Adaptive Evolution after Gene Duplication in Maize

Bin Wei

Sichuan Agricultural University   https://orcid.org/0000-0001-6853-8200

Hanmei Liu

Sichuan Agricultural University

Qianlin Xiao

Southwest University

Yongbin Wang

Sichuan Agricultural University

Junjie Zhang

Sichuan Agricultural University

Yufeng Hu

Sichuan Agricultural University

Yinghong Liu

Sichuan Agricultural University

Guowu Yu

Sichuan Agricultural University

Yangping Li

Sichuan Agricultural University

Yubi Huang (✉ yubihuang@sohu.com)

https://orcid.org/0000-0003-4323-7596

Adaptive Evolution after Gene Duplication in Maize

Bin Wei[1], Hanmei Liu[2], Qianlin Xiao[1,3], Yongbin Wang[1], Junjie Zhang[2], Yufeng Hu[1], Yinghong Liu[4], Guowu Yu[4], Yangping Li[1] and Yubi Huang[1*]

1 State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu 611130, China

2 College of Life Science, Sichuan Agricultural University, Ya'an 625014, China

3 College of Agronomy and Biotechnology, Southwest University, Chongqing 400716, China

4 College of Agronomy, Sichuan Agricultural University, Chengdu 611130, China

*Author for Correspondence: Yubi Huang, State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu 611130, China, yubihuang@sohu.com

Abstract

Background: Gene duplication can provide genetic basis for the evolution. The neo/sub-functionalization of redundant copies can promote plant growth and development, improve environmental adaptability, and even form new species.

Methods: We systematically identified the duplication genes and their origin of maize by comprehensively utilizing methods of homologous clustering, chromosomal collinearity, Ks analysis and phylogenetic analysis. The distinction of duplicated genes was analyzed by comparing the gene structure, sequence composition, expression, and functional differentiation.

Results: More than 70% of the genes in the maize genome have been found to be duplication genes, and mainly derived from the whole-genome duplication (WGD). The gene structure of the WGD genes is more complex, with rich components and long CDS. The GC content of WGD genes contributes to the bimodal distribution, indicating that a part of the duplicated gene contains abundant GC bases. The expression level of WGD genes is high and possesses tissue-specificity. The functionally differentiated duplication genes are mainly involved in the growth and stress response of maize.

Conclusions: The results of this study indicate that duplication genes produced by WGD provide new adaptability to maize growth, morphogenesis, and biotic and abiotic stress resistance, which are important to the evolution of maize.

Keywords

Gene duplication, Whole-genome duplication, Adaptive evolution, Maize

Background

Gene duplicates (GDs) often mainly result from chromosomal unequal fragment exchange, transposition and whole-genome duplication (WGD) [1–4]. Three mechanisms account for different proportions of GDs in different species, and WGD is more prevalent in plant kingdom [5]. GDs provide raw materials for evolution, which potentially alter gene dosages and function by neo/sub-functionalization. GDs with novel gene functions could be retained due to drift and selection. Like genetic mutations, gene duplication occurs first in individuals, and it has only a small probability to be retained when a duplication gene is neutral before duplication [6]. Even if the duplication gene remains in the population, it is still straightforward to lose unless it is beneficial to the organism [7, 8]. The probability of being retained in the population is based on its suitability. Unless the duplication gene is useful to the organism and its losing will result in a decrease in fitness, otherwise it cannot be stably retained in the genome [4, 9–15].

GDs are present in almost all eukaryotes [16–19], which often evolve discrepant function or expression patterns among copies [20–25]. WGD generates a copy of all genes, with greater potential than single gene duplication for the evolutionary success of enhancing organismal adaption to environmental challenges and appearing novel traits. Analysis of expression data in multiple species revealed differences in expression between duplication genes [26–29]. Also, the study analyzed the differences in microRNA regulation between paralogs in Arabidopsis, rice, and canola, and proposed their contribution to the difference in expression between duplication genes [30–32]. Duplication genes were involved in the evolution of important functions of speciation, and the existing studies had found that the duplication genes of angiosperms led to the diversification of seed and flower development regulating genes, which had a great influence on the angiosperm rise [33]. In the past 350 million years, three whole-genome duplications in Arabidopsis directly led to an increase in transcription factors, signal transducers, and development-related genes [34]. The evolution of C4 photosynthesis in Sorghum bicolor (sorghum), and Zea mays (maize) involved the duplication of the C3 gene and subsequent functional differences [35, 36]. Through the differential differentiation of expression and function, the retained duplication genes played a key role in improving species fitness [37–40].

Maize shares an ancient ancestor 50-85 million years ago (MYA) between other grasses [41, 42], and the sorghum which is more closely related to evolution, is presumed to be divided into 12MYA years ago [43]. Maize genome is very complex, with rich repeated sequences, transposons, and paralog genes. Previous studies speculated that maize has undergone at least

three whole-genome duplication events, including before the divergence of the monocotyledonous plant, before the divergence of the grass, and after the formation of the maize. The WGD events occurred about 110 MYA, 50 MYA and 12 MYA, respectively [44–46]. Previous research focused on the identification of WGD events, without a comprehensive analysis of retaining maize gene duplicates and their impact on development and adaptive evolution. We used maize, sorghum, Oryza sativa (rice) and Arabidopsis genes for homology clustering, collinearity analysis, and phylogenetic analysis to identify duplication genes in the maize genome. After that, differences in sequence, expression, and function between duplication genes were analyzed. The effect of maize duplication genes on the adaptive evolution of maize was further explored.

Results

Identification of duplication genes in maize

All the proteins of maize, sorghum, rice and Arabidopsis were collected as the subject database, and BLASTP of each protein was performed to obtain similar sequences. 7 Methods were used to cluster all the similar sequences, and 1,119,261 collections were obtained at last (Fig. 1a). Because both mcl and OrthoMCL are based on Markov model, the number of them are similar, which were 23,052 and 22,114 collections, respectively. The number of collections of Hcluster_sg, Hifix, MC-UPGMA, SiLix and SPICi methods are similar, which were 13,691, 14,619, 13,445, 14,611 and 11,826 collections, respectively. In these collections, a total of 17,137,340 gene pairs were obtained. It was found that 49% of all the gene pairs were identified by two or more clustering method, among then, 13% gene pairs were identified by two methods, and 12% gene pairs were identified by three methods (Fig. S1).

Gene pairs were identified by at least two methods, which would be considered as candidate gene pairs for further analysis. For the candidate dataset of all the gene pairs, 33,157, 31,165, 26,918 and 24,240 genes were involved in maize, sorghum, rice and Arabidopsis, and accounting for 84%, 90%, 75% and 88% of total genes of each species, respectively (Fig.1b). Moreover, based on the candidate dataset, we reconstructed 13,388 new collections and count the number of gene involved in new collections between species (Fig. 1c). A total of 1,175 collections only have maize genes, which was more than that in the other three species. Results found most genes from four species were included in new collections, which indicated that four species were closely related. Genomic collinearity analysis was performed using MCScanX to determine which GDs originated from WGD. A total of 311,199 alignment results from previous blast were introduced, setting a threshold of 1e-5, and finally get 6,146 collinear blocks, of which 454 collinear blocks were obtained in maize, 403 blocks between Arabidopsis and maize, 839 blocks between rice and maize, and 817 collinear blocks in sorghum and maize

(Table S1). The 6,146 collinear blocks contained a total of 75,606 genes, accounting for 55% of the total gene count of 137,060 of the four species.

The distribution of Ks can be used to estimate the occurrence time of these WGD events. The Ks values between gene pairs in all collinear blocks were calculated in the analysis (Fig. 2). There were two distinct Ks peaks inside the maize, which was consistent with the results of the two WGDs mentioned in the previous study [44–46]. The first duplication occurred at about 50 MYA; the most recent large-scale duplication incident occurred about 12 MYA. It was speculated that it had experienced a relatively large scale before the separation of Arabidopsis and maize, sorghum, and rice about 110 MYA. The genes produced by this ancient duplication were partially retained in each species.

Finally, we classified all maize genes according to duplication events. There were 28,911 duplication genes, accounting for 73.2% of maize genes. Further, a total of 21,654 duplication gene have homology and collinear relationships, defined as the syntenic duplication genes, accounting for 54.9% of the total genes. Combining the phylogenetic analysis of these gene, maize syntenic duplication genes were classified into four categories Maize WGD, Sorghum-Maize, Pre-grasses WGD and Pre-monocots WGD. The number of genes in categories was 9,606, 3,309, 5,917, 2,922, respectively, accounting for 26.9%, 13.5%,15.3% and 44.4% (Fig. 3, Table S2).

Characteristic of WGD genes in maize

WGD genes with longer CDS and more exons

According to the gene structure, we compared GDs originated from different ways, including the length of gene, 5'UTR, 3'UTR, coding region, exon and intron, number of exons. The average length of the single-copy gene is 5,364 bp, and of GDs is 4,495 bp. The mean CDS length of all genes is 1,096 bp, and the values of single-copy gene and GDs are 1,314 bp and 1,206 bp, respectively. Similar results exist in the length of the UTR region, the length of the intron, and the length of the exon (Fig. S2). It can be seen that the maize syntenic duplication gene and the single copy gene have more gene component numbers or longer gene components. The Pre-monocots WGD genes had higher transcripts number, average gene length, and average CDS length and average intron length (Table 1, Table S3).

The maize WGD gene contributes to the bimodal GC distribution

The level of GC can affect the interaction of base stacks and thus play an important role in stabilizing the gene structure [61]. In earlier studies, the GC3 distribution of different organisms could be divided into two main types, unimodal and bimodal[62, 63]. In our analysis, the GC3 of the coding regions of different genes of maize were calculated. The GC3 content distribution of the syntenic duplication gene and the tandem duplication gene formed a distinct bimodality, indicating that the partially duplication gene had high GC3. The genes with other

homolog relations are all unimodal, and most of the genes have low GC3 content (Fig. 4a). The GC3 content distribution of the duplication genes generated by different WGD showed that except for the Pre-monocots WGD duplication gene, other WGD genes formed bimodal (Fig. 4b). Pre-grasses WGD produced the largest proportion of high GC3 genes. Some of the duplication genes in the Maize-Sorghum and Maize WGD stages are also high GC3. Since these genes may be derived from the high GC3 gene produced by Pre- grasses WGD, it is speculated that Pre-grasses WGD is the main source of high GC3 genes in the maize genome.

High average expression and tissue-specific expression in WGD genes

In addition to sequence differences, changes in transcription levels are also an important part of genetic differences. The transcriptional data for 60 different periods and locations of maize reference genome B73 were used to understand the differences in transcription levels between different types of duplication genes [64]. A comparison was made between genes of different homology relations and different WGD genes in maize, in which a total of 39,457 genes have transcripts. The average values of transcription of six tissues between different genes were counted, an average transcription value (log2 (FPKM)) is 4.1 in the WGD gene, 4.6 in the single copy gene; this value is 1.22, 2.04, 1.57 and 1.66 in the genes without homologous, maize-specific, tandem duplication, and other duplication type, respectively (Fig. 5a/b). The results indicate that single copy genes, as well as WGD genes, have higher levels of transcription in maize. Further into the genes between different WGD events of line comparison and found that the average gene expression levels between different WGD events produced no significant difference.

The differentially expressed genes were screened by calculating the difference between the expression levels of different tissues and the mean values of all tissues. A total of 16,207 genes with differential expression were obtained (Fig. 5c). Most of the tissue-specific genes were syntenic duplication genes (9,902), accounting for 61% of tissue-specific genes. From the number of syntenic duplication genes expressed in specific tissues, the expression genes in anther, root, stem, leaf, seed, embryo and endosperm were 3,254, 2,911, 2,080, 1,608, 2,768, 3,306 and 4,032, respectively. Among them, genes specifically expressed in the endosperm accounted for 41% of the syntenic duplication genes. Further, there were 3,709 tissue-specific genes accounting for 39% of the total number of Maize WGD genes. Other duplication genes related to Maize-Sorghum, Pre-grasses WGD, and Pre-Monocots WGD were 1,626, 3,307, and 1,206, respectively, accounting for 38%, 28%, and 9% (Fig. 5d). The results showed that genes retained after Maize WGD and Pre-grasses WGD had more obvious tissue specificity in expression.

Maize WGD genes are mainly involved in developmental and stress responses

The functional analysis includes gene ontology annotation and metabolic pathway analysis of genes. The ontology annotations of maize recent WGD genes enriched in plasma membrane, transcription factor activity, response to abiotic stimulus (Fig. S3, Table S4). The Pre-grasses WGD genes were enriched in a similar functional type to the maize recent WGD gene. Responsive to stress in biotic or abiotic stress, cell death, flower development, and pollen-pistil interaction in the biological process (Table S5).

KEGG metabolic pathways enrichment analysis of the different type of the duplicated genes in maize found, in the most recent Maize WGD genes, the metabolic pathway is enriched in MAPK signaling pathways, plant hormone signaling, autophagy, and circadian rhythms (Table 2). The Pre-grasses WGD genes are enriched in amino acid metabolism-related pathways, sugar and fatty acid metabolism-related pathways, starch and sucrose metabolism, and the plant-pathogen interaction pathway, which play important roles in plant development and response to biotic and abiotic stresses [65] (Table S6).

The MAPK signaling pathway was enriched in KEGG analysis, which played an important role in the perception of stimulation and adaptive responses. In plant, MAPK is a complex family, involved in signal transduction of response to pathogens, drought, salinity, cold, injury, ozone, ROS and hormone stress [66–69]. Previous studies found 19 MAPK genes in maize[31, 70], and we identified 22 MAPK genes in the same collection, 14 of which had undergone the most recent maize WGD, and the remaining 8 genes were retained after Pre-grasses WGD (Table 3, Fig. 6a). Among them, the most studied gene ZmMPK6-2 which has been assigned to function at low temperatures and ABA signaling [71, 72]. The gene ZmMPK6-1, which has a very close evolution relationship with ZmMPK6-2, was formed by the most recent maize WGD. The expression of MAPK genes was further analyzed (Fig. 6b). Pre-grasses WGD retaining genes were mainly differentially expressed in embryo, endosperm, and pollen, while Maize WGD retaining genes were widely expressed in maize roots, stems and leaves. This different event preserves the tissue differences in gene expression, suggests that maize duplication genes have an important impact on the enrichment and functional building of the MAPK family of genes.

Discussion

Comprehensive homology, collinearity, Ks distribution, and phylogenetic analysis contribute to accurately identify WGD genes

More than 80% of the sequences in the maize genome are transposons, and some transposons are extremely active [73]. In order to accurately identify WGD, further analysis of

homologous gene aggregation/cluster is required. WGD typically leaves some characteristic collinear fragments at the chromosome level and results in more paralogous genes than conventional genes. Using these features, Ks can be used to estimate duplication occurs time and the Ks-based approach have been widely used [7, 74, 75]. In our results, Ks analysis within the collection can relatively accurately identify maize WGD which occurred about 110 million years ago, 50 million years ago, and 12 million years ago, which is the same as the results of previous studies [44–46]. At the same time, method based on collinear relationship identification need to have a collinear relationship at the whole genome level, which is also widely used [46, 76, 77]. Base on the method, we have identified 6,146 collinear blocks, which contained a total of 75,606 genes, accounting for 55% of the total genes of four species. However, it is susceptible to high levels of genomic rearrangement or loss of gene duplication, which is widespread and rapidly changing [78–80]. Recent studies had found that saturation effects might affect changes of Ks [81], making Ks estimated inaccurate duplication times. Jiao et al. attempted to construct a more accurate WGD event map by means of phylogenetic trees [33]. This method has better universality than a single collinearity-based or Ks-based method and has the advantage of simultaneously estimating the time and phylogeny of the WGD. However, it is not clear how the selection of a particular node of a particular gene family affects the outcome, as all selected duplicated nodes must be located within a given time period. It was unclear what kind of duplication event background distribution should be used during this time, or how the branch length estimation error affected the background distribution [82]. In our analyses, we are used a combination of collinearity, Ks distribution, and phylogenetic analysis and final obtained 15,158 gene pair of maize with a syntenic duplication relationship. For PGDD, which was a database of gene and genome duplication in plants [44], this number was 13,408. Among them, 9984 gene pair relationships were detected in our results and PGDD, accounting for 74.5% of the PGDD gene pair set. Explain that the results obtained are credible. At the same time, the Ks distribution of all homologous gene pairs was analyzed, and the results were similar to the Ks distribution of the collinearity genes, indicating that the results obtained in this study are accurate and reliable. In addition, since the predicted maize-sorghum divergence time is close to the time of the most recent WGD of maize, it is speculated that the large-scale genome duplication of maize or its ancestors formed the divergence of maize and sorghum species.

WGD genes with more complex gene structure

Our work with gene composition shows that WGD genes in maize have a more complex structure, showing longer gene, CDS and UTR length, as well as a greater number of exons and transcripts. Studies of polyploid plant evolution also suggested that WGDs provides genomic novelty and complexity to enhance organisms' environmental challenges and improve their own adaptability [83, 84]. It was inferred that complex genes preferentially reserved after WGD, or that genes began to enrich their composition in order to adapt to the challenge environment and finally reserved. Longer CDS is correspond to long protein, providing more abundant protein secondary and tertiary structure variants, and a greater number of exons provides the basis for transcript variants. After WGD, the duplicated copies of single-copy

genes were not retained, the possible reason was the important and conserved in function, which are related to the basal metabolism and regulation of the organism, the dosage alteration would cause organisms to death. If the WGD is experienced during species evolution, it means that one of its duplicated genes has been lost in the process of evolution. The retention of genes with a significantly richer genetic composition also means that they are more favorable in evolution and are preserved. It is further proved that genes after WGD remain in evolution and gradually form more abundant gene composition.

Duplication genes with high GC contribute to bimodal GC distribution of maize

A large number of studies have analyzed the GC content of different species genomes [63, 85–88]. Previous studies have shown that there are two main types of gene GC distribution in different organisms, namely unimodal and bimodal. The double peak indicates that some of the genes in the genome have higher GC content, and the other part have lower GC content. The GC content of grasses plants, Musa, Zingiberaceae, and warm-blooded animals showed a bimodal distribution, and cold-blooded animals and other plants (including dicots) showed a unimodal distribution [63, 85]. This study found that the GC bimodal of maize is mainly derived from tandem duplication and syntenic duplication genes. In the WGD gene, the duplication gene of the Pre-grasses WGD has the greatest contribution to the GC bimodal. It is speculated that due to the Pre-grasses WGD, the GC content of the grass ancestral gene is bimodal, resulting in the GC content of most of the grasses being bimodal.

Previous studies have suggested that high and low GC differences in genes are affected by selection pressure [63, 85, 89], and many research evidence on animals supports GC-preferred gene transfer (gBGC) to play an important role in the GC-preferring process of forming species genes [90–92]. Glémin et al. proposed a simple model that under the gBGC hypothesis, strong 5'-3' recombination and gBGC gradient behavior can produce a bimodal distribution of GC content, and a slower gradient may result in a single peak distribution [93]. In our results, Pre-grasses WGD occurred at about 110MYA which in the late Cretaceous [94]. It can be speculated that the higher surface temperature at that time may be the cause of Pre-grasses WGD, because of at elevated temperatures, increased GC content can enhance base stacking forces, thereby increasing base stability [61]. And base on the gGBC model, it can be speculated that the WGD of the individual causes the change of gGBC effect before grasses divergence, and the evolution of the gene after duplication separates high GC content genes and extends to populations and species, and finally forming bimodal in GC. Therefore, Pre-grasses WGD has an important contribution to the bimodal distribution of GC content in ancestors of grasses plants and various grasses plants after divergence.

WGD genes contribute to enhance environmental adaptability in maize

Artificially increasing or deleting duplication genes demonstrates their impact on fitness. Gene in Salmonella enterica is amplified to a high copy number in real-time evolution system, and the duplicated gene accumulate mutations that provide enzymatic specialization of

different copies and faster growth [11]. In addition, compared to the deletion of one of the duplication genes in Schizosaccharomyces pombe, the deletion of the duplicated gene in Saccharomyces cerevisiae significantly reduced their adaptability [40]. Duplication genes can enhance adaptability by altering gene expression, the previous study in angiosperms suggested that the duplicated MADS-box genes have been implicated in the regulation of a variety of flower developmental processes, and finally affecting the evolution of angiosperms [95]. Our work found the syntenic duplication gene of maize had a higher average expression level, indicating that the large-scale genome duplication genes were more active in expression. Further analysis revealed that the expression characteristics of maize syntenic duplication genes not only showed an increase in expression but also showed strong tissue specificity. A similar result was found in Gossypium hirsutum, the expression pattern of duplication genes in polyploid G. hirsutum has specificity at the organ and developmental stages, which have shown respond to stress [96, 97].

Functional divergence is the most direct evidence of the adaptive impact of duplication genes. Our work found that the maize WGD gene was mainly involved in development and response to stress by gene ontology enrichment analysis, and the metabolic pathway was enriched in MAPK signaling pathways. These functions and pathways have an important impact on the adaptability of species [66–69]. Based on these results, we depict the effect of duplicating genes on maize adaptive evolution: First, some of the genes may be pseudo-geneticized due to factors such as duplication of function or disruption of the dose effect after duplication. Then, some of the duplicated genes show divergence in expression, especially tissue-specific expression, and functional differentiation to adapt to changing environment. Finally, under the effect of natural selection or artificial selection, copies that are beneficial to enhance individual fitness are retained.

Conclusions

The evolution of species is a complex process in which the duplication of genes plays an important role. We have comprehensively identified duplication genes in maize, particularly three whole genome duplication genes. Our results indicate that duplication genes produced by WGD provide new adaptability to maize growth, morphogenesis, and biotic and abiotic stress resistance, which are important to the evolution of maize. These effects were mainly achieved by improving genetic material, enriching gene components, the divergence of gene expression and function. These results provide new insight into the evolution of duplication genes and their role in the evolution, and duplication events should be considered in future genome and gene family studies to account for the species and gene evolution.

Methods

Maize, rice, sorghum, and Arabidopsis genes and genomic data were obtained from the Ensembl Plants database. The maize reference genome is the AGPv3 version, assembled in total length of 3,233,616,351 bp and annotated 39,469 coding genes (AGPv3 5b). In order to explore the evolutionary pathways of GDs in maize, sorghum, rice and Arabidopsis were selected as references (Fig. S4). The Arabidopsis reference genome was published in 2010 (TAIR10), with the total length of 135,670,229 bp and 27,416 annotation genes. The rice reference genome used IRGSP-1.0, with 374,424,240 bp in length and 35,679 genes. The sorghum reference genome version was Sorbi1, with 738, 540, 932 bp in length and 34,496 coding genes (Table S7). The longest and complete coding sequence (CDS) of each gene and its encoded protein sequence were used to analyze. CDS sequences that had corrected initial and termination codons were included in the dataset.

Gene similarity ratio Blast pairwise alignment using protein sequences. All the protein sequences of maize, sorghum, rice and Arabidopsis were collected as a target dataset, and BLASTP was performed using each protein as a query, with the E-value of 1e-5 and the maximum number of aligned sequences to select of 100. Filter the results of the comparison pair: (1) Filter the higher repeat alignment results. The adjusted E-value of the two sequences A and B is evaluated in the analysis by: Adjusted E-value (A, B) = E-value (A, B) * ((counts of blast / counts of genes) / (counts(A) + counts(B))), retaining results with Adjusted E-value less than 0.05. (2) Filter the poor alignment in the alignment results. In order to screen out the poor matches in the comparison results, recalculate the alignment scores, and calculate the new alignment scores using the A and B genes as examples: cscore (A, B) = score(A, B)/max(best score Of A, best score of B), and then remove the cscore (A, B) less than 0.5 alignment results.

Gene cluster analysis was performed on genes of maize, sorghum, rice, and Arabidopsis using a variety of cluster analysis method (Table S8). Including SiLix [47]: using graph algorithms, constructing similar "sequence families" based on disjoint sets; efficient clustering method Hifix built on SiLix [48]; Mcl [49] and OrthMCL [50] based on Markov clustering algorithm; MC-UPGMA [51] which used unweighted cluster method using arithmetic averages; fast and efficient clustering SPICi [52] based on heuristic algorithm; Hcluster_sg [53] based on the constructing sparse graph.

Genomic collinear block analysis using MCScanX [54]. Downstream analysis of collinear block analyzed by collinearity. The coding region sequences were obtained for all aligned gene pairs, Ks and 4DTv were calculated, and then possible concatenated collinear blocks were filtered. The non-syntenic duplication of genes could be further classified by the collinear results. The synonymous substitution did not cause changes in gene function，which was often seen as a random event, so the difference between the two coding sequences could be estimated by counting the synonymous substitution probability between the two sequences. At the same time, according to Gaut et al.'s calculation, the average base replacement rate per dot per year in

the gene sequence is 6.5×10-9 [55]. Maize-specific genes and genes with homologous could be obtained by homology analysis to analyze the evolutionary relationships between genes in the previous set of collections. First, the coding region sequence and protein sequence of each gene were obtained, and muscle was used to perform sequence alignment. Simultaneously calculate the Ks value and 4DTv value between the two genes; use treebest to optimize the sequence alignment results, build the gene tree and extract homologous relationships between genes. Gene cluster analysis could distinguish homologous genes into duplication genes and single-copy genes. The WGD gene was divided into the most recent, before the divergence of grass and before the divergence of monocotyledonous plants. The duplication gene generated by three WGD events was used as the basis for subsequent analysis.

In order to further exploring the evolution of duplication genes, the subsequent analysis is an in-depth analysis from two aspects: (1) Comparison between genes of different homolog relation. a), no homologous gene (No Paralog); b), no paralogous and orthologous single copy gene (Singleton); c), clustering only maize gene (All maize); e), Tandem duplication gene (Tandem); f), syntenic duplication gene (Syntenic Duplication), other types of duplication genes (Other Duplication). (2) Comparison between genes of different duplication events originating in syntenic duplication genes. a) the most recent WGD of maize (Maize WGD); b) Maize and sorghum divergence (Maize-Sorghum); c), the WGD of maize before the divergence of the grass (Pre-grasses WGD); d), WGD of maize before monocot divergence (Pre-Monocots WGD).

Gene component difference analysis of duplication genes used the transcript, and exon number, gene length, 5'UTR, 3'UTR, CDS, exon, and intron length of each gene were separately counted. Codon analysis of the gene uses the CDS sequence of the gene to calculate the GC and GC3 content of each gene and calculate the distribution of GC and GC3 content under the sliding window. The TATA Box and CAAT box were used to perform statistical analysis on the promoter region of 2 KB upstream of the gene. Using maize expression data published by Sekhon et al. in 2013 [56]. The differences between the genes of different origin maize genes and different WGD event genes were compared by statistical expression mean and expression levels of different tissues: roots, stems, leaves, seeds, embryos, pollen, filaments, and endosperm. Differentially expressed genes were analyzed by calculating fold changes. Functional annotations were performed using InterProScan [57], functional annotation of Gene Ontology (GO) [58], and enrichment analysis of gene ontology using BiNGO [59]. Metabolic pathway analysis was performed using the KEGG database [60].

List of abbreviations

WGD: Whole-genome duplication

Pre-grasses WGD: Whole-genome duplication before the divergence of grasses

Pre-Monocot WGD: whole-genome duplication before the divergence of monocotyledonous plants

MYA: million years ago

CDS: Coding sequence

UTRs: Untranslated regions

GO: Gene ontology

gBGC: GC-preferred gene transfer

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All relevant data are contained within the paper and its Additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

Author's contributions

References

1. Long M. Evolution of novel genes. Curr Opin Genet Dev. 2001;11:673–680.

2. Marques-Bonet T, Girirajan S, Eichler EE. The origins and impact of primate segmental duplications. Trends Genet. 2009;25:443–454.

3. Pan D, Zhang L. Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: a novel strategy to estimate gene duplication rates. Genome Biol. 2007;8:R158.

4. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol. 2003;18:292–8.

5. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci U S A. 2009;106:13875–9.

6. Kimura M. The neutral theory of molecular evolution: a review of recent evidence. Jpn J Genet. 1991;66:367–386.

7. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–1155.

8. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature. 2006;440:341–345.

9. Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999;151:1531–1545.

10. Hughes AL. The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B Biol Sci. 1994;256:119–124.

11. Näsvall J, Sun L, Roth JR, Andersson DI. Real-time evolution of new genes by innovation, amplification, and divergence. Science. 2012;338:384–387.

12. Ohno S. Evolution by gene duplication. Springer-Verlag; 1970.

13. Papp B, Pal C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. Nature. 2003;424:194–197.

14. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 2007;39:1256–1260.

15. Qian W, Zhang J. Gene dosage and gene duplicability. Genetics. 2008;179:2319–2324.

16. Bridges CB. The Bar" gene" a duplication. Sci N Y NY. 1936;83:210–211.

17. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 2005;3:1700.

18. Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. New Phytol. 2009;183:557–64.

19. Wolfe KH, Shields DC, others. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 1997;387:708–712.

20. Chen X, Zhang J. The Ortholog Conjecture Is Untestable by the Current Gene Ontology but Is Supported by RNA Sequencing Data. PLoS Comput Biol. 2012;8:e1002784.

21. Kaessmann H, Vinckenbosch N, Long M. RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet. 2009;10:19–31.

22. Makova KD, Li W-H. Divergence in the spatial pattern of gene expression between human duplicate genes. Genome Res. 2003;13:1638–1645.

23. Nei M, Niimura Y, Nozawa M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet. 2008;9:951–963.

24. Qian W, Liao B-Y, Chang AY-F, Zhang J. Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends Genet. 2010;26:425–430.

25. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon–intron structure. Proc Natl Acad Sci. 2012;109:1187–1192.

26. Ganko EW, Meyers BC, Vision TJ. Divergence in Expression between Duplicated Genes in Arabidopsis. Mol Biol Evol. 2007;24:2298–309.

27. Gu Z, Nicolae D, Lu HH, Li W-H. Rapid divergence in expression between duplicate genes inferred from microarray data. TRENDS Genet. 2002;18:609–613.

28. Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, et al. The fate of duplicated genes in a polyploid plant genome. Plant J. 2013;73:143–53.

29. Wang S, Chen Y. Fine-Tuning the Expression of Duplicate Genes by Translational Regulation in Arabidopsis and Maize. Front Plant Sci. 2019;10. doi:10.3389/fpls.2019.00534.

30. Guo X, Gui Y, Wang Y, Zhu Q-H, Helliwell C, Fan L. Selection and mutation on microRNA target sequences during rice evolution. BMC Genomics. 2008;9:454.

31. Sun C, Wu J, Liang J, Schnable JC, Yang W, Cheng F, et al. Impacts of Whole-Genome Triplication on MIRNA Evolution in Brassica rapa. Genome Biol Evol. 2015;7:3085–96.

32. Wang S, Adams KL. Duplicate Gene Divergence by Changes in MicroRNA Binding Sites in Arabidopsis and Brassica. Genome Biol Evol. 2015;7:646–55.

33. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473:97–100.

34. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, et al. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci. 2005;102:5454–9.

35. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The Sorghum bicolor genome and the diversification of grasses. Nature. 2009;457:551–6.

36. Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH, et al. Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. Genome Biol. 2009;10:R68.

37. Fischer I, Dainat J, Ranwez V, Glémin S, Dufayard J-F, Chantret N. Impact of recurrent gene duplication on adaptation of plant genomes. BMC Plant Biol. 2014;14:151.

38. Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc R Soc B Biol Sci. 2012;279:5048–57.

39. Panchy N, Lehti-Shiu M, Shiu S-H. Evolution of Gene Duplication in Plants. Plant Physiol. 2016;171:2294–316.

40. Qian W, Zhang J. Genomic evidence for adaptation by gene duplication. Genome Res. 2014;24:1356–62.

41. Christin P-A, Spriggs E, Osborne CP, Strömberg CAE, Salamin N, Edwards EJ. Molecular Dating, Evolutionary Rates, and the Age of the Grasses. Syst Biol. 2014;63:153–65.

42. Kellogg EA. The Grasses: A Case Study in Macroevolution. Annu Rev Ecol Syst. 2000;31:217–38.

43. Swigoňová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, et al. Close Split of Sorghum and Maize Genome Progenitors. Genome Res. 2004;14:1916–23.

44. Lee T-H, Tang H, Wang X, Paterson AH. PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res. 2013;41:D1152–8.

45. Lyons E, Freeling M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 2008;53:661–73.

46. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science. 2008;320:486–8.

47. Miele V, Penel S, Duret L. Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics. 2011;12:116.

48. Miele V, Penel S, Daubin V, Picard F, Kahn D, Duret L. High-quality sequence clustering guided by network topology and multiple alignment likelihood. Bioinformatics. 2012;28:1078–85.

49. Van Dongen SM. Graph clustering by flow simulation. 2001. http://dspace.library.uu.nl/handle/1874/848. Accessed 9 Mar 2017.

50. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.

51. Loewenstein Y, Portugaly E, Fromer M, Linial M. Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space. Bioinformatics. 2008;24:i41–9.

52. Jiang P, Singh M. SPICi: a fast clustering algorithm for large biological networks. Bioinformatics. 2010;26:1105–11.

53. Schreiber F, Patricio M, Muffato M, Pignatelli M, Bateman A. TreeFam v9: a new website, more species and orthology-on-the-fly. Nucleic Acids Res. 2014;42 Database issue:D922-925.

54. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49.

55. Gaut BS, Doebley JF. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci U S A. 1997;94:6809–14.

56. Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, et al. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PloS One. 2013;8:e61005.

57. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

58. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

59. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. Bioinformatics. 2005;21:3448–9.

60. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.

61. Yakovchuk P. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res. 2006;34:564–74.

62. Clément Y, Fustier M-A, Nabholz B, Glémin S. The Bimodal Distribution of Genic GC Content Is Ancestral to Monocot Species. Genome Biol Evol. 2014;7:336–48.

63. Tatarinova TV, Alexandrov NN, Bouck JB, Feldmann KA. GC3 biology in corn, rice, sorghum and other grasses. BMC Genomics. 2010;11:308.

64. Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, et al. Genome-wide atlas of transcription during maize development. Plant J Cell Mol Biol. 2011;66:553–63.

65. Heilmann I. Phosphoinositide signaling in plant development. Development. 2016;143:2044–55.

66. Bigeard J, Hirt H. Nuclear Signaling of Plant MAPKs. Front Plant Sci. 2018;9. doi:10.3389/fpls.2018.00469.

67. Danquah A, de Zelicourt A, Colcombet J, Hirt H. The role of ABA and MAPK signaling pathways in plant abiotic stress responses. Biotechnol Adv. 2014;32:40–52.

68. Meng X, Zhang S. MAPK Cascades in Plant Disease Resistance Signaling. Annu Rev Phytopathol. 2013;51:245–66.

69. Taj G, Agarwal P, Grant M, Kumar A. MAPK machinery in plants. Plant Signal Behav. 2010;5:1370–8.

70. Liu Y, Zhang D, Wang L, Li D. Genome-Wide Analysis of Mitogen-Activated Protein Kinase Gene Family in Maize. Plant Mol Biol Report. 2013;31:1446–60.

71. Zhang A, Zhang J, Ye N, Cao J, Tan M, Zhang J, et al. ZmMPK5 is required for the NADPH oxidase-mediated self-propagation of apoplastic H2O2 in brassinosteroid-induced antioxidant defence in leaves of maize. J Exp Bot. 2010;61:4399–411.

72. Ding Y, Cao J, Ni L, Zhu Y, Zhang A, Tan M, et al. ZmCPK11 is involved in abscisic acid-induced antioxidant defence and functions upstream of ZmMPK5 in abscisic acid signalling in maize. J Exp Bot. 2013;64:871–84.

73. Wei F, Coe ED, Nelson W, Bharti AK, Engler F, Butler E, et al. Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet. 2007;3:e123.

74. Blanc G, Wolfe KH. Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. Plant Cell. 2004;16:1667–78.

75. Group GPW, Barker NP, Clark LG, Davis JI, Duvall MR, Guala GF, et al. Phylogeny and Subfamilial Classification of the Grasses (Poaceae). Ann Mo Bot Gard. 2001;88:373–457.

76. Grant D, Cregan P, Shoemaker RC. Genome organization in dicots: genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. Proc Natl Acad Sci U S A. 2000;97:4168–73.

77. Mayer K, Murphy G, Tarchini R, Wambutt R, Volckaert G, Pohl T, et al. Conservation of Microstructure between a Sequenced Region of the Genome of Rice and Multiple Segments of the Genome of Arabidopsis thaliana. Genome Res. 2001;11:1167–74.

78. Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr Opin Plant Biol. 2012;15:131–139.

79. Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang C-X, et al. Comparative Genomics of Plant Chromosomes. Plant Cell. 2000;12:1523–40.

80. Song K, Lu P, Tang K, Osborn TC. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc Natl Acad Sci U S A. 1995;92:7719–23.

81. Vanneste K, Van de Peer Y, Maere S. Inference of genome duplications from age distributions revisited. Mol Biol Evol. 2013;30:177–90.

82. Rabier C-E, Ta T, Ané C. Detecting and Locating Whole Genome Duplications on a Phylogeny: A Probabilistic Approach. Mol Biol Evol. 2014;31:750–62.

83. Hegarty MJ, Hiscock SJ. Genomic Clues to the Evolutionary Success of Polyploid Plants. Curr Biol. 2008;18:R435–44.

84. Barrier M, Baldwin BG, Robichaux RH, Purugganan MD. Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. Mol Biol Evol. 1999;16:1105–13.

85. Glémin S, Clément Y, David J, Ressayre A. GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. Trends Genet TIG. 2014;30:263–70.

86. Meunier J, Duret L. Recombination Drives the Evolution of GC-Content in the Human Genome. Mol Biol Evol. 2004;21:984–90.

87. Muyle A, Serres-Giardi L, Ressayre A, Escobar J, Glémin S. GC-Biased Gene Conversion and Selection Affect GC Content in the Oryza Genus (rice). Mol Biol Evol. 2011;28:2695–706.

88. Šmarda P, Bureš P, Horová L, Leitch IJ, Mucina L, Pacini E, et al. Ecological and evolutionary significance of genomic GC content diversity in monocots. Proc Natl Acad Sci. 2014;111:E4096–102.

89. Shi X, Wang X, Li Z, Zhu Q, Yang J, Ge S, et al. Evidence that Natural Selection is the Primary Cause of the Guanine-cytosine Content Variation in Rice Genes. J Integr Plant Biol. 2007;49:1393–9.

90. Clément Y, Arndt PF. Meiotic recombination strongly influences GC-content evolution in short regions in the mouse genome. Mol Biol Evol. 2013;30:2612–8.

91. Dreszer TR, Wall GD, Haussler D, Pollard KS. Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. Genome Res. 2007;17:1420–30.

92. Katzman S, Capra JA, Haussler D, Pollard KS. Ongoing GC-biased evolution is widespread in the human genome and enriched near recombination hot spots. Genome Biol Evol. 2011;3:614–26.

93. Glémin S, Clément Y, David J, Ressayre A. GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. Trends Genet. 2014;30:263–70.

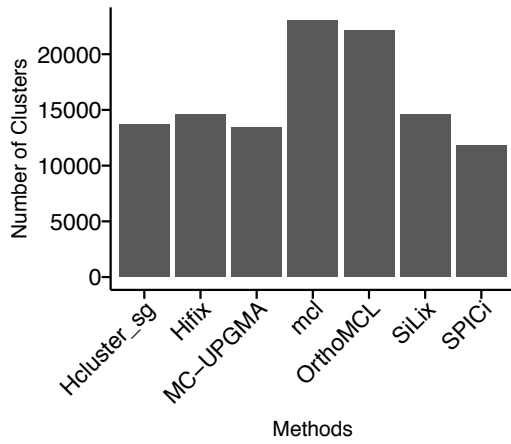94. Royer DL, Berner RA, Montañez IP, Tabor NJ, Beerling DJ. CO2 as a primary driver of

Phanerozoic climate. GSA Today. 2004;14:4.

95. Irish VF, Litt A. Flower development and evolution: gene duplication, diversification and redeployment. Curr Opin Genet Dev. 2005;15:454–60.
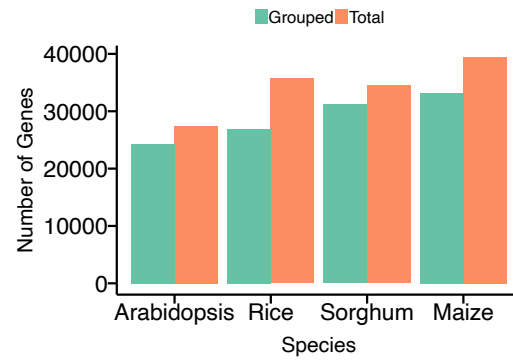
96. Adams KL. Evolution of Duplicate Gene Expression in Polyploid and Hybrid Plants. J Hered. 2007;98:136–41.

97. Dong S, Adams KL. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. New Phytol. 2011;190:1045–57.
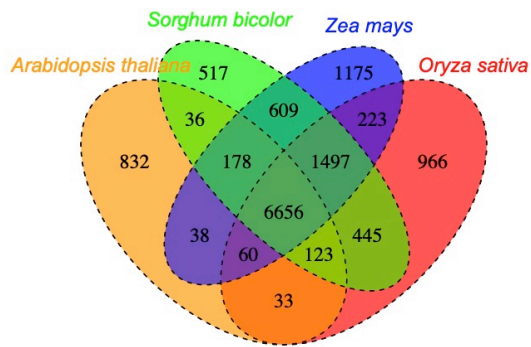
(a)

(b)

(c)

Figure 1. The number of collections of each clustering methods and the number and proportion of genes in the candidate dataset of each species. (a) The number of collections obtained by each clustering method, in which mcl and OrthoMCL obtained the most collections; (b) the number of genes in candidate dataset and total genes in each species; (c) The gene number of new collections distributed in species.
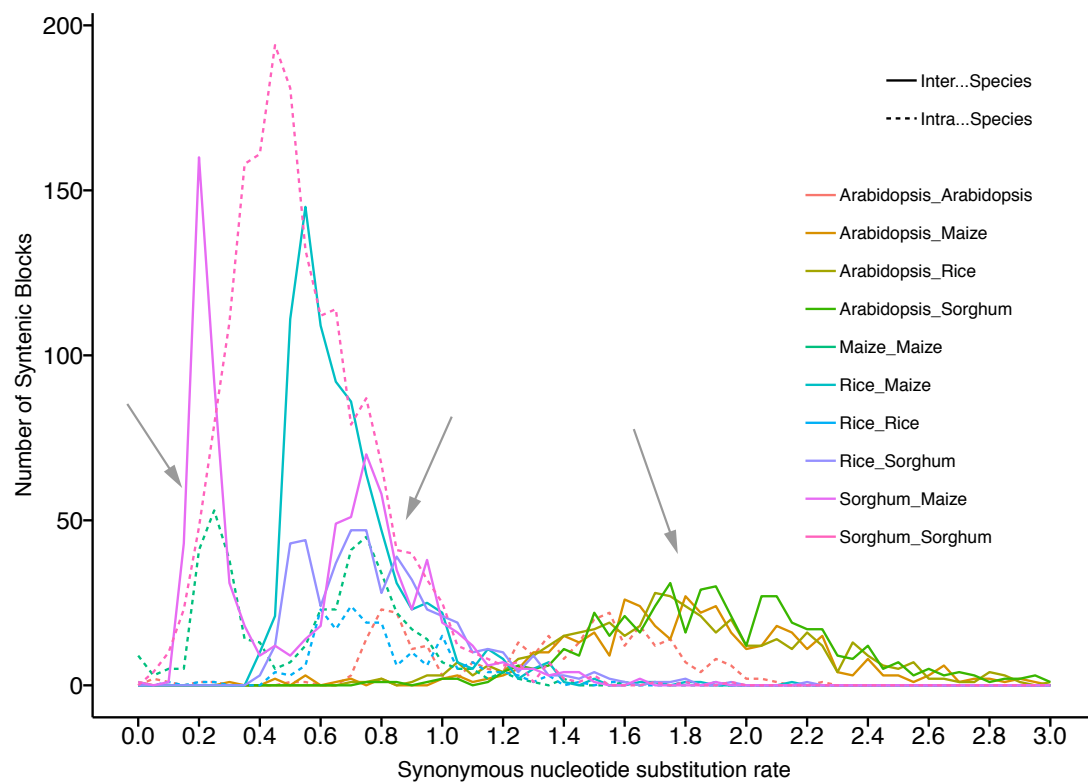
Figure 2. Distribution of Ks between collinear blocks of maize, sorghum, rice, and Arabidopsis. (Gray arrow points to speculative three genome duplication events in maize)
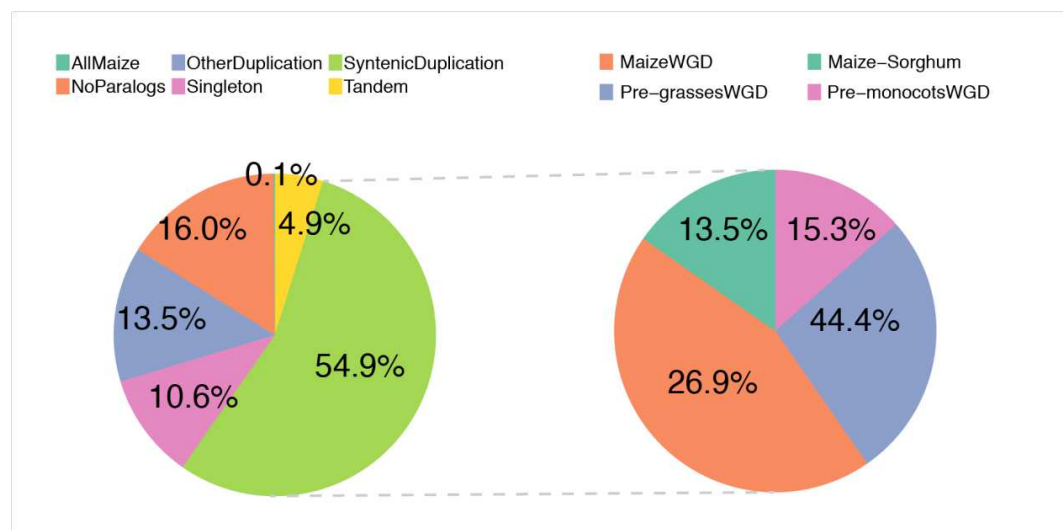

Figure 3. The proportion of identified maize genes with different origins and the proportion of genes involved in different whole duplication events.

Table 1. Gene components of different classes of genes in maize.

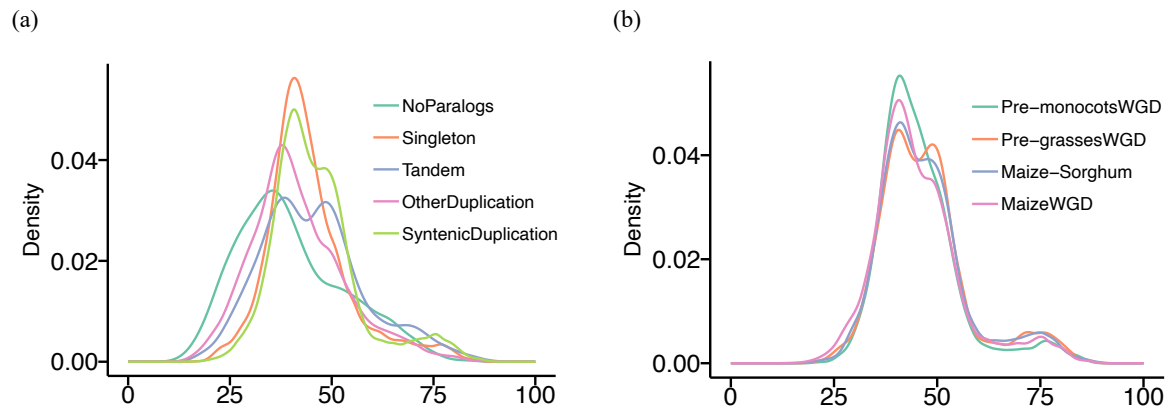| Type	Number of Transcripts	Length of Gene	Length of 5'UTR	Length of 3'UTR	Length of CDS	Number of exons	Length of exon	Length of intron
NoParalogs	1.2	2412	228	282	548	3	863	1549
OtherDuplication	1.3	3377	246	270	864	4.1	1197	2179
Singleton	1.8	5364	285	413	1206	5.9	1662	3702
SyntenicDuplication	1.8	4495	312	359	1314	5.5	1787	2707
Tandem	1.3	2333	205	322	849	2.7	1191	1142

(a)　　　　　　　　　　　　　　　　　(b)

Figure 4. Distribution of GC3/GC for maize genes. (a/b) Distribution of GC3 of different types of duplication genes in maize;

Table 2. Enrichment analysis of KEGG metabolic pathways of WGD genes in maize

Pathway	P-value	x	X	n	N	Description
zma00966	1.86E-03	5	1768	7	6672	Glucosinolate biosynthesis
zma03010	1.42E-07	176	1768	480	6672	Ribosome
zma03015	1.57E-06	75	1768	178	6672	mRNA surveillance pathway
zma03040	3.54E-04	97	1768	274	6672	Spliceosome
zma04016	3.64E-07	83	1768	196	6672	MAPK signaling pathway - plant
zma04075	5.98E-05	128	1768	363	6672	Plant hormone signal transduction
zma04136	4.54E-04	28	1768	62	6672	Autophagy - other
zma04141	1.85E-09	122	1768	291	6672	Protein processing in endoplasmic reticulum
zma04712	1.01E-06	39	1768	76	6672	Circadian rhythm - plant

Note: N: number of genes, n: number of genes involved in current metabolic pathways; X: number of Maize WGD genes, x: number of Maize WGD genes involved in current metabolic pathways.
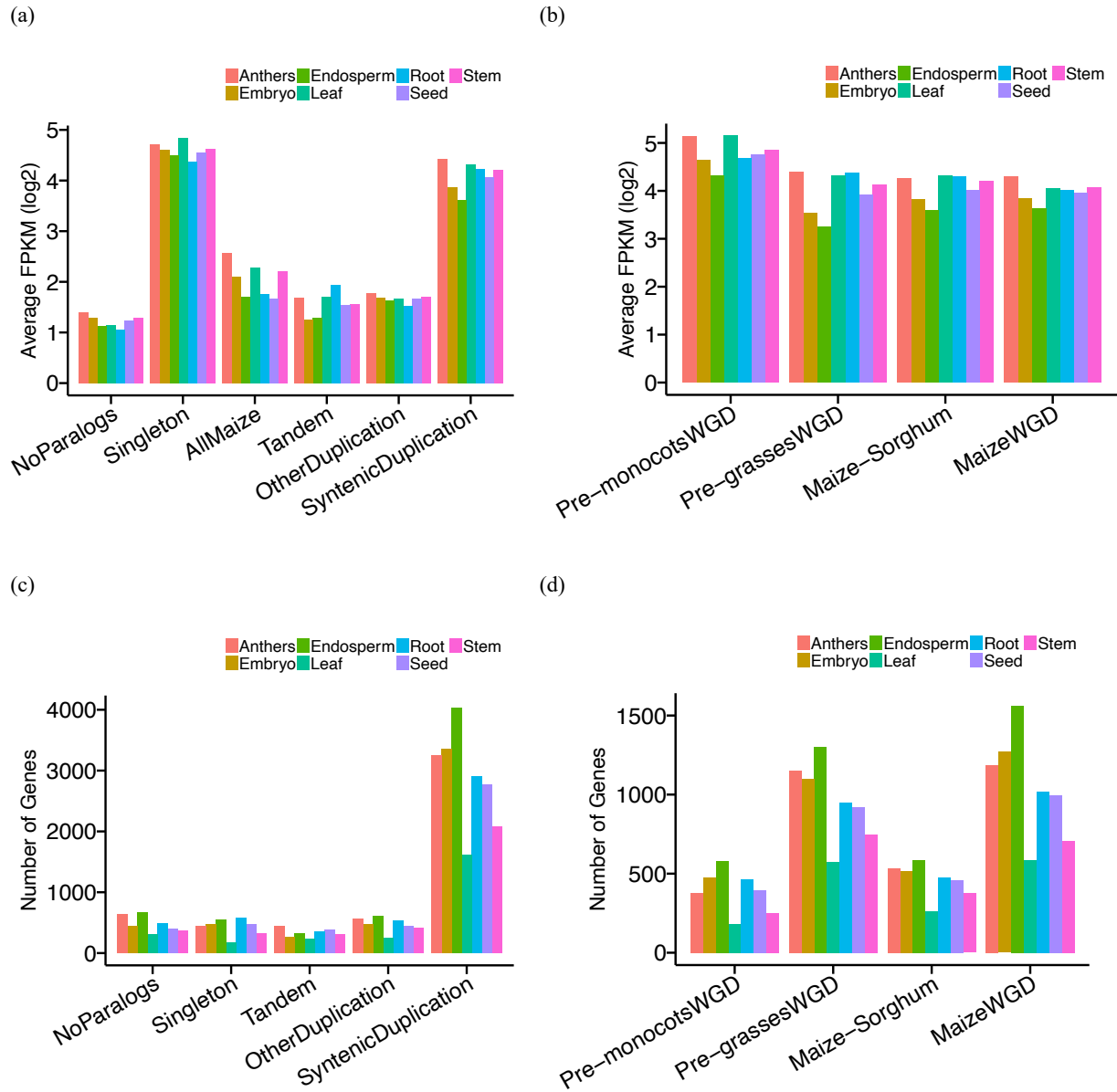
Fig. 5. Transcriptional level and the number of maize genes in different tissues. (a/b) Average expression of different types of duplication genes in maize in different tissues; (c/d) The number of differential expressions of different types of duplication genes in maize.

Table 3. Duplication type and GC content of maize MAPK genes.

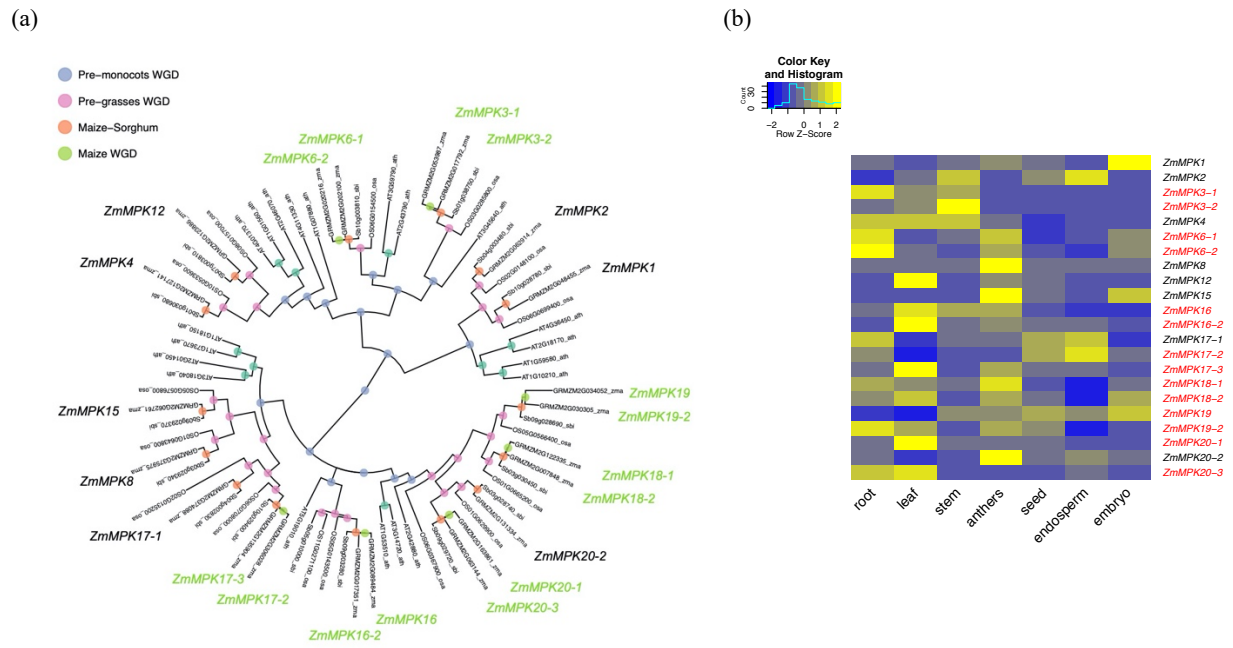| Gene ID	Common Name	Duplication Type	Length of CDS	GC3%	GC%
GRMZM2G053987	ZmMPK3-1	MaizeWGD	1125	59.3	62.9
GRMZM2G017792	ZmMPK3-2	MaizeWGD	1131	76.0	63.9
GRMZM2G002100	ZmMPK6-1	MaizeWGD	1197	47.3	49.2
GRMZM2G020216	ZmMPK6-2	MaizeWGD	1197	49.3	50.0
GRMZM2G089484	ZmMPK16	MaizeWGD	1674	39.1	45.5
GRMZM2G017351	ZmMPK16-2	MaizeWGD	933	39.2	45.4
GRMZM2G306028	ZmMPK17-2	MaizeWGD	1542	48.7	46.8
GRMZM2G135904	ZmMPK17-3	MaizeWGD	1335	42.1	49.9
GRMZM2G122335	ZmMPK18-1	MaizeWGD	1767	42.3	45.6
GRMZM2G007848	ZmMPK18-2	MaizeWGD	1770	43.0	44.9
GRMZM2G034052	ZmMPK19	MaizeWGD	1740	44.9	49.5
GRMZM2G030305	ZmMPK19-2	MaizeWGD	597	45.3	46.7
GRMZM2G163861	ZmMPK20-1	MaizeWGD	1809	40.3	48.9
GRMZM2G063144	ZmMPK20-3	MaizeWGD	1185	37.9	45.7
GRMZM2G048455	ZmMPK1	Pre-grassesWGD	1110	50.9	55.5
GRMZM2G062914	ZmMPK2	Pre-grassesWGD	1113	44.9	47.7
GRMZM2G127141	ZmMPK4	Pre-grassesWGD	1119	41.0	49.2
GRMZM2G375975	ZmMPK8	Pre-grassesWGD	1470	37.6	51.1
GRMZM2G123886	ZmMPK12	Pre-grassesWGD	1176	52.8	60.5
GRMZM2G062761	ZmMPK15	Pre-grassesWGD	1806	42.1	49.9
GRMZM2G374088	ZmMPK17-1	Pre-grassesWGD	1476	39.3	48.5
GRMZM2G131334	ZmMPK20-2	Pre-grassesWGD	1842	40.2	45.4

Figure 6. MAPK genes in maize. (a) Phylogenetic tree of MAPK genes; (b) Expression of maize MAPK genes in different tissues (Red indicates Maize WGD gene)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile.pdf