two isoforms). Significance of DE in **B)** and DS in **C)** is indicated by circle size, and we randomly down-sampled non-significant transcripts and events to improve visualization. Note concentration of DE transcripts and DS events within F_{st} peaks (e.g. pet11.01 inversion, putative *cis*-regulation) but also in regions of low F_{st} (e.g. chromosome 12, putative *trans*-regulation). **D**) Per-gene associations between F_{st} and expression divergence. **E**) Per-gene associations between F_{st} and splicing divergence. Per-gene F_{st} was estimated by using approximate cis-regulatory sliding windows with length of five consecutive genes (two upstream and two downstream). For panel **D**, if a particular gene had multiple AS events, we kept only the event with greatest absolute value ΔPSI value.

More overlap between DE and DS gene sets than expected by chance

More than 22% of differentially spliced genes were also differentially expressed: of the 5,103 DE and 1038 DS genes, 232 genes were both DE and DS (Fig. 4B). This overlap is significantly greater than expected due to chance, based on a hypergeometric test (p = 4.3e-09; representation factor = 1.42).

Putative functions of DE and DS genes

Our homology-based functional annotation (GO annotation) of the *H. annuus* Ha412HOv2 genome provided matches to *Arabidopsis thaliana* for 26,536 of the 32,3208 genes expressed in our study. We found 23 GO terms significantly enriched among genes up-regulated in the dune environment (10 after clustering), and these were primarily related to transcription, fatty acid biosynthesis, nutrient reservoir activity, and stress response processes, notably involving abscisic acid (Fig. 4A). For genes up-regulated in the non-dune ecotype, there were 152 significantly enriched GO terms (66 after clustering), and those involving translation, mRNA binding, embryo development, and photosynthetic processes/compartments were among the most significant (Fig. S7). Genes up-regulated in the dune ecotype were more likely to have a putative unique/unknown function compared to those up-regulated in the non-dune ecotype, based on fractions of either set that lacked a strong BLAST hit to *Arabidopsis* and thus lacked GO annotations (18% of dune versus 13% of non-dune up-regulated genes). This might partially explain the substantial difference in number of enriched GO terms between the two sets.

There were 34 (16 after clustering) significantly enriched GO terms among DS genes (97% of DS genes had GO annotations); these were primarily related to mRNA binding, photosynthesis, embryo development, and nitrogen assimilation (Fig. 4C). This represents an intriguing functional overlap with DE genes, specifically for those involved in embryo development and photosynthesis, which appear to be divergently regulated by both transcription and splicing mechanisms. Indeed, our GO enrichment analysis of the 232 genes that were both DE and DS recovered just four significantly enriched terms after clustering with GOMCL: chloroplast stroma, embryo development ending in seed dormancy, ATP binding, and metal ion binding (Fig. 4D).

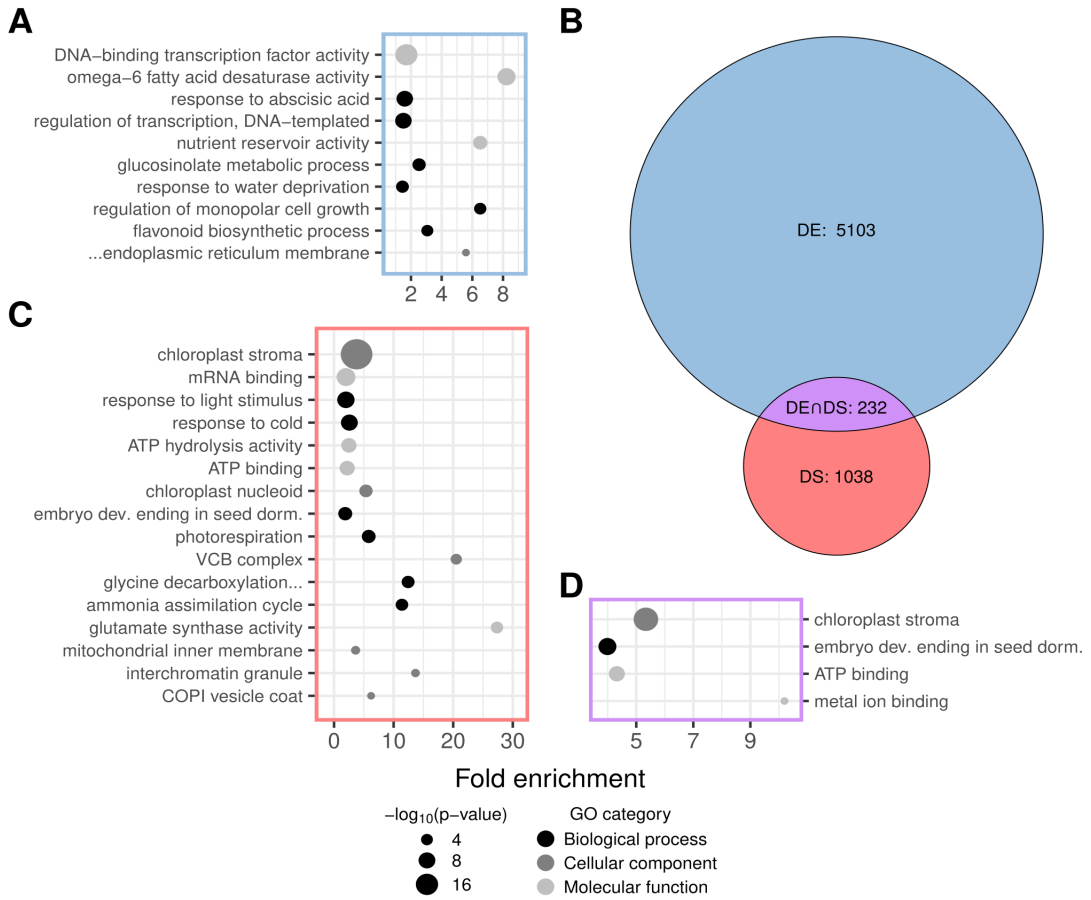

Figure 4. Functional enrichment of differentially regulated genes between ecotypes. A) GO terms enriched among DE genes up-regulated in the dune ecotype (N=2,480). **B)** Venn diagram showing number of transcripts that are differentially expressed (DE), differentially spliced (DS), or both (DE∩DS) between ecotypes (FDR < .05). Significance of overlap between DE genes and DS gene sets was assessed with a hypergeometric test ($p < .001$; Representation Factor, RF = 1.4; RF values > 1 indicate more overlap than expected due to chance). **C)** GO terms enriched among DS genes (N=1038). **D)** GO terms enriched among genes that are both DE and DS (N=232). GO terms are arranged top to bottom by decreasing enrichment test significance (decreasing dot size) in panels **A**, **C**, and **D**.

Beyond analysis of gene ontologies, we highlight individual genes based on magnitude and significance of divergence in expression and/or splicing. Three of the top four DE genes (ranked by adjusted p-value) are homologs of ATP synthase subunit beta (*ATPB*; ATCG00480) and were up-regulated in the non-dune ecotype (Table S1, Fig. S8). Homologs of four other chloroplast ATP synthase subunits (alpha, *ATPA,* ATCG00120; delta, *ATPD ,* AT4G09650; gamma, *ATPC1* , AT4G04640; epsilon, *ATPE,* ATCG00470) also tended to be up-regulated in the non-dune ecotype (Fig. S8). Two homologs of *ATPI*, a subunit of the ATP synthase proton pump complex CF0, were also among the most significant DE genes, and up-regulated in the non-dune ecotype (Fig. S8).

A homolog of the *A. thaliana* splicing factor, *ATO* (AT5G06160) was the 15th most differentially expressed transcript, ranked by $\log_2$ fold change (Ha412HOChr09g0395201; baseMean = 23, $\log_2$ fold change = 8.6, FDR = 1.7e-26; Fig. S9). We also identified homologs of splicing factor *SUS2* (AT1G80070) that were significantly up-regulated in the dune ecotype, with the most abundant being Ha412HOChr02g0050341 (baseMean = 213, $\log_2$ fold change = 1.5, FDR = 0.001; Fig. S9). Other homologs of both these splice factor genes were found in the Ha412HOv2 genome, which were not DE (Extended Data S1). Homologs of a third splicing factor, *CWC22* (AT1G80930), were also significantly up-regulated in the non-dune ecotype (Fig. S9).

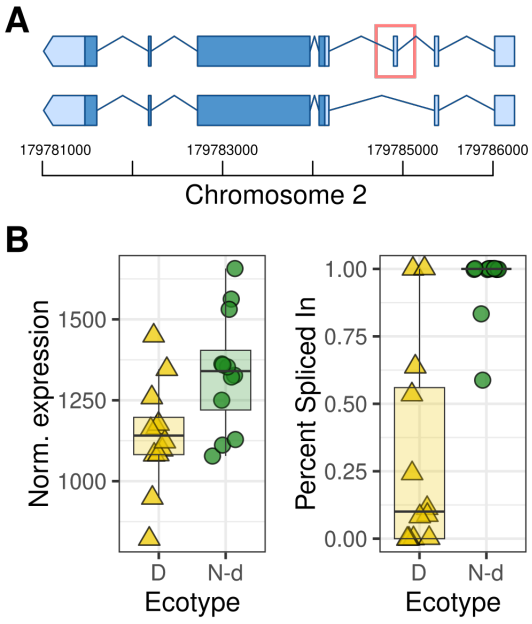

Figure 5. Differential regulation of *GLH17*, an example of a gene that is strongly differentially spliced and also differentially expressed between ecotypes. **A)** Structures of two alternative isoforms of *GLH17*, which constitute a case of exon skipping. The "inclusion" isoform is above the the "skipping" isoform, with the skipped exon marked by a red box. Dark blue blocks represent coding exons and light blue represents exons in untranslated regions (UTRs). The alternatively spliced (skipped) exon is within the 5' UTR. **B)** *GLH17 g*ene-level normalized expression differences between dune (D) and non-dune (N-d) ecotypes ($\log_2$ fold change = -4.3e-05, FDR < .05) and differences in alternative splicing (ΔPSI = -0.644, FDR << .001) for the exon skipping event. "Percent Spliced In" (PSI) refers to the proportion of reads supporting the inclusion isoform. Points represent individual plants.

Lastly we highlight one gene that overall showed the strongest splicing difference between ecotypes according to rMATS, was also DS according to the Smith et al approach, and was differentially expressed: Ha412HOChr02g0088581, a homolog of *A. thaliana* glycosyl hydrolase *GLH17* (AT3G13560). This gene harbored an exon skipping event with ΔPSI of -0.644, meaning the dune

466 ecotype more often expressed the alternative “skipping” isoform (Fig. 5). It had a reciprocal best match
467 to a Trinity DS gene and thus is among the highest confidence examples of differential splicing
468 (Extended Data S2). Intriguingly, the alternatively spliced (skipped) exon is in the 5’ untranslated region
469 of *GLH17* .
470

471 **DISCUSSION**

472 Extent of expression and splicing divergence between ecotypes

473 Previous research has demonstrated that strong allelic differences observed between the dune and
474 non-dune prairie sunflower ecotypes of GSD are adaptive, in other words, selection is driving allelic
475 differences (Goebl et al., 2022). Our results show the ecotypes also have evolved distinct patterns of
476 both gene expression levels and alternative splicing (Fig. 1C,D). Recent findings have been equivocal
477 as to which of these two processes evolves faster at time scales involving ecotypic adaptation and to
478 what extent they follow complementary versus independent trajectories (Jacobs & Elmer, 2021; Verta &
479 Jacobs, 2022). The first principal component for expression level, which also delineates ecotypes,
480 explains a substantially larger amount of overall variation compared to that of alternative splicing (Fig.
481 1), which is consistent with a larger role for expression level in local adaptation. We also observed
482 roughly five times more significantly deferentially expressed genes compared to differentially spliced
483 genes (Fig 3C), though we are cautious in taking this difference at face value because (1) the nature of
484 short read (and single end) sequencing data likely makes detection of alternative splicing more difficult
485 than expression (2) the relative proportion of alternatively spliced genes that are DS is similar to that of
486 expressed genes that are DE (around 15%), and (3) DE and DS are called using slightly different
487 thresholds, in this case LFC > 0 and ΔPSI > 0.01%. Still, a similar magnitude of difference in number of
488 DE versus DS genes exists between sexes of multiple bird species (Rogers et al., 2021), and analyses
489 of plastic differences in splicing and expression suggest a reduced role of AS compared to expression
490 level in environmentally determined phenotypes (Grantham & Brisson, 2018; Steward et al., 2022).
491 Comparisons of arctic charr ecotypes in contrast did not reveal consistently more DE than DS genes

(Jacobs & Elmer, 2021). Regardless of whether there is a meaningful difference in the importance of expression level vs alternative splicing in the evolution of GSD sunflower ecotypes, both processes are strongly associated with their divergence and thus might underlie adaptive traits. In terms of overlap between the two types of gene regulation, we found that ~22% of differentially spliced genes were differentially expressed, which was more often than expected by chance. The large majority of DS genes, however, were not differentially expressed, which is consistent with previous studies suggesting that the two regulatory processes mostly act independently (Jacobs & Elmer, 2021; Martín et al., 2021; Verta & Jacobs, 2022).

Disruption of gene coexpression networks

Gene coexpression also appears to have been dramatically restructured in the process of adaptive divergence: strong correlations in transcript level among genes are rarely preserved when comparing the coexpression network of one ecotype to the other (Fig. 2). Only eight modules showed significant preservation between networks, and the dune coexpression network also had a substantial number of unique modules, i.e. ones that had no close comparison in the non-dune network. Together these results suggest that, beyond the differential expression or splicing of particular genes, novel connections among genes could be important for—or a product of—adaptive divergence. The dramatic restructuring of gene coexpression is striking, since the ecotypes are so recently diverged. Comparison of wild versus domesticated cotton also revealed substantial restructuring of coexpression networks, marked by fewer modules with tighter connections (i.e. larger modules) in the domesticated variety (Gallagher et al., 2020). A similar scenario of coexpression rewiring has been reported for maize vs teosinte (Swanson-Wagner et al., 2012). It is interesting that each of these examples, including ours, all involve recent adaptive evolutionary shifts in phenotypes, ecology, and allelic variation, and also show substantial changes in gene coexpression.

Distribution of regulatory divergence across the genome

We were curious how gene expression level and splicing divergence is patterned across the genome in relation to sequence divergence, and what this might tell us about the relative contributions of *cis* versus *trans* regulation. We found that genes within chromosomal inversion regions—which correspond with high F_{st} peaks—were overall twice as likely to be DE and/or DS compared to genes in non-inversion areas of the genome. This indicates there are substantial *cis*-regulatory variants within these haploblocks that are contributing to regulatory divergence. However, we also found widespread expression and splicing divergence outside of inversions and within low F_{st} regions, which suggests either *trans* regulatory elements in the inverted/highly divergent regions modulate expression or splicing elsewhere in the genome, or additional *cis* regulation is happening outside of the inverted regions. Although we cannot rule out the possibility that our 2Mb sliding window F_{st} averages are masking high divergence at individual *cis* -regulatory SNPs (Fig 3A), we show that sequence divergence in *cis* explains only a small fraction of the total variation in expression level or splicing divergence (Fig. 3D,E).

We found that genes comprising transcription and splicing machinery were often differentially expressed or differentially spliced (Fig. 4A&C, Fig. S7), which is in agreement with previous findings (Jacobs & Elmer, 2021). GO terms such as DNA transcription factor activity, regulation of transcription, and mRNA binding were enriched in either DE or DS gene sets (Fig. 4A,C). A homolog of splicing factor *ATO* was one of the most divergently expressed genes between ecotypes, and we also found homologs of splicing factors SUS2 and *CWC22* that were up-regulated and down-regulated in the dune ecotype, respectively (Fig S9). Such genes represent strong candidates for *trans* regulatory loci: although their own expression and/or splicing may be controlled by *cis* variants, they could act to influence the transcription and/or splicing of genes elsewhere in the genome.

Together these results show that there are likely abundant *trans* regulatory variants contributing to variation in both expression level and alternative splicing. This supports previous work showing trans regulatory loci tend to contribute more to expression variation within species, while *cis* regulatory divergence becomes more impactful between species (Bao et al., 2019; Schaefke et al., 2013; Signor & Nuzhdin, 2018; Wittkopp et al., 2008). We speculate that increased divergence at trans regulatory loci

would be expected to result in more dramatic restructuring of coexpression networks, as described above, since these genes often influence expression of multiple other genes [i.e. they are more pleiotropic (Vande Zande et al., 2022)].

Divergent regulatory evolution highlights potential mechanisms of adaptation

Several traits and environmental variables differ between the GSD ecotypes/habitats over short geographic distance, including seed size, stem thickness, seedling development rate, nitrogen availability, exposure i.e. to light/temperature/wind, and soil stability (Andrew et al., 2012; Ostevik et al., 2016). Of these, larger seed size and tolerance of low nitrogen levels have been most clearly implicated as adaptive traits in the dune ecotype (K. Huang et al., 2020; Ostevik et al., 2016; Todesco et al., 2020). Results from gene ontology enrichment analyses suggest gene expression and alternative splicing evolution may influence these traits. They also point towards additional genes/pathways that may contribute to adaptation in the dunes via divergent regulation.

We found embryo and seed development-related GO terms (e.g. ‘embryo development ending in seed dormancy’) enriched among both DE and DS genes (Fig. 3, Fig. S7). Whether—or precisely how—divergent regulation of this category of genes gives rise to larger seeds in the dune ecotype remains unclear. Intriguingly, however, this same functional category was prevalent among genes differentially spliced between wild and domesticated common sunflower (*H. annuus*) genotypes (Smith et al., 2018). These results are consistent with developmental genes being enriched for alternative splicing (Bush et al., 2017) and implicate such genes as recurring targets of regulatory changes associated with rapid divergence.

Multiple nitrogen assimilation-related GO terms were enriched among differentially spliced genes (‘ammonia assimilation cycle’ and ‘glutamate synthase activity’; Fig. 3C), which underscores tolerance to low nitrogen availability as an important trait in the GSD system. It would be interesting to further explore how alternative splicing contributes to the underlying physiological mechanism of nutrient usage in the dunes. Adaptation to the dunes also appears to involve constitutive up-regulation

570 of abiotic stress response genes (Fig. 3A), including those involved in ‘response to abscisic acid’ (ABA),
571 which is a central plant stress response signaling hormone (Cutler et al., 2010). Overall, the Dune
572 habitat has been shown to favor a different growth and reproductive strategy than the non-dune habitat.
573 Consistent differences in gene expression and splicing of chloroplast and mitochondria-targeted genes
574 between ecotypes is an intriguing pattern in this context (Fig. 3, Fig. S7, Fig. S8) and raises the
575 possibility that cyto-nuclear interactions are an important factor in the divergence of GSD ecotypes.
576 Photosynthesis and metabolism-related genes have been previously unexplored in the GSD system,
577 though a recent study reported a similar pattern of down-regulation for such genes in domesticated
578 sunflower in response to drought treatments (Lee et al., 2022). We hypothesize that divergent regulation
579 of photosynthetic machinery and/or abiotic stress response pathways might be an additional way in
580 which growth and stress tolerance is changed in the dunes.

581 As an example of a gene with strongly differentiated splicing and significantly different
582 expression level between ecotypes, the sunflower homolog of *A. thaliana GLH17* illustrates a way that
583 alternative splicing may interact with expression level to affect proteome composition (Fig. 5).
584 Specifically, our results show that this gene undergoes skipping of an exon within the 5’ untranslated
585 region (UTR), which are well known to influence several aspects of gene expression, including rate of
586 translation, post translational modification, and targeting of the protein (Srivastava et al., 2018);
587 alternative splicing of 5’ UTRs likely influences translational efficiency (Chaudhary et al., 2019). *GLH17*
588 is involved in the auxin mediated initiation of lateral root emergence (Swarup et al., 2008). Given the
589 different soil nutrient (and potentially water) availability between the dune and non-dune habitats, root
590 structure would be an intriguing phenotype to investigate further.

591

592 <u>Conclusion</u>

593 Understanding the molecular processes that contribute to adaptation and divergence is a key goal of
594 evolutionary biology, but alternative splicing has been understudied in this regard, with just a handful of
595 well documented examples to date (Singh et al., 2017; Verta & Jacobs, 2022). Given that alternative

splicing is a important mechanism for plant development and plastic stress response, among other processes, we hypothesized it may underlie adaptive changes as well, especially in extreme habitats such as the Great Sand Dunes. We found that differences in splicing and expression level between a sand dune-adapted prairie sunflower ecotype and its neighboring non-dune ecotype were widespread throughout the genome at the seedling stage in a common environment. Overall, our results represent one of the first clear examples of genome-wide alternative splicing divergence within a non-crop plant species.

Although we do not specifically test the adaptive impact of alternative splicing and expression variation in the dunes, given that gene flow is ongoing between the ecotypes (Andrew et al., 2012, 2013), neutral divergence is expected to be limited in this system, with divergence of genes and traits driven mainly by selection (Goebl et al., 2022; Nosil et al., 2009; Yeaman & Whitlock, 2011). This raises the possibility that gene expression and alternative splicing are either indirectly or directly under selection, which is bolstered by the enrichment among DE and DS genes of functional annotations (e.g. seed development and nitrogen assimilation) seemingly related to known adaptive traits in the dune ecotype. Thus, we conclude that variation in alternative splicing and gene expression level are both likely contributing to adaptive divergence of ecotypes.

ACKNOWLEDGMENTS

PAI would like to thank: Luke Evans for helpful discussion regarding data analysis, Christopher Pauli for providing early guidance on transcriptomic analysis, and Scott Taylor and members of the Taylor Lab for feedback on figures and an earlier version of the manuscript. Seeds were collected at Great Sand Dunes National Park under permit #GRSA-2015-SCI-0008. The CU BioFrontiers Interdisciplinary Quantitative Biology (IQ Bio) program has been an important source of fellowship funding and support for PAI, AMG, CCRS, and KR. The authors would also like to acknowledge that GSDNP is within the traditional lands of the Ute, Arapaho, and Jicarilla-Apache tribes.

AUTHOR CONTRIBUTION STATEMENT

NCK conceived of the study and supervised the work. AG collected seeds. PAI and AG raised plants and performed RNA extractions. PAI conducted all analyses; KR performed the WGCNA analysis and drafted corresponding sections of the manuscript. CCRS provided code and assistance for the ‘Smith et al’ differential splicing analysis. NCK provided guidance on all analyses. PAI wrote the initial draft and designed all figures. All authors provided feedback on the draft.

CONFLICTS OF INTEREST

None

DATA ARCHIVING

The raw RNA sequencing reads will be deposited in the NCBI Sequence Read Archive upon manuscript acceptance.

REFERENCES

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrew, R. L., Kane, N. C., Baute, G. J., Grassa, C. J., & Rieseberg, L. H. (2013). Recent nonhybrid origin of sunflower ecotypes in a novel habitat. *Molecular Ecology*, *22*(3), 799–813. https://doi.org/10.1111/mec.12038

Andrew, R. L., Ostevik, K. L., Ebert, D. P., & Rieseberg, L. H. (2012). Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, *21*(9), 2078–2091. https://doi.org/10.1111/j.1365-294X.2012.05454.x

Andrew, R. L., & Rieseberg, L. H. (2013). Divergence Is Focused on Few Genomic Regions Early in Speciation: Incipient Speciation of Sunflower Ecotypes. *Evolution*, *67*(9), 2468–2482. https://doi.org/10.1111/evo.12106

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., Lelandais-Brière, C., Owens, G. L., Carrère, S., Mayjonade, B., Legrand, L., Gill, N., Kane, N. C., Bowers, J. E., Hubner, S., Bellec, A., Bérard, A., Bergès, H., Blanchet, N., … Langlade, N. B. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, *546*(7656), 148—152. https://doi.org/10.1038/nature22380

Bao, Y., Hu, G., Grover, C. E., Conover, J., Yuan, D., & Wendel, J. F. (2019). Unraveling cis and trans regulatory evolution during cotton domestication. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-13386-w

Berardini, T. Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L. A., Yoon, J., Doyle, A., Lander, G., Moseyko, N., Yoo, D., Xu, I., Zoeckler, B., Montoya, M., Miller, N., Weems, D., & Rhee, S. Y. (2004). Functional Annotation of the Arabidopsis Genome Using Controlled Vocabularies. *Plant Physiology*, *135*(2), 745–755. https://doi.org/10.1104/pp.104.040071

Bush, S. J., Chen, L., Tovar-Corona, J. M., & Urrutia, A. O. (2017). Alternative splicing and the evolution of phenotypic novelty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1713), 20150474. https://doi.org/10.1098/rstb.2015.0474

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

Chaudhary, S., Jabre, I., Reddy, A. S. N., Staiger, D., & Syed, N. H. (2019). Perspective on Alternative Splicing and Proteome Complexity in Plants. *Trends in Plant Science*, *24*(6), 496–506. https://doi.org/10.1016/j.tplants.2019.02.006

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: A complete reannotation of the Arabidopsis thaliana reference genome. *The Plant Journal*, *89*(4), 789–804. https://doi.org/10.1111/tpj.13415

Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R., & Abrams, S. R. (2010). Abscisic Acid: Emergence of a Core Signaling Network. *Annual Review of Plant Biology*, *61*(1), 651–679. https://doi.org/10.1146/annurev-arplant-042809-112122

Dainat, J. (n.d.). *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format* (v0.8.0). Zenodo. https://www.doi.org/10.5281/zenodo.3552717

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Edgar, R. C. (2021). *MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping*. bioRxiv.

Ehrenreich, I. M., & Pfennig, D. W. (2016). Genetic assimilation: A review of its potential proximate causes and evolutionary consequences. *Annals of Botany*, *117*(5), 769–779. https://doi.org/10.1093/aob/mcv130

Filichkin, S. A., & Mockler, T. C. (2012). Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes. *Biology Direct*, *7*(1), 20. https://doi.org/10.1186/1745-6150-7-20

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Gallagher, J. P., Grover, C. E., Hu, G., Jareczek, J. J., & Wendel, J. F. (2020). Conservation and Divergence in Duplicated Fiber Coexpression Networks Accompanying Domestication of the Polyploid Gossypium hirsutum L. *G3 Genes|Genomes|Genetics*, *10*(8), 2879–2892. https://doi.org/10.1534/g3.120.401362

Goebl, A. M., Kane, N. C., Doak, D. F., Rieseberg, L. H., & Ostevik, K. L. (2022). Adaptation to distinct habitats is maintained by contrasting selection at different life stages in sunflower ecotypes. *Molecular Ecology*, *n/a*(n/a). https://doi.org/10.1111/mec.16785

Göhring, J., Jacak, J., & Barta, A. (2014). Imaging of Endogenous Messenger RNA Splice Variants in Living Cells Reveals Nuclear Retention of Transcripts Inaccessible to Nonsense-Mediated Decay in Arabidopsis. *The Plant Cell*, *26*(2), 754–764. https://doi.org/10.1105/tpc.113.118075

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), Article 7. https://doi.org/10.1038/nbt.1883

Grantham, M. E., & Brisson, J. A. (2018). Extensive Differential Splicing Underlies Phenotypically Plastic Aphid Morphs. *Molecular Biology and Evolution*, *35*(8), 1934–1946. https://doi.org/10.1093/molbev/msy095

Hill, M. S., Vande Zande, P., & Wittkopp, P. J. (2021). Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews Genetics*, *22*(4), Article 4. https://doi.org/10.1038/s41576-020-00304-w

Howes, T. R., Summers, B. R., & Kingsley, D. M. (2017). Dorsal spine evolution in threespine sticklebacks via a splicing change in MSX2A. *BMC Biology*, *15*(1), 115. https://doi.org/10.1186/s12915-017-0456-5

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, *29*(14), 2535–2549. https://doi.org/10.1111/mec.15428

Huang, K., Jahani, M., Gouzy, J., Legendre, A., Carrere, S., Lázaro-Guevara, J. M., Segovia, E. G. G., Todesco, M., Mayjonade, B., Rodde, N., Cauet, S., Dufau, I., Staton, S. E., Pouilly, N., Boniface, M.-C., Tapy, C., Mangin, B., Duhnen, A., Gautier, V., … Rieseberg, L. H. (2022). *The genomics of linkage drag in sunflower* (p. 2022.06.07.495047). bioRxiv. https://doi.org/10.1101/2022.06.07.495047

Huang, Y., Lack, J. B., Hoppel, G. T., & Pool, J. E. (2021). Parallel and population-specific gene regulatory evolution in cold-adapted fly populations. *Genetics*, *218*(3), iyab077. https://doi.org/10.1093/genetics/iyab077

Jacobs, A., & Elmer, K. R. (2021). Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Molecular Ecology*, *30*(20), 4955–4969. https://doi.org/10.1111/mec.15817

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), Article 7392. https://doi.org/10.1038/nature10944

Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., Dinh, H. Q., Barta, A., & Brown, J. W. S. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Research*, *40*(6), 2454–2469. https://doi.org/10.1093/nar/gkr932

Khokhar, W., Hassan, M. A., Reddy, A. S. N., Chaudhary, S., Jabre, I., Byrne, L. J., & Syed, N. H. (2019). Genome-Wide Identification of Splicing Quantitative Trait Loci (sQTLs) in Diverse

Ecotypes of Arabidopsis thaliana. *Frontiers in Plant Science*, *10*. https://www.frontiersin.org/articles/10.3389/fpls.2019.01160

Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C. J., Yunes, J. M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., & Tang, H. (2018). GOATOOLS: A Python library for Gene Ontology analyses. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-28948-z

Laloum, T., Martín, G., & Duque, P. (2018). Alternative Splicing Control of Abiotic Stress Responses. *Trends in Plant Science*, *23*(2), 140–150. https://doi.org/10.1016/j.tplants.2017.09.019

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Lee, J. S., Jahani, M., Huang, K., Mandel, J. R., Marek, L. F., Burke, J. M., Langlade, N. B., Owens, G. L., & Rieseberg, L. H. (2022). Expression complementation of gene presence/absence polymorphisms in hybrids contributes importantly to heterosis in sunflower. *Journal of Advanced Research*, *42*, 83–98. https://doi.org/10.1016/j.jare.2022.04.008

Levis, N. A., & Pfennig, D. W. (2016). Evaluating 'Plasticity-First' Evolution in Nature: Key Criteria and Empirical Approaches. *Trends in Ecology & Evolution*, *31*(7), 563–574. https://doi.org/10.1016/j.tree.2016.03.012

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Lin, A., Ma, J., Xu, F., Xu, W., Jiang, H., Zhang, H., Qu, C., Wei, L., & Li, J. (2020). Differences in Alternative Splicing between Yellow and Black-Seeded Rapeseed. *Plants*, *9*(8), Article 8. https://doi.org/10.3390/plants9080977

Liu, J., Sun, N., Liu, M., Liu, J., Du, B., Wang, X., & Qi, X. (2013). An Autoregulatory Loop Controlling Arabidopsis HsfA2 Expression: Role of Heat Shock-Induced Alternative Splicing. *Plant Physiology*, *162*(1), 512–521. https://doi.org/10.1104/pp.112.205864

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lutz, U., Posé, D., Pfeifer, M., Gundlach, H., Hagmann, J., Wang, C., Weigel, D., Mayer, K. F. X., Schmid, M., & Schwechheimer, C. (2015). Modulation of Ambient Temperature-Dependent Flowering in Arabidopsis thaliana by Natural Variation of FLOWERING LOCUS M. *PLOS Genetics*, *11*(10), e1005588. https://doi.org/10.1371/journal.pgen.1005588

Mallarino, R., Linden, T. A., Linnen, C. R., & Hoekstra, H. E. (2017). The role of isoforms in the evolution of cryptic coloration in Peromyscus mice. *Molecular Ecology*, *26*(1), 245–258. https://doi.org/10.1111/mec.13663

Martin, A., & Orgogozo, V. (2013). The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. *Evolution*, *67*(5), 1235–1250. https://doi.org/10.1111/evo.12081

Martín, G., Márquez, Y., Mantica, F., Duque, P., & Irimia, M. (2021). Alternative splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. *Genome Biology*, *22*(1), 35. https://doi.org/10.1186/s13059-020-02258-y

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Miles, A., R, M., Ralph, P., Harding, N., Pisupati, R., Rae, S., & Millar, T. (2021). *Scikit-allel* (1.3.3). Zenodo. https://doi.org/10.5281/zenodo.4759368

Ner-Gaon, H., Leviatan, N., Rubin, E., & Fluhr, R. (2007). Comparative Cross-Species Alternative Splicing in Plants. *Plant Physiology*, *144*(3), 1632–1641. https://doi.org/10.1104/pp.107.098640

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). *vegan:*

Community Ecology Package (R package version 2.5-7). https://CRAN.R-project.org/package=vegan

Ostevik, K. L., Andrew, R. L., Otto, S. P., & Rieseberg, L. H. (2016). Multiple reproductive barriers separate recently diverged sunflower ecotypes. *Evolution*, *70*(10), 2322–2335. https://doi.org/10.1111/evo.13027

Pavey, S. A., Collin, H., Nosil, P., & Rogers, S. M. (2010). The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, *1206*(1), 110–129. https://doi.org/10.1111/j.1749-6632.2010.05765.x

Petrillo, E. (2023). Don't panic: An intron-centric guide to alternative splicing. *The Plant Cell*, koad009. https://doi.org/10.1093/plcell/koad009

Posé, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G. C., Immink, R. G. H., & Schmid, M. (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature*, *503*(7476), Article 7476. https://doi.org/10.1038/nature12633

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rogers, T. F., Palmer, D. H., & Wright, A. E. (2021). Sex-Specific Selection Drives the Evolution of Alternative Splicing in Birds. *Molecular Biology and Evolution*, *38*(2), 519–530. https://doi.org/10.1093/molbev/msaa242

Schaefke, B., Emerson, J. J., Wang, T.-Y., Lu, M.-Y. J., Hsieh, L.-C., & Li, W.-H. (2013). Inheritance of Gene Expression Level and Selective Constraints on Trans- and Cis-Regulatory Changes in Yeast. *Molecular Biology and Evolution*, *30*(9), 2121–2133. https://doi.org/10.1093/molbev/mst114

Signor, S. A., & Nuzhdin, S. V. (2018). The Evolution of Gene Expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

Singh, P., & Ahi, E. P. (2022). The importance of alternative splicing in adaptive evolution. *Molecular Ecology*, *31*(7), 1928–1938. https://doi.org/10.1111/mec.16377

Singh, P., Börger, C., More, H., & Sturmbauer, C. (2017). The Role of Alternative Splicing and Differential Gene Expression in Cichlid Adaptive Radiation. *Genome Biology and Evolution*, *9*(10), 2764–2781. https://doi.org/10.1093/gbe/evx204

Smith, C. C. R., Rieseberg, L. H., Hulke, B. S., & Kane, N. C. (2021). Aberrant RNA splicing due to genetic incompatibilities in sunflower hybrids. *Evolution*, *75*(11), 2747–2758. https://doi.org/10.1111/evo.14360

Smith, C. C. R., Tittes, S., Mendieta, J. P., Collier-zans, E., Rowe, H. C., Rieseberg, L. H., & Kane, N. C. (2018). Genetics of alternative splicing evolution during sunflower domestication. *Proceedings of the National Academy of Sciences*, *115*(26), 6768–6773. https://doi.org/10.1073/pnas.1803361115

Srivastava, A. K., Lu, Y., Zinta, G., Lang, Z., & Zhu, J.-K. (2018). UTR-Dependent Control of Gene Expression in Plants. *Trends in Plant Science*, *23*(3), 248–259. https://doi.org/10.1016/j.tplants.2017.11.003

Steward, R. A., de Jong, M. A., Oostra, V., & Wheat, C. W. (2022). Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-28306-8

Sugliani, M., Brambilla, V., Clerkx, E. J. M., Koornneef, M., & Soppe, W. J. J. (2010). The Conserved Splicing Factor SUA Controls Alternative Splicing of the Developmental Regulator ABI3 in Arabidopsis. *The Plant Cell*, *22*(6), 1936–1946. https://doi.org/10.1105/tpc.110.074674

Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M. B., Ross-Ibarra, J., Myers, C. L., Tiffin, P., & Springer, N. M. (2012). Reshaping of the maize transcriptome by domestication. *Proceedings of the National Academy of Sciences*, *109*(29), 11878–11883. https://doi.org/10.1073/pnas.1201961109

Swarup, K., Benková, E., Swarup, R., Casimiro, I., Péret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., Levesque, M. P., Carrier, D., James, N., Calvo, V., Ljung, K., Kramer, E., Roberts, R., Graham, N., Marillonnet, S., … Bennett, M. J. (2008). The auxin influx carrier LAX3 promotes lateral root emergence. *Nature Cell Biology*, *10*(8), Article 8. https://doi.org/10.1038/ncb1754

Szakonyi, D., & Duque, P. (2018). Alternative Splicing as a Regulator of Early Plant Development. *Frontiers in Plant Science*, *9*. https://www.frontiersin.org/articles/10.3389/fpls.2018.01174

Thatcher, S. R., Zhou, W., Leonard, A., Wang, B.-B., Beatty, M., Zastrow-Hayes, G., Zhao, X., Baumgarten, A., & Li, B. (2014). Genome-Wide Analysis of Alternative Splicing in Zea mays: Landscape and Genetic Regulation. *The Plant Cell*, *26*(9), 3472–3487. https://doi.org/10.1105/tpc.114.130773

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, *584*(7822), 602–607. https://doi.org/10.1038/s41586-020-2467-6

Tognacca, R. S., Rodríguez, F. S., Aballay, F. E., Cartagena, C. M., Servi, L., & Petrillo, E. (2022). Alternative splicing in plants: Current knowledge and future directions for assessing the biological relevance of splice variants. *Journal of Experimental Botany*, erac431. https://doi.org/10.1093/jxb/erac431

Vande Zande, P., Hill, M. S., & Wittkopp, P. J. (2022). Pleiotropic effects of trans-regulatory mutations on fitness and gene expression. *Science*, *377*(6601), 105–109. https://doi.org/10.1126/science.abj7185

Verta, J.-P., & Jacobs, A. (2022). The role of alternative splicing in adaptation and evolution. *Trends in Ecology & Evolution*, *37*(4), 299–308. https://doi.org/10.1016/j.tree.2021.11.010

Vitulo, N., Forcato, C., Carpinelli, E. C., Telatin, A., Campagna, D., D'Angelo, M., Zimbello, R., Corso, M., Vannozzi, A., Bonghi, C., Lucchin, M., & Valle, G. (2014). A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biology*, *14*(1), 99. https://doi.org/10.1186/1471-2229-14-99

Wang, X., Yang, M., Ren, D., Terzaghi, W., Deng, X.-W., & He, G. (2019). Cis-regulated alternative splicing divergence and its potential contribution to environmental responses in Arabidopsis. *The Plant Journal*, *97*(3), 555–570. https://doi.org/10.1111/tpj.14142

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Whitehead, A., & Crawford, D. L. (2006). Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences*, *103*(14), 5425–5430. https://doi.org/10.1073/pnas.0507648103

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Regulatory changes underlying expression differences within and between Drosophila species. *Nature Genetics*, *40*(3), Article 3. https://doi.org/10.1038/ng.77

Wright, C. J., Smith, C. W. J., & Jiggins, C. D. (2022). Alternative splicing as a source of phenotypic diversity. *Nature Reviews Genetics*, *23*(11), Article 11. https://doi.org/10.1038/s41576-022-00514-4

Yeaman, S., & Whitlock, M. C. (2011). THE GENETIC ARCHITECTURE OF ADAPTATION UNDER MIGRATION–SELECTION BALANCE. *Evolution*, *65*(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

Zhang, Z., & Xiao, B. (2018). Comparative alternative splicing analysis of two contrasting rice cultivars under drought stress and association of differential splicing genes with drought response QTLs. *Euphytica*, *214*(4), 73. https://doi.org/10.1007/s10681-018-2152-0

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics*, *35*(12), 2084–2092. https://doi.org/10.1093/bioinformatics/bty895