

RESEARCH ARTICLE

Multi-Phase US Spread and Habitat Switching of a Post-Columbian Invasive, *Sorghum halepense*

U. Uzay Sezen¹, Jacob N. Barney², Daniel Z. Atwater², Gary A. Pederson³, Jeffrey F. Pederson⁴, J. Mike Chandler⁵, T. Stan Cox⁶, Sheila Cox⁶, Peter Dotray⁷, David Kopec⁸, Steven E. Smith⁸, Jill Schroeder⁹, Steven D. Wright¹⁰, Yuannian Jiao¹¹, Wenqian Kong¹, Valorie Goff¹, Susan Auckland¹, Lisa K. Rainville¹, Gary J. Pierce¹, Cornelia Lemke¹, Rosana Compton¹, Christine Phillips¹, Alexandra Kerr¹, Matthew Mettler¹, Andrew H. Paterson^{1*}

1 Plant Genome Mapping Lab., Athens, GA, United States of America, **2** Virginia Tech, Blacksburg, VA, United States of America, **3** USDA-ARS, PGRCU, Griffin, GA, United States of America, **4** USDA-ARS and University of Nebraska-Lincoln, Lincoln, NE, United States of America, **5** Texas A&M University, College Station, TX, United States of America, **6** The Land Institute, Salina, KS, United States of America, **7** Texas Tech University, Lubbock, TX, United States of America, **8** University of Arizona, Tucson, AZ, United States of America, **9** New Mexico State University, Las Cruces, NM, United States of America, **10** University of California Davis, Davis CA, United States of America, **11** State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, China Academy of Sciences, Beijing, China

* paterson@uga.edu (706) 5830162



OPEN ACCESS

Citation: Sezen UU, Barney JN, Atwater DZ, Pederson GA, Pederson JF, Chandler JM, et al. (2016) Multi-Phase US Spread and Habitat Switching of a Post-Columbian Invasive, *Sorghum halepense*. PLoS ONE 11(10): e0164584. doi:10.1371/journal.pone.0164584

Editor: Peter Prentis, Queensland University of Technology, AUSTRALIA

Received: June 15, 2016

Accepted: September 27, 2016

Published: October 18, 2016

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Funding was provided by the United States Department of Agriculture Biotechnology Risk Assessment Program (08-39211-19568 to AHP, 2012-33522-19790 to AHP and TSC), and the United States Department of Agriculture Controlling Weedy and Invasive Plants program (2013-67013-21306 to JNB and AHP).

Abstract

Johnsongrass (*Sorghum halepense*) is a striking example of a post-Columbian founder event. This natural experiment within ecological time-scales provides a unique opportunity for understanding patterns of continent-wide genetic diversity following range expansion. Microsatellite markers were used for population genetic analyses including leaf-optimized Neighbor-Joining tree, pairwise F_{ST}, mismatch analysis, principle coordinate analysis, Tajima's D, Fu's F and Bayesian clusterings of population structure. Evidence indicates two geographically distant introductions of divergent genotypes, which spread across much of the US in <200 years. Based on geophylogeny, gene flow patterns can be inferred to have involved five phases. Centers of genetic diversity have shifted from two introduction sites separated by ~2000 miles toward the middle of the range, consistent with admixture between genotypes from the respective introductions. Genotyping provides evidence for a 'habitat switch' from agricultural to non-agricultural systems and may contribute to both Johnsongrass ubiquity and aggressiveness. Despite lower and more structured diversity at the invasion front, Johnsongrass continues to advance northward into cooler and drier habitats. Association genetic approaches may permit identification of alleles contributing to the habitat switch or other traits important to weed/invasive management and/or crop improvement.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

From noble beginnings as a promising forage, Johnsongrass (*Sorghum halepense*) has become one of the most noxious agricultural weeds globally [1], costing US farmers tens of millions of dollars annually in management costs and yield losses [2]. In fact, the first US federal appropriation for weed control research targeted Johnsongrass (HB#121, 1900), yet its spread remains unchecked.

Johnsongrass ($2n = 40$) formed by natural hybridization between *S. bicolor*, (sorghum, $2n = 20$) an annual native of Africa, and *S. propinquum* ($2n = 20$), a wild perennial native to wet subtropical habitats in southeast Asia that diverged from *S. bicolor* ~1–2 million years ago [3]. Johnsongrass is used as a forage crop in many countries, and for food (seed/flour) in some. However, it more commonly occurs as a weed, having spread from its west Asian center of diversity across much of Asia, Africa, Europe, North and South America, and Australia [1]. Its post-Columbian establishment in the US is likely typical of its spread on other continents, being introduced as a prospective forage and/or unintentionally as a contaminant of sorghum seedlots [4]. The economic impact of Johnsongrass is further increased by its frequent crossing with crop-sorghum (*S. bicolor*) [5], virtually precluding commercial use of transgenic sorghum [6] that has high potential to reduce crop losses to insects, diseases, weeds, and abiotic stresses improving food security particularly in low-rainfall agro-ecosystems.

Unlike most other agricultural weeds, Johnsongrass is also an aggressive invader of natural or minimally managed habitats. Johnsongrass is a state-listed noxious weed in 20 states and an invasive species in 16 states [7]. This tremendous success can be partially attributed to the ‘invasive syndrome’ of Johnsongrass including robust spreading rhizomes, shattering inflorescences, rapid growth rates, high seed dormancy, large and extensive annual seed production, impressive disturbance tolerance, potential allelochemicals, and associations with nitrogen-fixing bacterial endophytes [8–11]. Additionally, Johnsongrass has a very large climate niche, with favorable growing conditions occurring across all non-Antarctic continents [12]. In many invasive plants the size of invasive ranges is positively correlated with native range size [13], and Johnsongrass is no exception.

A compelling question is whether the spread of Johnsongrass is because it is a remarkably effective ‘general purpose genotype’ [14] or reflects rapid adaptation to new conditions that might also favor alleles of value for transfer to sorghum. Johnsongrass has now expanded into non-agricultural habitats from its initial populations that have become established in agricultural fields displaying genotypically and phenotypically distinct responses to these new environments [15]. Under replicated common garden conditions these non-agricultural accessions manifest fitness traits that are harbinger of climatic niche expansion for this most threatening invasive plant. Other examples exist of rapid evolution of exotic species following introduction to novel conditions [16, 17]; but rarely do exotic species become so successful across such broad geographies and habitats as Johnsongrass has in the US. Interestingly, as Prentis, Wilson [18] outline, the sources of the genetic variation underlying these evolutionary changes remain unknown in most invasive species.

Rapidly spreading populations contain demographic signatures due to serial founder effects. Global and large continent-wide genetic data from human and *Arabidopsis* populations have demonstrated reduced genetic diversity and increased linkage disequilibrium patterns as populations migrate away from their centers of origin. Range expansions from a limited genetic stock have significant levels of background relatedness, but have rarely been explored in invasive plants, which are modern rapid globetrotters [19].

As a foundation for investigating principles underlying the post-Colombian spread of Johnsongrass across the USA, we study Simple Sequence Repeat (SSR)-based genotypes to elucidate

colonization and diversity patterns of geo-referenced populations sampling most of its continental United States range. This natural experiment provides a unique opportunity to track 'blank slate' colonization occurring within ecological time-scales, and provides a new perspective by broadening the timescale of invasion occurring over two centuries. A blank slate colonization may follow three trajectories: Founders may go extinct, maintain a restricted local population, or spread to all available habitats. These trajectories are shaped by the resultant vector of major co-acting evolutionary forces such as genetic drift, gene flow and natural selection. As a tetraploid with genomes inherited from two grasses respectively adapted to dry and tropical moist conditions, Johnsongrass is an outstanding experimental system in which to investigate aggressive invasion conforming to the third population trajectory.

DNA genotyping also provides a means to investigate whether patterns of genetic variation are related to the habitat or climate from which samples were collected. Findings complement and enhance historical information about the origins and spread of Johnsongrass, in particular indicating two independent invasions in very different climates separated by about 2000 miles. Finally, DNA fingerprint data provides the basis for selection and early characterization of a 'diversity panel' of US Johnsongrass, potentially suitable for use by association genetic approaches [20] to identify specific genes or small genomic regions contributing to spread of Johnsongrass and/or conferring adaptations of potential value for improvement of cultivated sorghum. These approaches are generalizable to investigating other weedy and/or invasive populations.

Methods

A total of 599 genotypes were studied, with 1–32 from each of 70 different collection sites in 12 states (Figure A in [S1 Fig](#), [S2 Table](#)). At most sites, seeds were sampled from plants at least 10 m apart to avoid clonal propagules [5]. Plants were grown in Watkinsville GA in 2012–13 to obtain tissue for DNA extraction [21]. Individuals were genotyped using 19 microsatellite markers designed to sit on arms of each of the 10 chromosomes [22] ([S1 Table](#)). PCR products were scored on 10% silver stained PAGE. Due to the tetraploidy 97 loci generated were treated as biallelic. The dissimilarity matrix and 100 bootstrapped N-J tree ([S1 File](#)) using one of the progenitor species *S. propinquum* as the out group were calculated by PowerMarker v3.25 [23]. N-J Tree was drawn by MEGA v4 [24]. Ordering of the leaves of the N-J tree were optimized based on the geographical context and a significance test with 1000 permutations was carried out using the software GenGIS [25]. The digital elevation model of the continental United States was downloaded from Oak Ridge National Labs' Shuttle Radar Topography Mission (SRTM) records. A chi-squared contingency test of habitat by phase was carried out using R version 3.1.1 [26]. To test for sampling bias, two multinomial logistic regressions of habitat type against 1) annual mean temperature and annual precipitation; and 2) latitude and longitude were made. P values were estimated by Z-tests on the regression coefficients for "roadside" and "disturbed" with "agricultural" as the reference habitat. Principal Coordinate Analysis (PCoA) was done by DARWIN based on the calculated dissimilarity matrix [27]. Population genetic analysis (mismatch distribution, Tajima's D , Fu's F , pairwise F_{st}) were carried out by Arlequin 3.5 [28]. Tests for neutrality and linkage disequilibrium were performed after 10000 simulations using the program POPGENE[29]. Population genetic structure was analyzed by two programs employing Bayesian statistics STRUCTURE v2.3.3 [30] and BAPS6 [31]. A range of inferred cluster (K) values from 1 to 30 was explored using 15 replications with 80000 burn in and 800000 simulation steps. Results were plotted using STRUCTURE HARVESTER web server to visualize highest deltaK values corresponding to real population partitioning [32]. Output from STRUCTURE HARVESTER (indfile, popfile) were permuted in CLUMPP

[33] and turned into a graphic using DISTRUCT [34]. Spatial structure analyses based on Voronoi tessellation were done by BAPS6 using individual and admixture models with the same range of inferred cluster values.

Results

Neutrality test results after 10000 simulations have shown that most marker loci employed in the study are neutral. Only 7 out of 97 loci are found outside the 95% confidence intervals (Figure F in [S1 Fig](#)). Test for Linkage Disequilibrium for non-random association between pairs of loci in finite subdivided populations revealed a small total variance value (0.0847) indicating large human-assisted migration rates on a continental scale (among states) reducing LD ([S3 Table](#), [S4 Table](#)).

Principal Coordinate Analysis (PCoA) funneled the 599 genotypes into 231 groupings (Figure A in [S1 Fig](#), Figure B in [S1 Fig](#)). Some of these groupings consisted of individuals sampled from many locations and were not geographically distinct even after they were partitioned according to the 12 states of origin (Figure C in [S1 Fig](#)), presumably reflecting very high levels of intra-population diversity.

A Neighbor-Joining (N-J) tree with optimized leaf ordering visualized phylogeographic relations along the east–west axis ([S1 File](#)). The tree has its deepest branches in South Carolina (SC), providing genetic evidence for the US center of origin. The N-J tree clusters a portion of samples from southeastern states such as Georgia (GA) and Alabama (AL) with SC and extends into Texas (TX) within the first few basal nodes ([Fig 1](#) Phase 1). Colonization from SC continued northward to at least Virginia (VA) and westward to at least New Mexico (NM) ([Fig 1](#) Phase 3).

The N-J tree provides evidence for a second introduction, in Arizona (AZ). The founder cluster from AZ shows a very high correlation with the geographical axis. There are at least two distinct episodes of colonization from AZ into TX and then reaching to Florida (FL) and GA. ([Fig 1](#) Phase 2, Phase 4).

The remainder of the N-J tree reflects admixture of genotypes following spread from AZ and SC, respectively. On this tree genotypes form distinct clusters for California (CA), TX, Nebraska (NE) and Kansas (KS) ([Fig 1](#) Phases 5abc). The CA-dominated branch of the tree harbors genotypes from GA, FL, NE, Kentucky (KY) and VA on its leaves suggesting eastward gene flow from CA (Phase 5a). This pattern repeats in another clade of similar branch length (thus perhaps similar age) harboring genotypes from CA, AZ, VA, TX and GA (Phase 5b). Curious intercalation of AZ and VA genotypes may reflect long-distance gene flow between the two states after trans-continental railroad connection and eastward grain or livestock transportation. Concurrent with the rest of Phase 5, TX, KS and NE form their individual clades with no significant gene flow among themselves but occasionally flow into AL, GA and FL (Phase 5c).

Genotypes from FL, KY and VA appear at the ‘tips’ of branches indicating the most recent colonization events. As a peninsula FL has limited connectivity for gene flow. Besides, more than half of it is covered by swamps and provides fewer opportunities for population expansion with quite different agricultural practices. Indeed, FL appears to be a genetic sink rather than a source, being re-stocked at least six more times after the initial colonization ([Fig 1](#) Phase 4, Phase 5a, Phase 5c).

Strikingly, 80% of the early colonist genotypes (Phases 1–3) were found in agricultural habitats. Genotypes from disturbed habitats were mostly found in the more recent Phase 4 and 5 clusters, and genotypes from roadside habitats were largely restricted to the Phase 5 clusters ([Table 1](#)). In these most distal clusters the agricultural habitat was underrepresented,

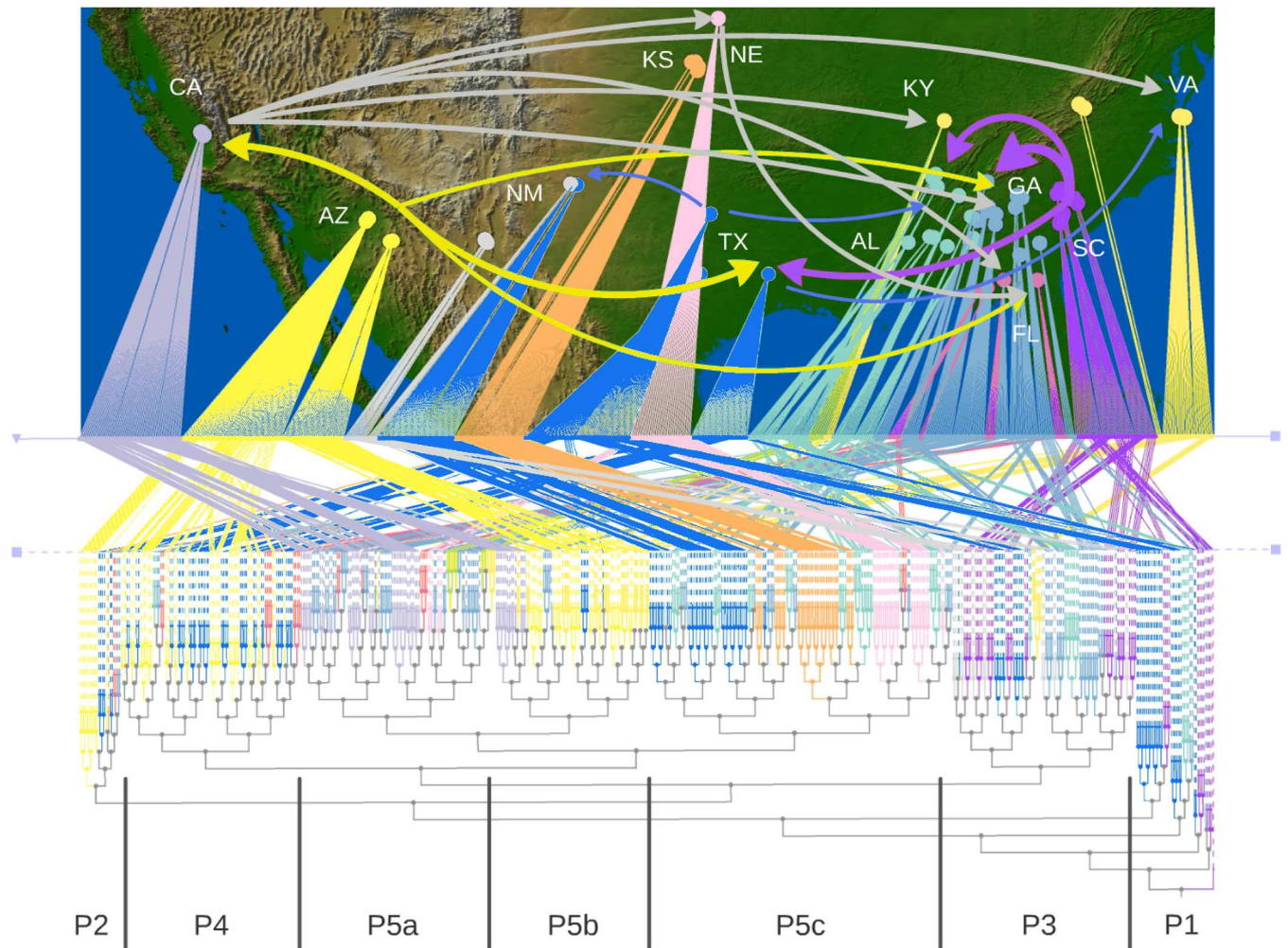


Fig 1. Map of sampling sites and N-J tree with an optimized leaf ordering along east-west geographical axis for *Sorghum halepense* genotypes. One of the progenitor species, *S. propinquum*, is used as outgroup. Colonization is outlined in 5 phases (P1-P5). Major gene flow pathways are shown using arrows in four colors (red, green, blue, gray). Initial colonization from southeastern US starting from SC (red arrows) are followed by the second introduction from AZ (S2, S4 green arrows). Gene flow from TX into NM, GA and VA (P3, blue arrows) happens concurrent with local gene flow among GA-AL-SC (P3 arrows not shown for clarity). From CA, there is a massive eastward radiation into NE, KY, FL and VA (P5a, gray arrows). While there is no detectable gene flow among KS, NE and TX (P5c) there is southbound gene flow from NE to FL (S5c, gray arrow) and KS into AL and GA (P5c, arrows not shown for clarity).

doi:10.1371/journal.pone.0164584.g001

Table 1. Contingency table showing observed frequency of *Sorghum halepense* accessions belonging to each genetic cluster, found in each habitat type ($X^2 = 121.79$; $df = 12$; $P < 0.0001$).

	Agricultural	Disturbed	Roadside
Phase 1	22	1	3
Phase 2	18	0	0
Phase 3	14	4	6
Phase 4	47	17	1
Phase 5a	24	10	19
Phase 5b	13	21	32
Phase 5c	0	12	26

doi:10.1371/journal.pone.0164584.t001

suggesting that genetic differentiation in Johnsongrass accompanied a habitat switch away from agricultural habitats and toward disturbed and roadside habitats. Reintroductions of Johnsongrass to the Southeastern states reflect these changes in habitat; later colonizers were found primarily in disturbed and roadside habitats, while the more ancestral genotypes were found in agricultural habitats.

These results suggest that initial expansion of Johnsongrass occurred primarily in agricultural habitats, with a subsequent transition of a cluster of closely related genotypes to non-agricultural habitat throughout the entire US range. This pattern was not influenced by sampling bias across geographic clines. Multinomial logistic regressions show that the three habitat types were equally represented along latitudinal, longitudinal, and rainfall clines ($P > 0.213$; multinomial logistic regression, Z-test), although agricultural habitat was less common in areas with high mean annual temperatures ($P < 0.031$). However, such sampling bias would have countered the trends we observed, as it would have caused us to be more, and not less, likely to sample non-agricultural habitat in the warmer ancestral populations.

Mismatch analysis [35], shows populations from all 12 states to exhibit demographic and spatial expansion with unimodal distributions (Figure C in S1 Fig, Figure D in S1 Fig). KY and FL showed a slowdown in expansion, with Harpending's raggedness index two orders of magnitudes larger than in GA and TX (Table 2). This slowdown may be explained in FL by geography and in KY may be an artifact of low sample number. Statistical tests of selective neutrality also suggested population expansion. Tajima's D values were all positive indicating a past bottleneck but p-values were not significant (Table 1). Fu's Fs, a more sensitive measure of population fluctuations, showed large negative values supported with significant p-values indicating a history of population expansion.

Pairwise comparisons of allelic diversity within populations identify GA and TX as the states now harboring the greatest Johnsongrass diversity. SC, KS and CA harbored the least within-population diversity (Fig 2A). A pairwise Fst matrix detected patches of low population structure centered around GA and TX, with increases toward the edges of the continent-wide colonization. The most distinct state is VA, showing the highest Fst values in almost every pairwise comparison, the exceptions being with GA and TX. The pattern is mirrored at the other end of the continent where NE, CA and AZ show high Fst values when compared with VA and KY. High Fst's drop when VA and KY are compared with GA and TX. On the other hand, TX/GA, AZ/NM and SC/AL have the lowest Fst values. These findings indicate TX/GA to represent the current center of diversity, but SC/AL and AZ/NM to have been centers of origins whose diversity are relatively lower (Fig 2B).

Two Bayesian analysis programs, STRUCTURE [30] and BAPS6 [31], were used to investigate population structure and probe composition of ancestral subpopulation clusters. These

Table 2. Summary of sample sizes, number of polymorphic sites, Harpending's raggedness index, Theta pi measure, Tajima's D, Fu's Fs values of *Sorghum halepense* populations according to the sampling locations (by state). The two progenitor genotypes *S. propinquum*, *S. bicolor* and the laboratory standard *S. halepense* are grouped as PBH.

	PBH	VA	KY	SC	GA	FL	AL	TX	NE	KS	NM	AZ	CA	mean	s.d.
Sample Size	3	29	12	42	85	22	54	127	32	36	17	86	54	46.08	35.15
polymorphic site #	47	89	73	85	96	84	88	95	83	84	75	87	83	82.2	12.41
Raggedness	na	0.01	0.017	0	0.0009	0.0079	0.0029	0.0005	0.0055	0.0043	0.007	0.0014	0.001	na	na
Theta pi	31.3	32.9	30.12	30	35.99	31.48	31.33	36.86	30.69	29.36	30	30.58	28.81	31.49	2.42
Tajima's D	0	1.72	1.14	1.87	2.94	1.47	2.18	3.54	1.84	1.66	1.48	2.54	2.03	1.88	0.86
p-value	0.32	0.93	0.88	0.95	0.98	0.92	0.96	0.99	0.95	0.93	0.93	0.97	0.95	0.9	0.17
Fu's Fs	2.32	-9.8	-1.8	-20.6	-24.08	-6.07	-24.15	-23.93	-12.41	-15.9	-3.85	-24.02	-24.13	-14.49	9.79
p-values	0.55	0	0.12	0	0.0003	0.014	0.0001	0.0034	0.0009	0.0004	0.038	0.0004	0	0.056	0.151

doi:10.1371/journal.pone.0164584.t002

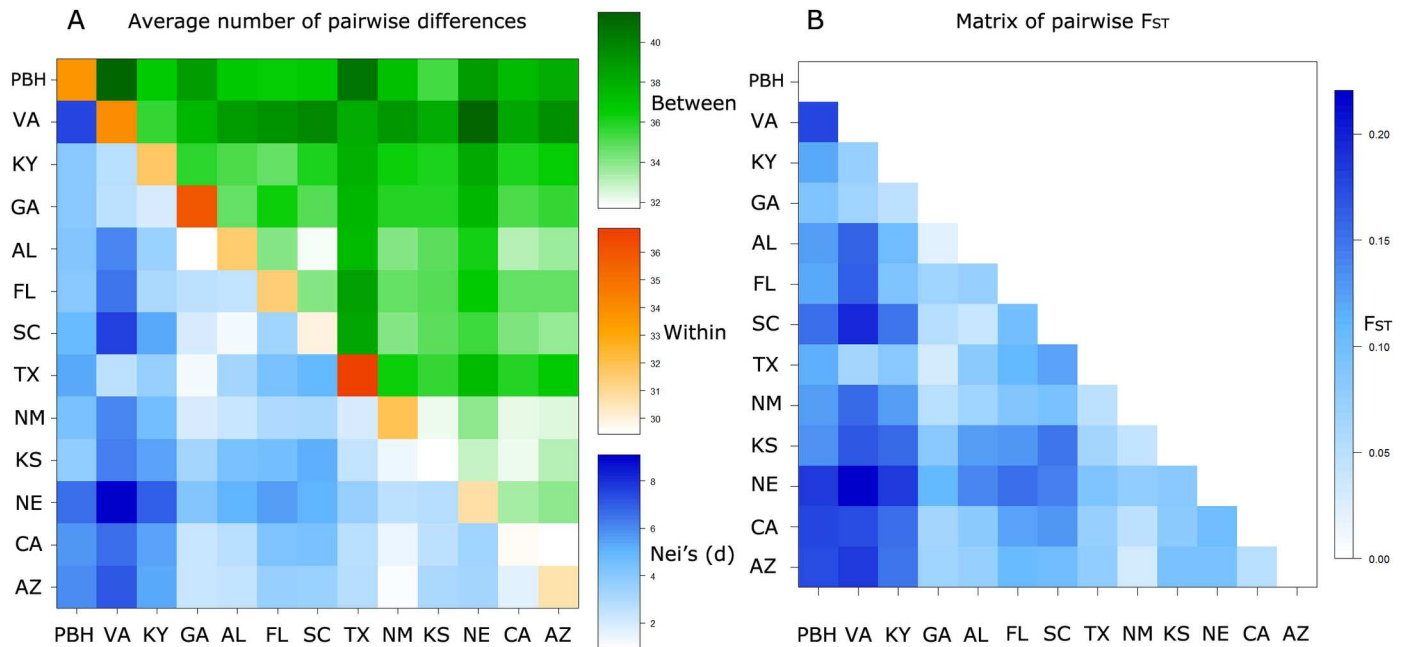


Fig 2. (a) Pairwise comparisons of Nei's distances (net and raw distances) among (lower/upper diagonal) and within (along diagonal) *Sorghum halepense* populations. **(b)** Pairwise comparisons of F_{ST} among populations. Populations diverge as they get farther away from GA and TX. The two progenitor genotypes *S. propinquum*, *S. bicolor* and the laboratory standard *S. halepense* are grouped as PBH.

doi:10.1371/journal.pone.0164584.g002

programs assume that there are K clusters and the true value of K is estimated by a continuous sweep of replicated runs based on multilocus genotype data. Using the Evanno method, STRUCTURE detected between 15 and 26 population clusters [32] (Fig 3; Fig 4). Population cluster numbers remained bound within those values at very high K ($K = 80$, Figure E in S1 Fig). Further permutation of results using CLUMPP grouped most ancestral population blocks around southeastern states except one block from Arizona which represents a second separate introduction confirming the pattern painted by the N-J tree (Fig 1) [33]. Spatial analysis using BAPS6 also detected this ancestral population block starting from $K = 5$, predicting a maximum of 15 distinct populations at present (Fig 4).

Discussion

Historical demography of a population is imprinted on present day genetic structure, and genetic trees can be quite informative to understand the past. Researchers have successfully reconstructed pre-Columbian dispersal of sweet potato (*Ipomea batatas*) into Oceania [36]. Similarly, prehistoric origins of cultivation and human-assisted dispersal of coconut palm (*Cocos nucifera*) around the World has been reconstructed [37]. Here in this study, a N-J tree of Johnsongrass genotypes with optimized leaf ordering demonstrates a correlation with geography. The most basal branches in this tree originate from SC and spread into GA, AL and TX (Fig 1, Phase 1). According to a historical account [38], John Means of SC introduced Johnsongrass in contaminated hemp seed from Egypt shortly after the Revolutionary War. His daughter married the eponymous Colonel Johnson and moved to Alabama, where Col. Johnson wrote an 1847 letter resulting in the common name Johnsongrass [39]. The Arizona Gazette reports farmers complaints about Johnsongrass as early as 1890 [38] in concordance with a second introduction suggested by the N-J tree (Fig 1, Phase 2).

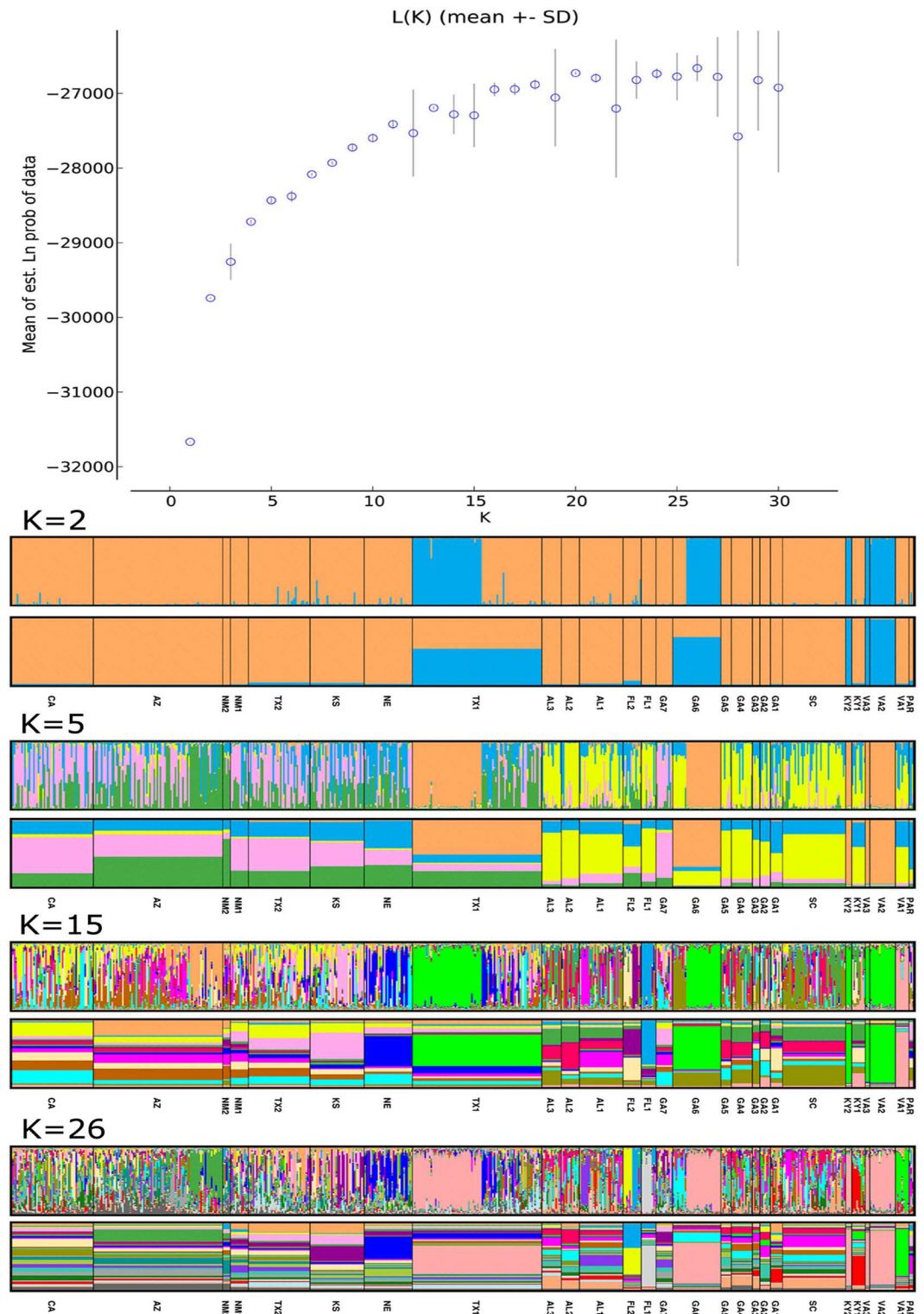


Fig 3. Saturation plot of *Sorghum halepense* genotypes after STRUCTURE runs based on Evanno Method (top). K values reach an asymptote between 15 and 26. DISTRUCT bar graph visualization of results after permuted by CLUMPP at four different K cluster assumptions (K = 2, 5, 15, 26). Clustering based on population averages (lower bars) and individual genotypes (upper bars). The two parental species *Sorghum bicolor* and *Sorghum propinquum* are labeled as PAR.

doi:10.1371/journal.pone.0164584.g003

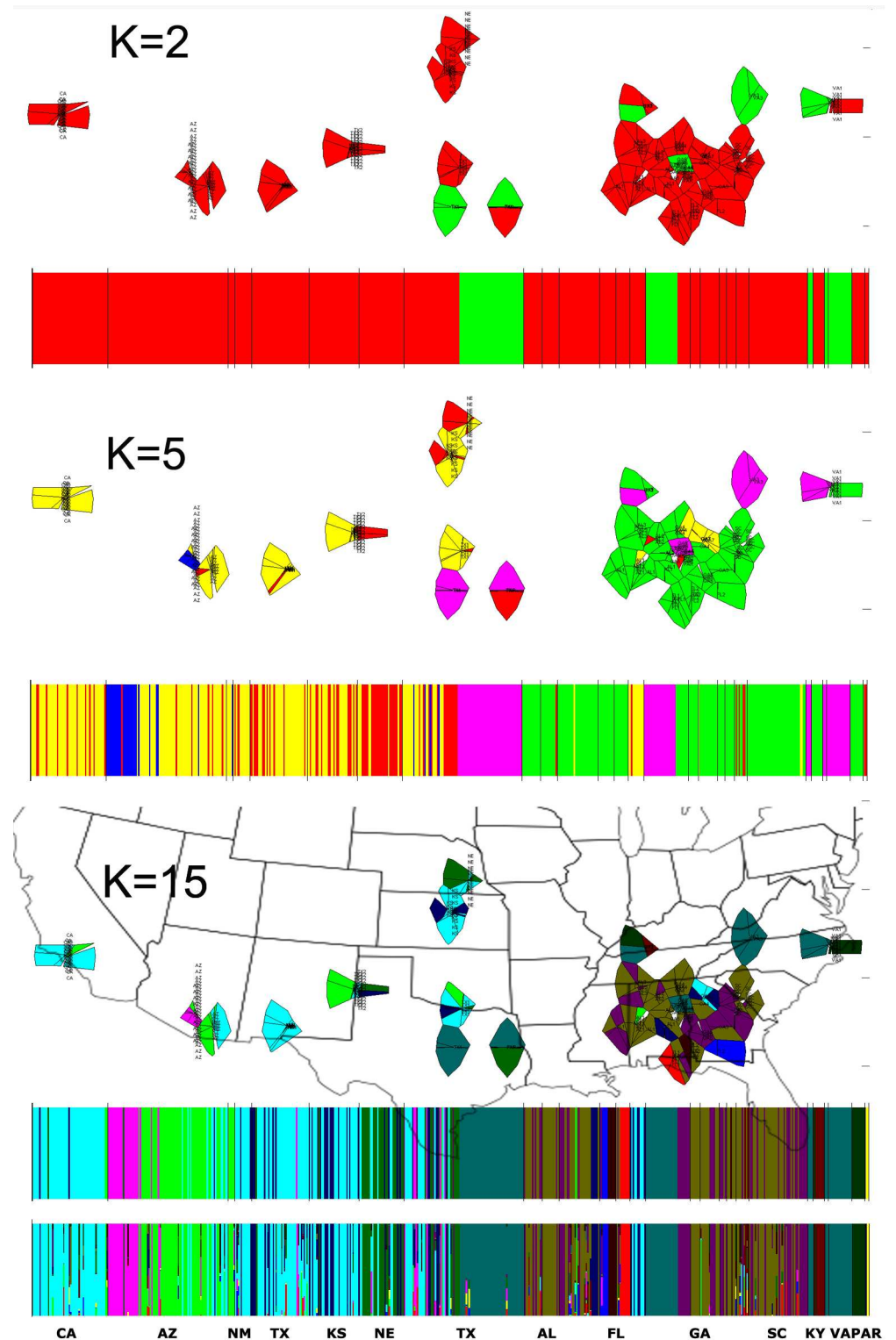


Fig 4. BAPS6 geographical clustering using Voronoi tessellations and bar representations of genotypes at three K values based on population averages (K = 2, 5, 15). State boundaries overlaid for K = 15 for visual guidance. Clustering based on individual genotypes (lower bars) is included for K = 15. The two parental species *Sorghum bicolor* and *Sorghum propinquum* are labeled as PAR.

doi:10.1371/journal.pone.0164584.g004

Johnsongrass is a striking example of continent-wide colonization following a post-Columbian founder event. This natural experiment in a largely self-pollinating perennial grass provides an opportunity to study population genetic dynamics operating at an ecological time-scale. Population genetic theory predicts that founders have limited genetic diversity with the potential for allele frequencies to drift erratically compared to their large panmictic source population [40]. Theory also predicts that initial sites of colonization should harbor the largest genetic diversity, with less diversity and more genetic structure at the margins. These predictions are based on the assumption of a single source for colonization. Here, multiple introductions from two ends of a geographical range appear to generate richer diversity in the middle ground than at the initial sites of colonization. At the same time, margins can still have high genetic structure in accordance with classic predictions. A review of 80 exotic animals, plants, and fungi found that multiple introductions and long post-introduction residence time are associated with increased gene flow and increased diversity, which is often lost in exotic species [41].

Pairwise comparisons of F_{st} and Nei's distances of AZ and SC show reduced within-population and increased among-population variation (Fig 2). AZ poses exceptionally dry conditions that may impose selection for drought responsive traits, the *S. bicolor* progenitor being noted for drought tolerance although *S. propinquum* is native to wet habitats. Reduced within-population genetic diversity observed in inferred centers of introduction could be attributed to both founder effects and possible selection for characteristics optimal for conditions as different as dry AZ and humid SC. The observed increase in genetic diversity in TX and GA might be a result of admixture between divergent genotypes from the two different introductions.

Although useful to describe the present state of populations [42] an ordination method (PCoA) was insufficient to resolve a founding Johnsongrass population, most probably because multiple points of introduction create a diffusion pattern different from classic single source population expansions. Among 231 groupings (Figure A in S1 Fig), many were geographically indistinct, perhaps reflecting the remarkably high level of DNA polymorphism in Johnsongrass. Allelic richness of 182 genetically-mapped RFLP loci averaged 6.13 alleles per locus in 13 Johnsongrass and 5 *S. alnum* [backcross derivatives of natural crosses between Johnsongrass and *S. bicolor*: [43]] accessions from the USA (8), Australia (2), Algeria, Argentina, Chile, India, Kazakhstan, New Zealand, South Africa, and the former USSR; versus 3.39 for a worldwide sample of 55 landrace and wild sorghum accessions; and 1.9 for 16 F1 hybrid sorghums from eight US commercial breeding programs [5]. However, PCoA provided a convenient first pass to choose a reduced sample of genotypes for further characterization as a 'diversity panel' of US *S. halepense*, potentially suitable for use by forward genetic approaches to identify alleles responsible for specific adaptations [20] (Figure A in S1 Fig, Figure B in S1 Fig).

The gradient observed in pairwise F_{st} comparisons of CA with other states is one of the most visually distinct features on the matrix, where F_{st} reaches its highest value in VA (Fig 2B) and shows a steady decline to a low at GA. In both pairwise comparisons VA stands out as the most divergent population, demonstrating peak values in comparisons with NE, SC and AZ. Range expansion leads to a reduction in genetic diversity with increasing distance from the origins of the expansion and we observe lowest within population genetic diversity at the edges of the sampled area such as Kansas and CA (Fig 2A).

Expansion of Johnsongrass throughout the Southeastern US corresponded to a striking habitat switch from agricultural (e.g., corn fields) to non-agricultural (i.e., roadside or disturbed systems) habitats in a multi-phase colonization process [15]. Ancestral genotypes of Johnsongrass are now found almost exclusively in agricultural habitats, and more derived genotypes (Phase 4 and 5) are found almost exclusively in non-agricultural habitats (Table 1). This suggests that initial expansion of Johnsongrass was in primarily agricultural habitats followed by a

secondary phase of invasion that occurred when a family of non-agricultural specialists evolved and rapidly colonized non-agricultural habitats throughout the entire Southern US. This second phase of invasion may reflect a period of rapid railroad and highway expansion that occurred in the late-nineteenth and early-twentieth centuries. Concomitantly with transportation development was a dramatic shift in agricultural weed management that moved away from mechanical control (e.g., cultivation) to chemical control, which may have facilitated this putative ecotypic shift. If this habitat switch represents a niche expansion for Johnsongrass, it may help to explain the particular invasiveness demonstrated by Johnsongrass, which is nearly ubiquitous in the farms, fields, prairies, and ditches of the Southern U.S [15].

The finding that DNA genotyping is diagnostic of a habitat switch suggests that specific alleles or allele combinations may be responsible for the switch, and that their identification might lead to novel means of mitigating the spread of Johnsongrass, at least to non-agricultural habitats. The limited sampling of SSRs that was sufficient for this study is inadequate to scan the entire genome for associations with important Johnsongrass traits, but reduced-representation resequencing of a diverse subset of the accessions studied here may provide the required resolution. For example, extensive spread and admixture may have seeded diverse geographies with identical-by-descent alleles for herbicide resistance, that appear to comprise independent evolutions when subsequent conditions such as increased herbicide usage favor their increase in frequency [44]. Such a genome-wide association approach may make it possible to dissect the genetic basis of the habitat switch, however more specific traits that differentiate the respective habitat types will first need to be identified and quantified in this tetraploid weed [15].

Despite lower and more structured diversity at the invasion front, Johnsongrass continues to advance northward [45], with predictions of the damage niche (conditions in which a plant becomes an agricultural pest) to also move northward with climate change [46]. Other invaders evolve more rapidly on the leading edge than those in the interior of their distribution [47]; and Johnsongrass is poised to continue this advance with more habitats available to colonize.

The richness of alleles in Johnsongrass and its spread to environments beyond the reach of its progenitors could offer novel and valuable variants for improvement of sorghum for resistance to drought, cold, disease and other stresses. Our improved understanding of population structure, together with an incipient 'diversity panel' of US *S. halepense*, provides for association genetic approaches to identify alleles responsible for specific traits of importance to weed/invasive management, and/or to crop improvement [20].

Supporting Information

S1 Fig. Figure A. Map of sampling sites and Principle Coordinate Analysis (PCoA) distribution of samples. **Figure B.** Principle Coordinate Analysis and locations of five groups at two extremes shown with number of genotypes in each group. **Figure C.** Mismatch distribution profiles (demographic expansion) from 12. **Figure D.** Mismatch distribution profiles (spatial expansion) from 12 states. **Figure E.** STRUCTURE analysis of genotypes at $K = 80$. **Figure F.** Neutrality test results for genetic markers employed in the study after 10000 simulations. (DOCX)

S1 File. Newick format file of N-J tree with bootstrap values.
(NWK)

S1 Table. Genetic marker information.
(DOCX)

S2 Table. GPS locations of samples.
(CSV)

S3 Table. Overall averages of variance components in test for Linkage Disequilibrium for non-random association between pairs of loci in finite subdivided populations.

(DOCX)

S4 Table. Variances for every pairwise comparison of loci in Linkage Disequilibrium test for non-random association between pairs of loci in finite subdivided populations.

(DOCX)

Acknowledgments

The efforts of Cheryl Fiore (J. Schroeder lab), Carl Bell, and John Roncoroni in identifying sites and contributing to germplasm collection are acknowledged.

Author Contributions

Conceptualization: AHP JNB.

Data curation: UUS AHP.

Formal analysis: UUS DZA YJ.

Funding acquisition: AHP JNB.

Investigation: UUS AHP.

Methodology: AHP JNB.

Project administration: AHP.

Resources: GAP JFP JMC TSC SC PD DK SS JS SDW WK VG SA LKR GJP CL RC CP AK
MM AHP JNB.

Supervision: AHP JNB.

Validation: UUS.

Visualization: UUS.

Writing – original draft: UUS.

Writing – review & editing: UUS JNB DZA AHP.

References

1. Holm L, Doll J, Holm E, Pancho J, Herberger J. World Weeds: Natural Histories and Distribution. New York: John Wiley and Sons; 1997. 1129 p.
2. Burke IC, Wilcut J, Cranmer J. Cross-Resistance of a Johnsongrass (*Sorghum halepense*) Biotype to Aryloxyphenoxypropionate and Cyclohexanedione Herbicides. Weed Technol. 2006; 20:571–5.
3. Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH. An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. Genome Research. 2004; 14(9):1812–9. PMID: [ISI:000223675900017](#). doi: [10.1101/gr.2479404](#)
4. McWhorter CG. Introduction and spread of Johnsongrass in the United States. Weed Science. 1971; 19(5):496. PubMed PMID: WOS:A1971K516800005.
5. Morrell PL, Williams-Coplin D, Bowers JE, Chandler JM, Paterson AH. Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. Molecular Ecology. 2005; 14:2143–54. doi: [10.1111/j.1365-294X.2005.02579.x](#) PMID: [15910333](#)

6. Casas AM, Kononowicz AK, Zehr UB, Tomes DT, Axtell JD, Butler LG, et al. Transgenic sorghum plants via microprojectile bombardment. *Proc Natl Acad Sci U S A*. 1993; 90(23):11212–6. Epub 1993/12/01. PMID: [8248230](#); PubMed Central PMCID: PMC47952.
7. Quinn L, Barney JN, McCubbins J, Endres A. Navigating the “Noxious” and “Invasive” Regulatory Landscape: Suggestions for Improved Regulation. *BioScience*. 2013; 63(2):124–31.
8. Rout M, Chrzanowski T. The invasive *Sorghum halepense* harbors endophytic N2-fixing bacteria and alters soil biogeochemistry. *Plant Soil*. 2009; 315:163–72.
9. Paterson A, Schertz K, Lin Y, Liu S, Chang Y. The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. *Proc Natl Acad Sci USA*. 1995; 92:6127–31. PMID: [11607551](#)
10. Monaghan N. The biology of Johnson grass (*Sorghum halepense*). *Weed Res*. 1979; 19:261–7.
11. Liu Y, Zhang C, Wei S, Cui H, Huan H. Compounds from the subterranean part of johnsongrass and their allelopathic potential. *Weed Biol Manage*. 2011; 11(3):160–6.
12. Barney JN, DiTomaso JM. Global climate niche estimates for bioenergy crops and invasive species of agronomic origin: potential problems and opportunities. *PLoS ONE*. 2011; 6(3):e17222 doi: [10.1371/journal.pone.0017222](#) PMID: [21408056](#)
13. Lavoie C, Shah MA, Bergeron A, Villeneuve P, van Kleunen M. Explaining invasiveness from the extent of native range: new insights from plant atlases and herbarium specimens. *Divers Distrib*. 2013; 19(1):98–105. doi: [10.1111/ddi.12014](#)
14. Baker HG. Characteristics and modes of origin of weeds. In: Baker HG, Stebbins GL, editors. *The genetics of colonizing species*. New York, NY: Academic Press; 1965. p. 147–72.
15. Atwater DZ, Sezen UU, Goff V, Kong W, Paterson AH, Barney JN. Reconstructing changes in the genotype, phenotype, and climatic niche of an introduced species. *Ecography*. 2015:n/a-n/a. doi: [10.1111/ecog.02031](#)
16. Maron JL, Vila M, Bommarco R, Elmendorf S, Beardsley P. Rapid evolution of an invasive plant. *Ecol Monogr*. 2004; 74(2):261–80.
17. Lavergne S, Molofsky J. Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc Natl Acad Sci USA*. 2007; 104(10):3883. doi: [10.1073/pnas.0607324104](#) PMID: [17360447](#)
18. Prentis PJ, Wilson JR, Dormontt EE, Richardson DM, Lowe AJ. Adaptive evolution in invasive species. *Trends Plant Sci*. 2008; 13(6):288–94. doi: [10.1016/j.tplants.2008.03.004](#) PMID: [18467157](#)
19. Stohlgren TJ, Pylovsek P, Kartesz J, Nishino M, Pauchard A, Winter M, et al. Widespread plant species: natives versus aliens in our changing world. *Biol Invasions*. 2011; 13:1931–44. doi: [10.1007/s10530-011-0024-9](#)
20. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doeblay J, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98(20):11479–84. PMID: [WOS:000171237100090](#). doi: [10.1073/pnas.201394398](#)
21. Aljanabi SM, Forget L, Dookun A. An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. *Plant Molecular Biology Reporter*. 1999; 17(3):281–. doi: [10.1023/A:1007692929505](#) PMID: [ISI:000084371200010](#).
22. Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, et al. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. *Theor Appl Genet*. 2005; 111(1):23–30. doi: [10.1007/s00122-005-1952-5](#) PMID: [WOS:000230141700004](#).
23. Liu KJ, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005; 21(9):2128–9. doi: [10.1093/bioinformatics/bti282](#) PMID: [ISI:000228783000061](#).
24. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 2007; 24(8):1596–9. doi: [10.1093/molbev/msm092](#) PMID: [ISI:000248848400005](#).
25. Parks DH, Beiko RG. Quantitative Visualizations of Hierarchically Organized Data in a Geographic Context. 2009 17th International Conference on Geoinformatics, Vols 1 and 2. 2009:316–21. ISI:000277622400060.
26. R-Team. R, a language and environment for statistical computing, version 3.1.1. 2014.
27. Perrier X, Jacquemond-Collet JP. DARwin Software. 2006.
28. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 2010; 10(3):564–7. doi: [10.1111/j.1755-0998.2010.02847.x](#) PMID: [ISI:000276407300020](#).

29. Yeh FC, Yang R-C, Boyle TBJ, Ye Z-H, Mao JX. POPGENE, the user-friendly shareware for population genetic analysis. University of Alberta: Molecular Biology and Biotechnology Centre, University of Alberta, Canada; 1997.
30. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*. 2007; 7(4):574–8. doi: [10.1111/j.1471-8286.2007.01758.x](https://doi.org/10.1111/j.1471-8286.2007.01758.x) PMID: [ISI:000247758100008](https://pubmed.ncbi.nlm.nih.gov/18126160/).
31. Corander J, Siren J, Arjas E. Bayesian spatial modeling of genetic population structure. *Computational Statistics*. 2008; 23(1):11–29. doi: [10.1007/s00180-007-0072-x](https://doi.org/10.1007/s00180-007-0072-x) PMID: [ISI:000252294000007](https://pubmed.ncbi.nlm.nih.gov/18126160/).
32. Earl DA, Vonholdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. 2012; 4(2):359–61. doi: [10.1007/s12686-011-9548-7](https://doi.org/10.1007/s12686-011-9548-7) PMID: [ISI:000303536400036](https://pubmed.ncbi.nlm.nih.gov/21731660/).
33. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007; 23(14):1801–6. Epub 2007/05/09. btm233 [pii] 10.1093/bioinformatics/btm233. PMID: [17485429](https://pubmed.ncbi.nlm.nih.gov/17485429/). doi: [10.1093/bioinformatics/btm233](https://doi.org/10.1093/bioinformatics/btm233)
34. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*. 2004; 4(1):137–8. doi: [10.1046/j.1471-8286.2003.00566.x](https://doi.org/10.1046/j.1471-8286.2003.00566.x) PMID: [ISI:000189159500042](https://pubmed.ncbi.nlm.nih.gov/18126160/).
35. Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST. Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95(4):1961–7. doi: [10.1073/pnas.95.4.1961](https://doi.org/10.1073/pnas.95.4.1961) PMID: [ISI:000072115900108](https://pubmed.ncbi.nlm.nih.gov/18126160/).
36. Roullier C, Benoit L, McKey DB, Lebot V. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. *Proceedings of the National Academy of Sciences*. 2013; 110(6):2205–10. doi: [10.1073/pnas.1211049110](https://doi.org/10.1073/pnas.1211049110)
37. Gunn BF, Baudouin L, Olsen KM. Independent Origins of Cultivated Coconut (*Cocos nucifera* L.) in the Old World Tropics. *Plos One*. 2011; 6(6):e21143. doi: [10.1371/journal.pone.0021143](https://doi.org/10.1371/journal.pone.0021143) PMID: [21731660](https://pubmed.ncbi.nlm.nih.gov/21731660/)
38. Tellman B, editor Stowaways and invited guests: How some exotic plants reached the American southwest. California Exotic Pest Plant Council; 1996 1996; California. California1996.
39. McWhorter CG. Introduction and Spread of Johnsongrass in United States. *Weed Science*. 1971; 19(5):496–500. PubMed PMID: [ISI:A1971K516800005](https://pubmed.ncbi.nlm.nih.gov/18126160/).
40. Excoffier L, Foll M, Petit RJ. Genetic Consequences of Range Expansions. *Annual Review of Ecology Evolution and Systematics*. 2009; 40:481–501. doi: [10.1146/annurev.ecolsys.39.110707.173414](https://doi.org/10.1146/annurev.ecolsys.39.110707.173414) PMID: [ISI:000272455700023](https://pubmed.ncbi.nlm.nih.gov/18126160/).
41. Dlugosch KM, Parker IM. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol Ecol*. 2008; 17(1):431–49. doi: [10.1111/j.1365-294X.2007.03538.x](https://doi.org/10.1111/j.1365-294X.2007.03538.x) PMID: [17908213](https://pubmed.ncbi.nlm.nih.gov/17908213/)
42. Novembre J, Stephens M. Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*. 2008; 40(5):646–9. doi: [10.1038/Ng.139](https://doi.org/10.1038/Ng.139) PMID: [ISI:000255366700034](https://pubmed.ncbi.nlm.nih.gov/18126160/).
43. Parodi LR. Una nueva especie de Sorghum cultivada en la Argentina. *Rev Argent Agron*. 1943; 10:361–72.
44. Fernandez L, de Haro LA, Distefano AJ, Carolina Martinez M, Lia V, Papa JC, et al. Population genetics structure of glyphosate-resistant Johnsongrass (*Sorghum halepense* L. Pers) does not support a single origin of the resistance. *Ecology and evolution*. 2013; 3(10):3388–400. doi: [10.1002/ece3.671](https://doi.org/10.1002/ece3.671) PMID: [24223277](https://pubmed.ncbi.nlm.nih.gov/24223277/); PubMed Central PMCID: [PMC3797486](https://pubmed.ncbi.nlm.nih.gov/PMC3797486/).
45. Warwick SI, Phillips D, Andrews C. Rhizome Depth—the Critical Factor in Winter Survival of Sorghum-Halepense (L) Pers (Johnson Grass). *Weed Research*. 1986; 26(6):381–7. doi: [10.1111/j.1365-3180.1986.tb00721.x](https://doi.org/10.1111/j.1365-3180.1986.tb00721.x) PMID: [ISI:A1986E974500002](https://pubmed.ncbi.nlm.nih.gov/18126160/).
46. McDonald A, Riha S, DiTommaso A, DeGaetano A. Climate change and the geography of weed damage: Analysis of US maize systems suggests the potential for significant range transformations. *Agriculture Ecosystems & Environment*. 2009; 130(3–4):131–40. doi: [10.1016/j.agee.2008.12.007](https://doi.org/10.1016/j.agee.2008.12.007) PMID: [ISI:000264677100008](https://pubmed.ncbi.nlm.nih.gov/18126160/).
47. Phillips BL, Brown GP, Shine R. Life-history evolution in range-shifting populations. *Ecology*. 2010; 91(6):1617–27. doi: [10.1890/09-0910.1](https://doi.org/10.1890/09-0910.1) PMID: [ISI:000278859700008](https://pubmed.ncbi.nlm.nih.gov/18126160/).