

RESEARCH ARTICLE

Open Access

# Relationship, evolutionary fate and function of two maize co-orthologs of rice *GW2* associated with kernel size and weight

Qing Li<sup>1</sup>, Lin Li<sup>1</sup>, Xiaohong Yang<sup>1</sup>, Marilyn L Warburton<sup>2</sup>, Guanghong Bai<sup>1,3</sup>, Jingrui Dai<sup>1</sup>, Jiansheng Li<sup>1\*</sup>, Jianbing Yan<sup>1,4\*</sup>

## Abstract

**Background:** In rice, the *GW2* gene, found on chromosome 2, controls grain width and weight. Two homologs of this gene, *ZmGW2-CHR4* and *ZmGW2-CHR5*, have been found in maize. In this study, we investigated the relationship, evolutionary fate and putative function of these two maize genes.

**Results:** The two genes are located on duplicated maize chromosomal regions that show co-orthologous relationships with the rice region containing *GW2*. *ZmGW2-CHR5* is more closely related to the sorghum counterpart than to *ZmGW2-CHR4*. Sequence comparisons between the two genes in eight diverse maize inbred lines revealed that the functional protein domain of both genes is completely conserved, with no non-synonymous polymorphisms identified. This suggests that both genes may have conserved functions, a hypothesis that was further confirmed through linkage, association, and expression analyses. Linkage analysis showed that *ZmGW2-CHR4* is located within a consistent quantitative trait locus (QTL) for one-hundred kernel weight (HKW). Association analysis with a diverse panel of 121 maize inbred lines identified one single nucleotide polymorphism (SNP) in the promoter region of *ZmGW2-CHR4* that was significantly associated with kernel width (KW) and HKW across all three field experiments examined in this study. SNPs or insertion/deletion polymorphisms (InDels) in other regions of *ZmGW2-CHR4* and *ZmGW2-CHR5* were also found to be significantly associated with at least one of the four yield-related traits (kernel length (KL), kernel thickness (KT), KW and HKW). None of the polymorphisms in either maize gene are similar to each other or to the 1 bp InDel causing phenotypic variation in rice. Expression levels of both maize genes vary over ear and kernel developmental stages, and the expression level of *ZmGW2-CHR4* is significantly negatively correlated with KW.

**Conclusions:** The sequence, linkage, association and expression analyses collectively showed that the two maize genes represent chromosomal duplicates, both of which function to control some of the phenotypic variation for kernel size and weight in maize, as does their counterpart in rice. However, the different polymorphisms identified in the two maize genes and in the rice gene indicate that they may cause phenotypic variation through different mechanisms.

## Background

The genetic improvement of grain yield in major cereals has traditionally been one of the most important contributions to an increased global food supply. Gains via phenotypic selection have been steady but slow over the years, and the ever-growing world population makes it

necessary to increase the rate of gain in grain yields over what has been achieved in the past. Grain size and weight are important components of grain yield, but the genetic basis of these traits in maize (*Zea mays*), one of the most important global food staples, is insufficiently understood. To date, several loci have been shown to affect maize kernel development through mutant analysis [1-5], but only one gene (*GS3*) has been found to affect kernel size and weight in a natural population of maize [6]. With the successful completion of the B73

\* Correspondence: lijiansheng@cau.edu.cn; yjianbing@gmail.com

<sup>1</sup>National Maize Improvement Center of China, Key Laboratory of Crop Genomics and Genetic Improvement (Ministry of Agriculture), China Agricultural University, 100193 Beijing, China

genome sequencing project, more than 32,000 genes have been identified [7]; this provides a good opportunity for QTL cloning and function verification. However, the identification of genes related to grain yield is still a great challenge because of complexity of this trait. QTLs cloned to date suggest that genes with various functions can affect grain yield, including genes involved in protein degradation [8,9], hormone metabolism [10,11], and other processes [12,13]. So many genetic factors affect final grain yield that it has even been suggested that the entire genome may be involved [14].

Comparative QTL mapping studies have shown that some QTL for many traits, including grain yield, are located on collinear chromosomes in different species [15-18]. This suggests that mutations in orthologous genes contribute to similar trait variation. If the current, similar function of the genes was gained following divergence of the different species being compared, the mutations will be completely independent, and may be dissimilar in nature. This has been seen in the well-known "green revolution" genes *Rht* in wheat, *GAI* in *Arabidopsis* and *Dwarf8* in maize, which all contribute to short plant stature [19].

The availability of the rice genomic sequences [20,21] facilitated the identification of many genes controlling grain size and weight [9,22-25]. The *GW2* gene was the first gene controlling grain width to be cloned in rice. It encodes a RING-type protein with E3 ubiquitin ligase activity, and functions as a negative regulator of grain width and weight. A 1 bp deletion in the fourth exon resulting in a premature stop codon and a truncation of 310 amino acids causes the increase in rice grain width and weight [9]. Only one copy of *GW2* exists in rice, but a search of the latest B73 genomic sequence identified two homologous gene sequences [7]. Whether both maize genes represent co-orthologs of the rice *GW2* gene has not been tested. Besides, based on the conserved function found for various other genes, it is reasonable to suppose that the maize orthologs of rice *GW2* may also affect grain size and weight, which needs to be tested further.

Association analysis can identify genetic polymorphisms that are associated with phenotypic variation. Compared with linkage analysis, it is time- and cost-effective. More importantly, it can investigate more than two alleles at the same time and can reach extraordinarily high resolution in species with rapid linkage disequilibrium (LD) decay [26]. Maize contains abundant genetic diversity and LD decays within 2 kb in diverse material [27-29]. This rapid LD decay pattern makes maize an ideal plant for association analysis to identify causal polymorphisms or closely linked polymorphisms, which can then be used to develop functional markers for marker-assisted selection [30]. To date, association analysis

has been widely used in maize to dissect the genetic basis of complex traits, such as kernel carotenoid content [30], kernel starch content [31], kernel quality traits [32] and flowering time [33].

The objectives of this study were to clarify the relationship of the two maize genes that were found to be homologous to the rice *GW2* gene; to investigate their evolutionary fate following the duplication of these genes in maize; and to characterize the contribution and putative function of these two genes in maize grain yield-related traits.

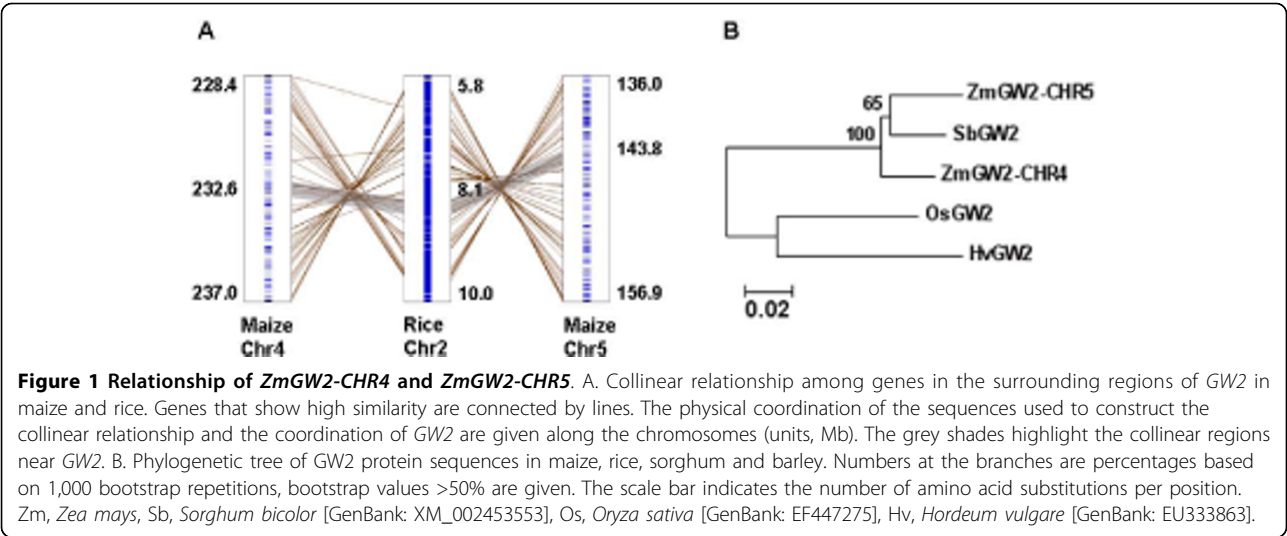
## Results

### The rice *GW2* gene has two co-orthologs located on duplicated chromosomes in maize

Blast searches with the rice *GW2* protein sequence [GenBank: ABO31101] against the maize high throughput genomic sequence database [34] identified two maize bacterial artificial chromosome (BAC) clones, AC212189 on chromosome 4 and AC211190 on chromosome 5, that contain sequences showing high similarity to the rice protein. The structures of these two genes were determined using three maize complementary DNA (cDNA) clones from GenBank, EU968771, FJ573211 and EU962093, which showed high sequence similarity to the two BAC clones. Both genes consist of eight exons, with an overall sequence similarity of 94% to each other and 93% to the rice *GW2* gene across the coding region. They were named *ZmGW2-CHR4* and *ZmGW2-CHR5*, based on their locations on the maize chromosomes (See additional file 1: Similarity between *ZmGW2-CHR4* and *ZmGW2-CHR5* across the cDNA region).

Because the maize genome has been shown to be replete with duplicated chromosomal regions [7,35], we investigated if the two homologs represent duplicated genes. As shown in Figure 1A, other gene sequences in the vicinity of the two maize genes also show high similarity, as would arise following an ancient chromosome duplication event. Comparison of the regions around both maize genes with the region containing rice *GW2* showed that both maize regions are collinear with the rice region, indicating that the two maize genes are co-orthologs of the rice *GW2* gene. Besides, compared to the rice region containing *GW2*, both maize regions contain an inversion (Figure 1A).

Phylogenetic analysis of the duplicated maize genes and their corresponding counterparts in sorghum, rice and barley showed that *ZmGW2-CHR5* is more closely related to the sorghum counterpart than to the other maize copy (Figure 1B), indicating that sequence divergence has occurred between the two maize genes. To identify the fixed polymorphic sites between the maize genes, we sequenced the coding regions of both genes



in eight diverse maize lines belonging to five heterotic groups. In total, we found 70 fixed sites, of which 44 were synonymous mutations (Table 1). All polymorphisms but two were SNPs. The 70 sites are not evenly distributed across the entire coding region, as 60% of the polymorphisms occurred in exon 8, and no polymorphisms were found in exon 2 (Table 1). Only three SNPs were found in the RING domain, all of which were synonymous mutations, suggesting that this domain may have a conserved function in both genes.

***ZmGW2-CHR4* is located within a QTL for HKW**

*ZmGW2-CHR4* and *ZmGW2-CHR5* were located in maize chromosomal bins 4.09 and 5.04, respectively. Previous studies have mapped many QTL for kernel weight to these two regions (See additional file 2: QTL for grain yield mapped in previous studies in maize bins 4.09 and 5.04). In particular, in an F<sub>2:3</sub> population [37] and an immortalized F<sub>2</sub> (IF<sub>2</sub>) population developed by our lab at the China Agricultural University [40,41], a QTL for HKW was mapped to bin 4.09 (Figure 2A; Table 2). *ZmGW2-CHR4* was mapped to 248 cM on chromosome 4, between simple sequence repeat (SSR) markers bnlg292 and umc1173, and within the QTL confidence interval for HKW, near the left border of the

QTL (Figure 2A; Table 2). This QTL was identified stably across four seasons and explained 3.2% to 6.9% of the phenotypic variation in the IF<sub>2</sub> population derived from inbred lines Zong3 and 87-1. The Zong3 allele can increase HKW from 0.5 to 1.3 g (Table 2). Although the large QTL confidence interval (10.4 ~ 18.7 cM) contained lots of genes, the involvement of *GW2* in HKW in rice and the co-location of *ZmGW2-CHR4* with this QTL indicated that *ZmGW2-CHR4* is a good candidate for this QTL and may be involved in grain weight variation.

**Genetic diversity and LD of the two genes across maize inbred lines**

The genetic diversity within the RING domain was not analyzed because it is well conserved across the two maize genes (Table 1). Instead, we analyzed three regions of comparable sizes in corresponding regions of the two maize genes (Figure 3; Figure 4). The average levels of nucleotide diversity ( $\pi$ ) in the two genes are comparable, both having about 4 nucleotide differences per 1,000 sites between two random sequences across the entire genes (Table 3). However, the nucleotide diversity is not evenly distributed across both genes. While the 5' end showed the most abundant diversity in

**Table 1 Distribution of sequence polymorphisms between *ZmGW2-CHR4* and *ZmGW2-CHR5***

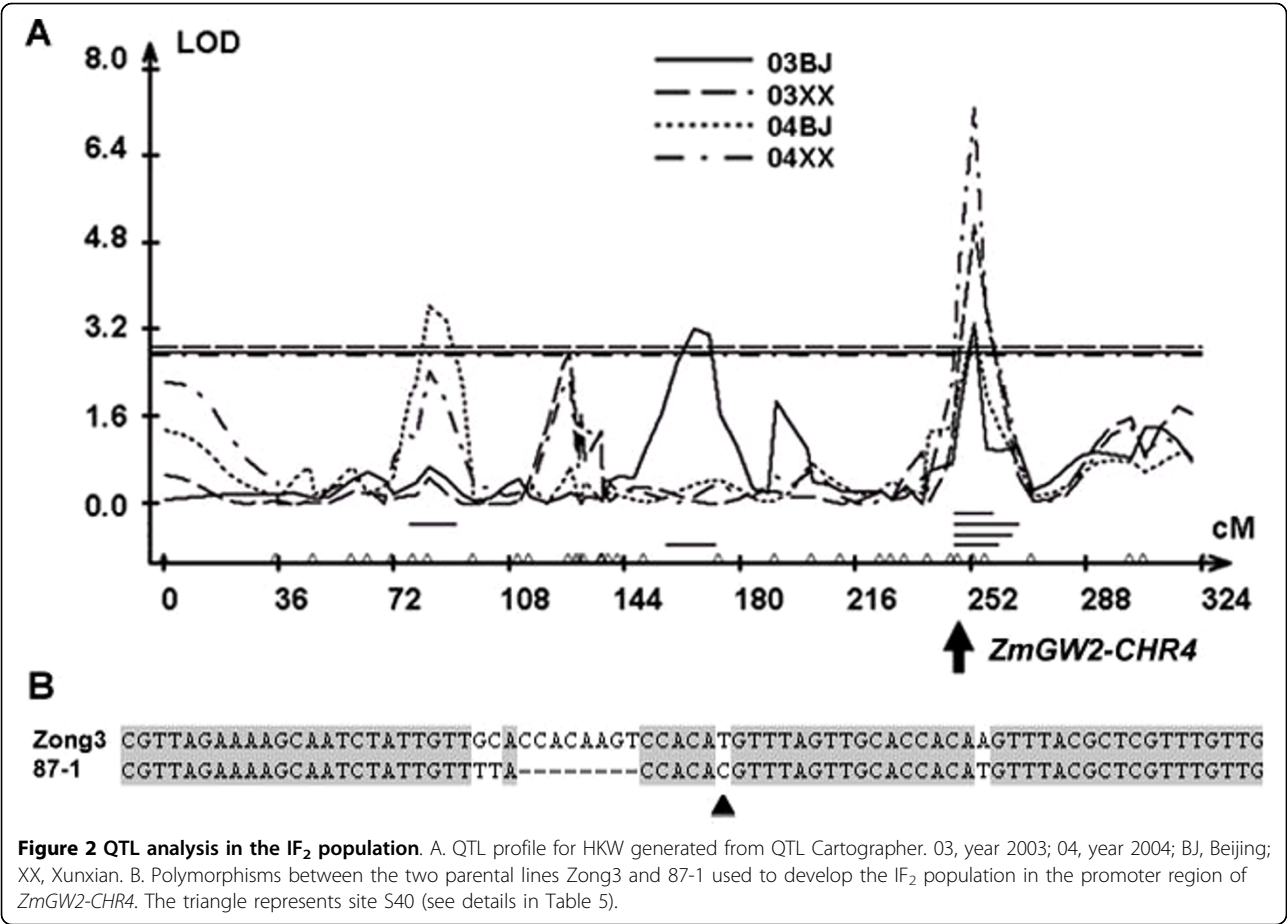
Exons	E1-1 <sup>a</sup>	E1-2 <sup>b</sup>	E2 <sup>b</sup>	E3 <sup>b</sup>	E4-1 <sup>b</sup>	E4-2	E5	E6	E7	E8 <sup>c</sup>	Total
Synonymous	6	1	0	2	0	2	6	2	0	25	44
Non-synonymous	4	0	0	0	0	2	0	1	2	17	26
Total	10	1	0	2	0	4	6	3	2	42	70

E, exon

<sup>a</sup> Including a 6 bp InDel

<sup>b</sup> The location of the RING domain, which begins from within the end of exon 1 (E1-2) and extends to within the beginning of exon 4 (E4-1)

<sup>c</sup> Including a 3 bp InDel



*ZmGW2-CHR4* ( $\pi = 7.7 \times 10^{-3}$ ), this region showed the fewest polymorphisms in *ZmGW2-CHR5* ( $\pi = 2.6 \times 10^{-3}$ ). The middle portion of *ZmGW2-CHR4* has only two polymorphisms, while the corresponding region of *ZmGW2-CHR5* has 28 polymorphisms and the highest level of genetic diversity. Tajima's *D* statistic was calculated to determine whether the two genes were subjected to selective constraints. As shown in Table 3, it appears that neither gene has been the subject of natural selection when the entire gene sequences were analyzed. Because the selection effect may not extend throughout the entire gene [42], we calculated Tajima's *D* statistic

separately for the three regions in each gene. The results showed that the middle portion of *ZmGW2-CHR4* (Figure 3) has a significant positive Tajima's *D* value, suggesting the presence of selection at this region, which is consistent with the observed low level of nucleotide diversity (Table 3).

The LD decay patterns of the two genes shown in Figures 3 and 4 indicate that both genes contain discrete LD blocks. In *ZmGW2-CHR4*, a large LD block was observed at the 5' end (Figure 3), and in *ZmGW2-CHR5*, a large LD block was observed in the middle portion (Figure 4). The only two polymorphisms in the

**Table 2 QTL for HKW mapped in the IF<sub>2</sub> population**

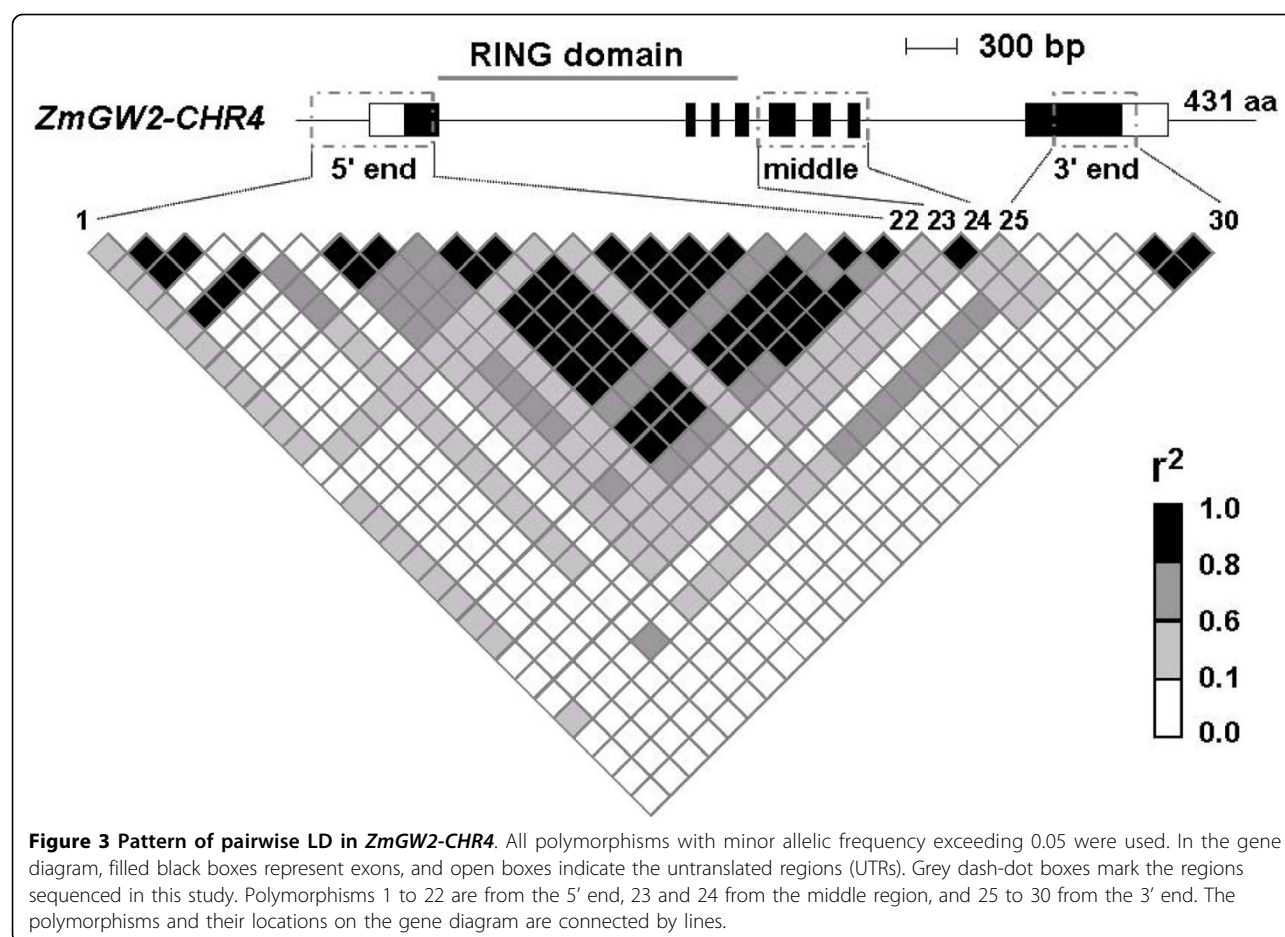
Environments	Position (cM)	Confidence interval (cM)	LOD	Additive <sup>a</sup>	Dominant <sup>b</sup>	R <sup>2</sup> (%) <sup>c</sup>
03BJ	252.5	248.0-260.4	3.3	0.7	0.5	3.3
03XX	252.5	248.0-264.3	5.1	1.2	1.1	5.1
04BJ	252.5	248.0-266.7	3.3	0.5	0.6	3.2
04XX	252.5	248.0-258.4	7.3	1.3	0.9	6.9

03, year 2003; 04, year 2004; BJ, Beijing; XX, Xunxian

<sup>a</sup> Positive values indicate that the Zong3 allele increases HKW

<sup>b</sup> Positive values indicate that the heterozygote has higher phenotypic value than the mean of the two homozygotes

<sup>c</sup> Phenotypic variation explained by each QTL



middle portion of *ZmGW2-CHR4* are in complete LD; thus, only two haplotypes were observed (Table 3; Figure 3), consistent with the observed selection and reduced nucleotide diversity in this region. Although LD extends in each of the three regions in both genes (less than 800 bp), LD among the three regions within both genes was within 2,000 bp (Figure 3; Figure 4), consistent with previous results [27-29]. We further investigated LD between the two genes, only 1.4% of  $r^2$  values for all pairs of polymorphisms was greater than 0.10, and the largest  $r^2$  was less than 0.16.

#### Association analysis of four yield-related traits

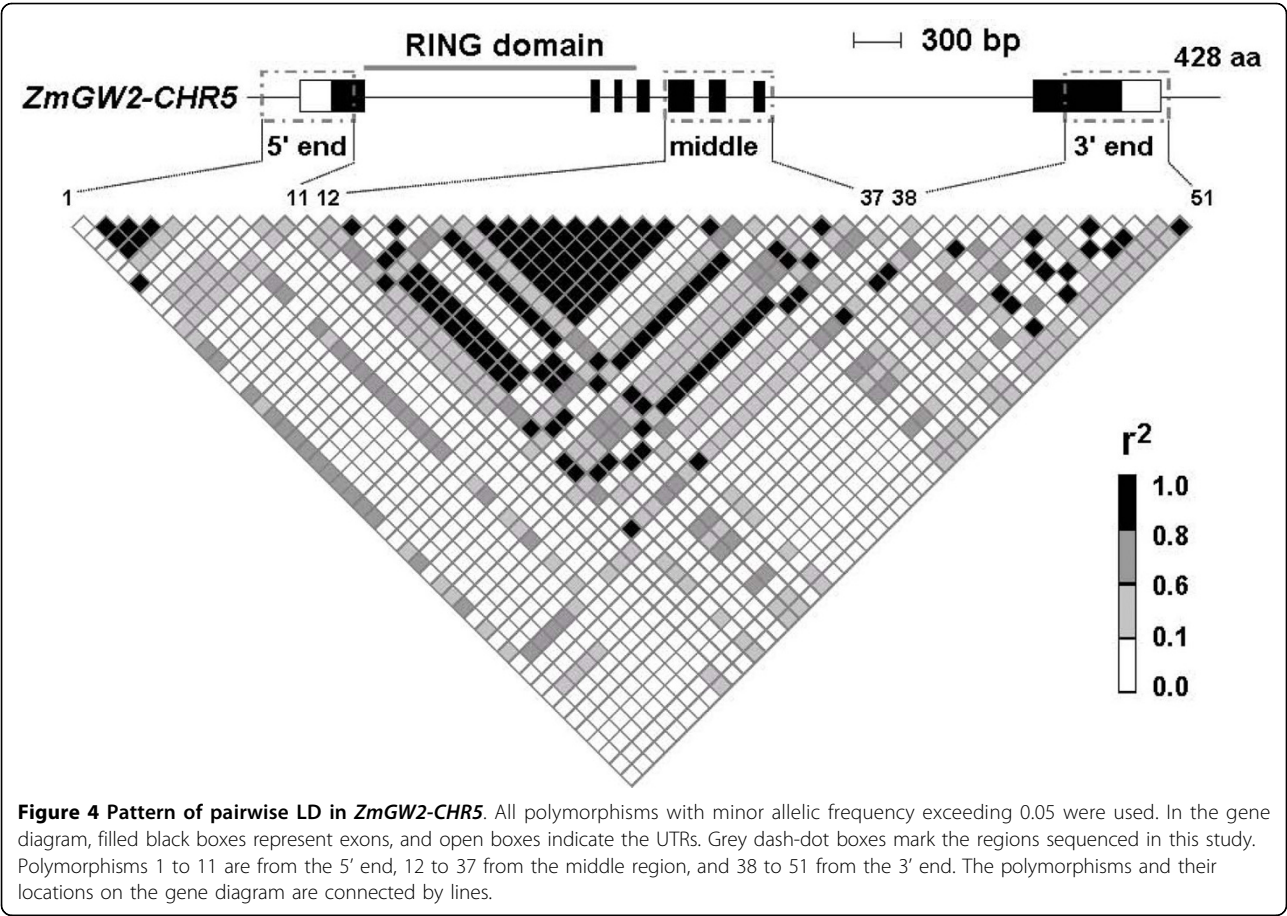
Analysis of variance (ANOVA) showed significant phenotypic variation for all four yield-related traits among the maize lines studied (Table 4), indicating that the assembled panel is suitable for association analysis. A significant year by genotype effect was observed for all four traits. The phenotypic distributions ranged from 6.48 to 11.58 mm for KL, 5.65 to 10.57 mm for KW, 3.14 to 6.73 mm for KT and 7.87 to 42.86 g for HKW; with an average of 8.90 mm, 8.12 mm, 4.70 mm and 22.50 g, respectively (Table 4). Significant positive

phenotypic and genetic correlations between KW and KT, and between kernel size traits (KL, KW and KT) and HKW were observed (Table 4), indicating that an increase in any of the three kernel size traits can increase HKW, and thus perhaps grain yield.

The mixed model controlling for population structure (Q) and kinship (K) as estimated using molecular markers [43] was employed to test associations between the four yield-related traits and polymorphisms from the two maize genes. Out of 360 (30 sites  $\times$  4 traits  $\times$  3 years) and 612 possible associations (51 sites  $\times$  4 traits  $\times$  3 years), 81 and 49 associations were significant for *ZmGW2-CHR4* and *ZmGW2-CHR5*, respectively, at  $P \leq 0.05$ ; while at  $P \leq 0.01$ , 33 and 22 associations remained significant, indicating that the observed associations were not expected only by chance.

Taking the LD ( $r^2 > 0.8$ ) level among sites into account, seven and five sites from *ZmGW2-CHR4* and *ZmGW2-CHR5*, respectively, were significantly associated with at least one of the four yield-related traits at  $P \leq 0.01$ . Information on the location, genotype, frequency and probability value for each site can be found in Tables 5 and 6. The S40 site from *ZmGW2-CHR4*





**Table 3 Nucleotide diversity and neutrality test of *ZmGW2-CHR4* and *ZmGW2-CHR5***

Gene	Region	N <sup>a</sup>	Size (bp) <sup>b</sup>	S <sup>c</sup>	H <sup>d</sup>	$\pi$ ( $\times 10^{-3}$ ) <sup>e</sup>	Tajima's D <sup>f</sup>
<i>ZmGW2-CHR4</i>	5' end	116	689	22	11	7.7	0.80
	middle	118	665	2	2	1.5	2.51*
	3' end	115	495	5	6	1.8	-0.08
	Total	112	1849	29	21	3.9	0.97
<i>ZmGW2-CHR5</i>	5' end	111	585	9	9	2.6	-0.47
	middle	117	694	28	7	6.3	-0.51
	3' end	120	658	12	7	3.4	0.01
	Total	109	1937	46	19	4.2	-0.27

\*, P < 0.05

<sup>a</sup> The number of analyzed maize inbred lines

<sup>b</sup> The total number of sites (excluding sites with gaps/missing data)

<sup>c</sup> The number of segregating sites (excluding InDels)

<sup>d</sup> The number of haplotypes

<sup>e</sup> The number of nucleotide differences per site between two randomly chosen sequences

<sup>f</sup> The D value calculated according to the polymorphic sites

was of great interest because it showed associations with KW and HKW across all three field experiments (Table 5), and explained 8.0% and 11.4% of the phenotypic variation for KW and HKW, respectively in Beijing in 2007. This site also segregated in the IF<sub>2</sub> population and mapped to a region where a QTL for HKW was identified (Figure 2B). Three other sites from *ZmGW2-CHR4* (S27, S304 and S1730) and one site from *ZmGW2-CHR5* (S1789) were significantly associated with KW in two out of the three field experiments; and sites S628 from *ZmGW2-CHR4* and S1632 from *ZmGW2-CHR5* showed significant effects on KT in two out of the three field experiments. Other trait/site associations identified in this study were significant in only one field experiment and could not be repeated across years and locations (Table 5; Table 6).

We further investigated whether the directions of effects in the association population were the same as what were predicted by the QTL mapping analysis of

**Table 4 Mean squares of ANOVA and descriptive statistics and correlation coefficients for four yield-related traits**

Category	Source of variation	DF	KL (mm)	KW (mm)	KT (mm)	HKW (g)
ANOVA	Year	2	52.09**	0.45**	33.78**	755.90**
	Genotype	120	1.84**	2.40**	0.87**	72.02**
	Replication (Year)	3	0.08	0.18	0.13	84.25**
	Year × Genotype	222 (215) <sup>a</sup>	0.41**	0.23**	0.18**	16.59**
	Error	287 (298) <sup>a</sup>	0.11	0.09	0.07	6.45
Descriptive statistics	Range		6.48-11.58	5.65-10.57	3.14-6.73	7.87-42.86
	Mean ± SD		8.90 ± 0.84	8.12 ± 0.76	4.70 ± 0.60	22.50 ± 4.99
Correlation coefficients <sup>b</sup>	KL (mm)			0.16	-0.15	0.37**
	KW (mm)		0.17**		0.51**	0.78**
	KT (mm)		-0.09*	0.36**		0.62**
	HKW (g)		0.26**	0.68**	0.58**	

DF, degree of freedom; HKW, one-hundred kernel weight; KL, kernel length; KT, kernel thickness; KW, kernel width; \*\*, P < 0.01; \*, P < 0.05

<sup>a</sup> Because the phenotypic values for some lines were missing either for kernel size or for kernel weight, the DF are different. The numbers in parentheses indicate the DF for HKW and the numbers outside parentheses indicate the DF for KL, KT and KW

<sup>b</sup> The numbers above the diagonal are genetic correlation coefficients and the numbers below the diagonal are phenotypic correlation coefficients

the IF<sub>2</sub> population (Figure 2; Table 2). Of the seven significant polymorphisms associated with HKW from *ZmGW2-CHR4*, four segregated in the IF<sub>2</sub> population and could be tested. In the association panel, all four favourable alleles are from the inbred line 87-1 (Table 5), while in the IF<sub>2</sub> population, the allele that increases

HKW is from the other parent, Zong3 (Table 2). For *ZmGW2-CHR5*, we did not detect any QTL for HKW in its vicinity in the IF<sub>2</sub> population. Consistent with this result, the only polymorphism that showed significant association with HKW in the association panel did not segregate in the IF<sub>2</sub> population (Table 6).

**Table 5 Associations between yield-related traits and the polymorphisms of *ZmGW2-CHR4***

Site	Location	Genotype <sup>a</sup>	Frequency	Env	KW <sup>b</sup>	KL <sup>b</sup>	KT <sup>b</sup>	HKW <sup>b</sup>	Zong3/87-1 <sup>c</sup>
S27	Promoter	0/8	22/95	07BJ 07HN 08HN	1.07E-04 0.0090			0.0029	0/8
S40	Promoter	C/T	18/99	07BJ 07HN 08HN	3.32E-04 0.0137 0.0086	0.0353	0.0226 0.0054	2.24E-04 0.0498 0.0242	T/C
S304	Promoter	C/T	19/97	07BJ 07HN 08HN	0.0021 0.0196	0.0058		0.0037	T/C
S338	Promoter	0/1	26/90	07BJ 07HN 08HN			0.0088	0.0352	1/0
S628	Exon 1	C/T	95/21	07BJ 07HN 08HN	0.0356		0.0102 0.0078	0.0127	C/C
S1730	Exon 8	C/T	95/20	07BJ 07HN 08HN	0.0066 0.0087	0.0446	0.0180	4.22E-04	C/C
S1865	3' UTR	0/1	95/20	07BJ 07HN 08HN	0.0198	0.0050	0.0292	0.0075	0/0

Env, environment; HKW, one-hundred kernel weight; KL, kernel length; KT, kernel thickness; KW, kernel width; 07BJ, year 2007 Beijing; 07HN, year 2007 Hainan; 08HN, year 2008 Hainan

<sup>a</sup> The numbers indicate the number of deleted nucleotides, and the letters indicate nucleotides. Numbers and letters in bold indicate the favourable alleles with the exception of the S1865/07BJ-KL association

<sup>b</sup> Significant probabilities

<sup>c</sup> The allele of inbred lines Zong3 and 87-1 which were used to develop an IF<sub>2</sub> population for QTL mapping [40,41]. Letters and numbers in bold indicate the favourable alleles

**Table 6 Associations between yield-related traits and the polymorphisms of *ZmGW2-CHR5***

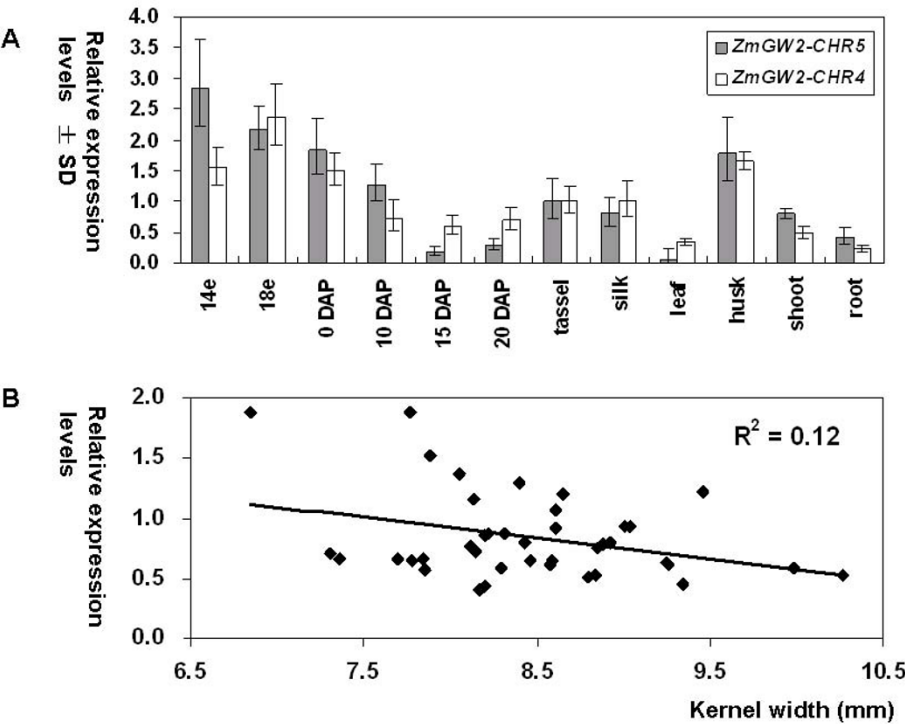
Site	Location	Genotype <sup>a</sup>	Frequency	Env	KW <sup>b</sup>	KL <sup>b</sup>	KT <sup>b</sup>	HKW <sup>b</sup>	Zong3/87-1 <sup>c</sup>
S908	Intron 5	0/1	101/16	07BJ 07HN 08HN	0.0054	0.0024			1/0
S1601	Exon 8	A/T	22/98	07BJ 07HN 08HN	0.0055	0.0018			A/T
S1632	Exon 8	A/C	14/106	07BJ 07HN 08HN			0.0464 0.0087		C/C
S1789	3' UTR	0/17	105/15	07BJ 07HN 08HN	0.0394				17/0
S2051	3' flanking	0/13	27/93	07BJ 07HN 08HN	0.0032 0.0067	0.0043		0.0363	0/0

Env, environment; HKW, one-hundred kernel weight; KL, kernel length; KT, kernel thickness; KW, kernel width; 07BJ, year 2007 Beijing; 07HN, year 2007 Hainan; 08HN, year 2008 Hainan

<sup>a</sup> The numbers indicate the number of deleted nucleotides, and the letters indicate nucleotides. Numbers and letters in bold indicate the favourable alleles

<sup>b</sup> Significant probabilities

<sup>c</sup> The allele of inbred lines Zong3 and 87-1 which were used to develop an IF2 population for QTL mapping [40,41]



**Figure 5 Expression analysis of *ZmGW2-CHR4* and *ZmGW2-CHR5*.** A. Tissue expression patterns. 14e, 18e, ears from plants with 14 and 18 expanded leaves, respectively. DAP, days after pollination. The height of each column represents the relative expression level of the corresponding gene. B. Correlation between the expression levels of *ZmGW2-CHR4* and kernel width. Kernels collected at 0 DAP were used for expression analysis.



### Expression analyses of the two maize genes

Real-time quantitative reverse transcription PCR (qRT-PCR) of 12 different tissues was performed to address whether the two maize genes had diverged expression patterns and whether these patterns were associated with a role in kernel development. As shown in Figure 5A, expression trends of the two genes were quite similar, with a correlation coefficient of 0.84, indicating no divergence between the two genes in the tissues examined. The highest expression levels were observed in immature ears for both genes, and expression levels were reduced in kernels after pollination, suggesting a role in kernel development. Correlation analysis between the expression levels of the two genes and the four traits across 41 maize inbred lines showed that *ZmGW2-CHR4* transcript abundance was negatively correlated with KW (Figure 5B,  $N = 41$ ,  $P = 0.03$ ,  $R^2 = 0.12$ ). However, despite the correlation between KW and the expression level of *ZmGW2-CHR4*, and the association between KW and six *ZmGW2-CHR4* polymorphisms (Table 5), none of these polymorphisms affected the expression level of *ZmGW2-CHR4*. No significant correlations were observed between *ZmGW2-CHR5* transcript levels and any of the four traits.

## Discussion

### Evolution of the two maize genes

The collinear relationship between the rice region containing *GW2* and the maize regions containing *ZmGW2-CHR4* and *ZmGW2-CHR5* suggested that the two maize genes are duplicated genes and both of them are co-orthologs of rice *GW2* (Figure 1A). Previous studies consistently showed that maize has a segmental allotetraploid origin, in which the maize genome was thought to have arisen through hybridization of two ancestral diploids whose genomes had partially diverged, one of which shares a more recent common ancestor with the sorghum genome [44,45]. To clarify the relationship of the *GW2* genes in maize and sorghum, we performed phylogenetic analysis with the *GW2* protein sequences, which showed that *ZmGW2-CHR5* is more closely related to its counterpart in sorghum than to *ZmGW2-CHR4* (Figure 1B). This finding supports the segmental allotetraploid origin of maize, and indicates that the two maize genes may have evolved independently for a period of time [44,45].

Three processes have been proposed to explain the evolutionary fates of duplicated genes within a species: non-functionalization, neo-functionalization and sub-functionalization. In non-functionalization, one of the duplicates accumulates deleterious mutations and eventually degenerates to a pseudogene or is lost from the genome [46-48]. Occasionally, mutations in regulatory or coding regions can lead to novel gene function (neo-

functionalization) [48,49]. Alternatively, both genes may experience some degeneration and lose partial functionality, but can complement each other (sub-functionalization). This can occur through partition of either protein domains or regulatory elements of the ancestral gene [50-52]. Sequence analyses showed that *ZmGW2-CHR4* and *ZmGW2-CHR5* are highly conserved, with an overall similarity of 94% across the coding region. None of the mutations in either gene led to truncated proteins, and no non-synonymous nucleotide changes were observed in the RING domain. Additionally, both genes are expressed across various tissues. Thus, neither of the genes has experienced a non-functionalization process.

To assess if neo-functionalization and sub-functionalization have occurred, previous knowledge of the ancestral gene's function is required, which is usually unavailable. Besides, both processes can occur through changes in regulatory or coding regions. Therefore, we tested and distinguished the two processes by investigating whether reciprocal degenerations have occurred. For example, if one gene was expressed in roots and silenced in leaves, sub-functionalization requires that the other gene must be silenced in roots and expressed in leaves (occurrence of reciprocal degeneration); while neo-functionalization requires that the other gene must be silenced in both tissues so that the expression in roots could be assumed to be a new function. In *ZmGW2-CHR4* and *ZmGW2-CHR5*, the exon number, protein length and protein domain are all well conserved (Table1; Figure 3; Figure 4), with no protein domains lost or acquired, indicating that neo-functionalization or sub-functionalization through changes in the coding region had not occurred. We further investigated the expression patterns of the two genes across 12 maize tissues (Figure 5A). Although the promoter regions of the two genes are highly divergent, both genes were expressed in all of the 12 tissues and their expression levels were highly correlated ( $r = 0.84$ ), indicating that neither neo-functionalization nor sub-functionalization through changes in regulatory regions had occurred in the tissues examined. However, we could not exclude the possibility that the two genes may show divergent expression patterns under various environmental conditions or in specific cell types which have not been tested.

### Genetic variations in *ZmGW2-CHR4* and *ZmGW2-CHR5* are associated with kernel size and weight in maize

The orthologous relationship and high sequence conservation between the two maize genes and the rice *GW2* suggest that *ZmGW2-CHR4* and *ZmGW2-CHR5* may perform similar functions as *GW2* in rice. Results from expression, linkage and association analyses corroborate this supposition. The expression levels of both maize

genes varied according to the developmental stages of the ear or kernel (Figure 5A), implying that they may be involved in ear or kernel development. This was further supported by association analysis. Although the small size of the association mapping panel did not guarantee optimal power of association tests [53], both genes were found to contain polymorphisms affecting at least one of the four yield-related traits (Table 5; Table 6). Two more lines of evidence were presented to support a role of *ZmGW2-CHR4* in kernel size and weight. One came from QTL mapping results (Figure 2; Table 2), which placed *ZmGW2-CHR4* within the confidence interval of a consistent QTL for HKW. Another came from the negative correlation between the expression levels of *ZmGW2-CHR4* and KW (Figure 5B), indicating that down-regulation of this gene may lead to elevated KW, and thus grain yield. This is consistent with the negative function of *GW2* in rice, where a truncated protein leads to an increase in grain size and weight [9].

Previous study with another yield-related gene, *GS3*, showed that different polymorphisms underlie similar phenotypes in rice and maize [6]. Here, we report a similar phenomenon. In rice, a 1 bp deletion in the fourth exon leads to a premature truncated protein and causes enhanced grain width and weight [9], while in maize, mutations in other regions of *ZmGW2-CHR4* and *ZmGW2-CHR5* cause those phenotypes (Table 5; Table 6). However, we cannot exclude the possibility that the 1 bp deletion found in the rice *GW2* gene also exists in maize, and would have been found if more (and more diverse) lines are sequenced for these genes. Moreover, the genetic polymorphisms that were associated with phenotypic variation for kernel size and weight are different between the two maize genes, indicating that they probably affect kernel size and weight through different mechanisms.

In QTL analysis, the favourable allele for HKW is from the inbred line Zong3 (Table 2). In our association panel, the favourable alleles of all four polymorphisms that segregated in the QTL mapping population were from the other parental inbred line, 87-1 (Table 5). Three possible explanations may explain the discrepancy in the direction of allelic effects. One is that the associations are false positives created due to the possible presence of population structure and individual relatedness. The best statistical model to account for the effect of population structure and individual relatedness and control the false positive rate is the mixed linear model [43,53]. This was the model used in this study, with which we found that one polymorphism was consistently associated with HKW across three different environments, and the direction of the allelic effect was the same across these environments (Table 5). For complex traits, such as HKW, the consistent detection of a

significant association across various environments using a well-performed statistical model implies that the associations are not false positives.

A second possibility for the discrepancy is that we may not have identified the actual functional polymorphism, but rather only one linked to it. Our evidence shows that *ZmGW2-CHR4* and *ZmGW2-CHR5* are associated with kernel size and weight in maize. However, it is still unclear what the actual causal polymorphism(s) is (are) in each case because association analysis based on LD can identify neutral polymorphisms (hereafter referred to as “the detected polymorphism”) in high LD with the actual functional polymorphism (See additional file 3: LD between sites significantly associated with kernel size and weight in *ZmGW2-CHR4*; additional file 4: LD between sites significantly associated with kernel size and weight in *ZmGW2-CHR5*). When the detected polymorphism is in complete LD with the functional polymorphism, the favourable allele at the detected polymorphism can represent the favourable allele at the functional polymorphism completely. However, if recombination occurred between the two polymorphisms, some inbred lines would show a favourable allele at the detected polymorphism, but actually contain the unfavourable allele at the functional polymorphism. This cannot be detected in association analysis where the mean effects of all lines are measured, but can be detected in linkage analysis where only two lines (this is, lines where recombination occurred between the functional polymorphism and the detected polymorphism) are involved.

A third possibility is the presence of other functional polymorphisms that we did not detect in the present study. The interval of a QTL usually spans 10-30 cM [54], and can contain more than one gene involved in the expression of the same trait in some cases [10,55]. Thus, the QTL effect actually represents the combined effect of all functional polymorphisms within the region. In the association analysis that included Zong3 and 87-1, we detected favourable alleles in *ZmGW2-CHR4* from 87-1, but if there were other unfavourable functional polymorphisms in *ZmGW2-CHR4* or other possible genes within the QTL that were not measured in the association analysis, the mean effect would identify 87-1 as the unfavourable parent in the QTL mapping population. This is quite possible, because our correlation analysis also pointed to the presence of other functional polymorphisms. In this study, we found a significant negative correlation between the expression levels of *ZmGW2-CHR4* and KW, but this correlation cannot be explained by the identified significant polymorphisms in the association analysis, implying the presence of other causal polymorphisms. These polymorphisms could be *cis*-acting elements far upstream of the genes, as in the

case of *tb1* [56]. Alternatively, they could be unknown *trans*-acting elements hidden in the genome, as have been reported in the maize genome [57]. However, it is very difficult to identify upstream *cis*- or *trans*-acting elements using the candidate gene association analysis strategy. An alternative method to identify these elements and to further explore the genetic basis of complex quantitative traits is genome-wide association studies with high density marker coverage [58].

## Conclusions

This study investigated the relationship, evolutionary fate and function of two maize genes involved in kernel size and weight. The two genes represent chromosomal duplicates that are co-orthologs of *GW2* in rice. The sequences of both genes are well conserved, with no mutations leading to a pseudo-molecule, and no new protein motifs were found, suggesting that both genes may have conserved functions in maize. Expression and candidate gene-based association analyses suggested that both genes play a role in kernel size and weight variation, as does rice *GW2*. However, the identified polymorphisms that contribute to phenotypic variation are different between the maize and rice genes and between the two maize genes, suggesting that the three genes may cause phenotypic variation through different mechanisms. Mutant or transformation experiments would shed more light on this hypothesis. The conservation of function (all associated with variation in kernel size and weight) together with the diversification of mechanism (different identified polymorphisms) among the three genes can help us to understand the similarities as well as the differences in the genetic basis of grain yield in rice and maize.

## Methods

### Determining the relationship between the two maize genes

Sequences in the vicinity of the two maize genes and the rice *GW2* gene were used to generate the collinear relationships among them. Because the three regions have different degrees of chromosomal expansion, different lengths of sequences were used. Specifically, a stretch of 8.6 Mb on maize chromosome 4 (B73 genome, 228.4-237.0 Mb), 20.9 Mb on maize chromosome 5 (B73 genome, 136.0-156.9 Mb) and 4.2 Mb on rice chromosome 2 (*Oryza sativa japonica* TIGR5, 5.8-10.0 Mb) were used. The online program SyMAP v3.0 [59] was used to draw and display the collinear relationship among the three regions with default settings.

The homologous sequences of the rice *GW2* gene from other species, including maize, sorghum and barley, were obtained via BLAST analysis in NCBI [34] and PlantGDB [60]. The phylogenetic tree was generated

using MEGA, version 3.1 [61] with the neighbor-joining method, Kimura two-parameter distance and pairwise deletion analysis. Robustness of the constructed phylogenetic tree was tested with 1,000 bootstrap repetitions of the informative polymorphisms.

### Genetic mapping and QTL analysis

A primer pair, M9 (See additional file 5: Primers used in this study) was designed from the sequences of *ZmGW2-CHR4* and used to map the gene in the recombinant inbred line (RIL) population derived from Zong3 and 87-1 [62] using MAPMAKER/EXP 3.0 [63]. The RIL population was used to develop an IF<sub>2</sub> population consisting of 441 crosses to map QTL and heterotic loci for yield-related traits [40,41]. The IF<sub>2</sub> genotypes were deduced according to the marker genotypes of their RIL parents, and the detailed method and the IF<sub>2</sub> design can be found in a previous study [41]. The composite interval mapping model [64] implemented in QTL Cartographer 2.5 [65] was used to map QTL for HKW in the IF<sub>2</sub> population following addition of the M9 marker. A window size of 10 cM, 5 control markers and the forward regression method were used. The threshold for declaring a QTL was determined by 1,000 random permutations with a significance level of 0.05.

### Sequencing and analyses

Coding regions of the two genes in eight maize inbred lines were sequenced in order to identify the fixed sites, which were the sites showing polymorphisms between the two maize genes but showing no variation within each gene across a panel of diverse lines. These lines were from the five main heterotic groups in China [66], including three lines from the TangSPT group (K12, Hai014 and S22), two from the Reid group (Shen5003 and 812) and one each from the Lancaster (4F1), Temp-tropic (SW1611) and Zi330 groups (5311). The heterotic groups were classified according to the genetic distance calculated using SSR markers. The genetic distance between lines within each group is closer than between lines from different groups. For the nucleotide diversity and LD analyses, three segments of each gene were sequenced, corresponding to the 5' end, the middle portion and the 3' end, across 121 lines with good agronomic performance among the assembled association mapping panel consisting of 155 diverse lines [53]. The primer pairs were designed in corresponding regions of the two homologous genes so that comparisons between them could be performed. Detailed information on the primers can be found in Additional file 5: Primers used in this study. Direct PCR products from the inbred lines in this study, which are almost completely homozygous across the whole genome, were sequenced. For ambiguous chromatograms, the products were sequenced again

in the reverse direction, or the DNA was re-amplified and sequencing was repeated.

After sequencing, an initial alignment was performed with the multiple sequence alignment program MUSCLE [67] to detect singletons, which are polymorphisms that are found only once among the sequenced materials. Lines in which singletons were found were analyzed again until they were confirmed as correct. MUSCLE was then used again to align the confirmed sequences, which were subsequently refined manually in BioEdit [68]. Three parameters implemented in the DnaSP, version 4.00 [69] were used to measure genetic diversity: the average pairwise nucleotide difference per site -  $\pi$ , the number of segregating sites -  $S$ ; and the number of haplotypes -  $h$ . Tajima's  $D$  statistic [70] was also calculated to investigate evidence for past selection. The LD level between sites with allelic frequency > 0.05 was calculated using TASSEL 2.0.1 [71].

#### Field design and statistical analyses

A maize association mapping panel with 155 diverse inbred lines developed by Yang et al. [53] was used to find associations between the DNA polymorphisms in the two maize genes and grain yield components. Population structure ( $Q$ ) and relative kinship ( $K$ ) were reported in a previous study [53]. Briefly, population structure was inferred using 82 SSR markers in STRUCTURE 2.2 [72,73] with five independent runs at each  $k$  (number of populations, which was set from 1 to 10). Results indicated the presence of two sub-populations [53], which were incorporated into the  $Q$  matrix. The kinship matrix  $K$  was calculated using 884 SNP markers in SPAGeDi [74], and negative values between individuals were set to 0 [43]. Of the 155 inbred lines, 121 lines with good agronomic performance were phenotyped in the present study. Two of the field experiments, Beijing and Hainan in 2007, have been reported previously [6]. An additional field experiment with two replications in Hainan was performed in 2008 using similar field design and management as in Li et al. [6]. KL, KW, KT and HKW were measured as described by Li et al. [6]. The analysis of variance, descriptive statistics and correlation analysis for the four yield-related traits were performed using the SAS system (version 8.02, SAS Institute Inc., Cary, NC, USA). Association analysis was performed with TASSEL 2.0.1 [71] using the mixed model,  $Q+K$  [43].

#### RNA extraction and qRT-PCR

Plant materials used to perform expression pattern analysis were described in a previous study [6]. Briefly, twelve tissues were collected from the inbred line 87-1, including seedling shoot, seedling root, mature leaf, tas-

sel from plants with 15 expanded leaves, silk and husk from ears 0 days after pollination (DAP), immature ears from plants with 14 and 18 expanded leaves, and kernels harvested at four time points after pollination (0 DAP, 10 DAP, 15 DAP and 20 DAP). In addition, kernels at 0 DAP were collected from 41 inbred lines (See additional file 6: Materials used for correlation analysis) which were grown in a field in Shangzhuang, Beijing in 2007 to perform correlation analysis between expression levels and yield-related traits. TRIzol (Invitrogen, Carlsbad, California, USA) and RNase-free DNase (Promega, Madison, Wisconsin, USA) were used to prepare the total RNA, which was then used to synthesize the cDNA with the MMLV retro-transcriptase and an oligo (dT) primer (Promega). qRT-PCR was performed with the Ex Taq premix (Takara Shuzo, Kyoto, Japan) and primers listed in Additional file 5: Primers used in this study. These specific primers were designed based on sequence differences between the two maize genes, so that each primer pair will amplify only one of the genes. All experiments were carried out following the manufacturers' instructions. The  $2^{-\Delta\Delta C_T}$  method [75] was employed to calculate relative expression levels with the housekeeping gene ubiquitin as an endogenous control. Tassel and 0 DAP kernels from the inbred line 87-1 were used as the reference tissues in the expression pattern and correlation analyses, respectively. Three replicates were performed to calculate the average and standard deviation of expression levels for each sample.

#### Additional material

**Additional file 1: Similarity between *ZmGW2-CHR4* and *ZmGW2-CHR5* across the cDNA region.** This is a figure showing the similarity between *ZmGW2-CHR4* and *ZmGW2-CHR5* across the cDNA region. The coding region is depicted using a filled grey box. The similarity was averaged over every 10 aligned nucleotides. From this figure, we can see that the coding region is well-conserved, while the 5' UTR and 3' UTR diverge between the two genes.

**Additional file 2: QTL for grain yield mapped in previous studies in maize bins 4.09 and 5.04.** This is a table showing the QTL for grain yield mapped in previous studies in maize bins 4.09 and 5.04.

**Additional file 3: LD between sites significantly associated with kernel size and weight in *ZmGW2-CHR4*.** This is a table. It shows the LD level between sites significantly associated with kernel size and weight in *ZmGW2-CHR4*.

**Additional file 4: LD between sites significantly associated with kernel size and weight in *ZmGW2-CHR5*.** This is a table. It shows the LD level between sites significantly associated with kernel size and weight in *ZmGW2-CHR5*.

**Additional file 5: Primers used in this study.** This is a table. It shows the primers used in this study.

**Additional file 6: Materials used for correlation analysis.** This is a table. It shows the lines used for correlation analysis between the expression levels of *ZmGW2-CHR4* and *ZmGW2-CHR5* and the four yield-related traits.

## Abbreviations

ANOVA: analysis of variance; BAC: bacterial artificial chromosome; cDNA: complementary DNA; DAP: days after pollination; HKW: one-hundred kernel weight; IF<sub>2</sub>: immortalized F<sub>2</sub>; InDels: insertion/deletion polymorphisms; KL: kernel length; KT: kernel thickness; KW: kernel width; LD: linkage disequilibrium; qRT-PCR: real time quantitative reverse transcription PCR; QTL: quantitative trait locus; RIL: recombinant inbred line; SNP: single nucleotide polymorphism; SSR: simple sequence repeat; UTRs: untranslated regions.

## Acknowledgements

This research was supported by the National Hi-Tech Research and Development Program of China (2006AA102183, 2006AA10A107) and National Basic Research and Development Program of China (2007CB10900). Thanks would be given to Dr. Jihua Tang, who is a previous member in our lab and now works in Henan Agricultural University, for sharing the IF<sub>2</sub> design and relevant data.

## Author details

<sup>1</sup>National Maize Improvement Center of China, Key Laboratory of Crop Genomics and Genetic Improvement (Ministry of Agriculture), China Agricultural University, 100193 Beijing, China. <sup>2</sup>USDA-ARS Corn Host Plant Resistance Research Unit Box 9555 Mississippi State, MS 39762. <sup>3</sup>College of Agriculture, Xinjiang Agricultural University, Urumqi, 830052 Xinjiang, China. <sup>4</sup>International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico.

## Authors' contributions

QL carried out the sequence, linkage, expression and association analyses and wrote the manuscript. LL performed the relationship analysis of the genes among species. GHB participated in field experiments and trait evaluation. XHY participated in association analysis. LL and XHY helped to prepare the materials. MLW gave critical suggestions to the interpretation of the results and helped to revise the manuscript. JBY designed the study. JRD and JSL participated in its design and coordination. JBY and JSL helped to draft the manuscript. All authors read and approved the final manuscript.

Received: 31 January 2010 Accepted: 14 July 2010

Published: 14 July 2010

## References

- Miller ME, Chourey PS: **The maize invertase-deficient *miniature-1* seed mutation is associated with aberrant pedicel and endosperm development.** *Plant Cell* 1992, **4**:297-305.
- Maitz M, Santandrea G, Zhang Z, Lal S, Hannah LC, Salamini F, Thompson RD: ***rgf1*, a mutation reducing grain filling in maize through effects on basal endosperm and pedicel development.** *Plant J* 2000, **23**:29-42.
- Fu S, Meeley R, Scanlon MJ: ***empty pericarp2* encodes a negative regulator of the heat shock response and is required for maize embryogenesis.** *Plant Cell* 2002, **14**:3119-3132.
- Gutiérrez-Marcos JF, Prà MD, Giulini A, Costa LM, Gavazzi G, Cordelier S, Sellam O, Tatout C, Paul W, Perez P, Dickinson HG, Consonni G: ***empty pericarp4* encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize.** *Plant Cell* 2007, **19**:196-210.
- Martin A, Lee J, Kichey T, Gerentes D, Zivy M, Tatout C, Dubois F, Balliau T, Valot B, Davanture M, Tercé-Laforgue T, Quilleré I, Coque M, Gallais A, Gonzalez-Moro M, Bethencourt L, Habash DZ, Lea PJ, Charcosset A, Perez P, Murigneux A, Sakakibara H, Edwards KJ, Hirel B: **Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production.** *Plant Cell* 2006, **18**:3252-3274.
- Li Q, Yang X, Bai G, Warburton ML, Mahuku G, Gore M, Dai J, Li J, Yan J: **Cloning and characterization of a putative *GS3* ortholog involved in maize kernel development.** *Theo Appl Genet* 2010, **120**(4):753-763.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reilly AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Buren VP, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK: **The B73 maize genome: complexity, diversity, and dynamics.** *Science* 2009, **326**:1112-1115.
- Disch S, Anastasiou E, Sharma VK, Laux T, Fletcher JC, Lenhard M: **The E3 ubiquitin ligase BIG BROTHER controls *Arabidopsis* organ size in a dosage-dependent manner.** *Curr Biol* 2006, **16**:272-279.
- Song XJ, Huang W, Shi M, Zhu MZ, Lin HX: **A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase.** *Nat Genet* 2007, **39**(5):623-630.
- Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M: **Cytokinin oxidase regulates rice grain production.** *Science* 2005, **309**(5735):741-745.
- Riefler M, Novak O, Strnad M, Schmülling T: ***Arabidopsis* cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism.** *Plant Cell* 2006, **18**:40-54.
- Frary A, Nesbitt TC, Frary A, Grandillo S, Knaap van der E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD: ***fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size.** *Science* 2000, **289**(5476):85-88.
- Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A: ***MINISEED3 (MIN3)*, a *WRKY* family gene, and *HAIRY2 (H2)*, a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*.** *Proc Natl Acad Sci USA* 2005, **102**(48):17531-17536.
- Zhu C, Gore M, Buckler ES, Yu J: **Status and prospects of association mapping in plants.** *The Plant genome* 2008, **1**:5-20.
- Chen H, Wang S, Xing Y, Xu C, Hayes PM, Zhang Q: **Comparative analyses of genomic locations and race specificities of loci for quantitative resistance to *Pyricularia grisea* in rice and barley.** *Proc Natl Acad Sci USA* 2003, **100**(5):2544-2549.
- Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, McNally K, Khush GS, Paterson AH, Li ZK: **Convergent evolution of perennality in rice and sorghum.** *Proc Natl Acad Sci USA* 2003, **100**(7):4050-4054.
- Magalhaes JV, Garvin DF, Wang Y, Sorrells ME, Klein PE, Schaffert RE, Li L, Kochian LV: **Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae.** *Genetics* 2004, **167**:1905-1914.
- Paterson AH, Lin YR, Li ZK, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE: **Convergent domestication of cereal crops by independent mutations at corresponding genetic loci.** *Science* 1995, **269**(5231):1714-1718.
- Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP: **'Green revolution' genes encode mutant gibberellin response modulators.** *Nature* 1999, **400**:256-261.
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macaluso T, Oliphant A, Briggs S: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*).** *Science* 2002, **296**(5565):92-100.
- Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L,



- Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 2002, **296**(5565):79-92.**
22. Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X: **Natural variation at the *DEP1* locus enhances grain yield in rice. *Nat Genet* 2009, **41**(4):494-497.**
23. Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M: **Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet* 2008, **40**(8):1023-1028.**
24. Wan X, Weng J, Zhai H, Wang J, Lei C, Liu X, Guo T, Jiang L, Su N, Wan J: **Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele *gw-5* in a recombination hotspot region on chromosome 5. *Genetics* 2008, **179**:2239-2252.**
25. Weng J, Gu S, Wan X, Gao H, Guo T, Su N, Lei C, Zhang X, Cheng Z, Guo X, Wang J, Jiang L, Zhai H, Wan J: **Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res* 2008, **18**:1199-1209.**
26. Yu J, Buckler ES: **Genetic association mapping and genome organization of maize. *Curr Opin Biotech* 2006, **17**:155-160.**
27. Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES: **A first-generation haplotype map of maize. *Science* 2009, **326**:1115-1117.**
28. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES: **Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA* 2001, **98**:11479-11484.**
29. Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS: **Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA* 2001, **98**:9161-9166.**
30. Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES: **Natural genetic variation in *lycopen epsilon cyclase* tapped for maize biofortification. *Science* 2008, **319**:330-333.**
31. Wilson LM, Whitt SR, Ibáñez AM, Rocheford TR, Goodman MM, Buckler IV ES: **Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 2004, **16**:2719-2733.**
32. Manicacci D, Camus-Kulandaivelu L, Fourmann M, Arar C, Barrault S, Rousselet A, Feminias N, Consoli L, Francès L, Méchin V, Murigneux A, Prioul JL, Charcosset A, Damerval C: **Epistatic interactions between *opaque2* transcriptional activator and its target gene *CyPPDK1* control kernel trait variation in maize. *Plant Physiol* 2009, **150**:506-520.**
33. Pressoir G, Brown PJ, Zhu W, Upadaya N, Rocheford T, Buckler ES, Kresovich S: **Natural variation in maize architecture is mediated by allelic differences at the PINOID co-ortholog *barren inflorescence2*. *Plant J* 2009, **58**:618-628.**
34. National Center for Biotechnology Information. [http://www.ncbi.nlm.nih.gov].
35. Gaut BS: **Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. *Genome Res* 2001, **11**:55-66.**
36. Melchinger AE, Utz HF, Schön CC: **Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 1998, **149**(1):383-403.**
37. Yan JB, Tang H, Huang YQ, Zheng YL, Li JS: **Quantitative trait loci mapping and epistatic analysis for grain yield and yield components using molecular markers with an elite maize hybrid. *Euphytica* 2006, **149**:121-131.**
38. Abler BSB, Edwards MD, Stuber CW: **Isoenzymatic identification of quantitative trait loci in crosses of elite maize inbreds. *Crop Sci* 1991, **31**(2):267-274.**
39. Lübberstedt T, Melchinger AE, Schön CC, Utz HF, Klein D: **QTL mapping in testcrosses of European flint lines of maize: I. Comparison of different testers for forage yield traits. *Crop Sci* 1997, **37**(3):921-931.**
40. Tang J: **Genetic dissection on the basis of grain yield and its components as well as heterosis in an elite hybrid Yuyu22. *Postdoc thesis* China Agricultural University, Department of Plant Genetics and Breeding 2005.**
41. Tang J, Yan J, Ma X, Teng W, Wu W, Dai J, Dhillon BS, Melchinger AE, Li J: **Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F<sub>2</sub> population. *Theor Appl Genet* 2010, **120**:333-340.**
42. Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, Fallier M, Bombles K, Lukens L, Doebley JF: **The origin of the naked grains of maize. *Nature* 2005, **436**:714-719.**
43. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 2006, **38**(2):203-208.**
44. Gaut BS, Doebley JF: **DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc Natl Acad Sci USA* 1997, **94**:6809-6814.**
45. White S, Doebley J: **Of genes and genomes and the origin of maize. *Trends Genet* 1998, **14**(8):327-332.**
46. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes. *Science* 2000, **290**(5494):1151-1155.**
47. Kellogg EA: **What happens to genes in duplicated genomes. *Proc Natl Acad Sci USA* 2003, **100**(8):4369-4371.**
48. Prince VE, Pickett FB: **Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet* 2002, **3**:827-837.**
49. Ohno S: **Ancient linkage groups and frozen accidents. *Nature* 1973, **244**(5414):259-262.**
50. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999, **151**:1531-1545.**
51. Lynch M, Force A: **The probability of duplicate gene preservation by subfunctionalization. *Genetics* 2000, **154**:459-473.**
52. Adams KL, Cronn R, Percifield R, Wendel JF: **Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci USA* 2003, **100**(8):4649-4654.**
53. Yang X, Yan J, Shah T, Warburton ML, Li Q, Li L, Gao Y, Chai Y, Fu Z, Zhou Y, Xu S, Bai G, Meng Y, Zheng Y, Li J: **Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theor Appl Genet*.**
54. Salvi S, Tuberosa R: **To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci* 2005, **10**(6):297-304.**
55. Monna L, Lin HX, Kojima S, Sasaki T, Yano M: **Genetic dissection of a genomic region for a quantitative trait locus, *Hd3*, into two loci, *Hd3a* and *Hd3b*, controlling heading date in rice. *Theor Appl Genet* 2002, **104**:772-778.**
56. Clark RM, Wagler TN, Quijada P, Doebley J: **A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* 2006, **38**(5):594-597.**
57. Swanson-Wagner RA, DeCook R, Jia Y, Bancroft T, Ji T, Zhao X, Nettleton D, Schnable PS: **Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids. *Science* 2009, **326**:1118-1120.**
58. Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES: **Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 2009, **21**:2194-2202.**
59. Soderlund C, Nelson W, Shoemaker A, Paterson A: **SyMAP: a system for discovering and viewing syntenic regions of FPC maps. *Genome Res* 2006, **16**:1159-1168.**
60. Resources for Plant Comparative Genomics. [http://www.plantgdb.org].
61. Kumar S, Tamura K, Nei M: **MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 2004, **5**(2):150-163.**
62. Ma XQ, Tang JH, Teng WT, Yan JB, Meng YJ, Li JS: **Epistatic interaction is an important genetic basis of grain yield and its components in maize. *Mol Breeding* 2007, **20**(1):41-51.**
63. Lincoln S, Daly M, Lander E: **Constructing genetic maps with MAPMAKER/EXP 3.0. *Whitehead Institute Technical Report* Cambridge MA 1992.**
64. Zeng ZB: **Precision mapping of quantitative trait loci. *Genetics* 1994, **136**:1457-1468.**
65. Wang S, Basten CJ, Zeng ZB: **Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC 2007.**

66. Teng WT, Cao JS, Chen YH, Liu XH, Jing XQ, Zhang FJ, Li JS: **Analysis of maize heterotic groups and patterns during past decade in China.** *Scientia Agricultura Sinica* 2004, **37**(12):1804-1811.
67. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucl Acids Res* 2004, **32**(5):1792-1797.
68. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Ser* 1999, **41**:95-98.
69. Rozas J, Rozas R: **DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis.** *Bioinformatics* 1999, **15**(2):174-175.
70. Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123**:585-595.
71. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES: **TASSEL: software for association mapping of complex traits in diverse samples.** *Bioinformatics* 2007, **23**(19):2633-2635.
72. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**:945-959.
73. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.** *Genetics* 2003, **164**:1567-1587.
74. Hardy OJ, Vekemans X: **SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels.** *Mol Ecol Notes* 2002, **2**:618-620.
75. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method.** *Methods* 2001, **25**:402-408.

doi:10.1186/1471-2229-10-143

**Cite this article as:** Li et al.: Relationship, evolutionary fate and function of two maize co-orthologs of rice *GW2* associated with kernel size and weight. *BMC Plant Biology* 2010 **10**:143.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

