Allopolyploidy, Diversification, and the Miocene Grassland Expansion

Matt Estep
Michael McKain
Dilys Diaz
Jinshun Zhong
*University of Missouri–St. Louis*
John Hodge

See next page for additional authors

Follow this and additional works at: [https://irl.umsl.edu/biology-faculty](https://irl.umsl.edu/biology-faculty)

Part of the Biology Commons, and the Botany Commons

**Recommended Citation**

Estep, Matt; McKain, Michael; Diaz, Dilys; Zhong, Jinshun; Hodge, John; Hodkinson, Trevor; Layton, Daniel; Malcomber, Simon; Pasquet, Rémy; and Kellogg, Elizabeth, "Allopolyploidy, Diversification, and the Miocene Grassland Expansion" (2014). *Biology Department Faculty Works*. 57.

DOI: [https://doi.org/10.1073/pnas.1404177111](https://doi.org/10.1073/pnas.1404177111)

Available at: [https://irl.umsl.edu/biology-faculty/57](https://irl.umsl.edu/biology-faculty/57)

This Article is brought to you for free and open access by the Biology at IRL @ UMSL. It has been accepted for inclusion in Biology Department Faculty Works by an authorized administrator of IRL @ UMSL. For more information, please contact marvinh@umsl.edu.
Authors
Matt Estep, Michael McKain, Dilys Diaz, Jinshun Zhong, John Hodge, Trevor Hodkinson, Daniel Layton,
Simon Malcomber, Rémy Pasquet, and Elizabeth Kellogg

This article is available at IRL @ UMSL: https://irl.umsl.edu/biology-faculty/57
Allopolyploidy, diversification, and the Miocene grassland expansion

Matt C. Estepa,b,1, Michael R. McKainb,c,1, Dílys Vela Díazad,1, Jinhun Zhonga,e, John G. Hodgee,3, Trevor R. Hodkinson4, Daniel J. Laytonb,c, Simon T. Malcomber4, Rémy Pasquetb,1, and Elizabeth A. Kellogg5,2

1Department of Biology, University of Missouri-St. Louis, St. Louis, MO 63121; 2Department of Biology, Appalachian State University, Boone, NC 28608; 3Donald Danforth Plant Science Center, St. Louis, MO 63132; 4Department of Biology, Washington University, St. Louis, MO 63130; 5Department of Plant Biology, University of Vermont, Burlington, VT 05405; 6Department of Botany, Trinity College Dublin, Dublin 2, Ireland; 7Division of Environmental Biology, National Science Foundation, Arlington, VA 22230; 8International Centre of Insect Physiology and Ecology (icipe), Nairobi, Kenya; 9Unité de Recherche Institut de Recherche pour le Développement 072, Laboratoire Évolution, Génomes et Spéciation, 91198 Gif-sur-Yvette, France; and 10Université Paris-Sud 11, 91400 Orsay, France

Edited* by John F. Doebley, University of Wisconsin-Madison, Madison, WI, and approved July 29, 2014 (received for review March 4, 2014)

The role of polyploidy, particularly allopolyploidy, in plant diversification is a subject of debate. Whole-genome duplications precede the origins of many major clades (e.g., angiosperms, Brassicaceae, Poaceae), suggesting that polyploidy drives diversification. However, theoretical arguments and empirical studies suggest that polyploid lineages may actually have lower speciation rates and higher extinction rates than diploid lineages. We focus here on the grass tribe Andropogoneae, an economically and ecologically important group of C3 species with a high frequency of polyploids. A phylogeny was constructed for ca. 10% of the species of the clade, based on sequences of four concatenated low-copy nuclear loci. Genetic allopolyploidy was documented using the characteristic pattern of double-labeled gene trees. At least 32% of the species sampled are the result of genetic allopolyploidy and result from 28 distinct tetraploidy events plus an additional six hexaploidy events. This number is a minimum, and the actual frequency could be considerably higher. The parental genomes of most Andropogoneae polyploids diverged in the Late Miocene coincident with the expansion of the major C3 grasslands that dominate the earth today. The well-documented whole-genome duplication in Zea mays ssp. mays occurred after the divergence of Zea and Sorghum. We find no evidence that polyploidization is followed by an increase in net diversification rate; nonetheless, allopolyploidy itself is a major mode of speciation.

Polyploidy (whole-genome duplication) is often linked with the acquisition of new traits and subsequent species diversification, particularly in plants (1, 2). Ancient polyploidy correlates with major land-plant radiations (3) and the origins of orders, large families, and major clades (4–7) although, in many cases, sharp changes in diversification rates are delayed for millions of years after the polyploidization event (1). This phylogenetic pattern has led to the hypothesis that polyploidy causes or promotes diversification. Good mechanistic reasons support such a hypothesis. Studies of naturally occurring and synthetic polyploids find changes in gene expression, gene loss, release of transposons, and changes in morphology and physiology immediately after polyploidy (4, 8–12). This pattern is particularly true for allopolyploids, which originate from a cross between genetically distinct parents often representing different species; in such cases, the biological changes seen after polyploidization may not reflect the effects of genome doubling per se, but rather the effect of hybridization between distantly related progenitors (4).

Despite the appeal of the hypothesis that polyploidy causes diversification, there is evidence to the contrary. As noted by Stebbins (13), “polyploidy has been important in the diversification of species and genera within families, but not in the origin of the families and orders themselves,” implying that polyploidy is only a minor force in diversification (see also ref. 14). Surveys of angiosperms and ferns have found no evidence for increased speciation after polyploidization (15, 16), supporting Stebbins’s hypothesis. However, because of the large scale of the analyses, these studies necessarily had to rely on inference of polyploidization events from chromosome numbers. Although unavoidable, inferences from chromosome numbers require assumptions about which numbers represent polyploids and when the polyploids arose.

Phylogenetic trees of nuclear genes can reliably identify allopolyploidization events (17–21) because they produce characteristic double-labeled tree topologies in which the polyploid species appears twice (Fig. S1). In such trees, allopolyploids are readily identified and can be discovered even when no chromosome counts are available (e.g., ref. 22). This tree-based method will recover genetic allopolyploids; following the terminology of Doyle and Egan (23), these polyploids may be descendants of two species (taxonomic allopolyploids) or one (taxonomic autopolyploids). The gene-tree approach thus provides unambiguous evidence of allopolyploidization and provides an estimate of relative timing that does not require calibration of a tree. The approach does require identification of all or most paralogues, which can be accomplished by investigating genes in categories that tend to be retained after whole-genome duplication, such as transcription factors.

The grass family is a powerful system in which to study allopolyploidization. About 80% of its species are polyploid (24), and Hunziker and Stebbins (25) have noted that the grasses are the “only large family in which high frequency of polyploids prevails throughout the family.” The polyploids that Stebbins refers to and that we analyzed data; and E.A.K. wrote the paper.

Author contributions: M.C.E., R.P., and E.A.K. designed research; M.C.E., M.R.M., D.V.D., and J.Z. contributed equally to this paper.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database. For a list of accession numbers, see Table S1.

1To whom correspondence should be addressed. Email: ekellogg@danforthcenter.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1404177111/-/DCSupplemental.

Significance

Duplication of genomes following hybridization (allopolyploidy) is common among flowering plants, particularly in the grasses that cover vast areas of the world and provide food and fuel. Here, we find that genome duplication has occurred at a remarkable rate, accounting for at least a third of all speciation events in a group of about 1,200 species. Much of this genome duplication occurred during the expansion of the C3 grasslands in the Late Miocene. We find no evidence that allopolyploidy leads directly to a change in the net rate of diversification or correlates with the origin of novel morphological characters. However, as a mode of speciation, the frequency of allopolyploidization is surprisingly high.

www.pnas.org/cgi/doi/10.1073/pnas.1404177111

PNAS | October 21, 2014 | vol. 111 | no. 42 | 15149–15154

10.1073/pnas.1404177111

*This Direct Submission article had a prearranged editor.

M.C.E., R.P., and E.A.K. designed research; M.C.E., M.R.M., D.V.D., and J.Z. contributed equally to this paper.

The authors declare no conflict of interest.

See Correction for this article

Edited* by John F. Doebley, University of Wisconsin-Madison, Madison, WI, and approved July 29, 2014 (received for review March 4, 2014)

Allopolyploidy is often linked with the acquisition of new traits and subsequent species diversification, particularly in plants (1, 2). Ancient polyploidy correlates with major land-plant radiations (3) and the origins of orders, large families, and major clades (4–7) although, in many cases, sharp changes in diversification rates are delayed for millions of years after the polyploidization event (1). This phylogenetic pattern has led to the hypothesis that polyploidy causes or promotes diversification. Good mechanistic reasons support such a hypothesis. Studies of naturally occurring and synthetic polyploids find changes in gene expression, gene loss, release of transposons, and changes in morphology and physiology immediately after polyploidy (4, 8–12). This pattern is particularly true for allopolyploids, which originate from a cross between genetically distinct parents often representing different species; in such cases, the biological changes seen after polyploidization may not reflect the effects of genome doubling per se, but rather the effect of hybridization between distantly related progenitors (4).

Despite the appeal of the hypothesis that polyploidy causes diversification, there is evidence to the contrary. As noted by Stebbins (13), “polyploidy has been important in the diversification of species and genera within families, but not in the origin of the families and orders themselves,” implying that polyploidy is only a minor force in diversification (see also ref. 14). Surveys of angiosperms and ferns have found no evidence for increased speciation after polyploidization (15, 16), supporting Stebbins’s hypothesis. However, because of the large scale of the analyses, these studies necessarily had to rely on
discuss in this paper all formed after the well-documented whole-genome duplication at the origin of the family (26, 27); the grass WGD (rho) occurred about 80–90 million years ago (Mya) (28, 29).

Here, we focus on the tribe Andropogoneae (subfamily Panicoideae), a group of about 1,200 species in 90 genera. The tribe is morphologically diverse and contains some of our most economically important crop plants [maize (Zea mays), sorghum, and sugarcane (Saccharum officinarum)]. In addition, members of the tribe are ecologically dominant species of temperate and tropical grasslands (e.g., Andropogon gerardii, Schizachyrium scoparium, and Sorghastrum nutans) of the North American tall grass prairie; Themeda triandra of East African and Australian grasslands), some of the most troublesome agricultural weeds (Imperata cylindrica, Sorghum halepense, Ischaemum spp.), many important forage crops (Dichanthium spp., Bothriochloa spp., Andropogon spp.), and some burgeoning biofuels (Miscanthus and Saccharum). In short, the Andropogoneae feed, and increasingly fuel, the planet.

Of the species listed in the preceding paragraph, all but sorghum are allopolyploid. We therefore sought to test whether each allopolyploidy event leads to one or a few species (the Stebbins hypothesis), in which case, the allopolyploids should appear in the phylogeny as the result of numerous independent polyploidy events. Alternatively, allopolyploidy could correlate with diversification of a major clade or even the entire tribe (the diversification hypothesis), in which case most allopolyploids would be the product of a single polyploidy event followed by radiation at the tetraploid or hexaploid level; we would see only a handful of polyploidization events deep in the tree (Fig. S1, second row).

Previous attempts to resolve relationships within Andropogoneae, using both chloroplast and nuclear markers, have produced phylogenies that are poorly supported, with interfamilial relationships unresolved (30–36). The pattern apparently indicates a rapid radiation. One reason for the difficulty of resolving the phylogeny may be a failure to account for polyploidy. Mixing of orthologues and paralogues in gene trees, combined with extensive reticulation in the history of the genera, could easily destroy any phylogenetic signal. The apparent rapid radiation of the tribe might also reflect diversification following a single polyploidy event.

In this paper, we used a phylogenetic approach, supported by data on chromosome number and genome size. We found a remarkable number of allopolyploidization events, none of which correlate with shifts in diversification rate or with the origin of unique morphological characters. Repeated formation of polyploids correlates with expansion of the C₄ grasslands in the Late Miocene.

Results: Allopolyploidy occurs repeatedly throughout the Andropogoneae, accounting for at least a third of the speciation events. Based on our phylogenetic criteria, 28 groups (numbered in Fig. 1 and Figs. S2 and S3) show the phylogenetic signature of allo-tetraploidization, with an additional six hexaploid events (letters in Fig. 2). Together, these events have given rise to 32 of the 100 sampled species, with a few species forming repeatedly, from more than one independent event.

Most allopolyploidization events have occurred within genera or species. For example, T. triandra is the product of an event (event 20) separate from the tetraploid that gave rise to T. villosa and T. anindinacea (event 19) (Fig. 1 and Fig. S3). Hyparrhenia hirta, Hyparrhenia rufa, and Hyparrhenia diplandra appear to be the products of four separate events (events 9–12, with additional rounds of polyploidization leading to hexaploids in H. hirta (events A and A') (Fig. 1 and Fig. S3). We looked for shifts in net diversification rate (speciation minus extinction) using Bayesian Analysis of Macrowevel Mixtures (BAMM) (37). A model with no shift of evolutionary rates had a slightly higher posterior probability than one with one shift (odds ratio 1.0395); in the model with one shift, the position of the shift cannot be determined with certainty (Fig. S4). However, there is no evidence of an increase in diversification in polyploid clades.

To test net diversification rates in diploids versus polyploids directly, we used a nonparametric comparison between the numbers of species in diploid clades and those in their sister polyploid clades and found no significant difference (P = 0.637). We then used the Binary State Speciation and Extinction (BiSSE) model (38) and found that net diversification is significantly higher in diploids, assuming a sampling fraction of 8% (likelihood ratio test, P < 0.02) (Fig. S5); the pattern is similar assuming 50% sampling, but the difference is not significant. Although estimated speciation rates are higher in polyploids, estimated extinction rates are also higher (K-S test, P < 2.2e−16), leading to lower net diversification.

Only in 2 of the 32 allopolyploid lineages can we be certain that polyploidy is followed by diversification to produce more than one or two species. Event 2 represents a hybridization and polyploidy event that led to Z. mays plus Tripsacum dactyloides: i.e., the allopolyploidy apparently preceded the origin of the genera, which
Bothriochloa species, because this clade includes all of the South American species that we sampled (in addition to a few species from the Old World) and because all South American species are thought to be hexaploid or higher, we suspect that all of the remaining 62% are diploid. We infer diploidy if our accession has a genome size less than ca. 1 Gbp, published counts for the species are all diploid, and we find only a single gene copy for each locus; all three criteria apply, for example, to Sorghum bicolor, Diheteropogon amplentecus, Pseudosorghum fasciculare, and Andropogon virginicus. The latter species is a member of a clade of four species that is sister to one of the big bluestem (Andropogon) genomes (Fig. S3); A. virginicus is, like A. gerardii, a native of the North American prairie and thus is a plausible model for one genome of big bluestem. For most other species, however, chromosome counts are unavailable or published counts include several ploidy levels and are often poorly documented. Genome-size estimates may be lacking for field-collected specimens and are in any case not definitive. Discovering only one gene copy is weak evidence because our PCR-based approach may not have uncovered all paralogues. We do not think gene loss is a major concern because we have always been able to locate paralogues in known polyploids given sufficient depth of sequencing and because we deliberately chose gene classes that are generally retained in duplicate after polyploidization events.

Most alloploidy formed in the Late Miocene or more recently. Of the documented polyploidization events, only three could have occurred more than 9 Mya (Table 1). Gene-tree data provide an estimate of the divergence time of the parental diploid genomes but cannot determine exactly when the hybridization and genome-doubling events occurred (23); dates in Table 1 are thus the oldest possible dates for polyploidization. However, if speciation has occurred after polyploidization, as in Themeda villosa-anudinacea, or Zea-Tripsacum, the date of the earliest tetraploid speciation event provides the most recent possible date. Thus, the Zea-Tripsacum event occurred after 13 Mya (the divergence of the parental genomes) but before 8.5 Mya (the earliest estimate for divergence of the two genera).

**Discussion**

The number of independent alloploidy events in Andropogoneae is remarkably and unexpectedly high. Thirty-four unambiguous alloploidy events can be identified easily from the tree, and this is certainly an underestimate of the true number. Although we do not know all mechanisms of speciation in the tribe, polyploidy may be as important as any other single mechanism. Major clades with distinctive morphology were established before the alloploidy events, which also do not correlate with obvious habitat shifts, or subsequent bursts of speciation. No polyploidization event
Table 1. Divergence dates of parental genomes of allopolyploids, and 95% highest posterior density (HPD) values

<table>
<thead>
<tr>
<th>Labeled event</th>
<th>Taxon</th>
<th>Divergence of parental genomes, Mya</th>
<th>95% HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chrysopogon gryllus</td>
<td>8.4</td>
<td>10.6 4.1</td>
</tr>
<tr>
<td>2</td>
<td>Zea-Tripsacum</td>
<td>13.6</td>
<td>18.2 9.1</td>
</tr>
<tr>
<td>3</td>
<td>Apocops intermedius</td>
<td>4.7</td>
<td>6.3 2.4</td>
</tr>
<tr>
<td>4</td>
<td>Miscanthus sinensis</td>
<td>7.4</td>
<td>9.4 3.7</td>
</tr>
<tr>
<td>5</td>
<td>Miscanthus-Saccharum</td>
<td>3.8</td>
<td>6.2 2.0</td>
</tr>
<tr>
<td>6</td>
<td>Polytrias indica</td>
<td>3.4</td>
<td>5.8 1.9</td>
</tr>
<tr>
<td>7</td>
<td>Ischaemum santapau</td>
<td>2.3</td>
<td>3.4 0.9</td>
</tr>
<tr>
<td>8</td>
<td>Dimeria fuscescens</td>
<td>4.2</td>
<td>7.9 3.4</td>
</tr>
<tr>
<td>9</td>
<td>Hyparrhenia diplandra</td>
<td>4.2</td>
<td>6.1 1.9</td>
</tr>
<tr>
<td>10</td>
<td>Hyparrhenia rufa (153)</td>
<td>7.0</td>
<td>8.3 3.8</td>
</tr>
<tr>
<td>11</td>
<td>Hyparrhenia rufa (108)</td>
<td>7.0</td>
<td>8.3 3.8</td>
</tr>
<tr>
<td>12</td>
<td>Hyparrhenia hirta (018, 185)</td>
<td>2.2</td>
<td>3.0 1.1</td>
</tr>
<tr>
<td>13</td>
<td>Schizachyrium sanguineum var. hirtiflorum</td>
<td>5.5</td>
<td>7.2 3.0</td>
</tr>
<tr>
<td>14</td>
<td>Schizachyrium sanguineum</td>
<td>4.3</td>
<td>6.3 2.4</td>
</tr>
<tr>
<td>15</td>
<td>Andropogon gerardii</td>
<td>8.6</td>
<td>10.2 4.7</td>
</tr>
<tr>
<td>16</td>
<td>Eulalia quadrirevis</td>
<td>3.8</td>
<td>8.6 2.9</td>
</tr>
<tr>
<td>17</td>
<td>Cymbopogon citratus</td>
<td>3.0</td>
<td>3.6 1.5</td>
</tr>
<tr>
<td>18</td>
<td>Cymbopogon parneri</td>
<td>4.2</td>
<td>4.7 1.9</td>
</tr>
<tr>
<td>19</td>
<td>Themeda villosa</td>
<td>3.3</td>
<td>3.8 0.8</td>
</tr>
<tr>
<td>20</td>
<td>Themeda triandra</td>
<td>2.0</td>
<td>3.3 1.0</td>
</tr>
<tr>
<td>21</td>
<td>Capillipedium parviflorum</td>
<td>3.1</td>
<td>3.6 1.2</td>
</tr>
<tr>
<td>22</td>
<td>Capillipedium spicigerum</td>
<td>1.4</td>
<td>1.6 0.4</td>
</tr>
<tr>
<td>23</td>
<td>Bothriochloa bladhii (093, 082, 076)</td>
<td>7.5</td>
<td>9.8 5.0</td>
</tr>
<tr>
<td>24</td>
<td>Dichanthium annulatum</td>
<td>1.4</td>
<td>2.2 0.8</td>
</tr>
<tr>
<td>25</td>
<td>Bothriochloa insculpta</td>
<td>2.8</td>
<td>4.6 1.8</td>
</tr>
<tr>
<td>26</td>
<td>Bothriochloa bladhii (087)</td>
<td>8.5</td>
<td>11.1 5.7</td>
</tr>
<tr>
<td>27</td>
<td>Bothriochloa ischadem (063)</td>
<td>1.5</td>
<td>2.7 0.7</td>
</tr>
<tr>
<td>28</td>
<td>Bothriochloa spp.</td>
<td>6.0</td>
<td>7.3 3.4</td>
</tr>
<tr>
<td>A</td>
<td>Hyparrhenia hirta (185, 189, 018)</td>
<td>3.6</td>
<td>5.4 2.3</td>
</tr>
<tr>
<td>A’</td>
<td>Hyparrhenia hirta (018)</td>
<td>3.0</td>
<td>4.1 1.7</td>
</tr>
<tr>
<td>B</td>
<td>Schizachyrium sanguineum var. hirtiflorum</td>
<td>6.5</td>
<td>8.3 3.5</td>
</tr>
<tr>
<td>C</td>
<td>S. sanguineum var. hirtiflorum</td>
<td>9.9</td>
<td>11.8 5.8</td>
</tr>
<tr>
<td>D</td>
<td>Cymbopogon citratus</td>
<td>6.6</td>
<td>7.9 3.3</td>
</tr>
<tr>
<td>E</td>
<td>Bothriochloa spp.</td>
<td>9.4</td>
<td>11.8 6.1</td>
</tr>
<tr>
<td>F</td>
<td>Bothriochloa bladhii (093)</td>
<td>8.5</td>
<td>11.1 5.7</td>
</tr>
</tbody>
</table>

Numbered events indicate allotetraploids; letters indicate allohexaploids. Accession numbers in parentheses are for species represented by more than one accession.

Encompasses the base of the tree; we thus reject the most extreme diversification hypothesis—that allopolyploidy precedes the radiation of the tribe. Instead, allopolyploidy events have occurred repeatedly postdiversification.

Simple inspection of the tree, plus results of the BAMM and BiSEE analyses, is consistent with the Stebbins hypothesis—that polyploidization is common but not accompanied by an increase in net speciation rate (13, 15, 16, 39). In fact, we found a modest net decrease in diversification, similar to that modeled by Arrigo and Barker (39). The polyploidization events we identify are old enough that there has been enough time for subsequent speciation (Table 1). The tree includes speciation events as young as 0.5 My. Generation time in these plants is 1–2 y, so 0.5 My represents at least 250,000 generations, presumably long enough for detectable evolutionary change.

Our data also fit with the Radiation Time-lag Hypothesis—that polyploidy has no immediate effect on diversification at all but rather that the effects are delayed (1). For example, nearly 20 My elapsed between the genome duplication in the common ancestor of the grasses and the major radiation of species (28), and a similar time lag occurred after the duplication at the origin of Brassicaceae (40). The time-lag hypothesis provides one possible way to reconcile the diversification hypothesis and the Stebbins hypothesis—polyploidy may have no effect that can be detected over the first 10–20 million years, and any burst of speciation occurs later. However, the time-lag hypothesis also posits that genome duplication correlates with the origin of a novel and possibly adaptive trait. If that has occurred in any of our taxa, we have not been able to identify it.

Unlike most previous studies, our estimate of polyploidization is drawn solely from gene-tree structure although we bolster the conclusion with available data on genome size and in a few cases chromosome number. Although we could compare the speciation rate in a clade of polyploids sister to a clade of diploids, the polyploids may have formed (as in Hyparrhenia) by repeated independent polyploidization events; thus, the comparison does not directly assess diversification after polyploidization.

Although many lineages have failed to diversify after allopolyploidy, there are exceptions. The 17 species in Zea-Tripsacum (event 2) and 13 species of New World Bothriochloa (events 28 and E) clearly diversified after polyploidization. Without exhaustively sampling all 1,200 species in the tribe, we cannot be certain exactly how many other such groups exist, but it is possible to estimate an upper bound on the number. For example, the three sampled species of Hyparrhenia result from at least four separate tetraploidization events (events 9–12). About 75% of the 54 species in the genus are polyploid, or about 40; if they all descended from one of the identified allopolyploid events, then the largest possible radiation would be about 37 species. However, it is unlikely that we accidentally found the only
Polyploidization events in the genus and that all remaining polyploids are descended from one of them, so the total number following a single allopolyploidization event is likely to be much lower. Similar arguments can be made regarding other genera.

Most of the allopolyploidization events identified here occurred in the Late Miocene, simultaneous with or following the well-documented expansion of the C4 grasslands (41). The dominant species of modern C4 grasslands are members of Andropogoneae, and most are allopolyploid. Unexpectedly, we found that many of these ecological dominants in the genera Andropogon, Schizachyrium, and Hyparrhenia form a single, strongly supported clade [Bayesian posterior probability (pp) 1.0, Maximum Likelihood bootstrap (ML) 100] whose origin is dated to about 10.5 [13.5–7.5, 95% highest posterior density (HPD)] Mya. This date correlates closely with the date when C4 species came to dominate grasslands in Africa and Southern Asia (Pakistan), also estimated about 10–11 Mya; the expansion in North America is dated about 7 Mya (41). Allopolyploidy is thus correlated with ecological success, whether or not it leads to diversification.

Polyploid formation and grassland expansion are correlated in time but may or may not be causally connected. Allopolyploids might tolerate more diverse environments than their diploid ancestors, such that repeated formation of polyploids could have helped drive the formation of the grasslands. Alternatively, expansion of the grasslands might have led to expansion of the range of progenitor diploids; if these diploids then came into contact with diverged species, hybridized and formed allopolyploids, then the causation is the opposite—expansion of the grasslands drove polyploidization. The third possibility, of course, is that the correlation of polyploidy and grassland expansion is simply fortuitous and neither caused the other.

The complex phylogenetic structure of the BCD clade confirms and extends data collected over 40 y ago by de Wet and Harlan (42) (Fig. 2B). In a substantial set of crossing and cytogenetic studies, they demonstrated that the genomes of the species that are easily distinguished morphologically, are intersterile at the diploid level. However, at the tetraploid level, fertility barriers are weak, and interfertile species and allopolyploids are rampant. Many of the species are also apomictic, which is known to correlate strongly with polyploidy (43). At the center of the complex is a species that was then called Bothriochloa intermedia, now B. bladhii. This species can be clearly identified as a Bothriochloa, despite the acquisition of genomes from Capillipedium and Dichanthium. Harlan and de Wet (44) coined the term “poliploides” for B. bladhii to indicate its ability to “absorb” genes from other genera.

Some accessions of B. bladhii indeed have a Bothriochloa genome plus another that is part of a clade with Capillipedium (event 23); other Bothriochloa accessions have Bothriochloa plus Dichanthium genomes (event 26). These accessions were all grown to flowering for this study and are unmistakably morphologically Bothriochloa despite their parentage. Thus, the B. bladhii morphology appears to be dominant in some way, despite the disparate genomes from the other two species.

Our data do not identify autopolyploids, which are defined taxonomically as polyploids in which the parents are the same species, or genetically as plants exhibiting polysomic inheritance (23, 45, 46). In most cases, the lack of diploids makes it impossible to assess whether the parents are conspecific with the polyploids or not. For example, Bothriochloa insculpta (event 25) and T. triandra (event 20) are tetraploids with two distinct genomes (genetic allopolyploids), but we do not know whether the genome donors would both be classified in the same species. Our data also suggest a need for caution in inference of taxonomic autopolyploidy. Each accession of B. bladhii includes genomes from Bothriochloa and from either Capillipedium or Dichanthium, but the latter genomes are morphologically “invisible.” Likewise, A. intermedius has one parent classified as Apocopy and one as Geumina. Going by morphology alone, plants of B. bladhii or A. intermedius with different chromosome numbers would be interpreted as taxonomically autopolyploid whereas, in fact, they are both taxonomically and genetically allopolyploid. Morphology gives little or no hint of the allopolyploid origin of the species.

The phylogeny presented here (Fig. 1 and Figs. S2 and S3) is consistent with previous studies of Andropogoneae but considers the results. Like previous workers, we found that the sister genus to the rest of the tribe is Arundinella [formerly in Arundinellae], that Andropogoneae is monophyletic, and that the early diversification appears to have been rapid. Novel results are too numerous to list here but include a moderately well supported clade (1.0 pp, 71% ML) that we call “core Andropogoneae,” which includes over half the species in the tribe and within which relationships are mostly strongly supported. This clade is considered to be subtribe Andropogoninae by Kellogg (47), an interpretation that is broader than that of Clayton and Renvoize (48). The ecologically dominant genera Andropogon, Schizachyrium, and Hyparrhenia together form a strongly supported clade. Saccharum and Miscanthus are closely related, consistent with previous work (49).

Zea is a well-documented allotetraploid, and the relationships of its genomes to sorghum have long been of interest (50, 51). A previous study (52) concluded that the divergence of sorghum and the maize genomes was nearly simultaneous and occurred about 11.9 Mya. We provide a slightly earlier date, along with strong evidence that maize allopolyploidy occurred after the divergence from sorghum. In addition, we confirm that polyploidy of Miscanthus sinensis also occurred after the Miscanthus–Sorghum divergence, consistent with recent mapping studies (53–55).

The methods used here provide a robust minimum estimate of the number of allopolyploidization events in the clade and do not require the same assumptions as other approaches. Specifically, assessment of polyploidy from chromosome number or genome size requires an assumption of what the diploid value is likely to be; because both chromosome number and genome size can increase by mechanisms unconnected to polyploidy, they are helpful but not definitive. Likewise, the frequency distribution of synony- mous sites (Ks plots) is another method of assessing allopolyploidy but does not have the power to determine whether near-simultaneous events are independent or not, or to distinguish allopolyploidy from autoploidy. As increasing numbers of nuclear genomes become available, we suggest that the gene-tree signature of allopolyploidy may provide valuable estimates of the precise timing and independence of allopolyploidization events.

Finally, we note that our formal tests for diversification, like those in the literature, assume an underlying bifurcating model of evolution whereas the question addressed here (and elsewhere) concerns the result of reticulation, the evolutionary history of the polyploids, and the appropriate null model (56). For the analyses in BAMM and BiSSE, we included only one genome of the polyploids, thereby forcing the tree to be bifurcating. In addition, BiSSE simply tests whether there is an elevated rate of net diversification in clades that contain polyploids, confounding the numbers of species that are formed by the polyploidization event itself and species formed by divergent evolution after the reticulation (Fig. S1). Rigorous determination of the role of polyploidy in subsequent diversification awaits development of new methods.

Summary, allopolyploidization is not clearly linked to increases in net diversification rate over a time scale of ca. 20 million y (the depth of the current phylogeny) although, in a few lineages, there has been some speciation following the polyploidization event. Nonetheless, allopolyploidization itself is surprisingly common and clearly produces new combinations of genomes repeatedly and frequently. Whatever their long-term fate, the resulting allopolyploids have come to dominate the major grassland ecosystems of the planet.

Materials and Methods

Taxon sampling included 114 accessions; 2 outgroup species, 2 species of Arundinella, and 100 species of Andropogoneae in 40 genera. Vouchers and sources of material are listed in Table S1. Trees were constructed from five regions of four nuclear loci, following Estep et al. (57) (SI Materials and Methods). We found no evidence that these loci are lost after polyploidization; in known
polyploids, paralogues are found consistently as long as sequence depth is sufficient. Sequences were concatenated to produce a single data matrix with a minimum of three out of five loci for each taxon. Concatenated trees were reconstructed using both maximum likelihood (ML) and Bayesian approaches. Divergence times were estimated using BEAST 1.7.5 (58). Tests for shifts in the underlying model of diversification were conducted using BAMM (37), on a tree constructed by pruning the BEAST tree to leave only a single paraglogue (genus) per species. Differences in speciation rate for polyploids versus diploids were estimated using the Bayesian State Speciation and Extinction (BSISE) model (38) as implemented in Mesquite (39) and diversitree (60). See SI Materials and Methods for more details. Genome size was measured by flow cytometry a minimum of four times using standard methods, and the values were averaged (61). To obtain a minimum estimate of the number of allopolyploid events, we looked for a clear phylogenetic signal of allopolyploidy, in the form of a multiple-labeled gene tree (Fig. S1).

ACKNOWLEDGMENTS. This study was supported National Science Foundation Grants DBI-0604923 and DEB-11456884 (to E.A.K.) and by the E. Desmond Lee and Family Endowment at the University of Missouri–St. Louis.