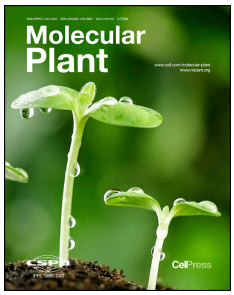
Journal Pre-proof



The genomes of the allohexaploid *Echinochloa crus-galli* and its progenitors provide insights into polyploidization-driven adaptation

Chu-Yu Ye, Dongya Wu, Lingfeng Mao, Lei Jia, Jie Qiu, Sangting Lao, Meihong Chen, Bowen Jiang, Wei Tang, Qiong Peng, Lang Pan, Lifeng Wang, Xiaoxiao Feng, Longbiao Guo, Chulong Zhang, Elizabeth A. Kellogg, Kenneth M. Olsen, Lianyang Bai, Longjiang Fan

PII: S1674-2052(20)30214-8
DOI: https://doi.org/10.1016/j.molp.2020.07.001
Reference: MOLP 954

To appear in: *MOLECULAR PLANT*
Accepted Date: 30 June 2020

Please cite this article as: **Ye C.-Y., Wu D., Mao L., Jia L., Qiu J., Lao S., Chen M., Jiang B., Tang W., Peng Q., Pan L., Wang L., Feng X., Guo L., Zhang C., Kellogg E.A., Olsen K.M., Bai L., and Fan L.** (2020). The genomes of the allohexaploid *Echinochloa crus-galli* and its progenitors provide insights into polyploidization-driven adaptation. Mol. Plant. doi: https://doi.org/10.1016/j.molp.2020.07.001.

The genomes of the allohexaploid *Echinochloa crus-galli* and its progenitors provide insights into polyploidization-driven adaptation

Chu-Yu Ye[1,#], Dongya Wu[1,#], Lingfeng Mao[1], Lei Jia[1], Jie Qiu[2], Sangting Lao[1], Meihong Chen[1], Bowen Jiang[1], Wei Tang[3], Qiong Peng[4], Lang Pan[4], Lifeng Wang[4], Xiaoxiao Feng[5], Longbiao Guo[3], Chulong Zhang[6], Elizabeth A. Kellogg[7], Kenneth M. Olsen[8], Lianyang Bai[4,*], Longjiang Fan[1,*]

[1]Institute of Crop Sciences & Institute of Bioinformatics, Zhejiang University, Hangzhou 310058, China

[2]Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, Shanghai 200235, China

[3]State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, China

[4]Hunan Weed Science Key Laboratory, Hunan Academy of Agriculture Science, Changsha, 410125, China

[5]Agricultural Experiment Station, Zhejiang University, Hangzhou, 310058, China

[6]Institute of Biotechnology, Zhejiang University, Hangzhou, 310058, China

[7]Donald Danforth Plant Science Center, St. Louis, MO 6132, United States

[8]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, United States

[#]These authors contributed equally to this work.

*Authors for correspondence:

Longjiang Fan (fanlj@zju.edu.cn) or Lianyang Bai (bailianyang2005@aliyun.com)

Short Summary

We generated high quality genome sequences of hexaploid *Echinochloa crus-galli* and its progenitor *E. oryzicola* (tetraploid) and a diploid species (*E. haploclada*). Gene family expansion, subgenome evolution and transcriptomic changes during hexaploidization were compared between *E. crus-galli* and bread wheat. The results illustrate different patterns of genome evolution during polyploidization of the agricultural weed and a crop.

Abstract

The hexaploid species *Echinochloa crus-galli* is one of the most detrimental weeds in crop fields, especially in rice paddies. Its evolutionary history is similar to that of bread wheat, arising through polyploidization after hybridization between a tetraploid and a diploid species. Here we generated and analyzed high quality genome sequences of diploid (*E. haploclada*), tetraploid (*E. oryzicola*) and hexaploid (*E. crus-galli*) *Echinochloa* species. Gene family analysis showed that disease resistance genes such as those containing the NB-ARC domain have been significantly lost during *Echinochloa* polyploidization, which is contrary to significant expansion of those genes during wheat polyploidization. The result suggests that natural selection might favor reduced investment in resistance in the weed to maximize its growth and reproduction. In contrast to the asymmetric patterns of genome evolution shown in wheat and other crops, no significant differences in selection pressure were detected between the subgenomes in *E. oryzicola* and *E. crus-galli*. Additionally, distinctive differences of transcriptomic dynamics in subgenome expression during hexaploidization were observed between *E. crus-galli* and bread wheat. This study documents genomic mechanisms for adaptation during polyploidization in a major agricultural weed and provides insights for crop breeding.

Key words: *Echinochloa* weeds; genome polyploidization; wheat; fitness cost

Introduction

The genus *Echinochloa* (Poaceae) includes numerous problematic weeds that cause reductions in crop yields (Aoki and Yamaguchi, 2008; Michael, 2001; Yabuno, 1966). Among *Echinochloa* species, *E. crus-galli* is the most prevalent weed, occurring widely in both rice paddies and other agricultural fields; its intense competition can reduce tillering in rice up to 50% (Guo et al., 2018; Juraimi et al., 2006). *Echinochloa crus-galli* is hexaploid (2n=6×=54) and is assumed to have arisen from hybridization between the tetraploid *E. oryzicola* (2n=4×=36) and an unknown diploid species (2n=2×=18) (Aoki and Yamaguchi, 2008; Yabuno, 1966). Compared to *E. crus-galli*, which occurs worldwide, the inferred progenitor species of *E. crus-galli* have limited geographical distributions. So far, the tetraploid *E. oryzicola* (also called *E. phyllopogon*, *E. crus-galli* var. *oryzicola*) has only been found in paddy fields, and diploid *Echinochloa* species are only found in Africa (Yabuno, 1983).

Polyploidy is very common in plants, with documented instances preceding the diversification of seed plants and the origin of angiosperms (Amborella Genome Project, 2013; Jiao et al., 2011), and whole genome duplication (WGD) events have occurred continuously in plants (Soltis et al., 2015). Polyploidy can be a major driver of plant species diversification and may play an important role in plant genome evolution (Soltis et al., 2015; Soltis and Soltis, 2009; Van de Peer et al., 2017). In addition, polyploidy may have increased the genetic variability of plant species and their adaptive plasticity (Chao et al., 2013; Freeling et al., 2015; Meimberg et al., 2009; Salman-Minkov et al., 2016; Soltis et al., 2015; te Beest et al., 2012; Wendel, 2015).

Some of our most important crop species are polyploids (Paterson and Wendel, 2015), such as cotton (tetraploid), oilseed rape (tetraploid) and bread wheat (hexaploid). In the case of bread wheat (*Triticum aestivum*), the crop is hexaploid with three subgenomes (i.e. A, B and D; genome formula AABBDD), which originated from hybridization between cultivated tetraploid emmer wheat (AABB, *T. dicoccoides*) and diploid goat grass (DD, *Aegilops tauschii*) approximately 8,000 years ago (Brenchley

et al., 2012). The allohexaploid genome of bread wheat could, in part, underly its ability to grow in diverse climates (IWGSC, 2014). With the availability of genome sequences of hexaploid and tetraploid wheat and two diploid progenitors, genome evolution and its contribution to agronomy traits during polyploidization have been partly revealed (Avni et al., 2017; Brenchley et al., 2012; IWGSC, 2014; IWGSC, 2018; Ling et al., 2018; Luo et al., 2017; Maccaferri et al., 2019; Zhao et al., 2017).

Currently, few polyploid agricultural weeds have been examined at the whole genome sequence level. The draft genome sequence of hexaploid *E. crus-galli* was previously reported by us (Guo et al., 2017). As a dominant weed species, its wide distribution and adaptation to a wide range of environments may be attributable in part to its polyploid genome. Like bread wheat and other allohexaploids, *E. crus-galli* is the result of hybridization of an allotetraploid with a diploid species. Comparison of the *E. crus-galli* and wheat genomes could help to elucidate the evolutionary consequences of hexaploidization, including genetic mechanisms underlying environmental adaptation. However, the extent to which such mechanisms are shared in the genomes of these two hexaploid grass species is unknown.

To examine the mechanisms of adaptation in *E. crus-galli* and patterns of genome evolution during polyploidization, we generated genome sequences of its progenitor *E. oryzicola* (tetraploid) and a diploid species (*E. haploclada*), which we have analyzed along with a significantly improved *E. crus-galli* genome sequence assembled from third-generation long reads. Gene family and subgenome evolution were further investigated, and the genomic and transcriptomic changes during hexaploidization were compared between *E. crus-galli* and bread wheat. The results illustrate different patterns of genome evolution during polyploidization of the agricultural weed and a crop.

Results

Sequencing, assembly and annotation of di-, tetra- and hexaploid *Echinochloa* genomes

We sequenced the genomes of *E. oryzicola* (accession ZJU2), the tetraploid progenitor of hexaploid *E. crus-galli*, and a diploid species *E. haploclada* (accession Pasquet 1083) and generated new data to improve the previous assembly of *E. crus-galli* (accession STB08) (Figure 1A; Table 1). Hexaploid *E. crus-galli* and tetraploid *E. oryzicola* have genome sizes of 1.4 and 1.0 Gb, respectively (Guo et al., 2017); the genome size of *E. haploclada* was estimated to be approximately 420 Mb based on *K*-mer analysis and flow cytometry (Supplementary Figure 1). These genome size estimates are consistent with their ploidy.

For the *E. haploclada* genome, we generated 92× PacBio long reads and 79× Illumina paired-end reads and assembled these reads into 1848 contigs, producing a ~440 Mb genome with a contig N50 length of 0.93 Mb (Table 1). The assembled genome size is slightly larger than the estimated size; this may be due to its high heterozygosity rate of ~1.8% and cases where allelic SNPs were misassembled as different loci (Supplementary Figure 1). Additionally, ~68 million valid Hi-C interacting unique pairs were generated successfully. With the aid of Hi-C sequence data, we anchored the assembly onto nine pseudochromosomes and obtained a nearly chromosome-level reference genome (scaffold N50 size = 48.75 Mb) (Table 1; Supplementary Figure 2). The percentage of conserved genes retrieved by the assembly (BUSCO score, v. 2) is 97.1%, which is comparable to that of other sequenced grass species. The *Echinochloa* genome shows high synteny with *Setaria italica*, a member of the same tribe as *Echinochloa* (Paniceae), further supporting our good assembly quality (Figure 1B). After genome annotation, 36,949 genes were predicted in the *E. haploclada* genome. Putative orthologs and paralogs were analyzed among *E. haploclada* and four other grass family members, including *Oryza sativa*, *Sorghum bicolor*, *S. italica* and *Zea mays*. We found 20 gene families containing 113 genes that appear to be *E. haploclada*–specific (Supplementary Figure 3).

For the genome of tetraploid *E. oryzicola*, we generated 63× PacBio and 120× Illumina reads. *De novo* assembly yielded a draft genome of 950 Mb, representing 92.8% of the *E. oryzicola* genome, with contig and scaffold N50 length of 1.87 Mb and 2.93 Mb, respectively (Table 1). A total of 66,521 protein-coding genes were predicted in the *E. oryzicola* genome (Table 1). Additionally, to improve our previous genome assembly of hexaploid *E. crus-galli* (Guo et al., 2017) we generated PacBio long reads representing ~86× coverage of the genome. The new assembly was dramatically improved, and its contig and scaffold N50 sizes were increased from 26 Kb and 1.80 Mb, respectively, in the previous version to 1.57 Mb and 4.09 Mb (Table 1). A total of 103,853 protein-coding genes were annotated in the new *E. crus-galli* genome assembly.

To further assess the assembly quality of the three *Echinochloa* species, we firstly mapped Illumina reads from paired-end libraries with 300 bp insertion size to each assembly. The results showed that very high percentage (93.5%, 98.5% and 98.0% in diploid, tetraploid and hexaploid *Echinochloa* genomes, respectively) of sequencing reads could be successfully mapped and observed insertion size is very close to expected size (Supplementary Table 1). Second, normal ratios of RNA-seq reads matching our assemblies were observed (87.6% for *E. oryzicola* and 84.6% for *E. crus-galli*), with a relatively lower value (72.2%) for *E. haploclada*, which was expected due to its high heterozygosity rate (Supplementary Tables 2-4). Third, LAI (LTR Assembly Index) was used to evaluate assembly continuity using LTR (long terminal repeat) elements (Ou et al., 2018). The results showed 24.38, 19.80 and 18.60 of LAI scores for *E. haploclada*, *E. oryzicola* and *E. crus-galli*, respectively, which are comparable to those of *O. sativa* (MSUv7) and Arabidopsis (TAIR10) (Supplementary Figure 4). Additionally, five publicly available sequences of fosmid clones generated by our previous study (Guo et al., 2017) were aligned to *E. crus-galli* assembly and the results showed high consistency (Supplementary Table 5). Taken together the results demonstrated high-quality assemblies of the three *Echinochloa* species.

When contigs from the tetraploid and hexaploid genomes were aligned to the nine pseudochromosomes of *E. haploclada*, two and three copies were evident for most genes, respectively, consistent with their ploidy (Figure 1C). Distribution of gene length also showed that the numbers of annotated genes generally have a ratio of 3:2:1 in the three *Echinochloa* genomes (Supplementary Figure 5). Repeat elements were identified in the three genomes, with relatively high levels in *E. oryzicola* (approximately 52% of the genome) and lower levels in *E. haploclada* (approximately 41% of the genome) (Table 1). Among the repetitive sequences, long terminal retrotransposons (LTRs) were the most abundant in all the three *Echinochloa* genomes, which is similar to other cereals (Supplementary Table 6).

Evolution of the *Echinochloa* genomes

We calculated the *Ks* value of orthologous gene pairs between diploid *E. haploclada* and five other grass species to estimate their divergence times. The *Ks* peak of *E. haploclada* and *S. italica* orthologous gene pairs has a value of 0.193, corresponding to an estimated divergence time of approximately 14.9 million years ago (mya) (Figure 2; Supplementary Figure 6) , which is similar to our previous result based on single copy genies in *E. crus-galli* and other grasses using a phylogenetic approach (Guo et al., 2017).

To estimate the *Echinochloa* speciation and subgenome divergence times, we first performed subgenome separation in tetraploid *E. oryzicola* and hexaploid *E. crus-galli*. Based on genomic synteny and genome similarity to the diploid *E. haploclada* (see details in Methods), we divided the *E. oryzicola* genome into its two component subgenomes: A_T (where "A" indicates the A *Echinochloa* subgenome and the subscript "T" indicates tetraploid species), containing 430.2 Mb genomic regions, and B_T, containing 378.7 Mb; together these cover 77.8% of the *E. oryzicola* assembly (Supplementary Figure 7). The A_T subgenome showed substantially lower genomic similarity with the *E. haploclada* genome than B_T (Supplementary Figure 8); this difference suggests an allopolyploid origin of *E. oryzicola*. In the same way, the

hexaploid *E. crus-galli* genome was divided into three subgenomes: A_H (399.8 Mb), B_H (367.9 Mb) and C_H (395.5 Mb), which together cover 83.1% of its genome assembly (Supplementary Figure 7). The subgenome C_H has significantly lower similarity with the tetraploid progenitor *E. oryzicola* genome than the other two subgenomes (Supplementary Figure 9).

To further examine divergence dates of *Echinochloa* subgenomes, *K*s values for homeologous gene pairs between subgenomes were calculated (Figure 2; Supplementary Figure 10). According to the *K*s peak, we dated the divergence time of the two subgenomes (A_T and B_T) of *E. oryzicola* at 4.5 mya (Figure 2). *Echinochloa oryzicola* speciation time (i.e., the timing of tetraploid formation) was further estimated based on the distribution of sequence divergence rates of transposable elements (TE) in the two subgenomes following the method suggested by Xu et al. (2019) (for details see Methods). The "bubble" peak in the TE divergence profile suggested the *E. oryzicola* speciation time at 1.9 mya (Supplementary Figure 11). Further, the divergence time of A_T from tetraploid *E. oryzicola* and A_H from hexaploid *E. crus-galli* was estimated at 0.31 mya, as well as that of the B_T and B_H; this indicates a recent formation of *E. crus-galli* (Figure 2; Supplementary Figure 10). Additionally, we found that our sequenced diploid *E. haploclada* genome is similar to the unknown diploid progenitor genome (C_H) of *E. crus-galli* with their divergence time estimated at 1.2 mya (Figure 2). The reads from diploid *E. haploclada* and tetraploid *E. oryzicola* aligned well to the hexaploid *E. crus-galli* genome, and appeared to match complementary subgenomes (Supplementary Figure 12). Accordingly, based on this similarity, diploid *E. haploclada* was used as a proxy for the progenitor genome in the analyses described below.

Mass loss of disease-resistance genes in *Echinochloa* genomes

Gene loss and gain are common during polyploidization. Gene family sizes were determined by protein domains in diploid, tetraploid and hexaploid *Echinochloa* and other grass genomes. We first compared gene family sizes between hexaploid *E.*

crus-galli and *O. sativa* / *S. italica* genomes using a dot matrix (Figure 3A; Supplementary Figure 13). The results showed that for the majority of gene families in *E. crus-galli*, gene family sizes are almost three times those of *O. sativa* and *S. italica*, consistent with *E. crus-galli* being a hexaploid and *O. sativa* and *S. italica* being diploids (Figure 3A; Supplementary Figure 13). This analysis also revealed that sizes of several gene families in *E. crus-galli* are much less than three times those in *O. sativa* and *S. italica*; these exceptions are predominantly genes with protein domains associated with disease resistance, e.g. NB-ARC genes (471 in *E. crus-galli* versus 469 and 400 in *O. sativa* and *S. italica*, respectively), D-mannose lectin genes (122 versus 129 and 122, respectively) and legume lectin genes (50 versus 91 and 61, respectively) (Lannoo and Van Damme, 2014; Meyers et al., 2005) (Figure 3A and 3B). Phylogenetic trees of NB-ARC genes in *Echinochloa* and *S. italica* genomes confirmed the occurrence of many *S. italica*-specific genes (Supplementary Figure 14).

Comparison of gene family sizes among the three *Echinochloa* species revealed that the sizes of most gene families in the hexaploid *E. crus-galli* are almost the same as the sum of the gene families in the diploid and tetraploid species (Figure 3A). This close correlation suggests that little gene loss has occurred for most gene families in tetraploid *E. oryzicola* and hexaploid *E. crus-galli* after polyploidization. Gene families showing this additive pattern include AP2 and GRAS domain proteins involved in abiotic stresses and cytochrome P450s and glutathione S-transferases associated with detoxification or non-target-site resistance to synthetic herbicides (Yu and Powles, 2014) (Figure 3B; Supplementary Figure 15). In contrast, NB-ARC disease-resistance genes were significantly lost in the tetraploid *E. oryzicola* after polyploidization (277 in *E. oryzicola* versus 240 in *E. haploclada*; $P < 0.0001$, Fisher's exact test) (Figure 3B). Similarly, in *E. crus-galli*, numbers of NB-ARC disease-resistance genes (471) are still less than the sum of those present in *E. oryzicola* and *E. haploclada* (517).

To further confirm this loss, we calculated synteny retention ratios of diploid *E.*

haploclada genes in tetraploid *E. oryzicola* and hexaploid *E. crus-galli* (i.e., within a gene family, the percentage of gene members keeping 1:(1:1):(1:1:1) syntenic relationship among six genomes/subgenomes including *E. haploclada*, A_T and B_T from tetraploid *E. oryzicola*, and A_H, B_H and C_H from hexaploid *E. crus-galli*) (Supplementary Figure 16). Across the genome, 46.4% of genes fit the 1:(1:1):(1:1:1) synteny retention ratio. In contrast, the synteny retention ratio for NB-ARC family genes (12.9%) is significantly lower ($P < 0.0001$, Fisher's exact test); this is also true for another well-known disease-resistance gene family, the wall-associated receptor kinases (18.3%; $P < 0.0001$) (Hurni et al., 2015) (Figure 3C). An example of a NB-ARC gene that deviates from the 1:(1:1):(1:1:1) syntenty retention ratio is illustrated in Figure 3d; in relation to the *E. haploclada* gene, only one homeologous copy is retained in *E. oryzicola* (within subgenome A_T), and in hexaploid *E. crus-galli* only the copy in subgenome C_H is retained (Figure 3D). In contrast, a GRAS gene in the same chromosomal region as the NB-ARC gene conforms to the 1:(1:1):(1:1:1) synteny retention ratio (Figure 3D).

To compare the patterns of gene family retention and loss between *Echinochloa* and wheat, we examined the wheat genomes using the same bioinformatic pipeline. Abiotic stress-related gene families such as AP2 and GRAS in the bread wheat genome have almost the same number of members as the sum of its two progenitor genomes. This pattern is similar to those in *E. crus-galli*, suggesting no obvious gene loss in both hexaploid species after hexaploidization (Figure 3B). Interestingly, for biotic stress related gene families, we found expansion in bread wheat after hexaploidization, e.g. many more NB-ARC domain proteins in bread wheat (a total of 2,210 genes) than the sum of members in tetraploid cultivated wheat (1,222) and diploid goat grass (501) ($P < 0.0001$, Fisher's exact test). This pattern differs from that of *E. crus-galli* where there has been gene loss in this family after polyploidization (Figure 3B). Additionally, we also found a significant expansion ($P < 0.05$, Fisher's exact test) of the NB-ARC gene family in cultivated tetraploid wheat relative to wild tetraploid wheat (Supplementary Figure 17).

Symmetric selection on subgenomes during *Echinochloa* polyploidization

To assess the selective pressure on subgenomes during *Echinochloa* polyploidization, we calculated *K*a/*K*s ratios of homeologous gene pairs. The ratio was not significantly different between the two subgenomes (A_T and B_T) in *E. oryzicola* (Figure 4A and 4B), and those two subgenomes (A_H and B_H) were still not different after hexaploidization (Figure 4A and 4B). To further confirm this observation, we calculated the pN/pS ratio based on SNPs in *E. crus-galli* populations collected in our previous work (Ye et al., 2019) (details in Methods), and also found no significant differences between the two subgenomes (Figure 4B). Taken together, these results suggest symmetric evolution of subgenomes under similar selection pressure during *Echinochloa* polyploidization.

The bread wheat genome evolved through a similar polyploidization process to *E. crus-galli* (Figure 4C). Therefore, the *K*a/*K*s ratio for subgenomes A and B in wild and cultivated tetraploid and hexaploid wheat were also examined using the same methods as for *Echinochloa*. These results showed that wheat subgenome A had a significantly higher ratio than subgenome B in both cultivated tetraploid and hexaploid wheat ($P < 0.0001$, Kruskal-Wallis test). However, no significant difference of *K*a/*K*s ratio between subgenomes A and B in wild tetraploid wheat was observed, showing the same lack of differentiation as in *Echinochloa* (Figure 4D). The results clearly demonstrated an asymmetric evolutionary pattern of subgenome evolution in hexaploid cultivated wheat but not *Echinochloa*.

Subgenome expression changes during *Echinochloa* hexaploidization

Homeologous triads (A_H, B_H and C_H homoeologs) in *E. crus-galli* were identified, and deviations from balanced expression characteristics were further analyzed following the approach used in bread wheat by Ramirez-Gonzalez et al. (2018) (Figure 5A). Genes with homeolog-specific dominant and suppressed expression accounted for 8.9 - 19.1% and 27.6 - 33.7% of all examined loci, respectively, in *E. crus-galli* (Figure

5A; Supplementary Table 7). For genes with expression dominance, no significant differences in frequency were found among three subgenomes in *E. crus-galli*, consistent with findings in bread wheat (Figure 5B; Supplementary Figure 18). For suppressed genes, a significantly higher proportion occurred in *E. crus-galli* subgenome A_H (donated by the tetraploid progenitor) than in B_H and C_H ($P < 0.0001$, paired *t*-test). Similarly, in bread wheat we observed that suppressed genes were significantly underrepresented in subgenome D (donated by the diploid progenitor) relative to subgenomes A ($P = 0.017$, paired *t*-test) and B ($P = 0.0002$, paired *t*-test) (Figure 5B; Supplementary Figure 18). Significantly more *Echinochloa* genes with dominant and suppressed expression were found in leaves than in roots, demonstrating tissue differences for subgenome expression patterns ($P < 0.01$, *t* test; Supplementary Table 7). This pattern is consistent with that of bread wheat (Supplementary Table 8). When investigating purifying selection on genes based on *K*a/*K*s ratios, we found that, for a given subgenome, dominant genes are under greater functional constraint whereas their homoeologous counterparts show evidence of relaxed selection (Supplementary Figure 19).

The transcriptomic profiles of *E. crus-galli* under biotic stress (infection with the blast *Pyricularia oryzae* and allelopathy by co-cultivation with rice) and abiotic stress (drought) were investigated. After stress treatments, three homeologous genes from each of approximately 50% of the triads in *E. crus-galli* showed balanced expression changes (either up-regulation or down-regulation for all the three genes in a triad, termed "consistent" expression change) (Supplementary Table 9). The ratio of genes with consistent expression change from subgenomes B_H and C_H (i.e. B_H/C_H consistent genes) was significantly higher than A_H/B_H and A_H/C_H consistent genes under three stress conditions ($P < 0.05$; paired *t*-test) (Supplementary Table 9). In comparison, bread wheat showed no significant differences among the three subgenomes' consistent genes (Supplementary Table 10). Most triads exhibited the same dominance status after stress treatments (79.6% - 86.9%) and only a few triads (0.66% - 1.10%) changed to the opposite status after treatment of different stresses (Figure 5A;

Supplementary Figures 20 and 21). To further estimate the influence of stress on subgenome expression in *E. crus-galli*, we calculated the relative contribution of a subgenome responsive to stress (termed response dominance of subgenome), which indicated that expression change of a gene in the subgenome is more dramatic than its homeologous genes in other subgenomes after stress treatments (details in Methods). Our results showed no significant differences for response dominance among subgenomes under all three kinds of stress treatments, which is similar to bread wheat (Supplementary Tables 11 and 12).

Discussion

In this study, we present high-quality genome sequences of diploid, tetraploid and hexaploid *Echinochloa* species and explore the genomic and transcriptomic evolution during polyploidization of the notorious agricultural weed *E. crus-galli*. The analogous hexaploidization processes between *E. crus-galli* and bread wheat led us to compare their genomic and transcriptomic responses to polyploidization. Our results show that *Echinochloa* weeds apparently acquired only a limited number of disease resistance genes and maintained largely symmetric selection on subgenomes. Additionally, we found a number of distinctive subgenome expression patterns.

Significant expansion of one set of disease resistance genes (NB-ARC genes) was found in cultivated wheat (IWGSC, 2018) and also confirmed in our reanalysis of the wheat genome (Figure 3B; Supplementary Figure 17). This was expected and may be the result of artificial selection, as disease resistance is a major crop breeding objective. However, disease resistance has well known fitness costs for energetic investment in growth and reproduction (Brown and Rant, 2013; Karasov et al., 2017; Kliebenstein, 2016; Nelson et al., 2018). The pioneering research on costs of resistance was related to resistance of potato to late blight (*Phytophthora infestans*) (Vanderplank, 1963). More research indeed revealed the costs of resistance for many genes in different plants; for example, the widely deployed *mlo* powdery mildew resistance gene in barley is associated with necrotic flecking and yield loss (Bergelson

and Purrington, 1996; Brown and Rant, 2013). In diploid *E. haploclada*, we unexpectedly found that resistance genes are much less than in *S. italica* and even through hexaploidization, the number of these genes (particularly NB-ARC genes) in *E. crus-galli* are close to or less than diploid *O. sativa* (Figure 3B). This pattern suggests that natural selection might favor reduced investment in resistance in the weed, which is then beneficial for maximizing growth and reproduction. Unlike a crop with strong resistance as a necessary agronomic trait, rapid growth and massive reproduction may be necessary for weediness. The result further implies that we may need to re-think our strategies for crop breeding programs (Weiner, 2019). For example, weed-like crop cultivars with stronger adaptation ability may be able to balance yield and environmental cost (input of chemicals).

Subgenome dominance is common in allopolyploids, in which one of the parental subgenomes often exhibits stronger purifying selection and significantly higher expression than those of the other subgenomes (Cheng et al., 2018; Edger et al., 2018; Yang et al., 2016; Zhang et al., 2015). We indeed also found significant differences in selection pressures between subgenomes A and B in both tetraploid and hexaploid cultivated wheat (Figure 4). However, this phenomenon was not observed in *Echinochloa* weeds and wild tetraploid wheat. Different selection forces (i.e. artificial and natural selection) might be one of the reasons causing the differences. The artificial selection may have imposed stronger pressure on target traits than previously thought, resulting in asymmetric evolution on subgenomes in a short time, similar to asymmetric domestication selection for fiber length observed in cotton (Wang et al., 2017).

Analysis of subgenome expression showed some common characteristics during polyploidization between the two hexaploids, such as (1) similar percentages of dominant and suppressed genes; (2) tissue differences for dominance expression; (3) more and less relaxed selection on dominant and suppressed genes, respectively; (4) no significant differences for response dominance among subgenomes under stress treatments; (5) changes of dominance status of triads under stress conditions.

Meanwhile, several findings of this study show novel subgenome expression patterns in the *Echinochloa* weeds. First, we found B_H/C_H consistent genes (with consistent expression change trend between genes in two subgenomes after stress) are significantly more common than those of A_H/C_H and B_H/C_H in *E. crus-galli* (Supplementary Table 9); this is expected because of that B_H and C_H are more recently diverged (Figure 2). However, no significant differences were found among A/B, A/D and B/D consistent genes, although B and D diverged more recently in bread wheat (Supplementary Table 10). Second, in *E. crus-galli*, a significantly higher proportion of suppressed genes occurred in subgenome A_H (donated by the tetraploid progenitor) than in B_H and C_H (Figure 5B). In hexaploid bread wheat, however, we observed significantly fewer suppressed genes in subgenome D (donated by the diploid progenitor). One possible reason is that subgenome D is critical for formation of important traits of bread wheat such as quality and disease resistance (Luo et al., 2017). Additionally, we found that expression patterns of the subgenomes of *E. crus-galli* are still similar to their ancestor progenitor genomes (i.e. A_H/B_H together with the tetraploid genome and C_H together with the diploid genome) based on clustering of the expression patterns of each subgenome (Supplementary file). In contrast, expression patterns of the subgenomes are similar to each other within a species in wheat after hexaploidization process (Supplementary file). We cannot provide the exact reasons currently, but artificial selection might be one possible factor shaping transcriptomic pattern in wheat but not in the weed and causing the differences between them.

Methods

Genome sequencing and assembly

The voucher specimens for the three sequenced species were deposited in the Herbarium of Zhejiang University (HZU), with herbarium accession No. HZU60206925 for *E. crus-galli* (STB08), HZU60206923 for *E. oryzicola* (ZJU2) and HZU60206921 for *E. haploclada* (Pasquet 1083).

Genomic DNA of *E. crus-galli* (STB08) was extracted from young leaves for sequencing library construction using CTAB method. DNA libraries for single molecule real-time (SMRT) PacBio genome sequencing were constructed following the standard protocols of the Pacific Biosciences company and sequenced on PacBio Sequel platforms (Pacific Biosciences). Raw reads of the *E. crus-galli* genome were corrected by using an error correction module embedded in Canu (version 1.8), and high-quality sub-reads were used for assembly by using Canu with parameter '-pacbio-raw minReadLength=2000' (Koren et al., 2017). Clean PabBio sub-reads were mapped to assembled contigs with minimap2 (Li, 2018). Consensus sequences were constructed and assembled contigs were corrected using racon (v1.4.0) (Vaser et al., 2017). Assembly was improved according the following steps for three rounds: (1) Assembled contigs were corrected by Illumina paired-end and mate-pair reads generated previously using Pilon (version 1.23) (Guo et al., 2017; Walker et al., 2014). (2) Corrected contigs were scaffolded with OPERA-LG (version 2.0.5) (Gao et al., 2016). (3) Gaps within scaffolds were filled with Illumina pair-end reads by using GapFiller (version 1.10) and clean PacBio sub-reads by using PBjelly incorporated in PBSuite (version 15.8.24) (Boetzer and Pirovano, 2012; English et al., 2014).

Genomic DNA of *E. oryzicola* (ZJU2) was extracted from young leaves for sequencing library construction in the CTAB method. For Illumina sequencing, paired-end and mate-pair libraries (insertion size ranges from 300 bp to 20 kb) were constructed following the manufacturer's instructions (Illumina, USA). PacBio long reads were generated, corrected and assembled into contigs using the same procedures for STB08 mentioned above. Corrected PacBio long sub-reads were further aligned to assembled contigs to generate consensus sequences by using pbalign (release 0.4.1) (https://github.com/PacificBiosciences/pbalign). Assembled contigs were improved by consensus sequences using arrow incorporated in SMRT Link (version 5.1) (www.pacb.com). Assembly was further improved for two rounds using the same procedure as for STB08 assembly improvement mentioned above.

The diploid *E. haploclada* (Pasquet 1083) was collected on 05 July 2011 by Rémy

Pasquet in coastal forest in Kenya, near Muhaka, 04°20.201 S 39°28.137 E. Taxonomic determination was performed by Elizabeth A. Kellogg; the voucher specimen is housed at the Missouri Botanical Garden (MO), and details are available at the following Tropicos URL: http://legacy.tropicos.org/Specimen/100668610. Genomic DNA extraction of *E. haploclada* and PacBio sequencing library construction followed procedures described previously. A Hi-C library was constructed using fresh young leaves of *E. haploclada* for pseudomolecule construction. The Hi-C experiments and sequencing procedures were similar to those described previously in cotton (Wang et al., 2017). PacBio long sub-reads were corrected and assembled into contigs using Canu (-pacbio-raw minReadLength=2000) and improved by corrected PacBio sub-reads by using arrow incorporated in SMRT Link and Illumina paired-end reads with Pilon. Improved contigs were further rebuilt into two sub-assemblies (ref and alt) with HaploMerger2 (Huang et al., 2017). Based on ref sub-assembly, clean Hi-C reads were analysed by using Juicer (version 1.6.2) and then 3D-DNA was used to scaffold contigs into pseudomolecule (Dudchenko et al., 2017; Durand et al., 2016). Additionally, the method *optimize* in AllHiC was also used to order and orientate contigs in each pseudo-chromosme cluster (Zhang et al., 2019). Based on the consistency between 3D-DNA and AllHiC assemblies and synteny to *S. italica*, we manually corrected some errors with discrete chromatin interaction patterns. BUSCO (version 2) was used to evaluate the completeness of assembled *Echinochloa* genomes (Simao et al., 2015).

Genome annotation

Repeat sequences of the three *Echinochloa* genomes were annotated using methods described previously (Guo et al., 2017).

For annotation of protein coding genes, the pipeline described previously (Guo et al., 2017) was adopted except for filtration for final gene sets. Briefly, a hybrid strategy combining ab initio predictions (i.e. GeneMark.hmm (Lukashin and Borodovsky, 1998), Fgenesh (Salamov and Solovyev, 2000) and Augustus (version 3.2.2) (Stanke et al., 2006)), homologous gene evidence and transcriptomic support (RNA-seq) was

applied in gene prediction. All gene structures predicted were integrated into consensus gene models by EVidenceModeler (version 1.1.1) (Haas et al., 2008). Based on EVidenceModeler integrated results, high confidence gene models were identified as those supported by homologous genes or transcript evidences or by at least two ab initio methods, and the remaining gene models were categorized as low confidence. High confidence gene models were further filtered to remove short gene models (less than 50 amino acids) and gene models with homology to sequence in the Repbase (E-value ≤ 1e-5, identity ≥ 30%, coverage ≥ 25%).

Subgenome construction

We employed a reference-guided method based on subgenome homology to construct subgenomes in *E. oryzicola* and *E. crus-galli*. For separation of two subgenomes in tetraploid *E. oryzicola*, the genes in contigs were used as markers and coding proteins were aligned mutually by BlastP with E-value cutoff of 1e-10 and DAGchainer (Haas et al., 2004) was used to identify syntenic blocks between contigs with at least 3 homologous gene pairs within 10 gene models. Two contigs whose marginal regions showed synteny to the left and right end of another contig would be linked and then formed homeologous scaffold pairs (Supplementary Figure 22B). Each homeologous scaffold pair was mapped to the diploid *E. haploclada* genome to place and orient on chromosomes (Supplementary Figure 22C). Subsequently, *E. haploclada* genome sequences were split into 100-mers, and these short reads were mapped to homeologous scaffold pairs of tetraploid *E. oryzicola*. Distinctly higher coverage depth was observed for one scaffold versus the other one in each of the homeologous pairs (Supplementary Figure 8). We then linked homeologous pairs according to depth and classified these scaffolds with higher and lower depth to two different subgenomes, respectively (Supplementary Figure 22D and 22E). For subgenome separation in hexaploid *E. crus-galli*, a similar approach was adopted. First, syntenic contig triads were identified and linked according to synteny to *E. haploclada* (Supplementary Figure 23B). Then tetraploid *E. oryzicola* genome sequences were split into 100-mers, and these reads were mapped to homeologous scaffold triads.

Distinctly higher coverage depth was found for two scaffolds of each homeologous triad as *E. oryzicola* is the progenitor of *E. crus-galli* (Aoki and Yamaguchi, 2008). Another scaffold of the triads with much lower overage depth was then linked and designated as the subgenome C_H of *E. crus-galli* (Supplementary Figure 23C). The scaffolds with high coverage depth were mapped to the two above identified subgenomes of *E. oryzicola* according to gene similarity and then these scaffolds were linked and divided into two subgenomes A_H and B_H of hexaploid *E. crus-galli* (Supplementary Figure 23D).

Divergence time estimation

Phylogenetic trees among *E. haploclada* and five other species (*S. italica*, *Z. mays*, *Sorghum bicolor*, *Brachypodium distachyon* and *O. sativa*) were built with RAxmL (Stamatakis, 2014) using 3,142 shared single-copy genes identified by OrthoFinder (Emms and Kelly, 2015) and visualized in iTOL (itol.embl.de) (Letunic and Bork, 2016). Protein sequences were aligned in MAFFT (version 7.310) (Katoh and Standley, 2013) under default parameters. We first aligned protein sequences manually among species or subgenomes by using BLASTP. DAGchainer was employed to determine syntenic blocks with at least 3 homologous gene pairs within a region of 10 gene models based on best-hit BlastP results. Estimation of divergence times was based on the non-synonymous substitution *K*s calculation of syntenic homologous gene pairs with the formula $T = Ks/2r$, where r is the rate of substitution. We used the nucleotide substitution rate of 6.5×10^{-9} mutations$\times$bp$^{-1}\times$generation^{-1} as a molecular clock (Molina et al., 2011). Kaks_calculator was used to calculate *K*s with the model of NG (Zhang et al., 2006). Transposable elements divergence was assessed by PercDivs (Percentage of substitutions in the matching region compared to the consensus) calculated in RepeatMasker. TE sequence divergence between two subgenomes of tetraploid *E. oryzicola* displaying a high degree of overlap suggested the consistency of TE evolutionary rate in two subgenomes (Supplementary Figure 11). The non-overlapped segregation region indicates the time frame from diploid progenitor divergence to genomes merging as tetraploid genome (Xu et al., 2019).

Gene family identification

Gene families were identified based on Pfam protein domains, which were identified by InterProScan (version 5.24-63.0) (Zdobnov and Apweiler, 2001). Besides *Echinochloa* genes annotated in this study, the others include genes from *S. italica* (v2.0) (Bennetzen et al., 2012), *S. bicolor* (v3.1.1) (Paterson et al., 2009), *O. sativa* (MSU v7.0) (Matsumoto et al., 2005), *B. distachyon* (v3.0) (Vogel et al., 2010), *A. tauschii* (Aet_v4.0) (Jia et al., 2013), *T. turgidum* (WEWSeq_v1.0 for wild emmer and Svevo.v1 for durum wheat) (Avni et al., 2017; Maccaferri et al., 2019) and *T. aestivum* (iwgsc_refseqv1.0) (IWGSC, 2014). Phylogenetic tree (Maximum likelihood) of NB-ARC genes was constructed using FastTree (Price et al., 2009).

Selection pressure analysis

*K*a/*K*s ratio of homologous gene pairs with those of *K*s < 0.001 excluded was calculated by Kaks_calculator. To assess the selection pressure of genes in a population level sample, pN/pS ratio was calculated based on whole-genome SNPs (Hao et al., 2018). pN is the number of nonsynonymous positions in each gene showing polymorphism divided by the total number of nonsynonymous changes possible in the gene. pS is the number of polymorphic synonymous positions divided by the number of possible synonymous sites. 30 *E. crus-galli* lines with genome re-sequencing data were selected for pN/pS calculation (Ye et al., 2019). SNP calling was performed using the method described previously with an exception that the new STB08 genome assembly was used (Ye et al., 2019).

Transcriptome sequencing and analysis

For drought stress treatment, the seedlings of three *Echinochloa* species at tillering stage were hydroponically cultivated and treated with 10% PEG6000 for 24 hours. The blast (*Pyricularia oryzae*) infection treatment on detached leaves of three *Echinochloa* species was similar to that described previously (Guo et al., 2017). Total RNA from roots and/or leaves was extracted for RNA sequencing. A total of three biological replicates for each treatment or control were performed. The transcriptomic

data of *E. crus-galli* co-cultured with rice (i.e. allelopathy treatment) were generated previously (Guo et al., 2017).

Illumina RNA-Seq libraries were prepared and sequenced on a HiSeq 4000 system following the manufacturer's instructions. Clean data were obtained by using NGSQCtoolkit (v2.3.3) to filter out low quality reads under default parameters. Reads were aligned to the *Echinochloa* genomes using Tophat (v2.0.9) (Trapnell et al., 2009). Expression values of genes were quantified by FPKM (fragments per kilobase of exon per million fragments mapped for each predicted transcript) using the Cufflinks toolkit (v2.2.1) and Cuffdiff was employed to do differential expression analysis (Trapnell et al., 2010).

Gene triads in hexaploid *E. crus-galli* were identified according to synteny among three homeologous subgenomes. Only triads with a summed expression of all three homeologs > 0.5 FPKM were kept for downstream analyses. Subgenome expression bias was determined following a method implemented in bread wheat RNA-seq analyses (Ramirez-Gonzalez et al., 2018). Briefly, we normalized the expression for each gene within the triad by calculating relative expression (relative expression equals to FPKM of each homeolog divided by summed FPKM of its corresponding triad). The relative expression of each homeolog within each triad determined the triad's position in the ternary plot for the global analysis. Seven expression bias categories (a balanced category with similar relative abundance of transcripts from the three homeologs, 3 dominant and 3 suppressed categories, classified on the basis of the higher and lower abundance of transcripts from a single homeolog with respect to those from the other two) were defined. We then calculated the Euclidean distance of the observed expression status of each triad and ideal expression status (see details in Supplementary Figure 24) of seven categories and expression bias category was assigned by the shortest distance for each triad (Supplementary Figure 24).

To further explore expression bias changes under stress, we measured the expression level changes of three homeologs within a triad. Triads with the same expression level change (up-regulated or down-regulated) of three homeologs were defined as

consistent response triad, and when response patterns of two homeologs were similar but different from the other homeolog, then we defined it as AB consistent, BC consistent or AC consistent response triads. For these triads with the same trends of response to stress, we calculated relative contribution (RC) of stress response as the following formula.

$$\mathrm{RC\,(Relative\,Contribution)} = \frac{\text{change of gene expression in one subgenome}}{\text{sum of gene expression change in all subgenomes}}$$

Similar to determining expression bias (Supplementary Figure 24), response bias categories were also assigned for each convergent response triad.

Wheat genomic data analysis

The same analysis pipelines for *Echinochloa* species were used in wheat omic data. These analyses including gene family identification, syntenic gene pair and triad identification, *K*a/*K*s calculation, gene expression quantification and subgenome expression pattern analyses. Genome sequences of *A. tauschii* (Aet_v4.0) and *T. turgidum* (WEWSeq_v1.0 for wild and Svevo.v1 for domesticated emmer) were obtained from Ensembl Plants release 46 (http://plants.ensembl.org). Genome sequence of *T. aestivum* (iwgsc_refseqv1.0) was obtained from Wheat@URGI portal (https://wheat-urgi.versailles.inra.fr/). Wheat transcriptomic data were obtained from previous studies (Supplementary Table 13).

Data availability

The genomic sequence and RNA-seq data included in this study were deposited into the BIG data centre (https://bigd.big.ac.cn/) under accession number PRJCA002334. The *Echinochloa* genome assemblies and the annotations are accessible at http://ibi.zju.edu.cn/RiceWeedomes/Echinochloa/.

Funding

This work was financially supported in part by the National Natural Science Foundation (9143511 and 31901899), Zhejiang Natural Science Foundation (LZ17C130001) and the Fundamental Research Funds for the Central Universities

(2020XZZX001).

Author Contributions

L.F. managed and organized the project. C.Y., D. W., W.T. and L.F. collected and cultivated plants and sequenced genomes. E. A. K. shared the *E. haploclada* material originally collected by collaborator Rémy Pasquet. D. W., C. Y., J. Q., S. L., M. C. and B. J. analyzed the data. L. M., L. J. and D. W. assembled the genomes. X. F. and C. Z. performed the *P. oryzae* infection experiment. L. B., Q. P., L. P., L. W., and L. G. discussed the data. K. M. O. and E. A. K. edited the manuscript. C.Y., D. W. and L. F. wrote the manuscript. All authors read and contributed the manuscript.

Acknowledgments

The authors thank the Jiangsu Collaborative Innovation Center for Modern Crop Production, 111 Project (B17039) and Zhejiang Key Laboratory of Crop Germplasm Resources.

640 **References**

641 Amborella Genome Project (2013). The *Amborella* genome and the evolution of
642 flowering plants. Science 342:1241089.
643 Aoki, D., and Yamaguchi, H. (2008). Genetic relationship between *Echinochloa*
644 *crus-galli* and *Echinochloa oryzicola* accessions inferred from internal
645 transcribed spacer and chloroplast DNA sequences. Weed Biol. Manag.
646 8:233-242.
647 Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I.,
648 Mascher, M., Spannagl, M., Wiebe, K., et al. (2017). Wild emmer genome
649 architecture and diversity elucidate wheat evolution and domestication.
650 Science 357:93-97.
651 Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C.,
652 Estep, M., Feng, L., Vaughn, J.N., Grimwood, J., et al. (2012). Reference
653 genome sequence of the model plant *Setaria*. Nat. Biotechnol. 30:555-561.
654 Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with
655 GapFiller. Genome Biol. 13:R56.
656 Bergelson J, Purrington CB (1996) Surveying patterns in the cost of resistance in
657 plants. Am Nat 148: 536-558.
658 Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M.,
659 McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., et al. (2012). Analysis
660 of the bread wheat genome using whole-genome shotgun sequencing. Nature
661 491:705-710.
662 Brown, J.K.M., and Rant, J.C. (2013). Fitness costs and trade-offs of disease
663 resistance and their consequences for breeding arable crops. Plant Pathol.
664 62:83-95.
665 Chao, D.Y., Dilkes, B., Luo, H., Douglas, A., Yakubova, E., Lahner, B., and Salt,
666 D.E. (2013). Polyploids exhibit higher potassium uptake and salinity tolerance
667 in *Arabidopsis*. Science 341:658-659.
668 Cheng, F., Wu, J., Cai, X., Liang, J.L., Freeling, M., and Wang, X.W. (2018). Gene

669 retention, fractionation and subgenome differences in polyploid plants. Nat.
670 Plants 4:258-268.
671 Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C.,
672 Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo
673 assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length
674 scaffolds. Science 356:92-95.
675 Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S.,
676 and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing
677 loop-resolution Hi-C experiments. Cell Syst. 3:95-98.
678 Edger, P.P., McKain, M.R., Bird, K.A., and VanBuren, R. (2018). Subgenome
679 assignment in allopolyploids: challenges and future directions. Curr. Opin.
680 Plant Biol. 42:76-80.
681 Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole
682 genome comparisons dramatically improves orthogroup inference accuracy.
683 Genome Biol. 16:157.
684 English, A.C., Salerno, W.J., and Reid, J.G. (2014). PBHoney: identifying genomic
685 variants via long-read discordance and interrupted mapping. BMC
686 Bioinformatics 15:180.
687 Freeling, M., Scanlon, M.J., and Fowler, J.E. (2015). Fractionation and
688 subfunctionalization following genome duplications: mechanisms that drive
689 gene content and their consequences. Curr. Opin. Genet. Dev. 35:110-118.
690 Gao, S., Bertrand, D., Chia, B.K., and Nagarajan, N. (2016). OPERA-LG: efficient
691 and exact scaffolding of large, repeat-rich eukaryotic genomes with
692 performance guarantees. Genome Biol. 17:102.
693 Guo, L., Qiu, J., Li, L.F., Lu, B., Olsen, K., and Fan, L. (2018). Genomic clues for
694 crop-weed interactions and evolution. Trends Plant Sci. 23:1102-1115.
695 Guo, L., Qiu, J., Ye, C., Jin, G., Mao, L., Zhang, H., Yang, X., Peng, Q., Wang, Y.,
696 Jia, L., et al. (2017). *Echinochloa crus-galli* genome analysis provides insight
697 into its adaptation and invasiveness as a weed. Nat. Commun. 8:1031.
698 Haas, B.J., Delcher, A.L., Wortman, J.R., and Salzberg, S.L. (2004). DAGchainer: a

699 tool for mining segmental genome duplications and synteny. Bioinformatics
700 20:3643-3646.
701 Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell,
702 C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure
703 annotation using EVidenceModeler and the program to assemble spliced
704 alignments. Genome Biol. 9:R7.
705 Hao, Y., Washburn, J.D., Rosenthal, J., Nielsen, B., Lyons, E., Edger, P.P., Pires,
706 J.C., and Conant, G.C. (2018). Patterns of population variation in two
707 paleopolyploid eudicot lineages suggest that dosage-based selection on
708 homeologs is long-lived. Genome Biol. Evol. 10:999-1011.
709 Huang, S.F., Kang, M.J., and Xu, A.L. (2017). HaploMerger2: rebuilding both
710 haploid sub-assemblies from high-heterozygosity diploid genome assembly.
711 Bioinformatics 33:2577-2579.
712 Hurni, S., Scheuermann, D., Krattinger, S.G., Kessel, B., Wicker, T., Herren, G.,
713 Fitze, M.N., Breen, J., Presterl, T., Ouzunova, M., et al. (2015). The maize
714 disease resistance gene Htn1 against northern corn leaf blight encodes a
715 wall-associated receptor-like kinase. Proc. Natl. Acad. Sci. U S A
716 112:8780-8785.
717 IWGSC. (2014). A chromosome-based draft sequence of the hexaploid bread wheat
718 (*Triticum aestivum*) genome. Science 345:1251788.
719 IWGSC. (2018). Shifting the limits in wheat research and breeding using a fully
720 annotated reference genome. Science 361:eaar7191.
721 Jia, J.Z., Zhao, S.C., Kong, X.Y., Li, Y.R., Zhao, G.Y., He, W.M., Appels, R., Pfeifer,
722 M., Tao, Y., Zhang, X.Y., et al. (2013). *Aegilops tauschii* draft genome
723 sequence reveals a gene repertoire for wheat adaptation. Nature 496:91-95.
724 Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph,
725 P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral
726 polyploidy in seed plants and angiosperms. Nature 473:97-100.
727 Juraimi, A.S., Tasrif, A., Kadir, J., Napis, S., and Sastroutomo, S.S. (2006).
728 Differential susceptibility of barnyard grass (*Echinochloa crus-galli* var.

crus-galli) ecotypes to *Exserohilum longirostratum*. Weed Biol. Manag. 6:125-130.

Karasov, T.L., Chae, E., Herman, J.J., and Bergelson, J. (2017). Mechanisms to mitigate the trade-off between growth and defense. Plant Cell 29:666-680.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772-780.

Kliebenstein, D.J. (2016). False idolatry of the mythical growth versus immunity tradeoff in molecular systems plant pathology. Physiol. Mol. Plant P. 95:55-59.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722-736.

Lannoo, N., and Van Damme, E.J.M. (2014). Lectin domains at the frontiers of plant defense. Front. Plant Sci. 5:397.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44:W242-W245.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100.

Ling, H.Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., Cao, Y., Gao, Q., Zheng, S., Li, Y., et al. (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. Nature 557:424-428.

Lukashin, A.V., and Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26:1107-1115.

Luo, M.C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T., Wang, L., Wang, Y., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature 551:498-502.

Maccaferri, M., Harris, N.S., Twardziok, S.O., Pasam, R.K., Gundlach, H., Spannagl, M., Ormanbekova, D., Lux, T., Prade, V.M., Milner, S.G., et al. (2019).

759 Durum wheat genome highlights past domestication signatures and future
760 improvement targets. Nat. Genet. 51:885-895.
761 Matsumoto, T., and Wu, J.Z., and Kanamori, H., and Katayose, Y., and Fujisawa, M.,
762 and Namiki, N., and Mizuno, H., and Yamamoto, K., and Antonio, B.A., and
763 Baba, T., et al. (2005). The map-based sequence of the rice genome. Nature
764 436:793-800.
765 Meimberg, H., Rice, K.J., Milan, N.F., Njoku, C.C., and McKay, J.K. (2009).
766 Multiple origins promote the ecological amplitude of allopolyploid *Aegilops*
767 (Poaceae). Am. J. Bot. 96:1262-1273.
768 Meyers, B.C., Kaushik, S., and Nandety, R.S. (2005). Evolving disease resistance
769 genes. Curr. Opin. Plant Biol. 8:129-134.
770 Michael, P. (2001). The taxonomy and distribution of *Echinochloa* species (barnyard
771 grasses) in the Asian-Pacific region, with a review of pertinent biological
772 studies. In: Proceedings of the 18th APWSS Conference (Beijing, China, 28
773 May-2 June 2001). Standard Press of China, Beijing, 57-66.
774 Molina, J., Sikora, M., Garud, N., Flowers, J.M., Rubinstein, S., Reynolds, A., Huang,
775 P., Jackson, S., Schaal, B.A., Bustamante, C.D., et al. (2011). Molecular
776 evidence for a single evolutionary origin of domesticated rice. Proc. Natl.
777 Acad. Sci. U S A 108:8351-8356.
778 Nelson, R., Wiesner-Hanks, T., Wisser, R., and Balint-Kurti, P. (2018). Navigating
779 complexity to breed disease-resistant crops. Nat. Rev. Genet. 19:21-33.
780 Ou, S.J., Chen, J.F., and Jiang, N. (2018). Assessing genome assembly quality using
781 the LTR Assembly Index (LAI). Nucleic acids research 46:e126.
782 Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach,
783 H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., et al. (2009). The
784 *Sorghum bicolor* genome and the diversification of grasses. Nature
785 457:551-556.
786 Paterson, A.H., and Wendel, J.F. (2015). Unraveling the fabric of polyploidy. Nature
787 biotechnology 33:491-493.
788 Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum

789 evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol.
790 26:1641-1650.
791 Ramirez-Gonzalez, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J.,
792 Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., et al. (2018). The
793 transcriptional landscape of polyploid wheat. Science 361: eaar6089.
794 Salamov, A.A., and Solovyev, V.V. (2000). Ab initio gene finding in *Drosophila*
795 genomic DNA. Genome Res. 10:516-522.
796 Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M.
797 (2015). BUSCO: assessing genome assembly and annotation completeness
798 with single-copy orthologs. Bioinformatics 31:3210-3212.
799 Soltis, P.S., Marchant, D.B., Van de Peer, Y., and Soltis, D.E. (2015). Polyploidy and
800 genome evolution in plants. Curr. Opin. Genet. Dev. 35:119-125.
801 Soltis, P.S., and Soltis, D.E. (2009). The role of hybridization in plant speciation.
802 Annu. Rev. Plant Biol. 60:561-588.
803 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and
804 post-analysis of large phylogenies. Bioinformatics 30:1312-1313.
805 Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006).
806 AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids
807 Res. 34:W435-439.
808 te Beest, M., Le Roux, J.J., Richardson, D.M., Brysting, A.K., Suda, J., Kubesova,
809 M., and Pysek, P. (2012). The more the better? The role of polyploidy in
810 facilitating plant invasions. Ann. Bot. 109:19-45.
811 Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice
812 junctions with RNA-Seq. Bioinformatics 25:1105-1111.
813 Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J.,
814 Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and
815 quantification by RNA-Seq reveals unannotated transcripts and isoform
816 switching during cell differentiation. Nat. Biotechnol. 28:511-515.
817 Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance
818 of polyploidy. Nat. Rev. Genet. 18:411-424.

819 Vanderplank JE (1963) Plant diseases: Epidemics and control. New York, USA:
820 Academic Press.
821 Vaser, R., Sovic, I., Nagarajan, N., and Sikic, M. (2017). Fast and accurate de novo
822 genome assembly from long uncorrected reads. Genome Res. 27:737-746.
823 Vogel, J.P., and Garvin, D.F., and Mockler, T.C., and Schmutz, J., and Rokhsar, D.,
824 and Bevan, M.W., and Barry, K., and Lucas, S., and Harmon-Smith, M., and
825 Lail, K., et al. (2010). Genome sequencing and analysis of the model grass
826 *Brachypodium distachyon*. Nature 463:763-768.
827 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo,
828 C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated
829 tool for comprehensive microbial variant detection and genome assembly
830 improvement. PloS one 9:e112963.
831 Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z., Shen, C., Li, J.,
832 Zhang, L., et al. (2017). Asymmetric subgenome selection and cis-regulatory
833 divergence during cotton domestication. Nat. Genet. 49:579-587.
834 Weiner, J. (2019). Looking in the wrong direction for higher-yielding crop genotypes.
835 Trends Plant Sci 24:927-933.
836 Wendel, J.F. (2015). The wondrous cycles of polyploidy in plants. Am. J. Bot.
837 102:1753-1756.
838 Xu, P., Xu, J., Liu, G., Chen, L., Zhou, Z., Peng, W., Jiang, Y., Zhao, Z., Jia, Z., Sun,
839 Y., et al. (2019). The allotetraploid origin and asymmetrical genome evolution
840 of the common carp *Cyprinus carpio*. Nat. Commun. 10:4625.
841 Yabuno, T. (1966). Biosystematic study of the genus *Echinochloa*. Jpn. J. Bot.
842 19:277-323.
843 Yabuno, T. (1983). Cytogenetical studies on the hybrids of *Echinochloa oryzicola*
844 Vasing. and the Thai tetraploid strain of *E. stagnina* (Retz.) Beauv. with the
845 West African species *E. obtusiflora* Stapf. Cytologia 48:597-604.
846 Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., Hu, Z., Chen, S., Pental, D., Ju,
847 Y., et al. (2016). The genome sequence of allopolyploid *Brassica juncea* and
848 analysis of differential homoeolog gene expression influencing selection. Nat.

Genet. 48:1225-1232.

Ye, C.Y., Tang, W., Wu, D., Jia, L., Qiu, J., Chen, M., Mao, L., Lin, F., Xu, H., Yu, X., et al. (2019). Genomic evidence of human selection on Vavilovian mimicry. Nat. Ecol. Evol. 3:1474-1482.

Yu, Q., and Powles, S. (2014). Metabolism-based herbicide resistance and cross-resistance in crop weeds: a threat to herbicide sustainability and global crop production. Plant Physiol. 166:1106-1118.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847-848.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat. Biotechnol. 33:531-537.

Zhang, X.T., Zhang, S.C., Zhao, Q., Ming, R., and Tang, H.B. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat Plants 5:833-845.

Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics, proteomics & bioinformatics 4:259-263.

Zhao, G., Zou, C., Li, K., Wang, K., Li, T., Gao, L., Zhang, X., Wang, H., Yang, Z., Liu, X., et al. (2017). The *Aegilops tauschii* genome reveals multiple impacts of transposons. Nat. Plants 3:946-955.

873 **Table 1.** Summary of genome sequencing, assembly and annotation of three
874 *Echinocloa* species

Species	*E. haploclada*	*E. oryzicola*	*E. crus-galli*
Ploidy	2n=2X=18	2n=4X=36	2n=6X=54
Estimated genome size, Gb	0.42	1.00	1.40
Sequencing platform and coverage	Pacbio (92×) +Illumina (79×) +HiC (276×)	Pacbio (63×) +Illumina (120×)	Pacbio (86×) +Illumina (148×)
Assembly size, Gb	0.44	0.95	1.34
Contig N50, Mb	0.93	1.87	1.57
Scaffold N50, Mb	48.75	2.93	4.09
Genes annotated	36 946	66 521	103 853
BUSCO assessment (%)	97.1	97.7	97.5
Percentage of repeat element (%)	41.09	52.19	46.26
GC content (%)	46.05	46.05	45.91

875

876

877 **Figure legends**

878 **Figure 1. Genomes and phenotypes of diploid *Echinochloa haploclada,* tetraploid**
879 ***E. oryzicola* and hexaploid *E. crus-galli*. (A)** Phenotypes of three sequenced
880 *Echinochloa* species. From left to right, *E. haploclada* (Pasquet 1083), *E. oryzicola*
881 (ZJU2) and *E. crus-galli* (STB08). **(B)** circos plot of *E. haploclada* genome showing
882 different genomic features including synteny to *S. italica* (only blocks with at least 20
883 syntenic genes were shown; different colors represent chromosomes of *S. italica*),
884 repeat element, gene density, GC content and transcription landscape. **(C)** Alignments
885 of *E. oryzicola* (gray-blue) and *E. crus-galli* (red) contigs to *E. haploclada* genome.
886 Black histograms in nine chromosome bars represent gene density of *E. haploclada*.

887 **Figure 2. Divergence times of *Echinochloa* genomes. (A)** The synonymous
888 substitution rates (*K*s) distribution of homologous genes between *E. haploclada* and *S.*
889 *italica*, and between subgenomes from two *Echinochloa* weeds. The subscript "T"
890 and "H" indicate subgenomes from tetraploid (A_T and B_T in *E. oryzicola*) and
891 hexaploid (A_H, B_H and C_H in *E. crus-galli*), respectively. Individual peak was
892 separately shown in Supplementary Figure 10. **(B)** Brief diagram showing the
893 evolution of *Echinochloa* species and estimation of polyploidization times. Gray lines
894 with square nodes represent three unknown ancestral diploid progenitors. Numbers
895 beside each branch point show divergence or hybridization time. mya, million years
896 ago.

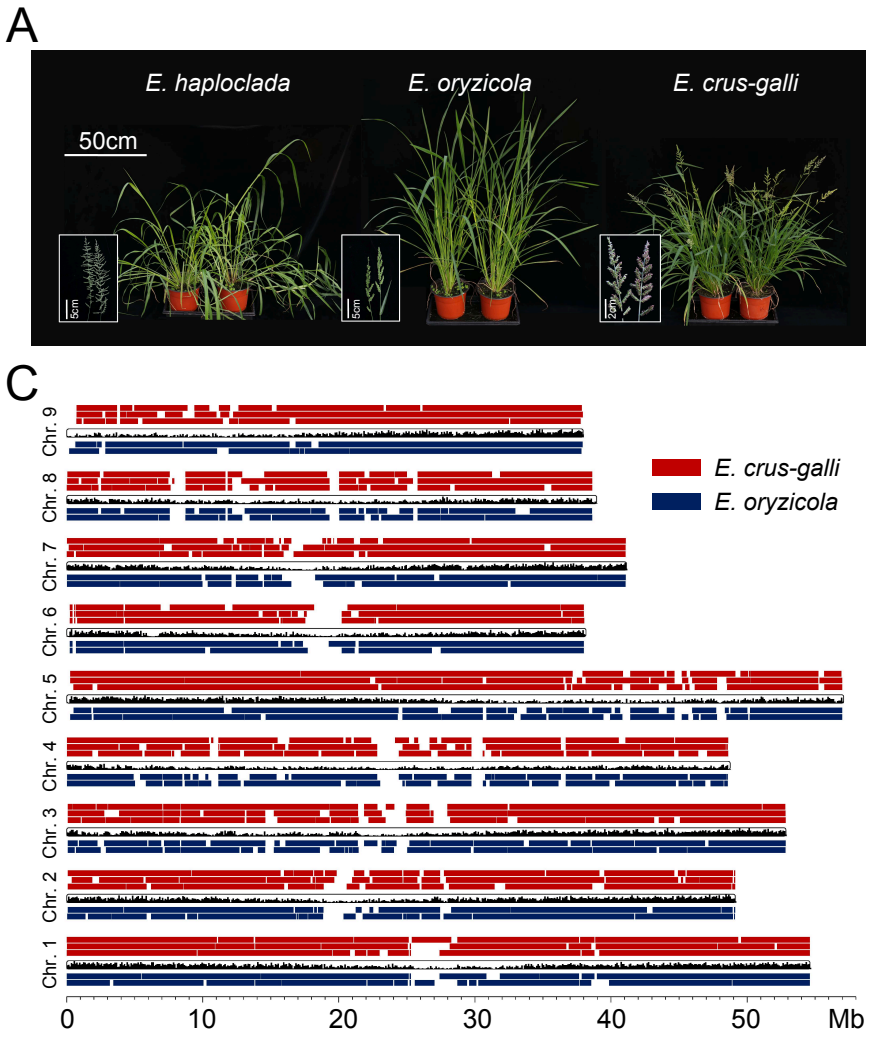
897 **Figure 3. Changes of gene family size during *Echinochloa* polyploidization. (A)**
898 Dot matrix and distribution of fold changes (lower right) of gene family sizes of *E.*
899 *crus-galli* with *O. sativa* and the sum of tetraploid and diploid *Echinochloa* species. In
900 the distribution of fold changes of gene family size, majority of gene families in *E.*
901 *crus-galli* is about three times those of *O. sativa* (left) and the same as the sum of the
902 diploid and tetraploid *Echinochloa* species (right) in size. **(B)** Comparison of abiotic
903 and biotic stress-related gene family sizes among *Echinochloa* species and other
904 grasses. *T. turgidum* refers to durum wheat. Topological relationship is derived from
905 the Timetree database (http://timetree.org/). For ploidy, one, two and three circles

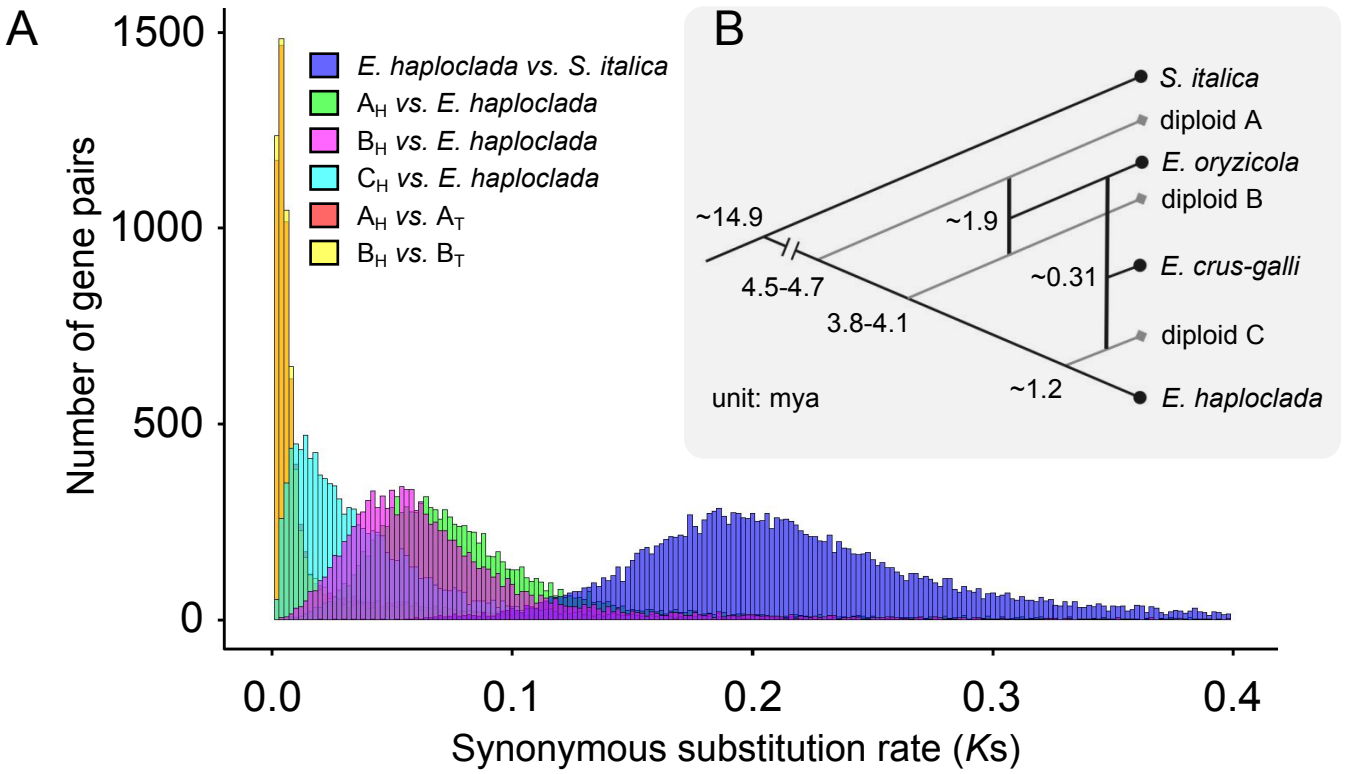
represent diploid, tetraploid and hexaploid, respectively. +/-, increase/decrease in size relative to corresponding outgroup species. *, P<0.01; **, P<0.001; ***, P<0.0001, Fisher's exact test. **(C)** Synteny retention ratio of seven gene families including disease-resistance gene families NB-ARC and wall-associated receptor kinase (WAK). Across the genome, 46.4% (indicated by the red dashed line) of genes fit the 1:(1:1):(1:1:1) synteny retention ratio. ***, P<0.0001, Fisher's exact test. **(D)** An example of loss for a NB-ARC domain encoding gene and retention for a GRAS gene during *Echinochloa* polyploidization. Shades between two segments represent synteny. A_T and B_T represent two subgenomes in tetraploid *E. oryzicola* genome, and A_H, B_H and C_H represent three subgenomes in hexaploid *E. crus-galli* genome, respectively.

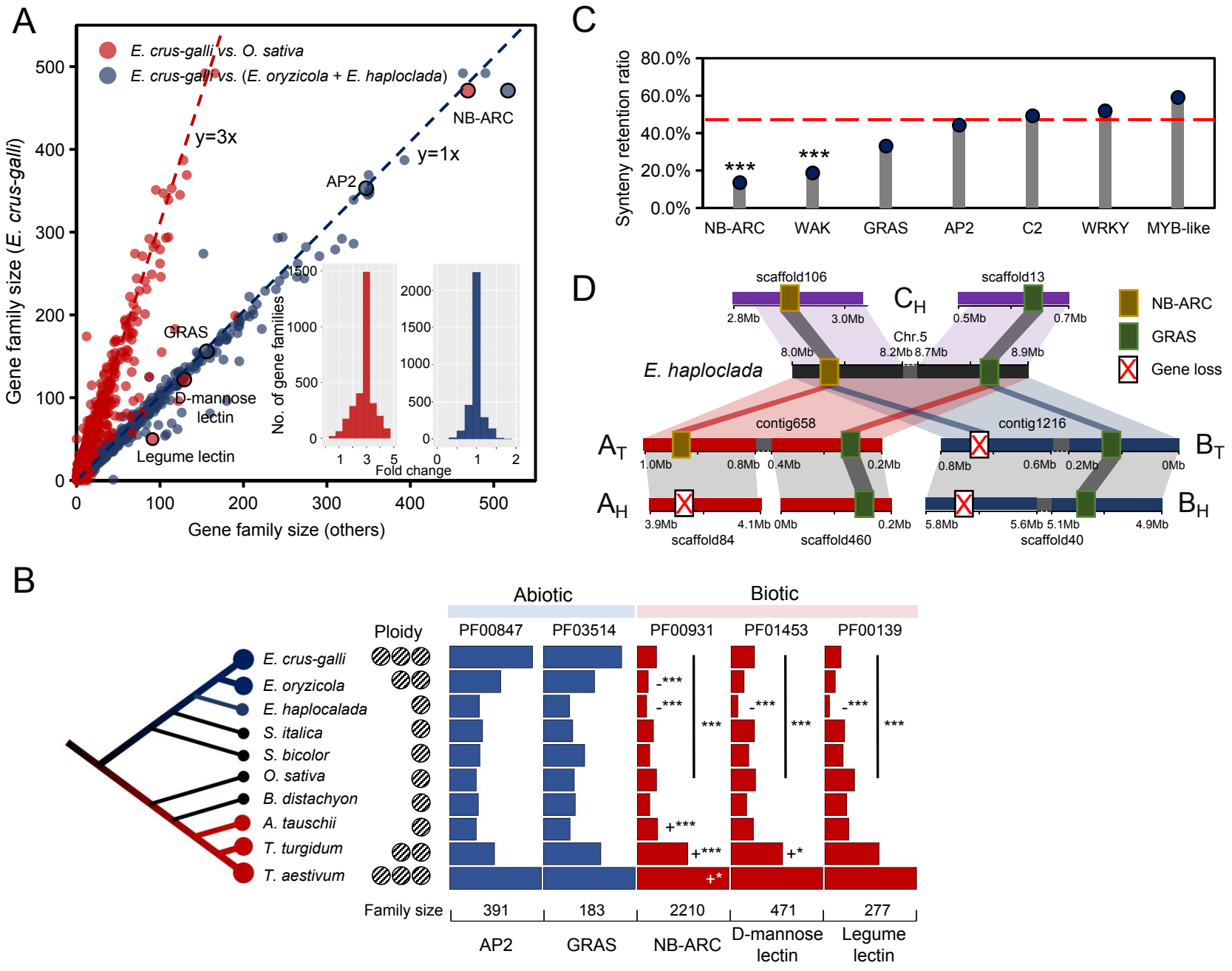
Figure 4. Comparison of selection pressure on subgenomes of *Echinochloa* weed and wheat during polyploidization. (A) The model of *Echinochloa* polyploidization process. #, *E. haploclda* was used as a proxy for the C_H progenitor genome. **(B)** Comparison of *K*a/*K*s ratio between subgenomes A_T and B_T in *E. oryzicola* (left) and A_H and B_H in *E. crus-galli* (middle) and comparison of pN/pS ratio of A_H and B_H in *E. crus-galli* populations (right). **(C)** The model of bread wheat polyploidization process. **(D)** Comparison of *K*a/*K*s ratio between subgenomes A and B in domesticated and wild tetraploid wheat (left) and hexaploid bread wheat (right) . In the box plots, the horizontal line shows the median value, and the whiskers show the 25% and 75% quartile values of *K*a/*K*s or Pn/Ps. Nonparametric Wilcoxon rank-sum test and Kruskal–Wallis pairwise comparison test were performed to evaluate significant differences for two and four samples, respectively. n.s., not significant; ***, P<0.0001.

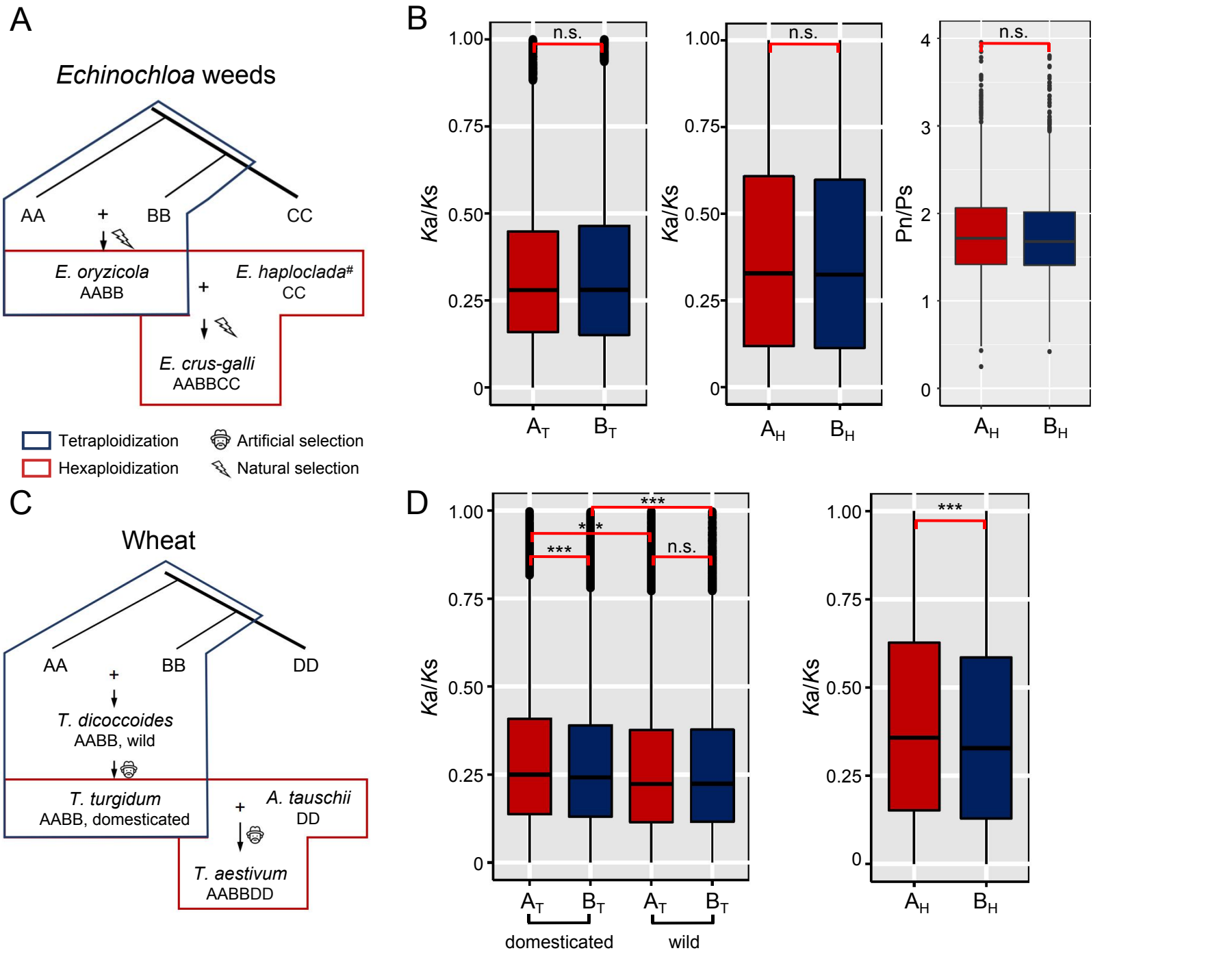
Figure 5. Subgenome expression profiling during *Echinochloa* hexaploidization. (A) Subgenome expression changes of *E. crus-galli* under drought stress by PEG (polyethylene glycol) treatment. Expression bias status for 13,619 triads of homeologous genes in roots under drought treatment (lower right triangle and bottom half of the middle circle) and control (upper left triangle and upper half of the middle circle) were shown. Summary of bias category changes was shown by the middle

chord diagram. **(B)** Relative expression of three subgenomes (i.e. expression value of one homoeolog divided by sum of expression values of three homoeologs in a triad) in *E. crus-galli* and bread wheat. Black verticle line marks the extremely average status with relative expression value of 0.33. A_H, B_H and C_H represent three subgenomes in hexaploid *E. crus-galli*, respectively. A, B and D represent three subgenomes in hexaploid wheat, respectively.


A
E. haploclada
E. oryzicola
E. crus-galli
50cm
B
E. haploclada
Chr.1
Chr.2
Chr.3
Chr.4
Chr.5
Chr.6
Chr.7
Chr.8
Chr.9
i
ii
iii
iv
v
vi
i. syntenic block to S. italica
ii. repeat element
iii. gene density
iv. GC content
v. root gene expression
vi. leaf gene expression
C
E. crus-galli
E. oryzicola
Chr. 1
Chr. 2
Chr. 3
Chr. 4
Chr. 5
Chr. 6
Chr. 7
Chr. 8
Chr. 9
0
10
20
30
40
50
Mb


A
1500
1000
500
0
Number of gene pairs
E. haploclada vs. S. italica
A_H vs. E. haploclada
B_H vs. E. haploclada
C_H vs. E. haploclada
A_H vs. A_T
B_H vs. B_T
0.0
0.1
0.2
0.3
0.4
Synonymous substitution rate (Ks)
B
S. italica
diploid A
E. oryzicola
diploid B
E. crus-galli
diploid C
E. haploclada
~14.9
4.5-4.7
3.8-4.1
~1.9
~0.31
~1.2
unit: mya


A
E. crus-galli vs. O. sativa
E. crus-galli vs. (E. oryzicola + E. haploclada)
y=3x
y=1x
NB-ARC
AP2
GRAS
D-mannose lectin
Legume lectin
Gene family size (E. crus-galli)
Gene family size (others)
No. of gene families
Fold change
C
Synteny retention ratio
80.0%
60.0%
40.0%
20.0%
0.0%
***
***
NB-ARC
WAK
GRAS
AP2
C2
WRKY
MYB-like
D
scaffold106
scaffold13
C_H
2.8Mb
3.0Mb
0.5Mb
0.7Mb
Chr.5
8.0Mb
8.2Mb 8.7Mb
8.9Mb
E. haploclada
NB-ARC
GRAS
Gene loss
contig658
contig1216
A_T
B_T
1.0Mb
0.8Mb 0.4Mb
0.2Mb
0.8Mb
0.6Mb 0.2Mb
0Mb
A_H
B_H
3.9Mb
4.1Mb 0Mb
0.2Mb
5.8Mb
5.6Mb 5.1Mb
4.9Mb
scaffold84
scaffold460
scaffold40
B
Abiotic
Biotic
Ploidy
PF00847
PF03514
PF00931
PF01453
PF00139
E. crus-galli
E. oryzicola
E. haplocalada
S. italica
S. bicolor
O. sativa
B. distachyon
A. tauschii
T. turgidum
T. aestivum
-***
-***
***
-***
***
-***
***
+***
+***
+*
+*
Family size
391
183
2210
471
277
AP2
GRAS
NB-ARC
D-mannose lectin
Legume lectin


A
Echinochloa weeds
AA
+
BB
CC
E. oryzicola
AABB
+
E. haploclada#
CC
+
E. crus-galli
AABBCC
Tetraploidization
Hexaploidization
Artificial selection
Natural selection
B
Ka/Ks
Pn/Ps
n.s.
1.00
0.75
0.50
0.25
0
4
3
2
1
A_T
B_T
A_H
B_H
C
Wheat
AA
+
BB
DD
T. dicoccoides
AABB, wild
T. turgidum
AABB, domesticated
+
A. tauschii
DD
T. aestivum
AABBDD
D
***
**
n.s.
domesticated
wild