A generalist–specialist trade-off between switchgrass cytotypes impacts climate adaptation and geographic range

Joseph D. Napier, Paul P. Grabowski, John T. Lovell, Jason Bonnette, Sujan Mamidi, Maria Jose Gomez-Hughes, Acer VanWallendael, Xiaoyu Weng, Lori H. Handley, Min K. Kim, Arvid R. Boe, Philip A. Fay, Felix B. Fritsch, Julie D. Jastrow, John Lloyd-Reiley, David B. Lowry, Roser Matamala, Robert B. Mitchell, Francis M. Rouquette Jr, Yanqi Wu, Jenell Webber, Teresa Jones, Kerrie Barry, Jane Grimwood, Jeremy Schmutz, and Thomas E. Juenger

Polyploidy results from whole-genome duplication and is a unique form of heritable variation with pronounced evolutionary implications. Different ploidy levels, or cytotypes, can exist within a single species, and such systems provide an opportunity to assess how ploidy variation alters phenotypic novelty, adaptability, and fitness, which can, in turn, drive the development of unique ecological niches that promote the coexistence of multiple cytotypes. Switchgrass, *Panicum virgatum*, is a widespread, perennial C4 grass in North America with multiple naturally occurring cytotypes, primarily tetraploids (4x) and octoploids (8x). Using a combination of genomic, quantitative genetic, landscape, and niche modeling approaches, we detect divergent levels of genetic admixture, evidence of niche differentiation, and differential environmental sensitivity between switchgrass cytotypes. Taken together, these findings support a generalist (8x)–specialist (4x) trade-off. Our results indicate that the 8x represent a unique combination of genetic variation that has allowed the expansion of switchgrass’ ecological niche and thus putatively represents a valuable breeding resource.

Significance

Polyploidy, which occurs in roughly half of all flowering plants and an even higher percentage of grasses, is thought to be a major driver of adaptation. Higher numbers of copies of each gene in polyploid genomes can increase genetic diversity, which could drive shifts in habitat preference, adaptability, and fitness. To test the effects of increased ploidy, we compared genomic diversity, environmental niche, and fitness responses across climatic gradients between tetraploid and octoploid switchgrass. We found that the octoploids contained novel combinations of the ancestral tetraploid genetic diversity, which was linked to the expansion of switchgrass into unsuitable habitats for tetraploid populations. Our experiments revealed evidence of niche divergence, differential fitness, and a generalist–specialist trade-off between cytotypes.

Author contributions: J.D.N., P.P.G., J.S., and T.E.J. designed research; J.B., X.W., L.H.H., M.K.K., A.R.B., P.A.F., F.B., J.D.J., J.L.R., D.B.L., R.M., R.B.M., F.M.R., Y.W., J.W., J.T.L., K.B., and J.G. performed research; J.D.N., P.P.G., J.T.L., S.M., M.J.G.-H., A.R.B., X.W., L.H.H., M.K.K., A.R.B., and T.E.J. wrote the paper.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2118879119/-/DCSupplemental.

Published April 4, 2022.
and fitness. Specifically, we discovered 1) recurrent and evolutionary distinct genesis of 8× populations containing novel combinations of genetic diversity; 2) similar morphological/ecotypic divisions within 4× and 8× cytotypes, but divergent cyotype fitness clines, indicating a generalist–specialist trade-off; and 3) niche evolution between 4× and 8× linked to climate adaptation. Combined, our results indicate that mixed-ploidy systems can be used as valuable tools to bolster the resilience of natural and agronomic systems by providing insight on how ploidy variation enables niche divergence, fitness trade-offs, and range expansion.

Results

Recurrent and Evolutionary Distinct Origin of 8× Switchgrass.

Previous genetic studies focused on switchgrass have mainly utilized the 4× cyotype, which is an allotetraploid with disomic inheritance (2n = 4x = 36) (22, 23), containing two copies each of the N and K subgenomes (4x = KKNN). Little is known about the meiotic behavior of 8× switchgrass (24), but previous analyses suggest that 8× samples are genetically similar to 4× samples (21). The 8× (2n = 8x = 72) switchgrass are generally considered auto-allopolyploids, behaving like an auto-tetraploid with four copies of each of the two subgenomes present in 4× switchgrass (8x = KKKKNNNNN) (20). However, the true inheritance patterns may be heterogeneous across the genome, as seen in other autopolyploids (25, 26). Furthermore, 4× and 8× switchgrass are reproductively isolated due to a postfertilization incompatibility system (27). To date, 8× switchgrass has been assumed to be predominantly composed of a single northern gene pool that is derived from 4× molecular diversity found in Midwestern upland switchgrass, but there has never been a comprehensive survey of the range-wide genetic diversity present in the 8× cyotype. Therefore, to explore the molecular population structure across the natural range of switchgrass, we used a total of 544 deeply resequenced 4× [previously published (18); cultivars and breeding material were omitted for this analysis] and 159 resequenced 8× samples from across the natural range of both cytotypes (Fig. 1A; SI Appendix, Table S1) genotyped at over 58 million single-nucleotide polymorphisms (SNPs).

We hypothesized that 8× switchgrass is derived from ancestral 4× genetic diversity. Analysis of the resequencing data supported the presence of three geographically distinct genetic subpopulations across both 4× and 8× cytotypes (referred to as Midwest, Gulf, and Atlantic). These subpopulations are consistent with previous findings that only utilized 4× samples (18), supporting our hypothesis that 8× genotypes are putatively derived from 4× molecular diversity. Due to the complexities of working with a mixed-cyotype system, we verified these patterns using a reference-free k-mer method, which accurately replicated the SNP-based results (Mantel test r = 0.97), indicating that polyploidy did not substantially bias the SNP-based population genetic estimation in this study. Compared to 4× switchgrass, 8× exhibited higher levels of admixture among the three subpopulations, notably between the Midwest and the other two subpopulations, Gulf and Atlantic (4×: 86% have >97% ancestry from a single gene pool, while 4% have >20% ancestry from two gene pools; 8×: 53% have >97% ancestry from a single gene pool, while 29% have >20% ancestry from two gene pools; Fig. 1B and C). The high levels of genetic admixture detected in the 8×, noticeably absent in the 4×, support the hypothesis that higher ploidy is associated with among-subpopulation mating (23). These results challenge a prevailing assertion that 8× is a secondary taxonomic division within the genetic lineages that compose the cold-adapted upland ecotype of switchgrass (e.g., ref. 28).

To further explore the process of 8× formation, we compared phylogenetic patterns found using nuclear markers to results based on molecular variation in chloroplast genomes, which generally show higher conservation compared to nuclear genomes and provide valuable tools for understanding the phylogenetic and evolutionary relationships between closely related taxa (29). Specifically, discordance between nuclear and chloroplast phylogenies within species or genera can provide evidence of chloroplast capture resulting from hybridization, though it can also arise from incomplete lineage sorting (30). For a subset of switchgrass genotypes of both cytotypes, we built phylogenetic trees from both nuclear variation and individual chloroplast assemblies to test for congruence between maternal and nuclear genomes between cytotypes (Fig. 2A; SI Appendix, Fig. S1). Among nuclear trees, 8× occur within the distinct clusters of all three gene pools, supporting repeated establishment of 8× lineages and indicating that 4× switchgrass are ancestral to the 8× (21). However, we observed significant discordance between the nuclear and chloroplast trees, whereby numerous individuals that occur within the Gulf and Atlantic clusters of the nuclear tree appear tightly grouped, with the chloroplast haplotypes dominated by individuals with a strong Midwest genetic background (Fig. 2A). Moreover, we estimated the divergence between the five main clades suggested by the tree built from chloroplast haplotypes using BEAST version 2.6.6 (31), with each clade containing both 4× and 8× samples. The estimated dates of the nodes for each of these clades range from 65 thousand years ago (kya) to 86 kya (SI Appendix, Table S1 and Fig. S2). The estimated age of divergence between switchgrass gene pools is estimated at >600 kya (18), indicating that the transition from 4× to 8× happened separately within each gene pool, potentially multiple times, well after the divergence of the gene pools. In fact, identical and near-identical chloroplast haplotypes are shared by 4× and 8× individuals, suggesting a very recent shift in ploidy in some populations (SI Appendix, Table S2). Combined, these observed patterns of population genetic structure, ancestry, and phylogenetic relationships indicate that 8× have emerged multiple times and most likely as the product of disparate parents, including individuals of divergent ancestry. Repeated establishment of 8× populations supports the idea that repeated emergence of 8× switchgrass might promote coexistence (32), which would increase 8× frequency in the overall switchgrass population, providing more chances to avoid minority cytype exclusion (33).

Morphological and Fitness Differentiation within Cytotypes of Switchgrass. Within 4× lines of switchgrass, documented climate-associated adaptation underscores phenotypic divergence between distinct switchgrass ecotypes (e.g., ref. 18), but the diversity in 8× switchgrass has been assumed to be far less dynamic and has primarily been treated as an exclusively ecotypically upland cytype (23, 28). However, our results indicate that cytype evolution is far more complex, where some 8× genotypes are highly admixed composites of upland ecotype and lowland ecotype 4× genetic variation, while others are closely related to either upland or lowland 4× genotypes. Given these disparate patterns of genetic relatedness, it is also possible that 8× phenotypes are far more diverse than previously expected. To explore overall morphological divisions and composition of both cytypes, we performed in silico classification

https://doi.org/10.1073/pnas.2118879119

pnas.org
using a suite of informative morphological data for 408 4X and 102 8X switchgrass samples with sufficient common-garden phenotypic data. Traits of target samples, originally collected from across the entire US geographic range of switchgrass, were measured at both northern and southern common-garden plantings (Austin, TX, and Kellogg Biological Research Station, MI). Assignment was conducted by using a Discriminant Analysis of Principal Components [DAPC; Fig. 3A (34, 35)], which

**Fig. 1.** Analyses of population genetic structure. (A) Spatial distributions of the switchgrass subpopulations. Pie charts show the proportion of an individual that was assigned to each subpopulation; 8x samples have a black collar around the pie chart. (B) Bar graphs represent the multilocus genetic assignment of each individual to each ADMIXTURE group. (C) PCA results for both cytotypes, where data points are shaded according to their ADMIXTURE genetic-cluster association.

**Fig. 2.** Phylogenetic relationships. (A) Chloroplast tree for a subset of switchgrass samples. Gray branches indicate less than 0.95 posterior probability support, and black branches indicate greater than 0.95 posterior probability support. Colors indicate the nuclear ancestry of the majority of samples in the assigned clade. (B) NeighborNet tree showing nuclear ancestry of all 4x and 8x samples.
placed the 4X into three broad groupings consistent with both the
long-standing division of upland (n = 232) and lowland (n = 96)
ectotypes (36–39), as well as emerging evidence of a coastal form
(n = 190) (18). Inconsistent with expectations from published lit-
erature, our results suggest different subsets of the 8X group, with
upland (n = 31), lowland (n = 9), and coastal (n = 62) 4X previously
classified through neural networks incorporating expert
knowledge (18), suggesting a previously undocumented high
degree of phenotypic variability across the octoploid cytotype.

Given our observations of extensive molecular and pheno-
typic diversity in the 8X cytotype, a pressing question is
whether naturally occurring switchgrass cytotypes exhibit diver-
gent climate adaptation and are associated with variable fitness
responses across a range of environments. Specifically, shifts in
ecological tolerance (i.e., shifts in the range of suitable climate
conditions) might have occurred from: 1) alterations in genetic
composition associated with higher ploidy levels and hybridiza-
tion; or 2) divergent adaptation in reproductively isolated cyto-
types (27, 40). Within switchgrass, biomass is considered a
proxy for plant fitness (17), capturing both aspects of vegetative
and reproductive vigor and output. This relationship is sup-
ported by the results from two field experiments conducted by
Palik et al. (41), which revealed a high correlation between bio-
mass and number of seeds per plant (R² = 0.83). To control
for the effects of underlying genetic structure and ensure suffi-
cient sample sizes, we compared the relative fitness, estimated
by biomass, of 4X and 8X in the Midwest subpopulation
(ancestry coefficient > 0.9) at 10 common-garden sites against
the climate distance from the origin climate of each collection
location (Fig. 3B, SI Appendix, Fig. S3). Our results showed
that, in common gardens with climates most similar to the cli-
mate of origin (the climate of the original collection sites), 4X
genotypes have higher relative fitness than 8X genotypes, but
fitness declines more slowly for 8X than 4X, with 8X eventually
having higher relative fitness than 4X in climates more different
from the climate of origin (ploidy and climate distance interac-
tion, F = 44.68, P = 3.23e-11).

Distinct Niches Linked to Genetic Variability Supports Cytotype
Coexistence. Ploidy increases may expand the diversity of suit-
able habitats potentially facilitating range expansion (42). To
test this hypothesis, we used environmental niche modeling
(ENM) based on occurrence data to explore potential niche dif-
terences between 8X and 4X overall and within the Midwest
subpopulation, where within-cytotype replication was high
enough to permit robust inference within subpopulations (43,
44). Although there are 8X within the Gulf and Atlantic sub-
populations, low representation of these samples in our dataset
(n < 15 for both) preclude a similar ENM approach for these
target groups. Using the identity test to compare niche similar-
ity (45), we observed statistically different niches in both the
overall set (D = 0.799, P = 0.002) and within the Midwest
subpopulation (D = 0.685, P = 0.002), suggesting different
environmental distributions of each switchgrass cytotype (Fig.
4). Higher habitat suitability for 8X in pronounced northeast-
ern/southeastern 4X range gaps suggests that increased ploidy
allowed colonization of habitat previously unsuitable to switch-
grass. ENM results from the Midwest indicate a wider breadth
of habitat suitability for the 8X (B2, a metric of niche breadth;
see ref. 46; of 8X = 0.676, B2 of 4X = 0.391), but, throughout
much of the central Midwest, 4X switchgrass have relatively
higher habitat suitability. These findings provide landscape-
scale support for the strong specialist vs. generalist trade-off we
observed in the common gardens, which may enable cytotype
coexistence.

It is possible that the broader environmental, phenotypic,
and ecological distributions of the 8X cytotype is, in part,
caused by wide crosses and reticulate admixture that may have
accompanied ploidy increases. Our population genomic analy-
ses indicated a putative among-subpopulation hybrid origin of
some 8X, which potentiates a role for increased genetic varia-
tion in facilitating ecological niche evolution between cytotypes
(40, 47). In the Midwest subpopulation, for instance, the
ancestry coefficients from ADMIXTURE are similar between
8X and 4X, but discordance between the chloroplast and
nuclear phylogenetic trees along with higher levels of admixture
(Figs. 1 and 2) between the Gulf and Midwest further south
suggest that the 8X Midwest subpopulation might harbor
alleles from the diverged Southern “Gulf” and “Atlantic” clades.
As observed in some 4X genotypes (18), admixture from locally
adapted gene pools could have enabled range expansion of
8X genotypes. To evaluate this potential mechanism, we
investigated signals of Atlantic and Gulf ancestry across the genomes of Midwest switchgrass using ancestry-informative alleles from SNPs with high $F_{st}$ (> 0.5) between the three gene pools (i.e., Gulf vs. Midwest and Atlantic vs. Midwest). The mean ancestry signal of both Gulf and Atlantic gene pools was higher in the Midwest 8x than 4x (MW8× Atlantic: 0.147, MW4× Atlantic: 0.0596; Mann–Whitney $u$ test $P$ value for Atlantic < 1.4 e-15; MW8× Gulf: 0.152, MW4× Gulf: 0.0552; Mann–Whitney $u$ test $P$ value for Gulf < 2.2e-16; SI Appendix, Fig. S4). Moreover, there is evidence of geographic variability in the ancestry signal, as Midwest 8x samples in the west demonstrate a higher ratio of Gulf to Atlantic ancestry signal compared to Midwest 8x samples in the east (Mann–Whitney $u$ test $P$ value < 1.7e-10; SI Appendix, Fig. S5).

To explicitly test if retained genomic introgression regions were putatively under selection in 8x populations, we used a series of targeted redundancy analyses (RDAs) to relate the presence of Gulf and Atlantic ancestry in 8x individuals with climatic and geographic factors. Gulf and Atlantic ancestry informative SNPs were >2x more strongly associated with climate (percentage of variance explained: Gulf = 42.1% and Atlantic = 41.3%) than geography (Gulf = 16.6% and Atlantic = 16.2%). When partitioning out the effect of geographic distance, for the Gulf and Atlantic, respectively, we identified 1,108 and 1,120 outlier candidate introgression SNPs that were strongly associated with aspects of climate. We compared these sets of SNPs in the Midwest 8x with the most pronounced shifts in allele frequencies away from the Midwest background and toward Atlantic ($n = 705$) or Gulf ($n = 753$) frequencies (Allele Frequency Shift Approach [AFS]). Of these, 126 SNPs in the Atlantic and 157 Gulf AFS outliers met the criteria to be included in the RDAs run using all Midwest samples, with 67 (53.2%) and 88 (56.1%) being significantly correlated with climate from the RDA approach. In comparison, only 4.7% and 5.0% of all SNPs in the RDA and within the same minor allele frequency (MAF) range have significant RDA correlations, an 11.3x and 11.2x enrichment in the AFS outliers. Many of the AFS outliers were on chromosomes 07K, 07N, and 08N (65.1% in Atlantic and 57.3% in Gulf), raising the possibility that these chromosomes are driving the enrichment pattern. However, even when these chromosomes are removed, 20 of 44 (45.5%) and 33 of 67 (49.3%) of AFS outliers have significant RDA results, compared to 4.7% and 5.3% of all SNPs in the RDA across the same chromosome set and similar MAF range, a 9.9x and 9.3x enrichment in the AFS outliers. This suggests that in the Midwest 8x, genomic regions containing SNPs with alleles at frequencies similar to those found in Atlantic and/or Gulf switchgrass play an outsized role in climate adaptation compared to the genome on the whole.

The SNPs in Midwest 8x with significant climate correlations from the RDA and allele frequency shifts toward Atlantic ($n = 67$) or Gulf ($n = 88$) demonstrated a strong relationship with primarily two climate variables: temperature seasonality (Bioclimatic variable 4) and precipitation in the driest quarter (Bioclimatic variable 17). These SNPs are distributed throughout the genome, but two regions, the left arm of Chr07 and the pericentromere of Chr08, are enriched for these overlapping candidate SNPs (Fig. 5). Specifically, on Chr08, there is a broad signal along a 16 Mb region of the pericentromere associated with precipitation in the driest quarter. We detected elevated linkage disequilibrium (LD) in this Midwest subpopulation in this 16 Mb region, where SNPs separated by 10 to 100 kb have LD values 3.7x greater than SNPs outside the pericentromeric region, compared to 2.5x and 2.6x in the Atlantic and Gulf subpopulations, respectively (SI Appendix, Fig. S6). Elevated LD in this region extends even farther, with SNPs separated by 1 to 10 Mb still having LD values 5.3x higher than background levels, compared to 2.8x and 1.7x for the Atlantic and Gulf subpopulations. The broad pattern on Chr08 is consistent with patterns expected from an inversion in Midwest 8x with origins from Southern switchgrass (i.e., Gulf and Atlantic).

**Discussion**

Examining ploidy variation within species provides a rare opportunity to examine patterns of ploidy establishment, coexistence of mixed cytotypes, and subsequent divergence (3) without the confounding effects of speciation. Using a combination of genomic, phenotypic, and niche modeling approaches, we discovered that the switchgrass 8x are not, as originally assumed, a uniform upland clade (e.g., refs. 23 and 28), but a genetically and phenotypically diverse polyphyletic group that arises within all known switchgrass ecotypes and gene pools (18). Moreover, compared to the 4x, 8x display higher...
performance stability over increasing climate distance and distinct ecological niches with suitable habitat in noticeable range gaps, indicating weaker signals of local adaptation and broader environmental tolerances. These results indicate that increases in switchgrass ploidy are correlated with stronger “general-purpose genotypes” (i.e., ref. 48), allowing them to persist in a wider range of environments: a classic generalist–specialist trade-off. When comparing diploids with higher ploidy levels, there has been a long-standing debate about whether shifts in tolerance are attributable to general-purpose genotypes (e.g., refs. 48–51) or if higher ploidy genotypes exist as a series of variable specialists (52, 53). Our results provide strong support that increases in ploidy in similar genetic backgrounds result in increased ecological tolerance that allows range expansion (54), but possibly at the cost of fitness or competitiveness under optimal conditions. Recent transcriptome analyses with another mixed-cytotype system (common reed) support such a trade-off, indicating that tetraploids completed their life cycle faster, but octoploids developed more stress tolerance (55).

Our results indicate higher levels of admixture of diverged gene pools in the 8X than 4X, suggesting that the occurrence of wide crosses might have promoted a shift to higher ploidy. For example, in the Midwest genetic subpopulation, 8X exhibit higher levels of Southern (Atlantic and Gulf subpopulations) switchgrass ancestry than the 4X, with many of the genomic regions containing Southern ancestry putatively under selection by the environment. This indicates that the elevated Southern switchgrass variation contained in the 8X, which is likely related to their evolutionary origins, served as adaptive introductions that allowed the expansion of the upland switchgrass range into habitat unsuitable for the 4X (56, 57). Specifically, when considering Chr08N, the LD patterns, high frequency of Southern switchgrass alleles, and correlation with climate variables raise the intriguing possibility that this region contains an inversion introduced by gene flow and acting as an “adaptive cassette” (58) and driven to near fixation by positive selection.

Fig. 5. Shifts in allele frequency from Midwest ancestry toward Gulf (A) and Atlantic (B) ancestry. The red SNPs are the overlapping outliers identified by both an environmental approach (RDA outliers) and the strongest shifts in allele frequency.

Overall, the genome-wide results indicate that elevated levels of Southern switchgrass variation in 8X Midwest switchgrass putatively promoted the evolution of the switchgrass range into previously unsuitable habitat for 4X switchgrass, emphasizing the beneficial effects of increased genetic variation resulting from higher ploidy levels (10, 47, 59). The signals detected here offer a mechanistic basis for colonization and subsequent coexistence of mixed-cytotype systems in their native ranges.

While the majority of biofuel switchgrass cultivars are 4X (60), our results challenge a traditional assumption that 4X span a considerably broader geographical range. Contrasting patterns of relative fitness and adaptation between the cytotypes across broad environmental gradients within the same gene pool suggest that increases in ploidy level have resulted in useful combinations of genetic diversity not present in the 4X. This is notable, as much of the genomic diversity associated with environmental selection in the switchgrass 4X was unique to each subpopulation (18), and, if some of the unique gene variants were to be combined, the resulting progeny would have adaptation traits allowing expansion to new ecological niches (61). Our results indicate that the 8X already represent such a unique combination of genetic variation that has allowed the expansion of switchgrass’ ecological niche and thus represent a valuable breeding resource. With the acceleration of anthropogenic climate change, our results suggest that careful examination of the genomic divergence present in systems with intraspecific ploidy variation could hold answers for utilizing natural diversity to ameliorate the effects of ongoing environmental shifts.

Materials and Methods

Flow Cytometry. To determine the ploidy level of target switchgrass samples, we utilized a LSRFortessa SORP Flow Cytometer (BD Biosciences). A total of 200 to 300 mg of young leaf tissue was macerated by using a razor blade and then treated with 1 mL of CyStain PI Absolute P nuclei extraction buffer (Sysmex Flow Cytometry) mixed with 1 μL of 2-mercaptoethanol for 15 min. Free nuclei were isolated by using a CellTrics 30-μm filter (Sysmex) followed by 20-min treatment with 2 mL of CyStain PI Absolute P staining buffer (Sysmex), 12 μL of propidium...
iodide, and 6 μL of RNase A for 20 min on wet ice. Subsequently, prepped samples were run on the flow cytometer to assess nuclei size, with a minimum of 10,000 nuclei analyzed per individual. Flow-cytometry results were analyzed by using FlowJo software (BD Biosciences), and average units of fluorescence per nucleus were used to bin samples into three categories. Using binning parameters established with flow-cytometry data from switchgrass samples of verified ploidy, ploidy level of the samples was assigned as 4x if the cell population had 40,000 to 80,000 units of fluorescence, 6x for 80,000 to 100,000 units, and 8x for 100,000 to 140,000 units.

Determining Ploidy with nQuire. We use the program nQuire (62) to determine the ploidy of samples in our dataset. We tested multiple coverage thresholds and found that requiring 20+ coverage in at least 500,000 variable positions across the genome provided ploidy calls that agree with flow-cytometry results and expectations based on previous population-structure results. Samples with fewer than 500,000 variable positions with 20+ coverage were removed from the analysis. Based on the nQuire results, the final dataset includes 544 4x, 3 6x, and 159 8x samples.

Ploidy Considerations in Mixed Dataset. A challenge of comparing between individuals of differing ploidy is generating genotypes that are comparable across cytotypes. Simulations show that simply using diploid genotypes for all cytotypes can generate biases when examining population structure (63). As such, we generated dosage genotypes for all samples reflecting the portion of the genotype that is the reference allele. The dosage genotypes ranged from 0 to 2, with 4x switchgrass containing dosage genotypes of 0, 1, or 2 reference alleles and 8x switchgrass containing dosage genotypes of 0, 0.5, 1, 1.5, and 2, reflecting 0 through 4 copies of the reference allele. We also generated diploid genotypes (three potential genotypes) for all samples to use with ADMIXTURE, which does not accept dosage genotypes.

SNP Calling. The samples were sequenced by using Illumina HiSeq ×10 and Illumina NovaSeq 6000 paired-end sequencing (2 × 150 bp) at HudsonAlpha Institute for Biotechnology, Huntsville, AL, and the Department of Energy Joint Genome Institute, Berkeley, CA. To account for variable library sizes, reads were pruned to a max coverage of 50X. Reads were mapped to the P. virgatum v5 assembly (18) by using bwa-mem (64). Duplicate reads were filtered by using Picard (http://broadinstitute.github.io/picard) and realigned around indels by using GATK 3.0 (65). Multisample SNP calling was done by using SAMTools mpileup (66) and Varscan V2.4.0 (67) with a minimum coverage of eight and a minimum alternate allele count of four. Alleles were determined by using a binomial test (hypothesized probability of success = 0.5). However, for the tetrasomic calls, a binomial test was performed for three hypothesized probabilities (0.25, 0.5, and 0.75) using the binom.test function in R based on a maximum P value and the contribution of reads that match reference: Five tetrasomic genotypes were determined, equivalent to AAAA, AAAB, AABB, AAAB, and BBBB (or 0/4, 1/3, 2/2, 3/1, and 4/0).

Sample Filtering. The unfiltered sample set includes the samples from Lovell et al. (18) and additional, primarily 8x resequenced samples to generate a sample set of 1,059 samples. We focused our analyses on natural populations, so any samples that originated from breeding populations or cultivars were removed. Any samples that were assigned using principal component analysis (PCA) or ADMIXTURE to a different gene pool from the rest of the samples from the same sampling source were removed. We calculated distance matrices and neighbor-joining (NJ) trees in R (68, 69), and samples assigned to a different clade in an NJ tree than the rest of the samples from the same sampling source were removed. Samples that looked to be from the same or closely related sampling source, but originated from genetically distant locations, were filtered based on the geographic patterns in that clade of the NJ tree. Any samples that did not fulfill the sequencing-coverage cutoffs for nQuire were also removed.

Population Genetic Approaches. We use the adegenet R package to run PCA on our genotypes (34). We used dosage genotypes to minimize biases in the PCA results based on ploidy differences (63), though the results for our analyses were consistent regardless of the use of disomic, tetrasomic, or polyploid genotypes (SI Appendix, Fig. S7).

We also used ADMIXTURE (70) using diploid genotypes to estimate the number of populations and the proportion of ancestry of the samples in the dataset. We ran PCA and ADMIXTURE on the entire dataset and separately for the inferred Midwest gene pools.

Neighbor Net Tree. Ancestral genotypes were estimated following ref. 18. Briefly, coding sequences (CDSs) were identified that share common ancestry with Panicum hallii and Sorghum bicolor. For each orthology network, two multiple sequence alignments were generated in mafft (71), one with and one without the switchgrass sequence, followed by extraction of marginal character states in Phangorn (72) using the maximum-likelihood algorithm. This process generated ancestral state genotypes at 10,874,224 SNPs in CDS regions, of which 100,000 were randomly selected, and a variant call format (VCF) file for the entire sample set and the ancestral genotype was generated for these SNPs. The VCF was converted to the NEXUS format by using vcftolyip (73). We used Splitstree (74) to generate a NeighborNet tree and the phangorn package in R (72) for visualization.

Chloroplast Analyses. Chloroplast assembly for each of the samples was performed by using NOVOpixly (75). The seed sequence used for assembly was identified by using BLAT (76) with Kanlow chloroplast sequence (GenBank accession no. HQ731441.1). The assemblies were aligned by using mafft v7.471 (71).

We used the chloroplast alignment and generated unweighted pair group method with arithmetic mean (UPGMA) and neighbor-joining trees and calculated bootstrap values using the ape package in R (77). For improved interpretation and visualization, we used these trees to select a subset of 218 samples (SI Appendix, Table S3), removing some closely related samples, particularly those with identical or near-identical haplotypes, but making sure to retain samples that clustered closely in the chloroplast trees, but had different ploidy or nuclear ancestry patterns. We then estimated the chloroplast tree using BEAST v2.6.6 (31), using the same parameters as detailed for “model 1” used below for estimating divergence date.

Using the UPGMA and neighbor-joining trees of all samples, we also selected paired 4x and 8x samples that belong to the same clade (bootstrap values > 0.9) that contain only samples from one gene pool or only samples from Southern switchgrass, according to population structure analysis of the nuclear genome. This process generated a set of 13.4x and 13.8x switchgrass samples along with P. hallii var. fl. We estimated the divergence dates between chloroplast haplotypes in this sample subset using BEAST v2.6.6 (31), using 8.35 million years ago (mya) as the prior for the divergence date between switchgrass and P. hallii (18).

We ran BEAST using three sets of parameters to test the robustness of the time estimates. For all tests, we used Site Model:Substitution Rate:estimate; Site Model:Substitution Rate:Gamma Category Count:4; Site Model:Shape:estimate; Site Model:Substitution Rate:Gamma:Alpha=0.001,Beta=1,000. Model 1 generated the smallest 95% intervals for the time estimates. For model 2, we used Site Model:Substitution Rate:Alpha=0.001, Beta=1,000. For all tests, we used Site Model:Substitution Rate:Alpha=0.001, Beta=1,000. Model 1 generated the smallest 95% intervals for the estimated parameters, but all three models generated highly similar results. The final tree contains five main clades with >99% posterior probability, each clade containing both 4x and 8x samples. “Southern” consists of individuals from the Gulf gene pool, though the overall clade also contains 4x samples with Atlantic ancestry. “Midwest A” and “Midwest B” contain samples from Midwest gene-pool clades. “Gulf” and “Atlantic” contain samples from clades with Gulf and Atlantic ancestry, respectively (SI Appendix, Table S4).

K-Mer Approach. To observe the effects of potential biases in calling 8x variation from a 4x reference, we employed a reference-free k-mer (sequences of DNA of length “k”) based approach to analyze the population structure and compare the overall patterns to the results obtained using SNPs. Reference-free and alignment-free methods of assessing population genetic structure have been shown to reduce some of the inherent biases involved in polyploid genetics (78). Here, we used the k-mer hashing method employed by the Mash program to confirm population genetic patterns derived from traditional SNP-based
analyses (79). In this analysis, raw sequence data from whole-genome sequencing are decomposed into k-mers, and then unique k-mers are tabulated for each individual in the dataset. The resulting tables are hashed by using the MinHash algorithm to facilitate computational comparison (80). The Jaccard distance between individual hashes then is a reliable estimate of the genetic distance between samples.

Our analysis followed the method used in VanWallendael and Álvarez (78) on 40 individuals randomly selected from the study. Briefly, we first trimmed fastq sequence files randomly to equal size to reduce library-size biases using fastx-tools (https://github.com/djones/fastx-tools). We then performed k-mer decomposition, sampling 100,000 random k-mers (−1 100,000) of length 21 bp (−k 21) and removing k-mers detected only once to avoid sequencing errors (−m 2). We used the inferred genetic distance matrix to perform Principal Coordinate Analysis and compared the results to SNP-based analyses using Mantel tests.

**Morphological/Ecotype Classifications.** To assign in silico morphological classifications to infer similarity to known switchgrass ecotypes, we utilized an assumption-free approach that reduces possible assignment biases by avoiding user-defined training groups. Specifically, a DAPC (34, 35) approach was run in adegenet (34) on the same set of 16 informative plant traits (leaf: length, width, length/width ratio, area, lamina thickness and lamina/midrib thickness ratio; whole plant: number of tillers, tiller height, product of tiller height × number, tiller height/count ratio, panicle height, panicle height/count ratio, leaf canopy height and tiller/leaf height ratio; phenotype: date of green-up and date of panicle emergence) used to call ecotypes in Lovell et al. (18), which were collected at a northern (KBSM in Hickory Corners, MI) and southern (PKLE in Austin, TX) common garden in 2019. Prior groups were determined by first transforming the phenotypic data using PCA, and then the first 10 principal components (PCs) were used in a k-means algorithm to classify individuals into three possible groupings to match the documented ecotypic variation present in the 4× (18), aiming to maximize the variation between groups. Next, DAPC was implemented on the 10 retained PCs to provide an efficient description of the morphological clusters using two synthetic variables, which are linear combinations of the original phenotypic variables that have the largest between-group variance and the smallest within-group variance (i.e., the discriminant functions). Since this approach is based on Discriminant Analysis, we also obtained posterior membership probabilities for each individual, thus allowing us to assess variation present in both cytotypes.

**Relative Fitness over Climate Distance.** For this analysis, a plant present at 1 of the 10 common-garden sites with a Midwestern genetic background was considered a replicate for either the 8× (n = 912) or 4× (n = 638) cytotype. At four of the common-garden sites (KBSM, FRMI, CLMB, and PKLE), some of the 8× genotypes were planted multiple times, but each of these were considered independent replicates, given that they each provide separate estimates of 8× Midwest fitness. Relative fitness for each plant was estimated by summing collected biomass values over 2 y (2019 and 2020) and dividing it by the maximum summed biomass value for each cytotype, respectively. Climate distance from collection origin to planting location was calculated for each replicate. Specifically, to assess climate distance, we calculated the outer distance between matrices of six environmental predictors from WorldClim (BIO1 = annual mean temperature, BIO2 = mean diurnal range, BIO3 = temperature seasonality, BIO5 = maximum temperature of warmest month, BIO16 = precipitation of wettest quarter, and BIO17 = precipitation of driest quarter). ENMs were then generated within the ENM tools package by using a generalized linear model algorithm. For each model, the occurrence data were coupled with 1,000 pseudoabsence data generated randomly within the modeled study area. Models were trained with 80% of the coupled occurrences and pseudoabsence data and tested with the remaining 20%. Each of the four ENMs was replicated 500 times, which were evaluated via the area under the receiver operating characteristic curve [AUC (82)]. Models with AUC values above 0.75 are considered potentially useful (83), and, in this case, all models had AUC values above this threshold (4× = 0.867, 8× = 0.800, 4× Midwest = 0.876, and 8× Midwest = 0.803). Estimates of niche breadth (82 based on ref. 46) were obtained for each ENM. Lastly, we conducted niche identity tests (n = 500 replications) for both the overall cytotype and Midwest models to determine, in each instance, if the two groups’ occurrences in the environment are random draws from a shared underlying distribution (45). Departure from the null hypothesis in this test could potentially indicate the occurrence of niche evolution among other possibilities (84). These analyses focused on the Midwestern subpopulation due to low sample size of 8× Gulf and Atlantic.

**Southern Switchgrass Ancestry Analyses.** We chose three training sets, one from each gene pool, composed of 40 4× individuals that were at the most extreme ends of the PCA distributions, had high ancestry to one of the subpopulations in the k = 3 ADMIXTURE analysis, and then filtered to include as many populations and as much geographic variation as possible. We then used vcftools (85) to identify SNPs with Fst > 0.5 between pairwise comparisons of gene pools, followed by slight LD pruning with plink (86) (LD < 0.99 in 1,000-bp windows). We filtered the Midwest-vs.-Gulf and Midwest-vs.-Atlantic high-Fst SNPs by missing data in all the Midwest samples (<20% missing data), resulting in 58,829 high-Fst SNPs for Midwest-vs.-Gulf and 59,119 high-Fst SNPs for Midwest-vs.-Atlantic. For each high-Fst SNP, one or both alleles can be informative about ancestry, and we called an allele as informative about ancestry if it was absent from one training set or its frequency in one training set was above 0.5 and 2.5× greater than the frequency in the other training set. An allele that is not ancestry-informative is present in both training sets at high enough frequency, so that it cannot be used to infer from which population it originated. Using these criteria, the Midwest-vs.-Gulf comparison has 20,273 (34.5%) SNPs, where both alleles are ancestry-informative, and 17,933 SNPs (30.5%) and 20,612 (35.0%) SNPs, where one allele is Midwest-informative or Gulf-informative, respectively, and the other allele is not ancestry-informative. For the Midwest-vs.-Atlantic comparison, 20,525 SNPs (34.7%) have both alleles as informative about ancestry, and we called them an allele as informative about ancestry if it was absent from one training set or its frequency in one training set was above 0.5 and 2.5× greater than the frequency in the other training set. An allele that is not ancestry-informative is present in both training sets at high enough frequency, so that it cannot be used to infer from which population it originated. Using these criteria, the Midwest-vs.-Atlantic comparison has 20,525 SNPs (34.7%) where both alleles are informative about ancestry, and 21,669 SNPs (36.7%) and 16,918 SNPs (28.6%) have Midwest-informative and Gulf-informative, respectively, for one allele, and the other allele is not ancestry-informative. We used dosage of Atlantic-informative and Gulf-informative alleles at each SNP for RDA and to estimate the amount of Southern switchgrass ancestry in each sample.

We estimated the shift in allele frequency (AFS) in a Midwest test population from the Midwest training set toward either the Gulf or Atlantic training sets by calculating the following: "test_range″ = allele frequency difference between the Atlantic or Gulf and the Midwest training sets; "test_range″ = allele frequency difference between test population and MW training set. Multiply any SNPs with a negative value in "test_range" by −1, so that a positive "test_range" value means a shift toward the allele frequency in the Atlantic/Gulf training set. "AFS score" = test_range/test_range and represents the percentage of the allele-frequency difference between the training sets that the population has shifted toward the Atlantic/Gulf frequency. AFS score = 1 means allele frequency is the same as Atlantic/Gulf training set; AFS score = 0 means allele frequency is the same as the Midwest training set. We used a cutoff of three SDs from the mean to identify candidate SNPs based on elevated AFS score.

**RDAs.** We implemented RDA in the R package vegan (87) to: 1) calculate the proportion of variation in the lowland ancestry SNPs that can be uniquely explained by climate, geography, and their joint effect (88); and 2) identify specific SNPs strongly associated with the environment (89, 90). To partition explainable variance attributable to climate and geography, we ran three models: 1) one full model with climate (BIO1 = annual mean temperature, BIO2 = average annual temperature, BIO3 = temperature seasonality, BIO5 = maximum temperature of warmest month, BIO16 = precipitation of wettest quarter, and BIO17 = precipitation of driest quarter). ENMs were then generated within the ENM tools package by using a generalized linear model algorithm. For each model, the occurrence data were coupled with 1,000 pseudoabsence data generated randomly within the modeled study area. Models were trained with 80% of the coupled occurrences and pseudoabsence data and tested with the remaining 20%. Each of the four ENMs was replicated 500 times, which were evaluated via the area under the receiver operating characteristic curve [AUC (82)]. Models with AUC values above 0.75 are considered potentially useful (83), and, in this case, all models had AUC values above this threshold (4× = 0.867, 8× = 0.800, 4× Midwest = 0.876, and 8× Midwest = 0.803). Estimates of niche breadth (82 based on ref. 46) were obtained for each ENM. Lastly, we conducted niche identity tests (n = 500 replications) for both the overall cytotype and Midwest models to determine, in each instance, if the two groups’ occurrences in the environment are random draws from a shared underlying distribution (45). Departure from the null hypothesis in this test could potentially indicate the occurrence of niche evolution among other possibilities (84). These analyses focused on the Midwestern subpopulation due to low sample size of 8× Gulf and Atlantic.
LD Analysis. For LD analysis, \( r^2 \) was calculated in PLINK v.1.9 (93) and VCFtools (85) between hiFst SNPs across all samples and within each of the three gene pools using diploid-only genotypes. LD heatmaps were visualized by using the LDheatmap R package (94).

Data Availability. Raw sequencing data for switchgrass included in our analyses are available at the NIH NCBI Sequence Read Archive (accession nos. all included in the supplemental information). Useful SNP sets and generalized versions of the code used in our analyses are publicly available at GitHub (https://github.com/jpnapiere/A-generalist-specialist-tradeoff-between-switchgrass-cytotypes-impacts-climate-adaptation-and-geogra) (95).

Acknowledgments. We thank J. Randall (North Carolina Botanical Garden), M. Casler (US Department of Agriculture-Agricultural Research Service (USDA-ARS)), A. Stottlemeyer (The Ohio State University Newark), and M. Harrison (Germplasm Resources Information Network) for generously sharing seed collections of switchgrass. We thank the HudsonAlpha Genomic Services Lab for loading Illumina X10 sequencing runs and the Joint Genome Institute (JGI) production group for sequencing library ASHOA. We are grateful for JGI sequencing capacity contributed to the community sequencing effort by the Bioenergy Research Centers through the Bioenergy Science Center led by the Oak Ridge National Laboratory. We thank Lisa Vormwald, Matthew Smith, Perla Dubeyen, John Sanley, Nick Ryan, Todd Bortnim, Tim Vugteveen, Scott Hoffman, Matt Donahue, and Justin Shih for propagating, planting, and managing switchgrass field plantings. This research was supported by the US Department of Energy, Office of Biological and Environmental Research, Genomic Science Program Grants DE-SC0014156 and DE-SC0021126 (to T.E.J.) and DE-SC0017883 (to D.B.L.). The work (Proposal 10.46936/10.25585/60001049) conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231. This material is based upon work supported in part by the Great Lakes Bioenergy Research Center, US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Contract DE-AC02-06CH11357. The field research was supported by the USDA-ARS, an equal opportunity employer.

Author affiliations: 1Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712; 2Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806; 3Department of Plant Biology, Michigan State University, East Lansing, MI 48824; 4Department of Agriculture–Natural Resources Conservation Service, Kingsville, TX 78363; 5Grassland, Soil, and Water Research Laboratory, US Department of Agriculture–Agricultural Research Service, Temple, TX 76502; 6Division of Plant Science and Technology, University of Missouri, Columbia, MO 65211; 7Environmental Science Division, Argonne National Laboratory, Lemont, IL 60439; 8Kika de la Garza Plant Materials Center, US Department of Agriculture–Natural Resources Conservation Service, Kingsville, TX 78363; 9Wheat, Sorghum, and Forage Research Unit, US Department of Agriculture–Agricultural Research Service, Lincoln, NE 68583; 10Texas A&M AgriLife Research and Extension Center, Texas A&M University, Overton, TX 75684; 11Department of Plant and Soil Sciences, Oklahoma State University, Stillwater, OK 74078; and 12Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.
