The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*

John T. Lovell\(^1\),\(^2\), Jerry Jenkins\(^1\), David B. Lowry\(^3\), Sujan Mamidi\(^1\), Avinash Sreedasyam\(^1\), Xiaoyu Weng\(^2\), Kerrie Barry\(^4\), Jason Bonnette\(^2\), Brandon Campitelli\(^2\), Chris Daum\(^4\), Sean P. Gordon\(^4\),\(^8\), Billie A. Gould\(^3\), Albina Khasanova\(^2\), Anna Lipzen\(^4\), Alice MacQueen\(^2\), Juan Diego Palacio-Mejía\(^2\), Christopher Plott\(^1\), Eugene V. Shakirov\(^2\),\(^5\), Shengqiang Shu\(^4\), Yuko Yoshinaga\(^4\), Matt Zane\(^4\), Dave Kudrna\(^6\), Jason D. Talag\(^6\), Daniel Rokhsar\(^7\), Jane Grimwood\(^1\), Jeremy Schmutz\(^1\),\(^4\) & Thomas E. Juenger\(^2\)

Environmental stress is a major driver of ecological community dynamics and agricultural productivity. This is especially true for soil water availability, because drought is the greatest abiotic inhibitor of worldwide crop yields. Here, we test the genetic basis of drought responses in the genetic model for C\(_4\) perennial grasses, *Panicum hallii*, through population genomics, field-scale gene-expression (eQTL) analysis, and comparison of two complete genomes. While gene expression networks are dominated by local cis-regulatory elements, we observe three genomic hotspots of unlinked trans-regulatory loci. These regulatory hubs are four times more drought responsive than the genome-wide average. Additionally, cis- and trans-regulatory networks are more likely to have opposing effects than expected under neutral evolution, supporting a strong influence of compensatory evolution and stabilizing selection. These results implicate trans-regulatory evolution as a driver of drought responses and demonstrate the potential for crop improvement in drought-prone regions through modification of gene regulatory networks.
Drought is the greatest abiotic determinant of agricultural yield\(^1\) and a key regulator of both ecological net productivity and carbon sequestration\(^2,3\). Unlike annual plant species, which can escape environmental stress through flowering time evolution, perennial plants must persist through periods of drought. While constitutive drought tolerance can lead to slower growth and lower yields, there is considerable genetic variation in the physiological nature and efficacy of facultative responses to soil moisture variation among plants\(^4,5\). Leveraging such genotype-by-environment interactions (GxE), especially in response to drought, is key to crop improvement.

Despite the central role that GxE plays in adaptation and plant productivity, the genetic basis of evolved differences in stress responses are poorly understood, except in a handful of laboratory model systems. However, it is clear that gene expression networks play particularly important roles in the evolution of physiological GxE as transcription factors and other regulatory elements are often environmentally induced\(^6\).

Regulatory elements fall into two main categories: distant trans-acting modifiers (e.g. transcription factors) and local cis-regulatory elements (e.g. promoter or coding sequence variants). Selection works most efficiently on traits that act in isolation; therefore, cis-elements, which typically regulate a single gene, may be particularly important in adaptive evolution\(^7\). Conversely, trans factors may cause correlated expression variation among many downstream genes, which should interfere among loci and reduce the adaptive potential of global trans-regulatory evolution\(^8,9\). To this end, many global regulatory elements evolve much more slowly than their target sequences\(^10,11\). Combined, cis- and trans-regulatory elements shape the gene expression landscape and form the basis of environmental stress responses. Therefore, defining the regulatory elements that lead to evolved differences in stress responses will improve our understanding of stress adaptation and the genetic basis of GxE.

To understand the genetic basis of drought stress tolerance in the context of perenniality, we have developed *Panicum hallii* as a genomic model for Panicoid grasses\(^12\). The Panicoidae is a diverse subfamily of predominantly perennial warm-season grasses with efficient \(\text{C}_4\) photosynthesis\(^14\), which imparts drought-tolerance and some of the highest biomass production among plants\(^15\). Not surprisingly, the Panicoidae encompasses the most promising bioenergy feedstock crops, including switchgrass, *Sorghum*, big bluestem, *Miscanthus*, and sugar cane. The geographic distribution of *P. hallii* spans a massive moisture availability gradient: the lowland variety (*P. hallii var. filipes*; hereon *filipes*) inhabits riparian and coastal sites with \(>100\) cm of annual precipitation while upland (*P. hallii var. hallii*; hereon *hallii*) populations are found across southwestern North America\(^16\) in both desert (\(<20\) cm annual precipitation) and semi-arid habitats. The extensive physiological diversity\(^13\) of *P. hallii* and its close evolutionary relationship to key crops make it an ideal genetic model for biotechnology development in perennial biofuel feedstocks.

Here we present a set of experiments and analyses that dissect the genetic basis of *P. hallii* drought responses, connecting DNA sequence variation to leaf-level physiology. We first assess the demographic history and population structure of the upland and lowland varieties of *P. hallii*. To understand the scale of genomic divergence between varieties, we compare complete de novo genome assemblies and annotations for a single genotype of each variety\(^1\). We then explore the genetic basis of evolved drought-responsive gene expression in an *F_2* mapping population. Through eQTL mapping and comparative genomics, we demonstrate that trans-regulatory elements and transcription factor binding site evolution are key contributors to molecular drought responses in *P. hallii*. Combined, these results and genetic resources elevate *P. hallii* among the elite plant genomes, provide the genetic model for perennial feedstocks, and begin to decipher the complex drought-responsive genetic networks that have diverged between upland and lowland ecotypes.

**Results**

**Demographic history of *P. hallii***. Habitat divergence between *filipes* and *hallii* is likely due to both adaptation and historical demographic processes associated with paleo-climatic change in North America. To explore the recent evolutionary history of these varieties, we deeply re-sequenced (35x median sequence coverage) 15 *filipes* and 78 *hallii* natural accessions (Supplementary Data 1) spanning the USA geographic range of each variety (Fig. 1a). Demographic models (Supplementary Figure 1) indicated that *hallii* and *filipes* began evolutionary divergence \(>1\) M years before present (ybp). This conclusion was confirmed by a molecular-clock based estimate of divergence of 1.08 M ybp. Furthermore, pairwise cross-coalescence, a measure of the degree of genetic divergence between populations\(^17\), between any two *hallii* and *filipes* subpopulations (Fig. 1b) all declined below 0.25 at least 500k ypb (Supplementary Figure 1). These data indicate that the two *P. hallii* varieties have maintained substantial reproductive isolation over a period of intense climatic fluctuations, including at least the last two glacial-interglacial cycles.

Effective population size \((N_e)\) of the largest two *hallii* subpopulations (inferred via genetic clustering, Fig. 1b) expanded following reduction of cross coalescence until 20–100k ybp, while \(N_e\) of the largest *filipes* subpopulation remained relatively stable and \(~50\)% smaller than any *hallii* subpopulations (Supplementary...
Figure 2; Supplementary Data 2) over this period. However, Ne of all subpopulations have contracted over the last ~50k years. Combined, these demographic patterns suggest that the current range of hallii is similar to, or smaller than its historical distribution and is not purely the product of a recent population expansion. Instead, between-variety habitat differentiation is consistent with historical range expansion of hallii into arid regions of the North American southwest.

Development and comparison of two complete de novo genomes. To test for a genetic signal of drought-associated evolution, we developed complete de novo reference genomes for the basal P. hallii accession (var. hallii HAL2) and a representative var. filipes accession (FIL2) (Figs. 1b, 2a-d). These Pacific Biosciences single molecule-based chromosomal assemblies contain 484.6 Mb (99.4% of assembled sequence in chromosomes, contig N/L50 = 15/8.3 Mb) and 508.0 Mb (94.8%, 117/1.1 Mb) of total HAL2 and FIL2 sequence, making them among the most complete plant genomes ever assembled (Supplementary Notes 1–2; Supplementary Figure 3; Supplementary Tables 1–4). Critically, these genomes exhibit near-perfect chromosome-scale synteny (Fig. 2d, Supplemental Note 3; Supplementary Figure 4, Supplementary Table 5; Supplementary Data 3). A total of 475 kb (<0.1%) of the HAL2 genomic assembly was not collinear with FIL2, including a 0.18 Mb translocation on Chr09 and a 0.295 Mb translocation between HAL2 Chr02 and FIL2 Chr04 (Supplementary Table 5). Additionally, a duplication was present on the proximate telomere and adjacent 2.22 Mb region of Chr03 and Chr08 in both genomes. This duplication is also present in Sorghum bicolor (Supplementary Figure 4), indicating that it is ancient and ancestral to Panicum.

It is possible to leverage such extensive synteny to infer sequence evolution and presence-absence variation (PAV) among orthologous clusters of genes. This analysis was performed via our GENESPACE pipeline (Supplementary Note 3), which employs a multi-species orthologous gene network construction approach constrained within collinear sequence blocks. GENESPACE allows for construction of outgroup-rooted gene networks and sequence alignments within duplicated (e.g. proximate Chr03 and Chr08) and single-copy regions. Critically, by informing orthology networks with outgroup sequences, we can determine whether gene annotations without orthologous sequences between HAL2 and FIL2 are derived or lost in each genome. Overall, the majority of high-confidence gene models (41,017 genes; 80.1%) existed as either single copy (‘1:1’) orthogroups or orthogroups with two or more members in at least one genome (‘1:2 +’, ‘2 +:1’ or ‘2 +:2 +’; Table 1; Supplementary Data 4). However, 10,176 genes (19.9% of all gene models) lacked
annotated orthologous sequence in either HAL2 or FIL2 (Table 1; Supplementary Data 5). These PAV genes may be the product of annotation support thresholding, pseudogenization or true sequence deletion in one genome.

To evaluate the evolution of these PAV genes, we compared DNA variation between annotated coding DNA sequence (CDS) of present genes with the unannotated assembly sequence of syntenic orthologous regions in the other genome. The majority of PAV genes were private to either HAL2 or FIL2 and never had orthologous sequences in any outgroups (Table 1, Supplementary Data 5). Other studies find that such private genes are generally of low quality and may not represent novel derived CDS, but instead are simply due to weak gene evidence that only survives thresholding in a single genome. Indeed, private genes were 9.6× more likely to be low confidence, with low homology or transcript support, than 1:1 orthogroup genes.

To understand sequence variation underlying the remaining PAV genes, which also had at least one outgroup sequence in the network, we categorized the degree of sequence variation between CDS and un-annotated DNA sequence of the alternative genome into un-annotated (>90% sequence identity), pseudogenized (90–10% identity), and deleted (< 10% identity) groups (Table 1).

The private gene models, the 296 un-annotated gene models, where a single gene model from each genome is represented in an orthofinder orthogroup. Other orthogroups contain two or more gene models from one or both P. hallii genomes. Private genes are found in single-gene orthogroups without representation of any outgroup (S. viridis and S. bicolor). Pseudogenes have >10% and ≤90% sequence identity, with low homology or transcript support. Gene models that did not satisfy these criteria were considered low support. To test for expression in leaf tissue in our field experiment, we counted transcript abundance in an HAL2-FIL2 F2 population. Expression had at least one count in ≥10% of the F2 population and mean counts >5, after excluding libraries with 0 counts.

Orthology, un-annotated, and presence/absence are inferred via the GENESPACE pipeline (Supplementary Note 3). The gene-annotation category (where HAL2 category precedes FIL2 follows) and number of genes found therein are presented in the first two columns. Annotation confidence score (Supplementary Note 2) is calculated via the degree of homology, gene expression, and PfAM support. Gene models that did not satisfy these criteria were considered low support. To test for expression in leaf tissue in our field experiment, we counted transcript abundance in an HAL2-FIL2 F2 population. Expression had at least one count in ≥10% of the F2 population and mean counts >5, after excluding libraries with 0 counts.

To dissect regulatory network evolution. To map the genetic basis of physiological divergence between varieties, we conducted a large-scale field drought experiment where 25 HAL2 and 34 FIL2 replicates and 243 FIL2xHAL2 F2 genotypes were subjected to a month-long natural drought. Half of the plants were watered 24 h prior to harvest (recovery treatment), while the remainder were harvested under existing drought conditions. At the physiological scale, midday leaf water potential (a measure of plant water status) was more responsive to re-watering among HAL2 than FIL2 plants ($\chi^2_{df} = 4.01, P = 0.045$, Supplementary Figure 5). This suggests that the arid-adapted hallii was more sensitive to temporally variable water resources, a trait that is likely advantageous in desert ecosystems marked by drought but punctuated by brief periods of high soil moisture.

Physiological responses like leaf water potential are driven in part by the evolution of gene expression regulatory sequences. To dissect the genetic basis of regulatory network divergence between hallii and filipes, we assayed total RNA abundance of the F2 population (Supplementary Data 7), constructed a genetic map (Supplementary Figure 6), and subsequently conducted gene expression quantitative trait locus (eQTL) mapping. To determine significance of constitutive (QTL) and treatment-responsive (QTL*E) QTL for the expression phenotype of each gene, we compared NULL (expression ~E; no QTL, only drought treatment), additive (expression ~QTL + E), and full (expression ~QTL + E + QTL*E) QTL models via likelihood ratio tests.

While PAV and multi-copy orthogroups are important aspects of evolution between HAL2 and FIL2, regulatory network inference of these genes is obscured by different copy numbers and large sequence deletions, represent the vast majority of molecular evolution between HAL2 and FIL2. Furthermore, protein coding DNA sequence (CDS) gain and loss is not common despite the >1 My divergence time between HAL2 and FIL2. Such conserved gene content and synteny despite significant sequence divergence (mean $\pi = 0.0195$) makes P. hallii an ideal system to test hypotheses about sequence evolution, selection, and molecular adaptation.
of the focal gene. For example, some F₂ genotypes will simply lack a PAV gene model, which would bias inference of the strength and presence of eQTL. Therefore, we limited our eQTL analysis to only those high-confidence genes (which have both homology and transcript evidence) with one-to-one reciprocal best hit (RBH) orthologs between the two genomes. This test allowed us to count the total expression of 21,227 putative RBH orthologs (Supplementary Data 3) by summing allele-specific and shared transcript counts; 80.4% (17,061) of these had a mean of >1 raw count across the F₂ population. The voom 23-normalized transcript counts from these genes were used as the independent phenotypic variables in our eQTL analysis. Combined, we found significant eQTL among 58% of the 17,061 genes with RBH orthologs between HAL2 and FIL2.

The genetic architecture of regulatory variants can have profound consequences on the evolutionary processes that produce heritable genetic diversity.7,24 Cis-elements in coding or promoter regions typically regulate the expression of a single, physically linked gene. In contrast, trans-elements, like transcription factors, can have global regulatory effects, which may constrain adaptation by increasing the likelihood of interference and limiting the efficacy of selection.9 Due to the potential antagonistic effects of trans-regulatory elements, we expected drought-adaptive regulatory networks in *P. hallii* to be dominated by cis variants. Indeed, the 9,088 significant cis-eQTL on average explained 30.2% of total gene expression variation while the 1,314 trans eQTL explained just 7.4% (Table 2).

To develop candidate sequence variants for cis-regulatory QTL, we conducted coding (Supplementary Data 9) and regulatory sequence (Supplementary Data 10) alignments of RBH orthologs between the HAL2 and FIL2 genomes. Cis-regulated genes were significantly more likely to have accumulated non-synonymous and other coding variants than genes without QTL (Table 2). This enrichment of potentially functional variants was much stronger within promoter regions, where genes with cis-eQTL were 1.5x more likely to have evolved differences in transcription factor binding affinity (TFBA) than genes without eQTL (Table 2). While it is likely that genes with differential expression are generally subject to weaker evolutionary constraint than those without evolved expression variation, these results show that cis-eQTL are generally driven by proximate regulatory loci, and not by genetically linked, but physically distant, regulatory loci.

Genes with only trans-eQTL are, by definition, not differentially regulated by local sequence variants. Consistent with such non-local effects, genes regulated by only trans-eQTL had slightly more conserved CDS (Fisher’s test odds = 1.2× more conserved, \( P = 0.08 \)) and similarly conserved TFBA (Fisher’s test odds = 1.08 ×, \( P = 0.1 \)) regions as genes without eQTL. Furthermore, since trans-regulated loci are targets of transcriptional regulatory elements, we expected transcription factor binding sites to be very conserved among genes with only trans-eQTL. Indeed, such genes had 1.45× greater transcription factor binding site sequence conservation (Fisher’s test \( P < 1 \times 10^{-5} \)) than genes with cis-eQTL, but less significantly conserved CDS regions (Fisher’s test odds = 1.32 ×, \( P = 0.003 \), Table 2). These patterns were even more significant among genes with GxE QTL (Table 2). These results suggest that non-synonymous variants are less responsible for the evolution of expression regulation than promoter sequences and demonstrate sequence conservation of transacting transcription factors binding sites.

### Exploring the causes of trans-eQTL hotspots

Differential adaptation across habitats is an example of a genotype-by-environment interaction (GxE) for fitness.4 Therefore, loci that contribute to adaptation may be disproportionately environmentally responsive and exhibit trade-offs.25 Despite their global rarity, trans-eQTL were more than four times as likely as cis-eQTL to have significant GxE effects (Fisher’s test odds = 4.40, \( P < 1 \times 10^{-16} \), Table 2), indicating a potentially adaptive role of trans-regulatory evolution between *P. hallii* varieties. Furthermore, there were three genomic hotspots where trans-eQTL were more common than cis-eQTL, which represented 3.1% of the physical genome sequence but 40.4% of all trans-eQTL (Fig. 3a–b, Table 3). Combined, these hotspots were responsible for 3.84 × (Fisher’s test \( P < 1 \times 10^{-16} \)) more GxE effects than all trans-eQTL outside this interval.

It is important to note that the eQTL hotspots we identified may not be caused by regulatory element evolution, but instead could be driven by the presence of a large physiologically QTL that altered plant water status and caused downstream gene expression variation of many water status-responsive genes. To test this hypothesis, we collected leaf water potential (LWP), which is the best available field-scale proxy for plant water status,26 from the entire F₂ population at both pre-dawn and midday at the same time as RNA sampling. There were no significant QTL (lowest empirically-derived \( P = 0.27 \), Supplementary Table 6) on any chromosome for either sampling period, indicating that genetic variation in plant water status does not in turn drive the physical clustering of trans-eQTL. Instead, the trans-eQTL hotspots and their significantly elevated environmental sensitivity suggest that pleiotropic mutations in transcription factors may be a major source of regulatory variation.

We searched for candidate genes by investigating the distribution of allelic effects among QTL mapping to each hotspot.27 For example, expression of all but two genes regulated by trans-eQTL in the *3a* primary hotspot were driven by plasticity of the FIL2, but not the HAL2 allele (Fig. 3c). Within this 964 kb interval, three genes had significant cis-eQTL (Supplementary Data 8) and elevated non-synonymous substitution rates (Supplementary Data 9). The most promising candidate, ABO3 (*P. hallii* ortholog of *A. thaliana* AT1G66600 - ABA Overly...
Supplementary Table 8. Source data for panels

Data

The FIL2 allele is 27.7x downregulated relative to the F1 allele, consistent with the extreme environmental sensitivities of non-functional ABO3 alleles in A. thaliana. It is important to note that eQTL analysis cannot definitively identify the causal loci driving trans-eQTL hotspots but instead provides candidates for future validation. However, the observed trans-eQTL colocalization clearly demonstrates the central role of global regulatory elements in evolved molecular drought responses.

Signatures of selection among cis–trans regulatory networks.

Elevated GxE among trans-eQTL, particularly within the hotspots, indicates possible selection on gene expression plasticity. To test the hypothesis of non-neutral regulatory network evolution, we examined the QTL effect distribution among the 684 genes with both cis- and trans-eQTL (Table 3). Under a neutral model, cis- and trans-regulatory elements should evolve independently, which produces a random distribution of trans-relative to cis-effect directionality (up/down regulated). However, directional and stabilizing selection produce correlated shifts among regulatory loci. For example, if directional selection favors up-regulation of a gene, adaptive evolution of both cis- and trans-regulatory loci would increase expression; such positive correlations among loci are known as reinforcing QTL effects. Alternatively, if stabilizing selection favors maintenance of ancestral expression levels, non-adaptive evolution of increased expression may be compensated for by selection for regulatory repression at another locus. Such compensatory evolution promotes antagonistic (negatively correlated) cis–trans regulatory networks.

We observed a significant genome-wide bias towards antagonistic effects (Fig. 3d, Fisher’s test odds $= 1.3, P = 0.015$) and a very strong signal in the drought treatment, where allelic effects were greater than two times more likely to be antagonistic than under neutral expectations (Fig. 3d, Fisher’s test odds $= 2.15, P < 1 \times 10^{-12}$). However, the relative strength of reinforcing and antagonistic effects was highly dependent on the position of the trans-eQTL (Table 3). For example, the *3a hotspot was enriched in reinforcing effects relative to all other trans-eQTL; however, the proportion of reinforcing effects at *3a was not significantly different than the neutral expectation of 1:1 (Table 3). In contrast, the *3b hotspot was disproportionately represented by genes with antagonistic cis-trans effects in both treatments relative to the background and neutral expectation (Table 3). These results indicate a genome-wide prevalence of antagonistic effects, which is typical of stabilizing selection on expression regulation.

Validating the physiological effects of a trans-eQTL hotspot.

Given the large number of genes regulated and the elevated rate of GxE, the trans-eQTL hotspots clearly drive variation in molecular responses within our drought-recovery experiment. In switchgrass, regulatory responses depended on the intensity and duration of drought, and our drought experiment is only a subsample of possible drought conditions. Therefore, we sought to validate the effects of the trans-eQTL hotspots in a broader context by exploring drought responses in a recombinant inbred line (RIL) population, derived from the F1 used in the eQTL study (Supplementary Table 8). We assayed leaf relative water content (RWC) and chlorophyll content (SPAD) and conducted allelic sensitivity tests on ABO3 backgrounds, this analysis permits inference of causality when testing the effects of allelic differences at a locus.

Allelic variation at the *3b hotspot never significantly affected any treatment-trait combination, and the *7 hotspot had only one significant association (RWC under drought, Supplementary Table 8). However, variation at the *3a hotspot marginally
affected RWC plasticity (t = −1.95, P = 0.052), RWC in the drought treatment (t = 1.55, P = 0.081), and significantly affected SPAD in both recovery (t = 2.42, P = 0.016) and drought treatments (t = 2.02, P = 0.044). Individually, each of these effects are not highly significant. Yet, when taken together we observe strong support for functional effects of allelic variation at the *3a hotspot with a Fisher’s combined test (χ^2_3a = 27.3, P = 0.0068, Supplementary Table 8). Combined, the allelic effects at the *3a hotspot are consistent with those of ABO3 (lower plasticity, higher constitutive RWCP and SPAD in HAL2) and demonstrate the predictive potential and multi-environment effect of trans-acting transcriptional regulatory elements.

Discussion

The intersection of physiology, quantitative genetics, and whole genome sequencing holds great promise for understanding the complex interaction between genetic variation and the environment. Using de novo whole genome assembly, large-scale field experimentation, and physiological genomic techniques in *P. hallii*, we were able to precisely map global regulatory loci and infer candidate variants. These experimental resources allow us to better infer the genetic basis of complex traits in plant adaptation, predict responses to current and future climatic stress, and develop a strategy for drought-responsive biotechnology in biofuel breeding programs.

Methods

**Genome sequencing assembly and annotation.** We sequenced the *Panicum hallii* var. *hallii* and var. *filipes* genotypes HAL2 and FIL2 using a whole genome shotgun strategy and standard sequencing protocols. Sequencing was conducted on both Illumina (HiSeq) and Pacific Biosciences (SEQUEL) platforms at the Department of Energy Joint Genome Institute (UC, Walnut Creek, CA, USA) and the HudsonAlpha Institute for Biotechnology (Huntsville AL, USA).

Sequencing effort for HAL2 (Supplementary Note 1) included one 800 bp insert 2 × 250 fragment library (150 x) and a total of 89.5x of PACBio reads (average subread length 11.8 kb). FIL2 sequencing was similar (Supplementary Note 1), with one 500 bp insert 2 × 150 fragment library (100 x) and a total of 95.8x of Pacific Biosciences reads (average subread length 9.6 kb). Both the FIL2 and HAL2 assemblies were performed using MECAT and polished using QUIVER.

We built a 325,613-marker map from shallow resequencing of the RIL population. This marker order was used to identify mis-joins in the assembly. A total of 115 mis-joins were identified in the FIL2 assembly and only 1 mis-join in HAL2. Scaffolds were then ordered, oriented, and joined together. A total of 1216 (FIL2) and 119 (HAL2) mis-joins were applied to the broken assemblies to form the final releases consisting of 9 chromosomes each. Short redundant sequence from contig ends were aligned to one another and collapsed when appropriate. A total of 18 (HAL2) and 450 (FIL2) repeated adjacent contig pairs were identified and collapsed. For the FIL2 release, a set of 30,315 (430.3 Mb) targeted clone sequences (Illumina), along with a set of 704,618 (3.74 Gb, 6.9x coverage) MOLECULE reads were used to patch 60 sequence gaps. Finally, Illumina reads were employed to correct any remaining MECAT consensus calling errors (e.g. homozygous SNPs and INDELs).

**PERRTRAN (Supplementary Note 2)** and PASA were used to produce 92,211 (HAL2) and 111,325 (FIL2) transcript assemblies from −1100 M (HAL2) and −1200 M (FIL2) 2 x 150 paired-end Illumina RNA-seq reads. Repetitive DNA elements were identified de novo with RepeatModeler. For mapping and genome comparisons, we soft-masked the genomes using repeatMasker, with ancestral repeats from RepeatMasker and the repeat annotations from repeatModeler output. Loci were determined by transcript assembly alignments and/or EXONERATE using transcript assemblies (https://github.com/nathanweeks/exonerate) against Arabidopsis thaliana, soybean, Kitaake rice, sorghum, foxtail millet, Brachypodium distachyon, grape and Swiss-Prot proteomes. Gene models were predicted by homology-based methods, FGENESH + / EST GenesCan, and AUGUSTUS using BRAKER. The best-supported predictions for each locus were selected using EST and protein support and penalized if overlapped with repeats. PASA was employed to add UTRs, splicing corrections, and alternative transcripts. High-confidence transcripts were called for loci with BLASTP Ccocrs >0.5, EST coverage, and <20% of CDS overlapping with repeats (if >20% overlap with repeats, only loci with Ccocrs >0.9 and homology coverage >0.70 were retained). Gene models with >30% TE domains (HOMER) were also called. Finally, gene models with a short single exon (<300 BP CDS), without protein domain or with weak expression evidence were removed.

Comparative genomics. We used comparative genomic approaches to accomplish the following goals: (1) identify orthologous pairs of genes, (2) define the scale and causes of presence absence variation among gene annotations, and (3) understand the scale of synteny between the HAL2 and FIL2 genomes. Given these goals and the highly-repetitive and less conserved intergenic regions in plant genomes, we used a gene-level approach to whole-genome alignments. We ignored regions that were not in proximity to annotated gene models.

The GENESPACE pipeline (Supplementary Note 3) is applied to a set of de novo genomic assemblies and annotations. In short, GENESPACE conducts standard inference of orthology using the orthofinder program but limits the search within known colinear (syntenic) blocks, generated by the multiple-collinearity inference program MCScan. This allows for the inference of orthology in duplicated chromosomal regions, as these appear as multiple distinct blocks in the alignments. In addition to pairwise peptide–peptide searches for orthologous gene groups, GENESPACE also conducts alignments against unannotated genomic sequences (via BLAT and EXONERATE) to discover the sequence identity of pseudogenized or otherwise un-annotated loci. The pipeline outputs alignments and some general sequence-divergence statistics for all orthogroup sequences among all genomes considered.

**F2 RNA sequencing and analysis.** HAL2, FIL2, and the F2 mapping population were exposed to a short term recovery drought experiment following Lovel et al. In short, all plants experienced a natural 30-day drought at the Ladybird Johnson Wildflower Center (Austin, TX; 30°19'N, 97°87' W). Drought treatment plants were harvested on 5 July 2013, while recovery treatment plants received 4 L of water on 7 July 2013 and were harvested on 8 July 2013. For each plant, we measured midday leaf water potential (LWP, Ψ_m) with a Scholander-type pressure bomb (PMS Instruments, model 1000) between 11:00 and 13:00. All plants reached anthesis by 5 July 2013.

Leaf tissue harvest was conducted on 5 and 8 July 2013 between 11:00 and 13:00, where the most recently fully emerged leaf was immediately flash frozen with liquid nitrogen. For total gene expression assays, RNA (3 μg, RIN ≥ 5) was extracted from 50–200 mg of homogenized (Geno/Grinder, Spex SamplePrep) and DNase I-treated leaf tissue with RNeasy Plant Mini kits (Qiagen). Total RNA libraries were prepared on a PerkinElmer ScIclone NGS robotic liquid handling system using Illumina’s TruSeq Stranded mRNA HT Sample Prep kit with 1 μg RNA per sample, and 10 cycles of library amplification PCR. Library quantification by KAPA Biosystem’s next-generation sequencing library qPCR was accomplished on a Roche LightCycler 480 real-time PCR instrument. Sequencing was performed on the Illumina HiSeq 2000 sequencer and a TruSeq SBS sequencing kit (200 cycles, v3, following a 2 x 150 indexed run recipe).

To quantitively expression, we mapped reads to concatenated genome assemblies and annotations of the HAL2 and FIL2 genomes. Uniquely mapping reads were
counted with STAR by definition these reads represent allele-specific expression (ASE). Shared (not allele-specific) reads were those that mapped to both orthologs; these were counted with the subread featureCount function. Total counts, calculated as the sum of ASE and shared counts, were used for eQTL analyses. High-confidence gene models with reciprocal best hit one-to-one orthologs and average total counts ≥1 were retained for further analysis. Total counts were voom normalized for eQTL analysis.

eQTL analysis. Expression quantitative locus mapping was performed in R/qtl. It is common to perform eQTL via one-way single QTL scans for each gene expression trait. However, there are several multi-locus correlated regions in our genetic map, where markers on one chromosome are in linkage disequilibrium (LD) with markers on another chromosome (Supplementary Figure 7). Therefore, genes with cis-eQTL that are physically within these LD regions may harbor spurious trans-effects of LD blocks. Since previous results suggested that eQTL were pervasive and of large effect in HAL2-FIL2 crosses, we scanned for QTL peaks, conditioning on the additive covariates of treatment and the cis-eQTL genotype. Both additive and QTL-by-treatment (G×E eQTL) scans were accomplished using the Haley-Knott and a Hidden Markov Model multipoint mapping procedure. The highest confidence gene models within an 8 Mb window were used to infer marker order in genetic mapping. These chains were split if gaps >500 kb existed, and re-chained if two blocks were adjacent. We parsed 30 large syntenic blocks, which contained 326k variable sites across all non-admixed libraries. A sliding window majority vote markers were processed to produce a saturated genetic map, where markers on one chromosome are in linkage disequilibrium (LD). Genes with cis-eQTL that are physically within these LD regions may harbor spurious trans-effects of LD blocks. Since cis-eQTL were not subject to genomic scans, we employed a standard method for inferring a genetic map with the Kosambi mapping function, we culled markers that minimized missing data and segregation distortion. The physical position was held constant, and the cumulative LOD score was generated using SAMtools mpileup utility. VarScan was used to call variants (SNPs and INDELS). These variants were then annotated using Snpeff. Bi-allelic SNPs and INDELs were annotated as synonymous, moderate effect, or large effect (e.g. premature stop) and were either retained or deleted. All genes were classified as either having or lacking significant CDS evolution. Significant genes had at least one moderate or larger effect SNP or INDEL, while non-significant genes were monomorphic, or had only low effect variants. We also tested the odds ratio against the overall prevalence of SNPs and INDELS. We re-sequenced 94 P. hallii genomes; 78 var. hallii, and 15 var. filipes. We selected more hallii samples than filipes because the geographic range of hallii is far larger than that of filipes. If identical sample sizes were used, we would be comparing geographically proximate filipes and geographically distant hallii. Since there is at least some isolation by-distance in P. hallii, we selected 38 hallii samples to choose sample sizes that generally reflect the relative range size of each variety.

Population genetic analyses. We resequenced 94 P. hallii genomes; 78 var. hallii, and 15 var. filipes. We selected more hallii samples than filipes as the geographic range of hallii is far larger than that of filipes. If identical sample sizes were used, we would be comparing geographically proximate filipes and geographically distant hallii. Since there is at least some isolation by-distance in P. hallii, we selected 38 hallii samples to choose sample sizes that generally reflect the relative range size of each variety.

 ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-018-07669-x | www.nature.com/naturecommunications 8

Promoter and coding DNA sequence variant annotation. Reciprocal best hit (RBH) orthologs are defined as those pairs of genes where each is the other best hit in the BLAST search. We selected the alternate gene for 1:1 RBH orthologs by finding the best HAL2(query):FIL2(target) and FIL2(query): HAL2(target) protein BLAT scores. The genes where both BLAT runs produced the same set of maximal gene pairings were the initial set of RBH orthologs. Given the known high degree of synteny, we then culled this set to include only pairings where both genes were on the same chromosome region, allowing for 90% of the total chromosome length as buffer.

To annotate coding DNA sequences (CDS), FIL2 CDS sequences were aligned against corresponding HAL2 sequences using Minimap2. The resulting alignment file was then subtracted to generate a subtracted CDS transcript file. A pileup-formatted file was generated using SAMtools mpileup utility. VarScan was used to call variants (SNPs and INDELS). These variants were then annotated using Snpeff. Bi-allelic SNPs and INDELs were annotated as synonymous, moderate effect, or high effect (e.g. premature stop), or insertion/deletion. All genes were classified as either having or lacking significant CDS evolution. Significant genes had at least one moderate or larger effect SNP or INDEL, while non-significant genes were monomorphic, or had only low effect variants.

We also tested the odds ratio against the overall prevalence of SNPs and INDELS. We re-sequenced 94 P. hallii genomes; 78 var. hallii, and 15 var. filipes. We selected more hallii samples than filipes as the geographic range of hallii is far larger than that of filipes. If identical sample sizes were used, we would be comparing geographically proximate filipes and geographically distant hallii. Since there is at least some isolation by-distance in P. hallii, we selected 38 hallii samples to choose sample sizes that generally reflect the relative range size of each variety.

We selected more hallii samples than filipes as the geographic range of hallii is far larger than that of filipes. If identical sample sizes were used, we would be comparing geographically proximate filipes and geographically distant hallii. Since there is at least some isolation by-distance in P. hallii, we selected 38 hallii samples to choose sample sizes that generally reflect the relative range size of each variety.

We selected more hallii samples than filipes as the geographic range of hallii is far larger than that of filipes. If identical sample sizes were used, we would be comparing geographically proximate filipes and geographically distant hallii. Since there is at least some isolation by-distance in P. hallii, we selected 38 hallii samples to choose sample sizes that generally reflect the relative range size of each variety.

We selected more hallii samples than filipes as the geographic range of hallii is far larger than that of filipes. If identical sample sizes were used, we would be comparing geographically proximate filipes and geographically distant hallii. Since there is at least some isolation by-distance in P. hallii, we selected 38 hallii samples to choose sample sizes that generally reflect the relative range size of each variety.
Dating divergence time. OrthoFinder was used with default parameters to identify clusters of single copy genes in 7 species along with Arabidopsis as out-group. Of the 3311 ortholog sets, we randomly chose 100 sets and aligned them in Clustal-omega. We used a HKY substitution model with 4 gamma rate categories and an uncorrelated log-normal relaxed clock model in BEAST2 to estimate the divergence times. A calibration time of 52 Mya was used for Brachypodium and Sorghum split. We performed four independent runs, each with 10 million generations and 5 million chain length, saving every 5000 chains. LogCombiner was used to combine trees that have an effective sample size of at least 200. The consensus trees were further estimated in TreeAnnotator using mean annotations are available through phytozome.

qPCR validation of RNA sequencing. For quantitative RT-PCR, we isolated total RNA from the leaves using an RNA extraction kit (TRIZol reagent, Invitrogen). About 0.5 μg total RNA was reverse-transcribed using the SuperScript III First-Strand Synthesis System (Invitrogen) in a volume of 20 μl to obtain cDNA. We used primers PhABO3q-F (5′-CCGGCATCTGATCCATCAAA-3′) and PhABO3q-R (5′-TCTGATGACCTGGGAAG-3′) for amplifying the transcript of PhABO3 and PhUBq-F (5′-CTCTGGATCTCCCTGTC-3′) and PhUBq-R (5′-TTCTGATTGATCTGATCAT-3′) for amplifying the ubiquitin-conjugating enzyme (PhUBq) at an internal control. We carried out quantitative RT-PCR in a total volume of 10 μl containing 2 μl of the reverse-transcribed product above, 0.25 μM gene-specific primers and 5 μl LightCycler 480 SYBR Green I Master (Roche) on a Roche LightCycler 480 II real-time PCR System according to the manufacturer's instructions. The measurements were obtained using the relative quantification method.

Recombinant inbred line field experiment. 294 F2 individuals of the HAL2xFIL2 recombinant inbred line (RIL) population were grown at Brackenridge Field Laboratory (BFL, Austin, TX, USA) in spring and summer of 2016. Seeds were germinated in the BFL greenhouse on 29-February and 3-March 2016. For each RIL line, ten seed coats for 40 seed were removed by scarifying the seed coat and placing it in wet aquarium sand on petri plates. Sand was rewetted as necessary until 14 March. Then, the seedlings were transplanted to the greenhouse to make a randomized split-plot design. The maximal position of each trans-eQTL hotspot was extracted and the consensus trees were further estimated in TreeAnnotator using mean and median node heights.

All RNA and resequencing reads have been deposited in the NCBI Short Read Archive (https://www.ncbi.nlm.nih.gov/sra). Bioproject, sample IDs, and metadata can be found in Supplementary Data 1 and 7. Both genome assemblies and annotations are available through phytozone (https://phytozone.jgi.doe.gov). The assemblies have also been deposited on Genbank (https://www.ncbi.nlm.nih.gov/genbank) under BioProjects PRJNA357853 (HAL2) and PRJNA352557 (FIL2). All statistical, QTL mapping, and visualization functions were implemented in R 3.4.3 and have been compiled into an R package stored on github (https://github.com/jtowell/qtlTools). Additional supporting data is provided as Supplementary Data. Details are provided in the Description of Additional Supplementary Files and the Reporting Summary. The source data underlying Figs 1A-1B, 2A-D, 3B-C and Supplementary Figure 3-7 are provided as a Source Data File. Source data for Fig 3A, D and Supplementary Figure 1-2 are found in Supplementary Data 7 and Supplementary Data 2, respectively. A reporting summary for this Article is available as a Supplementary Information file.

Received: 11 June 2018 Accepted: 16 November 2018
Published online: 06 December 2018

References
1. Ciais, P. et al. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437, 529–533 (2005).
2. Hover, D. L., Knapp, A. K. & Smith, M. D. Resistance and resilience of a grassland ecosystem to climate extremes. Ecology 85, 2652–2656 (2014).
3. Wilcox, K. R. et al. Asymmetric responses of primary productivity to precipitation extremes: a synthesis of grassland precipitation manipulation experiments. Glob. Chang. Biol. 23, 4376–4385 (2017).
4. Juenger, T. E. Natural variation and genetic constraints on drought tolerance. Curr. Opin. Plant Biol. 16, 274–281 (2013).
5. Bower, J. S. Plant productivity and environmental change. Science 218, 443–448 (1982).
6. Wagner, G. P. & Lynch, V. J. The gene regulatory logic of transcription factor evolution. Trends Ecol. Evol. 23, 377–385 (2008).
7. Schaecke, B. et al. Inheritance of gene expression level and selective constraints on trans- and cis-regulatory changes in yeast. Mol. Biol. Evol. 30, 2121–2133 (2013).
8. Keightley, P. D. & Hill, W. G. Variation maintained in quantitative traits with mutation-selection balance: pleiotropic side-effects on fitness traits. Proc. Biol. Sci. 242, 95–100 (1990).
9. Kondrashov, A. S. & Turelli, M. Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. Genetics 132, 603–618 (1992).
10. Matsukura, S. et al. Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. Mol. Genet. Genom. 283, 185–196 (2010).
11. DesMarais, D. L. et al. Physiologenic genomics of response to soil drying in diverse arabidopsis accessions. Plant Cell 24, 893–914 (2012).
12. Mizoi, J., Shinnozaki, K. & Yamasuchi, S. Arabidopsis AP2/ERF family transcription factors in plant abiotic stress responses. Biochim. Biophys. Acta 1819, 86–96 (2012).
13. Lowry, D. B. et al. The genetics of divergence and reproductive isolation between ecotypes of Panicum hallii. New Phytol. 205, 402–414 (2015).
14. Randle, J. & Boger, T. Evolution of the Poaceae (Grain plants). J. Systematics Evol. 53, 117–137 (2015).
15. Somerville, C., Youngs, H., Taylor, C., Davis, S. C. & Long, S. P. Seed stocks for lignocellulosic biofuels. Science 329, 790–792 (2010).
16. Lowry, D. B., Purmal, C. T. & Juenger, T. E. A population genetic transect of Panicum hallii (Poaceae). Am. J. Bot. 99, 592–601 (2013).
17. Schilling, S. & Durbin, R. The effect of variable population size and separation history from multiple genome sequences. Nat. Genet. 46, 919–925 (2014).
18. Grover, C. E. et al. Comparative genomics of an unusual biogeographic disjunction in the cotton tribe (Gossypieae) yields insights into genome downsizing. Genome Biol. Evol. 9, 3328–3344 (2017).
19. van Velzen, R. et al. Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proc. Natl Acad. Sci. USA 115, E4700–E4709 (2018).
20. Roy, J. & Mooney, H. A. Physiological adaptation and plasticity to water stress of coastal and desert populations of Heliotropium curassavicum L. Oecologia 52, 370–375 (1982).
21. Fedderke, M. E. & Hofmann, G. M. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annu. Rev. Physiol. 61, 243–282 (1999).
22. Rengel, D. et al. A gene-phenotype network based on genetic variability for drought tolerance traits. Plant Physiol. 169, 2826–2840 (2015).
23. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47–e47 (2015).
24. Emerson, J. J. et al. Natural selection on cis and trans regulation in yeasts. Genome Res. 20, 826–836 (2010).
25. Ågren, J., Oakley, C. G., McKay, J. K., Lovell, J. T. & Schemske, D. W. Genetic mapping of adaptation reveals fitness tradeoffs in Arabidopsis thaliana. Proc. Natl Acad. Sci. USA 110, 21077–21082 (2013).
26. Jones, H. G. Monitoring plant and soil water status: established and novel methods revisited and their relevance to studies of drought tolerance. J. Exp. Bot. 58, 119–130 (2007).
