

RESEARCH ARTICLE

Open Access

# Molecular evolution of the duplicated *TFIIA $\gamma$* genes in *Oryzaceae* and its relatives

Hong-Zheng Sun<sup>1,2</sup> and Song Ge<sup>\*1,2</sup>

## Abstract

**Background:** Gene duplication provides raw genetic materials for evolutionary novelty and adaptation. The evolutionary fate of duplicated transcription factor genes is less studied although transcription factor gene plays important roles in many biological processes. *TFIIA $\gamma$*  is a small subunit of *TFIIA* that is one of general transcription factors required by RNA polymerase II. Previous studies identified two *TFIIA $\gamma$* -like genes in rice genome and found that these genes either conferred resistance to rice bacterial blight or could be induced by pathogen invasion, raising the question as to their functional divergence and evolutionary fates after gene duplication.

**Results:** We reconstructed the evolutionary history of the *TFIIA $\gamma$*  genes from main lineages of angiosperms and demonstrated that two *TFIIA $\gamma$*  genes (*TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5*) arose from a whole genome duplication that happened in the common ancestor of grasses. Likelihood-based analyses with branch, codon, and branch-site models showed no evidence of positive selection but a signature of relaxed selective constraint after the *TFIIA $\gamma$*  duplication. In particular, we found that the nonsynonymous/synonymous rate ratio ( $\omega = d_N/d_S$ ) of the *TFIIA $\gamma$ 1* sequences was two times higher than that of *TFIIA $\gamma$ 5* sequences, indicating highly asymmetric rates of protein evolution in rice tribe and its relatives, with an accelerated rate of *TFIIA $\gamma$ 1* gene. Our expression data and EST database search further indicated that after whole genome duplication, the expression of *TFIIA $\gamma$ 1* gene was significantly reduced while *TFIIA $\gamma$ 5* remained constitutively expressed and maintained the ancestral role as a subunit of the *TFIIA* complex.

**Conclusion:** The evolutionary fate of *TFIIA $\gamma$*  duplicates is not consistent with the neofunctionalization model that predicts that one of the duplicated genes acquires a new function because of positive Darwinian selection. Instead, we suggest that subfunctionalization might be involved in *TFIIA $\gamma$*  evolution in grasses. The fact that both *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* genes were effectively involved in response to biotic or abiotic factors might be explained by either Dykhuizen-Hartl effect or buffering hypothesis.

## Background

Transcription factors are large families in the genome of most eukaryotic organism and often act as switches between discrete developmental programs [1] and play important roles in many biological processes in plants, such as developmental regulation, control of metabolic pathways, response to environment stimuli and harmful stress [2,3]. Unlike regulatory transcription factors, general transcription factors are conserved proteins that are used by organisms as diverse as human, rat, *Drosophila*, and yeast to initiate mRNA synthesis [4]. *TFIIA* is one of general eukaryotic transcription factors required by RNA

polymerase II and has been demonstrated to stimulate transcription by stabilizing TBP binding to the TATA box and by regulating TBP or *TFIID* dimerization to accelerate DNA binding [4,5]. All three polypeptides in *TFIIA* including the small subunits (*TFIIA $\gamma$* ) showed high sequence and structural conservation across different organisms, highlighting their significance in eukaryotic transcription [6,7]. Recent studies showed that there were two *TFIIA $\gamma$* -like genes in rice genome, in contrast to *Arabidopsis* where only one copy was found [8]. Sequence comparison indicated that two rice *TFIIA $\gamma$* -like genes had 85.5% identity at the amino acid level and shared high degrees of nucleotide and amino acid sequence similarity with the *Arabidopsis TFIIA $\gamma$* -like gene [7,8]. Interestingly, a mutant (V39E substitution) in the copy on rice chromosome 5 (*xa5*) was confirmed to confer resistance to rice

\* Correspondence: gesong@ibcas.ac.cn

<sup>1</sup> State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China  
Full list of author information is available at the end of the article

bacterial blight [7,8] and the other copy on chromosome 1 (*TFIIAγ1*) was found to be highly expressed when induced by pathogen invasion [9].

Gene duplication is widely recognized as a major evolutionary force shaping genome evolution, and provides raw genetic materials for evolutionary novelty and adaptation [10,11]. Duplication of transcription factor genes has been recently investigated, but almost all studies focused on regulatory transcription factors (e.g., [12-16]) and little is known about the evolution of basic transcription factor duplicates. The duplication and divergence of *TFIIAγ* gene in rice and their resistance reactions to rice bacterial blight raise a few of interesting questions. First, whether the new function of disease resistance is facilitated by the redundancy of *TFIIAγ* gene, as suggested by previous study [7]? Evidence showed that gene duplication might contribute to the ability of plants to obtain a defense response against disease and herbivory through the functional diversification of genes but empirical study is still scarce in plants [17,18]. Second, when the *TFIIAγ* duplication happened in history and what model fits the fate of the duplicated genes. The classic models of gene duplication predict that one of the duplicated genes is either lost by accumulation of deleterious mutations (pseudogenization or nonfunctionalization) [19,20] or acquires a new function because of positive Darwinian selection (neofunctionalization) [11,21]. Additional possible fates of the duplicated genes were also proposed, including maintenance of the ancestral function by both copies (redundancy) and subdivision of the ancestral function between copies (subfunctionalization and subneofunctionalization) [21-25]. Jiang et al. (2006) suggested that duplication of the *TFIIAγ* gene in rice gave rise to a new function for disease resistance during evolution. This hypothesis, however, remains to be justified by empirical molecular data. Molecular evolutionary analyses have been successfully used to test the alternative explanations for the retention and evolution of the duplicated genes (e.g., [14,16,26-29]). To reconstruct the phylogenetic relationships of *TFIIAγ* genes will help better elucidate the duplication history of two *TFIIAγ* and further reveal their evolutionary fates after the duplication.

Finally, we ask what role of selection plays on the evolution of duplicated *TFIIAγ* genes? Is there any change in the strength and mode of selection that have acted on the duplicate genes? What is the relative importance of relaxation of purifying selection and positive selection in the evolution of *TFIIAγ* genes? Previous studies often treated relaxation of purifying selection as the null hypothesis but positive selection after gene duplication has been well demonstrated (e.g., [28,30,31]). A few of current statistical methods provide effective ways to evaluate the role of positive selection following gene duplication and allow more specific cases can be addressed [28,32,33].

In the present study, we investigate the molecular evolution of the general transcription factor *TFIIAγ* in grasses, including a dense sampling of species of the rice tribe (Oryzeae). Based on the *TFIIAγ* gene phylogeny, we found that the duplication event giving rise to *TFIIAγ1* and *TFIIAγ5* happened in the common ancestor of extant grasses. Our molecular evolutionary analyses and likelihood ratio tests revealed the relaxation of selective constraint on *TFIIAγ* genes following gene duplication and an acceleration of *TFIIAγ1* gene evolution. In conjunction with expression data, we demonstrated that both *TFIIAγ* genes following the duplication were functional and under strong selection constraint in Oryzeae and its relatives, providing no evidence that either gene evolved new functions or became a pseudogene despite their long-term coexistence for at least 50 MYA. Instead, the evolutionary fates of two *TFIIAγ* genes could be explained either by the Dykhuizen-Hartl effect [34] which predicts that one of duplicate genes evolves under relaxed purifying selection and later convey a selective advantage under particular environments, or by the buffering hypothesis which suggests that selection for a buffering effect is a mechanism for duplicate gene preservation after whole genome duplication.

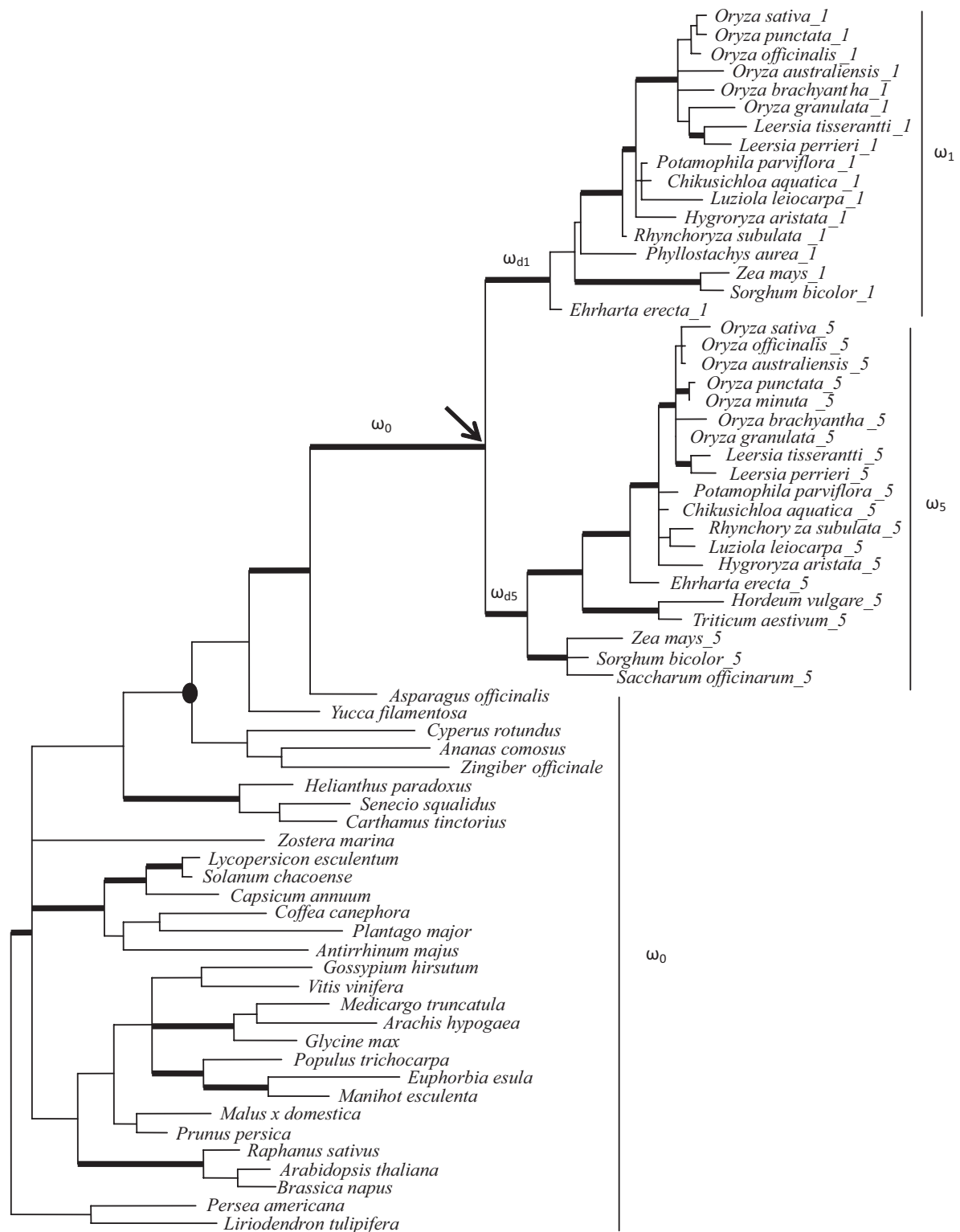
## Methods

### Species samples

The rice tribe (Oryzeae) includes approximately 12 genera and more than 70 species distributed across the tropical and temperate regions of the world [35,36]. In this study, we sampled 13 diploid species that represent the main lineages of Oryzeae, including six *Oryza* species, two *Leersia* species, and one each of other five genera in the tribe (Figure 1; Additional file 1). One species in the tribe Ehrhartoideae that is sister to Oryzeae, *Ehrharta erecta*, was used as an outgroup [35,37]. To infer the duplication event of the two *TFIIAγ* genes, we selected additional 12 monocots and 24 dicots to generate the phylogenetic tree of the *TFIIAγ* genes. In total, 30 sequences were isolated here and the remaining sequences were extracted from GenBank by BLAST searches [38]. Detailed information of the species and the sequences and their GenBank accession numbers is listed in additional file 1.

### Isolation and sequencing of *TFIIAγ* genes

On the basis of the *TFIIAγ*-like sequences from rice, wheat and maize [Additional file 1], we designed two pairs of universal PCR primers to amplify the *TFIIAγ* genes. They are the forward primers P1 (5'-TTCgAgCT-STACMggMggTC-3') or P3 (5'-ATggCCACCTTC-gAgCTSTA-3') and reverse primers P2 (5'-AggCCACRATCTTACCTTg-3') or P4 (5'-TCRCAggCACRATCTTAC-3'). The regions amplified and the



**Figure 1 Phylogeny of the TFIIA $\gamma$ -like genes.** Phylogeny was inferred by Bayesian inference under the GTR+I+G model. Bold branches are supported by the Bayesian posterior probability > 0.90. Solid circle indicates the monocot group and the arrow indicates the duplication event. Sequences in bold were included in the pruned tree on which different branch models of molecular evolution were tested using the PAML analysis.

locations of the PCR and internal primers (P7 and P8) are shown in additional file 2. Genomic DNA was extracted from fresh young leaves or silica-gel dried leaves using the CTAB methods as described in [25]. PCR amplification was performed in a volume of 25  $\mu$ l reaction using *ex*Taq polymerase (TaKaRa, Dalian, China). The cycling procedure was 35 cycles of denaturation at 94°C for 45 s, annealing at 56°C for 45 s and extension at 68°C for 8 min with 2 min of pre-denaturation and 10 min of final extension. PCR product was run on 1.2% agarose gel and all bands were excised under UV light, purified using Dinggou gel purification kit (Dingguo, Beijing, China), and sequenced using ET Terminator Kit (Amersham Pharmacia Biotech). All the PCR products were cloned into *pGEM* T-easy vectors (Promega, Madison, WI, USA) and at least 6 independent clones were sequenced. The purified fragments were also sequenced directly to make confirmation. If more than one copy was isolated in one species, we first construct a phylogeny including all the copies. If multiple copies from the same species clustered together, one copy was randomly selected in further analysis.

#### Characterization of expression by RT-PCR and EST database search

Reverse transcriptase-polymerase chain reaction (RT-PCR) was performed to investigate whether there is difference of expression between two *TFIIA $\gamma$*  genes. Total RNA was extracted from fresh leaves of eight species and young panicles of three species of *Oryzae* [Additional file 1] using Plant RNA Reagent (Invitrogen, Carlsbad, California, USA). The first strand cDNA was reverse-transcribed with oligo dT<sub>20</sub> primer. Subsequent detection was performed by PCR using up-stream primer P3 and low-stream specific primers P7 (5'-AYARWAACCTT-gCTCTTgACTTgg-3') and P8 (5'-gACNNTAACCTT-gCTCTTCACCTSA-3') (P7 for *TFIIA $\gamma$ 5* copy and P8 for *TFIIA $\gamma$ 1* copy). The *actin* gene was taken as control using primers ACT-59F (5'-AggCTggTTTCgCTggggATgATg-3') and ACTIN-764R (5'-ggACCTCggggCACCTgAACCTCT-3') [14]. The PCR procedure was 2 min of pre-denaturation at 94°C, 35 cycles of denaturation at 94°C for 30 sec, annealing at 54°C for 30 s, and extension at 72°C for 1 min, with a final extension of 10 min at 72°C. RT-PCR products were confirmed by sequencing.

In EST database search, all the hits of *Poaceae* species with e-value lower than  $1e^{-10}$  were collected. The sequences retrieved were aligned with rice *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* genes. By a neighbor-join phylogeny construction, all the sequences can be divided into two classes, corresponding to the *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* clades, respectively. We used the number of hits as an indicator of the expression level of the two copies, because a highly

expressed gene would have greater chance to be picked from cDNA library than a lowly expressed gene [39,40].

#### Sequence analysis

Sequences were aligned using a combination of methods implemented in BioEdit [41] and ClustalX 1.81 [42], with further manual refinements. The unalignable intron regions were excluded from the analyses. The GC content of all three codon positions and pairwise synonymous and nonsynonymous distances were calculated by MEGA3.1 [43]. Codon usage bias of the sequences was estimated by ENC (effective number of codons) that varies between 20 and 61, with the lower the value, the more biased codon usage [44]. We used Tajima's relative rate test [45], as implemented in MEGA3.1, to test for rate variation between two *TFIIA $\gamma$*  genes using *Ananas comosus* as outgroup. To visualize conservation and check the rate variation along the *TFIIA $\gamma$*  sequences, a sliding window analysis was performed by the K-estimator program [46]. Given relatively small length of *TFIIA $\gamma$*  genes, we used a window size of 10 amino acid (30 bp) and a step size of 3 amino acid (9 bp) in the sliding window analysis. *Poaceae* species were used in the sliding analysis. To avoid sampling bias, only *Oryza punctata* was used to represent the *Oryzae* species.

#### Phylogenetic analysis

Phylogenetic analyses were performed using maximum likelihood (ML) method, implemented in PAUP 4.0b10 [47], and Bayesian inference (BI) with MrBayes v.3.12 [48]. For ML, heuristic searches were run with random taxon addition, tree bisection reconnection swap for 100 replications. The reliability of branches was evaluated by 500 bootstrap replications. In each bootstrap heuristic search replication, the same parameter settings were used, except that number of heuristic search replications was set to 10. In ML and analyses the best nucleotide substitution models for each data set were selected using Modeltest 3.7 by corrected Akaike information criterion [49]. In BI with GTR+I+G model, Markov chain Monte Carlo (MCMC) analysis was run for 1,000,000 generations, sampled every 100<sup>th</sup> generations. The first 250,000 generations were set as burn-in.

We generated a phylogenetic tree of all *TFIIA $\gamma$*  or *TFIIA $\gamma$* -like sequences to explore the duplication history of the two duplicates. For this purpose, we used only coding sequences to construct the gene tree because the intron sequences between *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* were unalignable. The phylogenetic tree was rooted by the *TFIIA $\gamma$* -like genes of *Liriodendron tulipifera* and *Persea americana* that belong to two families (*Magnoliaceae* and *Lauraceae*) of the basal angiosperms [50].

### Tests for selection

The ratio of nonsynonymous to synonymous substitution sites ( $d_N/d_S$  or  $\omega$ ) is an effective measure to detect selection on a gene or gene region [33]. If the ratio is significantly less than 1 ( $\omega < 1$ ), purifying selection is inferred, while positive selection is evoked if the ratio is significantly greater than 1 ( $\omega > 1$ ). An estimate of the ratio close to 1 ( $\omega = 1$ ) indicates the presence of neutral evolution. To explore the selective processes acting on *TFIIA $\gamma$*  genes, we performed likelihood-based analyses using the codeml program of PAML version 4 [51]. We first tested whether the average  $\omega$  ratio differed among lineages of the gene tree by using the branch models that allow  $\omega$  to vary among lineages and assume different  $\omega$  ratios assigned to the branches before and after the duplication event. The one ratio model (M0) assumes a single  $\omega$  for all branches and all sites, whereas the other models allow for different  $\omega$  ratios among branches of the tree. The free ratio model (Mf) assumes an independent  $\omega$  ratio for each branch of the tree. The two ratio model M2r assumes one  $\omega$  ratio to all branches predating the duplication event ( $\omega_0$ ), and the other ratio to all branches postdating the duplication event ( $\omega_{d1} = \omega_{d5} = \omega_1 = \omega_5$ ). The three ratio model (M3r) assumes one ratio restricted to all branches predating the gene duplication ( $\omega_0$ ) and the other two to the branches of *TFIIA $\gamma$ 1* ( $\omega_{d1} = \omega_1$ ) and *TFIIA $\gamma$ 5* ( $\omega_{d5} = \omega_5$ ), respectively, following the duplication event. A more complex model, the four ratio model (M4r), assumes four independent  $\omega$  ratios: one ratio restricted to all branches predating the gene duplication ( $\omega_0$ ), one ratio to the branches immediately following the duplication ( $\omega_{d1} = \omega_{d5}$ ), and the last two assigned to the branches leading to *TFIIA $\gamma$ 1* ( $\omega_1$ ) and *TFIIA $\gamma$ 5* ( $\omega_5$ ) of grass species, respectively. Finally, the five ratio model (M5r) extends M4r to allow  $\omega$  ratios to differ between the *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* branches immediately postdating the duplication ( $\omega_{d1} \neq \omega_{d5}$ ) (Figure 1; Table 1). A likelihood ratio test (LRT) was conducted to determine whether there is statistically significant difference between two models. If the LRT is significant, the null hypothesis that two models are not significantly different is rejected, and the model with higher likelihood value is assumed to be a better model [28,52].

We next used site-specific models to examine whether particular amino acid residues were subject to positive selection because the  $\omega$  ratio is seldom detected greater than 1 if all the sites are averaged [53]. The nested codon models [28,54] were performed. In addition to one ratio model (M0), nearly neutral model (M1) classifies all the sites into 2 categories, one category under strict constraint ( $0 < \omega < 1$ ) and the other under neutral ( $\omega = 1$ ). Positive selection model (M2) is based on M1 and assumes a third category under positive selection ( $1 < \omega$ ).

The discrete model (M3) classifies all the sites into several categories, each with a different  $\omega$  ratio. Beta model (M7) assumes a beta distribution of the  $\omega$  ratios, and beta& $\omega$  model (M8) extends an independent ratio estimated by the data. Models assuming positive selection M8 and M2 are compared with null models M7 and M1, respectively. Positive selection is invoked if the LRT is significant and there is site with  $\omega > 1$  [28]. A comparison between M3 and M0 can tell whether the  $\omega$  ratio is homogeneous across different part of the gene.

We further performed the branch-site models A and B [55] to test for sites potentially under positive selection on *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* branches, respectively. Model A assumes  $0 < \omega_0 < 1$  and  $\omega_1 = 1$  and was compared with nearly neutral model (M1); while model B determines  $\omega_0$  and  $\omega_1$  as free parameters to be estimated and compared with discrete model (M3) [55].

## Results

### Cloning and characterization of two *TFIIA $\gamma$* genes

Using genomic DNA we cloned and sequenced two *TFIIA $\gamma$*  genes from all sampled *Oryzae* species except for *Leersia tisserantii* for which only *TFIIA $\gamma$ 1* was isolated, mainly because the second intron of *TFIIA $\gamma$ 5* in this species was too long to be amplified successfully by *exTaq* DNA polymerase. However, when using cDNA template, we obtained the coding region of *TFIIA $\gamma$ 5* and the first intron sequence using an internal primer for this species. Two *TFIIA $\gamma$*  copies were also isolated and sequenced from other Poaceae species, including *Ehrharta erecta*, *Zea mays* and *Sorghum bicolor*. Only single *TFIIA $\gamma$* -like gene was isolated from both *Cyperus rotundus* and *Zingiber officinale* despite different attempts have been tried, including optimization of PCR amplification, recombination of up and down stream primers. All the *TFIIA $\gamma$*  genes obtained in this study have three exons and two introns, with about 261 bp in coding sequence. The downloaded *TFIIA $\gamma$* -like sequences are cDNAs with full coding region. The *TFIIA $\gamma$ 1*-like sequences of rice, maize and sorghum were 327 bp in length and 9 bp (three codons) longer than the sequences of grass *TFIIA $\gamma$ 5*-like gene and those from the remaining species outside Poaceae. In *Oryzae*, sequence length ranged from 1.3 to 1.8 kb for *TFIIA $\gamma$ 1* and from 2.5 to 5.5 kb for *TFIIA $\gamma$ 5*. The first intron is about 70 ~100 bp in length for both genes, whereas the length of the second intron varied greatly [Additional file 2]. In coding regions, there is no indels between the two copies and can be aligned perfectly. We did not find the V39E substitution that lead to *TFIIA $\gamma$ 5* (*xa5*) to confer resistance to rice bacterial blight in all *Oryzae* species, indicating that such a mutation arises within *O. sativa*. The GC contents for the total and three individual codon positions were similar for the

**Table 1: Log likelihood values and parameter estimates under different branch models and tests of hypotheses**

Model	p	ln	Parameters for Branches	Models Compared	2ΔL
Mf: ω free	116	-4928.56	ω:0 ~ 0.513 <sup>a</sup>	Mf vs. M0	184.74***
M0: ω <sub>0</sub> = ω <sub>d1</sub> = ω <sub>d5</sub> = ω <sub>1</sub> = ω <sub>5</sub>	1	-5020.93	ω <sub>0</sub> = ω <sub>d1</sub> = ω <sub>d5</sub> = ω <sub>1</sub> = ω <sub>5</sub> = 0.055		
M2r: ω <sub>0</sub> ≠ ω <sub>d1</sub> = ω <sub>1</sub> = ω <sub>d5</sub> = ω <sub>5</sub>	2	-5015.74	ω <sub>0</sub> = 0.046 ω <sub>d1</sub> = ω <sub>d5</sub> = ω <sub>1</sub> = ω <sub>5</sub> = 0.077	M2r vs. M0	10.38**
M3r: ω <sub>0</sub> ≠ ω <sub>d1</sub> = ω <sub>1</sub> ≠ ω <sub>d5</sub> = ω <sub>5</sub>	3	-5009.32	ω <sub>0</sub> = 0.044 ω <sub>d1</sub> = ω <sub>1</sub> = 0.118 ω <sub>d5</sub> = ω <sub>5</sub> = 0.060	M3r vs. M2r	12.84***
M4r: ω <sub>0</sub> ≠ ω <sub>d1</sub> = ω <sub>d5</sub> ≠ ω <sub>1</sub> ≠ ω <sub>5</sub>	4	-5008.12	ω <sub>0</sub> = 0.044 ω <sub>d1</sub> = ω <sub>d5</sub> = 0.085 ω <sub>1</sub> = 0.122 ω <sub>5</sub> = 0.066	M4r vs. M3r	2.40
M5r: ω <sub>0</sub> ≠ ω <sub>d1</sub> ≠ ω <sub>d5</sub> ≠ ω <sub>1</sub> ≠ ω <sub>5</sub>	5	-5007.06	ω <sub>0</sub> = 0.044 ω <sub>d1</sub> = 0.043 ω <sub>d5</sub> = ∞ ω <sub>1</sub> = 0.121 ω <sub>5</sub> = 0.066	M5r vs. M4r	2.12

<sup>a</sup> zero *d*<sub>5</sub> branches are excluded.

\*\*Significant at *P* < 0.01 level; \*\*\* Significant at the *P* < 0.001 level.  
*p*, number of parameters.

same gene but those at the 3rd position (GC<sub>3</sub>) is higher in *TFIIAγ1* than in *TFIIAγ5* (75.9% vs. 70.1%, *P* < 0.001) [Additional file 3]. Estimates of the codon usage showed that *TFIIAγ5* had significantly lower ENC value than *TFIIAγ1* (42.9 vs. 48.5, *P* < 0.001), paralleling its higher expression level in grasses (see below).

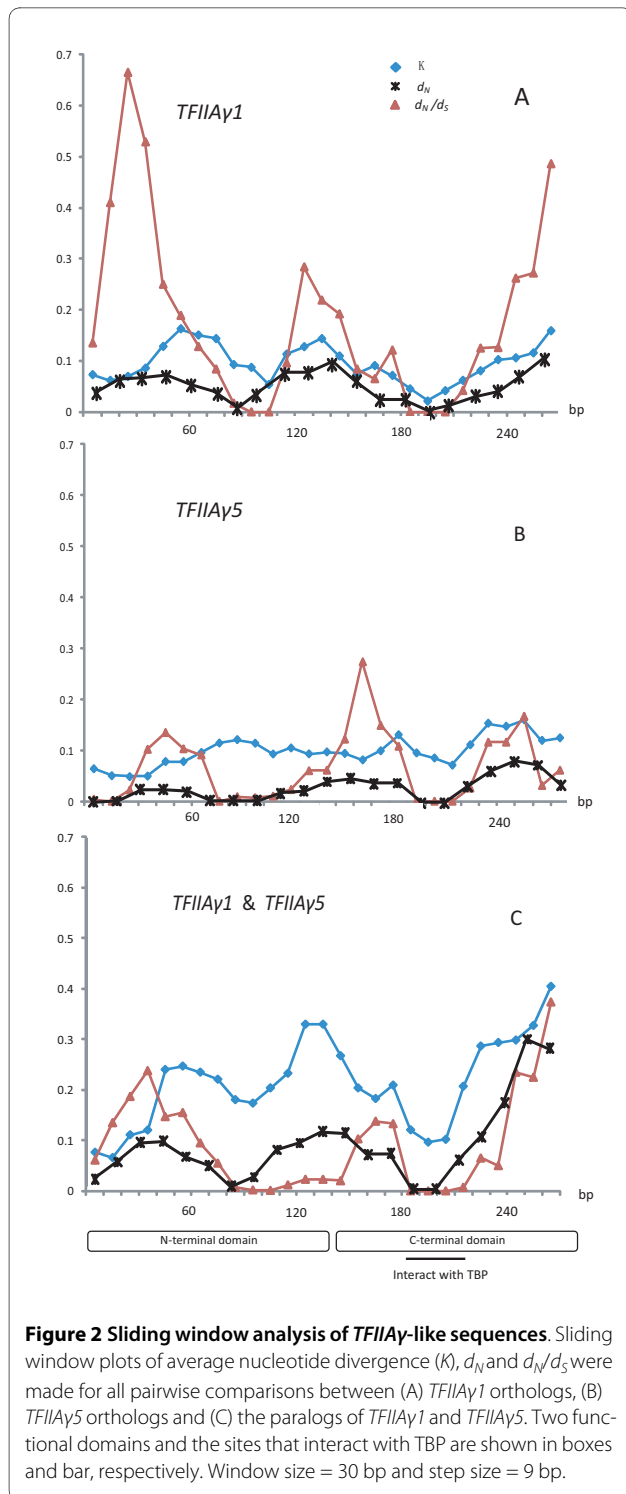
#### Phylogeny of *TFIIAγ* genes

The alignment of all the coding sequences was 318 bp in length including gaps. Of them, 152 sites were parsimony informative. A Bayesian phylogeny indicated that all monocot species except for *Zostera marina* of Zosteraceae formed a monophyletic group, which forms polytomy with the other angiosperm clades. Such unsolved relationship reflects our current understanding of angiosperm phylogeny on which monocots were not resolved fully with many other basal angiosperms [50]. It is noted that all the *TFIIAγ* sequences from the Poaceae species formed two clades supported by Bayesian posterior probability > 90, one consisting of *TFIIAγ1* homologs and the other *TFIIAγ5* homologs (Figure 1). All the Oryzeae species and most grass species outside Oryzeae have two distinct types of *TFIIAγ* sequences that fell into the two clades. In some grass species, only one *TFIIAγ*-like sequence was isolated, which formed a cluster with either *TFIIAγ1*-like or *TFIIAγ5*-like clade. In contrast, a single *TFIIAγ*-like copy was found in two species from the families closely related to Poaceae, *Cyperus rotundus* of

Cyperaceae and *Zingiber officinale* of Zingiberaceae. Moreover, the monocot clade is sister to the *TFIIAγ*-like sequences from the remaining angiosperm species (Figure 1). ML analyses produced similar tree topologies [Additional file 4]. These observations indicated that the duplication event giving rise to *TFIIAγ1* and *TFIIAγ5* occurred at the ancestors of Poaceae or before the divergence of Poaceae.

#### Sequence conservation and rate difference between two *TFIIAγ* genes

We performed a sliding window analysis by calculating the nucleotide divergence of the entire sequence with JC model (K), of nonsynonymous (*d*<sub>N</sub>) and synonymous substitution sites (*d*<sub>S</sub>). The *d*<sub>N</sub> values for both genes were lower than those of *d*<sub>S</sub> (*d*<sub>N</sub>/*d*<sub>S</sub> ≤ 0) in almost all sliding windows but all three parameters fluctuated across the genes (Figure 2). The conserved regions in *TFIIAγ1* are different from those in *TFIIAγ5* and some sites in *TFIIAγ1* might experience relaxation of selective constraints with elevated *d*<sub>N</sub>/*d*<sub>S</sub> values relative to those of *TFIIAγ5* (Figure 2A and 2B). In addition, both the K and *d*<sub>N</sub> values of *TFIIAγ1* were higher than those of *TFIIAγ5*, suggesting higher rate of evolution in *TFIIAγ1* genes. To detect the potential impact of intergenic conversion on molecular evolution [56], we further calculated the parameters between two paralogs (Figure 2C). We did not find significant difference in evolutionary rates between two domains in which



heterogeneity occurred across the sequences. It is evidence that low sequence differentiation was found around the functional regions (e.g., the region that interact with TBP), inconsistent with variation pattern of gene conversion that sequence divergence would occurred around the functional site [56].

Relative rate test was used to compare the *TFIIAγ1* and *TFIIAγ5* sequences from the main lineages in grasses in relations to the *TFIIAγ*-like sequence from *Ananas comosus* of the family Bromeliaceae that is closely related to Poaceae [50]. For all paralogs from 12 species tested, *TFIIAγ1* evolved 1.14 to 1.34 times faster than *TFIIAγ5* (Table 2). The tests were statistically significant or marginal significant for six out of 12 species. When more distinctly related species *Zingiber officinale* was used as an outgroup, the results were similar in that *TFIIAγ1* evolved faster than *TFIIAγ5* in all 12 species though the tests were not significant (Table 2). We calculated the synonymous and nonsynonymous substitution rates of *TFIIAγ1* and *TFIIAγ5* between the Oryzeae species and found that the average  $d_N$  value of *TFIIAγ1* was significantly higher than that of *TFIIAγ5* (0.033 vs. 0.011,  $P < 0.001$ ); the pairwise  $d_S$  values of *TFIIAγ1* and *TFIIAγ5* were also significant (0.155 vs. 0.131,  $P = 0.001$ ). The accelerated  $d_N$  in *TFIIAγ1* is obvious when we examined the amino acid alignments for the two genes, in which 21 sites had amino acid mutations in *TFIIAγ1* in contrast to 14 sites in *TFIIAγ5* [Additional file 5]. The overall  $\omega$  ( $d_N/d_S$ ) values for both genes were far below 1 (0.213 for *TFIIAγ1* and 0.084 for *TFIIAγ5*), indicating both genes were subjected to selection constraint, but the constraint on *TFIIAγ5* was stronger.

#### Selection constraints among lineages

We used different kinds of likelihood ratio tests to examine whether there was variation of  $\omega$  ratios on different lineages and, in particular, whether there is any increase in the  $\omega$  ratio after the *TFIIAγ* duplication. Free ratio (Mf) and two ratio (M2r) models both have significantly higher likelihood scores than one ratio model (M0), rejecting the null hypothesis that the *TFIIAγ*-like genes have evolved at constant rates along branches (Table 1). However, branch-specific  $\omega$  values under Mf model were all lower than one (ranging from 0 ~0.513), suggesting that purifying selection or constraint on amino acid sequence best explains the evolution of *TFIIAγ*-like genes in angiosperms. Two ratio model, with  $\omega_0 = 0.046$  for all branches before the *TFIIAγ* duplication and  $\omega_{d1} = \omega_{d5} = \omega_1 = \omega_5 = 0.077$  for the branches after the duplication, fits the data significantly better than one ratio models (M2r vs. M0,  $2\Delta L = 10.38$ ,  $P < 0.001$ ), indicative of a significant increase in  $\omega$  value following the duplication event. We further calculated the likelihood under comparison between models M3r and M2r to explore the assumption of the same selective constraints at two *TFIIAγ* genes after the duplication event. Likelihood of model M3r was significantly better than M2r ( $2\Delta L = 12.84$ ,  $P < 0.0001$ ), suggesting that different selective pressures occur in the two *TFIIAγ* genes with stronger purifying selection in

**Table 2: Tajima's relative rate tests for *TFIIAγ1/TFIIAγ5* duplicates using *Ananas comosus* and *Zingiber officinale* as outgroups**

Species	<i>Ananas comosus</i>			<i>Zingiber officinale</i>		
	$\gamma_1/\gamma_5^a$	$\chi^2$	P-value	$\gamma_1/\gamma_5$	$\chi^2$	P-value
<i>Oryza sativa</i>	1.14	1.32	0.250	1.18	1.60	0.206
<i>Oryza brachyantha</i>	1.19	2.31	0.128	1.09	0.42	0.516
<i>Oryza granulata</i>	1.14	1.26	0.262	1.07	0.23	0.631
<i>Leersia tisserantii</i>	1.14	1.09	0.297	1.07	0.20	0.655
<i>Potamophila parviflora</i>	1.29	4.83	<b>0.028*</b>	1.16	1.26	0.262
<i>Chikusichloa aquatica</i>	1.26	4.00	<b>0.046*</b>	1.11	0.64	0.423
<i>Rhynchoriza subulata</i>	1.21	2.78	0.096	1.14	0.95	0.330
<i>Hygroryza aristata</i>	1.20	5.77	<b>0.016*</b>	1.19	1.68	0.194
<i>Luziola leiocarpa</i>	1.34	2.19	0.139	1.19	1.52	0.217
<i>Ehrharta erecta</i>	1.22	3.57	0.059	1.20	2.61	0.106
<i>Zea mays</i>	1.27	4.33	<b>0.037*</b>	1.21	2.94	0.086
<i>Sorghum bicolor</i>	1.15	2.13	0.144	1.05	0.29	0.590

<sup>a</sup> The ratio of the genetic distance (Kimura 2-parameter) between *TFIIAγ1* and the outgroup over that between *TFIIAγ5* and the outgroup.

\* indicates significance at  $P < 0.05$  level.

*TFIIAγ5* ( $\omega_{d5} = \omega_5 = 0.060$ ) than in *TFIIAγ1* gene ( $\omega_{d1} = \omega_1 = 0.118$ ). Finally, the comparison between M5r and M4r indicated that the  $\omega$  ratios of the two branches immediately following the duplication event were not significantly different from each other ( $2\Delta L = 2.12$ ,  $P = 0.145$ ) (Table 1), implying that the asymmetric rates of *TFIIAγ* evolution occurred mainly after diversification of grasses.

#### Detecting positive selection in *TFIIAγ* genes

Given the fact that the selective constraints on *TFIIAγ* genes relaxed after duplication and conferred disease resistant or induced by pathogen in cultivated rice, it is interesting to ask whether any accelerated rate of relaxation happen and any amino acid residue is potentially under positive selection. Because the branch model test averages the  $\omega$  ratios across all sites and is a very conservative test of positive selection [33], we applied site-specific and branch-site models to *TFIIAγ* dataset. As shown in Table 3, site-specific models indicate that *TFIIAγ* genes were under strong purifying selection with  $\omega = 0.055$  in one-ratio model (M0). The discrete model (M3) was sig-

nificantly better than M0 ( $2\Delta L = 193.62$ ,  $P < 0.001$ ), indicating that the  $\omega$  ratio was not homogeneous among sites along the sequence. This is also obvious in the sliding window analysis (Figure 2) and the amino acid alignment of *TFIIAγ* genes [Additional file 5]. Models M2 and M8 assuming positive selection were not significantly better than the null models M1 and M7 (for M1 vs. M2,  $2\Delta L = 0.0$ ,  $P = 1.0$ ; for M7 vs. M8,  $2\Delta L = 0.0$ ,  $P = 1.0$ ), and no site was found to be under positive selection by Bayes Empirical Bayes (BEB) inference [32] using a probability criterion of 95%. Thus, the nearly neutral model was better to explain the data. In model M1, about 94% of the codons are under strict constraint ( $\omega = 0.030$ ), and the other 6% codons are under neutral evolution ( $\omega = 1.0$ ) (Table 3).

We further tested for evidence of positive selection on two *TFIIAγ* genes separately using branch-site models (Table 3). Branch-site models A and B specifying branch *TFIIAγ1* as the foreground branch were not significantly better than the null models M1 ( $2\Delta L = 0.1$ ,  $P = 0.95$ ) and M3 ( $2\Delta L = -44.92$ ,  $P = 1.0$ ). In analyses of the branch *TFIIAγ5*, however, model A was significantly better than the null model ( $2\Delta L = 13.66$ ,  $P < 0.001$ ) with  $\omega$  ratio greater



**Table 3: Parameters and likelihood scores of *TFIIA* genes under codon and branch-site models**

Model	<i>p</i>	ln	2ΔL	Estimate of parameters	Positively selective site
M0: one ratio	1	-5020.93	184.74**	$\omega$ :0.055	none
Codon model					
M1: nearly neutral	1	-4915.09	211.68**	$p_0 = 0.938, p_1 = 0.062$ $\omega_0 = 0.038, \omega_1 = 1.0$	not allowed
M2: positive selection	3	-4915.09	0	$p_0 = 0.938, p_1 = 0.044,$ $p_2 = 0.017$ $\omega_0 = 0.038, \omega_1 = 1.0,$ $\omega_2 = 1.0$	none
M3: discrete	5	-4818.28	193.62**	$p_0 = 0.714, p_1 = 0.224,$ $p_2 = 0.063$ $\omega_0 = 0.008, \omega_1 = 0.124,$ $\omega_2 = 0.480$	none
M7: beta	2	-4815.29		$p = 0.200, q = 2.073$	not allowed
M8: beta & $\omega$	4	-4815.29	0	$p_0 = 1.0, p = 0.200,$ $q = 2.073$ $p_1 = 0, \omega = 2.0$	none
Branch-site model					
Foreground: <i>TFIIA</i> $\gamma$ 1					
Model A	3	-4915.04	0.10	$p_0 = 0.910, p_1 = 0.060$ $(p_2 + p_3 = 0.030) \omega_2 = 1.0$	none
Model B	5	-4840.74	-44.92	$p_0 = 0.806, p_1 = 0.194 (p_2 + p_3 = 0)$ $\omega_0 = 0.015, \omega_1 = 0.259,$ $\omega_2 = 0$	none
Foreground: <i>TFIIA</i> $\gamma$ 5					
Model A	3	-4908.26	13.66**	$p_0 = 0.929, p_1 = 0.058$ $(p_2 + p_3 = 0.013) \omega_2 = \infty$	90T
Model B	5	-4840.74	-44.92	$p_0 = 0.806, p_1 = 0.194 (p_2 + p_3 = 0)$ $\omega_0 = 0.015, \omega_1 = 0.259,$ $\omega_2 = 0$	none

\*\*Significant at  $P < 0.01$  level; \*\*\* Significant at the  $P < 0.001$  level.  
*p*, number of parameters.

than 1, but model B was not significantly better than the null model (Table 3). We checked the inferred positive selection site (90T) across all protein sequences and found that it was fixed in both copies, with all *TFIIA* $\gamma$ 1 genes being T and *TFIIA* $\gamma$ 5 genes Q [Additional file 5]. This observation suggests it unlikely that positive selection occurs in either copy in grasses. Alternatively, this site might experience positive selection immediately after duplication of *TFIIA* $\gamma$  gene in ancestor of grasses and then fixed under strong purifying selection in grasses. It should be noted that the *TFIIA* $\gamma$ 5 gene was highly expressed with significantly lower ENC relative to *TFIIA* $\gamma$ 1 gene [Additional file 3]. Therefore, the  $\omega$  value greater than one at 90 site of *TFIIA* $\gamma$ 5 gene might be caused by low  $d_s$  value rather than positive selection

because synonymous sites are likely to be under negative selection in highly expressed genes due to codon usage bias [57].

#### Gene expression of the *TFIIA* genes

Two rounds of RT-PCR were performed to determine the expression of *TFIIA* $\gamma$ 1 and *TFIIA* $\gamma$ 5 genes in tribe Oryzeae species. In the first round, equal amount of template cDNA was added in the reaction of *TFIIA* $\gamma$ 1 and *TFIIA* $\gamma$ 5. The expression of *TFIIA* $\gamma$ 5 was detected in all the leaves and young panicles, while the expression of *TFIIA* $\gamma$ 1 was weaker than that of *TFIIA* $\gamma$ 5 for most expected bands, and were almost invisible in *O. officinalis*, *O. australiensis* and *Leersia tisserantii* (Figure 3). The weaker bands of *TFIIA* $\gamma$ 1 indicated that it was expressed

at lower level relative to *TFIIA $\gamma$ 5*. When a second round PCR was taken, the expected bands appeared in all the species. To avoid contamination, all RT-PCR products of *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* were confirmed by sequencing, and the resulting sequences were identical to the coding regions of genomic sequences in each species. These results showed that both copies were expressed in leaf and young panicle of Oryzaceae species, but the *TFIIA $\gamma$ 5* was expressed at higher level.

Different expression levels of two *TFIIA $\gamma$*  genes were further confirmed by the GenBank EST database search using rice *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* sequences. Both copies were found in rice, maize and sorghum, but the hits of *TFIIA $\gamma$ 5* far outnumbered those of the *TFIIA $\gamma$ 1* copy in rice and maize [Additional file 6]. In several other Poaceae species, only the *TFIIA $\gamma$ 5* copy was found. The low number of hits indicated that the *TFIIA $\gamma$ 1* expression was much lower than that of *TFIIA $\gamma$ 5*, consistent with our RT-PCR findings. In addition, the matches of *TFIIA $\gamma$ 5* expression were found in all types of cDNA libraries, including the callus, mature or immature tissue, stressed or unstressed and different developing stage libraries; whereas the *TFIIA $\gamma$ 1* hits appeared mainly in drought-stressed tissue, pollen, immature and meristematic and mixed libraries [Additional file 6]. These observations suggest that *TFIIA $\gamma$ 5* might be constitutively expressed and *TFIIA $\gamma$ 1* be expressed under stress induction or expressed in specific tissues.

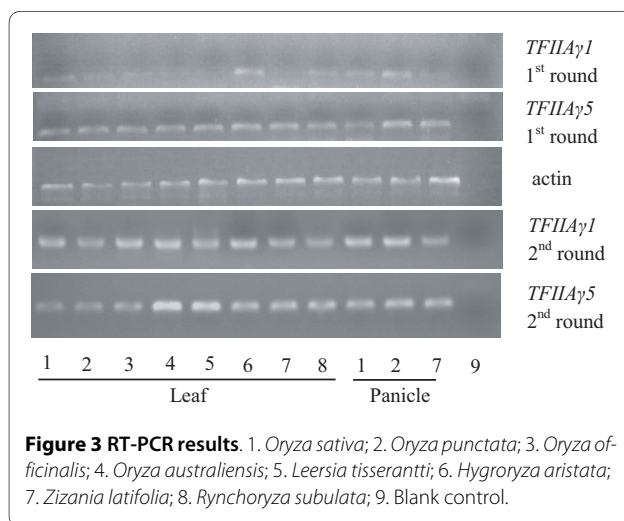
## Discussion

This study identified two *TFIIA $\gamma$*  genes for all Oryzaceae species and the representatives of grass species, which formed two monophyletic clades corresponding to the rice *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* genes; whereas only a single copy was found for the remaining monocots and angiosperm species. Phylogenetic analyses of all the *TFIIA $\gamma$* -like sequences indicated that the duplication of *TFIIA $\gamma$*

into *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* occurred before the divergence of rice and maize (Figure 1). This implies that the duplication event that gave rise to *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* genes might occur before the common ancestor of extant grasses because rice (subfamily Ehrhartoideae) and maize (subfamily Panicoideae) are two distinctly related lineages in the grass family [58,59].

It has been demonstrated that the rice genome experienced two large-scale duplications, one whole genome duplication occurred about 70 MYA, and an additional segmental duplication happened 5 ~ 21 MYA involving chromosomes 11 and 12 [60-62]. Previous studies found that the location of two rice *TFIIA $\gamma$*  genes corresponded to a large-scale duplication of a portion of rice chromosomes 1 and 5 [7,8]. To determine whether the timing of the duplication event leading to *TFIIA $\gamma$ 1/TFIIA $\gamma$ 5* is consistent with the whole genome duplication around 70 MYA, we calculated the synonymous distance ( $d_s$ ) between *TFIIA $\gamma$*  orthologs and paralogs for rice and maize by the method of Nei and Gojobori (1986). The  $d_s$  distances between the *TFIIA $\gamma$*  orthologs were 0.388 for rice and 0.457 for maize and those between the paralogs of rice and maize were 0.592 (*TFIIA $\gamma$ 1*) and 0.497 (*TFIIA $\gamma$ 5*), respectively. According to a molecular clock assuming rice and maize diverged 50 MYA [58], the *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* paralogs diverged about 54 ~ 76 MYA. This date coincides with the time scale that Poaceae diverged 55 ~ 77 MYA [58,59]. Wang et al. (2005) identified 10 large duplicated blocks arising from the whole genome duplication, including two blocks involving chromosomes 1 and 5. Our further search on rice genome found that two rice *TFIIA $\gamma$*  genes located on block 10 determined by Wang et al. (2005). Therefore, the *TFIIA $\gamma$*  duplication is within a large duplicated segment of rice genome and most likely to arise following a whole genome duplication event that was assumed to have occurred before the divergence of Poaceae [60-62].

Our timing of the *TFIIA $\gamma$*  duplication suggests that the *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* paralogs have been maintained in the grass genome for a considerable amount of time (at least 50 MYA). This implicates that selection rather than random drift is responsible for the retention of both *TFIIA $\gamma$*  activities during grass evolution because most gene duplicates have a short lifespan (within a few million years after duplication) before one copy was deleted (pseudogenization) [24]. It has been well established that gene duplication is often followed by an elevated rate of protein evolution and a large proportion of the duplicate pairs displayed asymmetric evolution, i.e., one of the duplicates evolves much faster than the other [19,29,63-65]. Conant and Wagner (2003) analyzed four completely sequenced genomes and found that 20% - 30% of duplicate gene pairs showed asymmetric evolution in the amino acid sequence, and particularly, the greater this



asymmetry, the greater the  $d_N/d_S$  ratio in a gene pair, indicating that most asymmetric divergence might be caused by relaxed selective constraints on one of the duplicates. In well agreement with previous studies, we found significantly higher  $\omega$  ratios for branches arising from the duplication event in rice tribe and its relatives, suggesting weaker purifying selection on the duplicate genes during diversification of grasses after the duplication event. Moreover, the  $\omega$  ratios of the *TFIIA $\gamma$ 1* sequences are two times higher than those of *TFIIA $\gamma$ 5* sequences, consistent with the results of relative rate tests in which *TFIIA $\gamma$ 1* evolved faster than *TFIIA $\gamma$ 5* (Table 2). Such an asymmetric evolution of the *TFIIA $\gamma$*  duplicates reflects an acceleration of evolutionary rate of *TFIIA $\gamma$ 1* relative to *TFIIA $\gamma$ 5*. Our likelihood-based analyses with both branch and codon models showed no evidence of positive selection but a signature of relaxed selective constraint after the *TFIIA $\gamma$*  duplication and subsequent acceleration of *TFIIA $\gamma$ 1* gene. The low  $\omega$  values (0.060 ~ 0.118) across the branches leading to both *TFIIA $\gamma$*  duplicates also suggest that strong selection constraints remain for the two copies after the duplication, with *TFIIA $\gamma$ 1* evolving under weaker selective constraint in grass species.

The fate of duplicated genes has been a hot debate since Ohno (1970), and several hypotheses have been proposed to interpret the preservation of both copies, including neofunctionalization [11], subfunctionalization [21,24], subneofunctionalization [66] and some other models (see review in Semon and Wolfe 2007). Based on sequence analyses and expression data, Iyer and McCouch (2004) found that the recessive mutation on *TFIIA $\gamma$ 5* locus for resistance to rice bacterial blight did not affect the essential function of *TFIIA $\gamma$*  gene and hypothesized that *TFIIA $\gamma$ 5* functioned both as a general transcription factor and as a resistance gene (*xa5*) in rice, which was further demonstrated by subsequent complementation test and 3-D structure prediction [7]. We conducted a secondary structure prediction of the *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* proteins of grass species and found little difference in the secondary structures between the two copies [Additional file 5]. These observations, in combination of our molecular evolutionary analyses (Tables 1 and 3), demonstrated that both *TFIIA $\gamma$*  genes were functional and under selection constraint in Oryzae and its relatives. Thus, pseudogenization is unlikely involved in *TFIIA $\gamma$*  evolution. Because extra amounts of protein or RNA products such as rRNAs and histones are in high demand [22], the retention of both *TFIIA $\gamma$*  copies might be attributed partly to the importance of *TFIIA $\gamma$*  as a component of TFIIA that is a general transcription factor needed in all polymerase II transcriptions [4,5].

Jiang et al (2006) investigated the expression patterns of two *TFIIA $\gamma$*  genes in rice and indicated that the *TFIIA $\gamma$ 1* gene was not expressed in young panicle, in contrast to

*TFIIA $\gamma$ 5* that expressed in all organs tested (leaf, stem, panicle, and root). In our study on *O. sativa*, *O. punctata* and *Z. latifolia*, however, the expression of *TFIIA $\gamma$ 1* was detected in both leaves and young panicles but the expression level was much lower relative to *TFIIA $\gamma$ 5* gene (Figure 3). These observations, in conjunction with our expression data, indicate that after whole genome duplication, the expression of *TFIIA $\gamma$ 1* copy was significantly reduced while *TFIIA $\gamma$ 5* remained constitutively expressed and maintained the ancestral role as a subunit of the TFIIA complex. Consequently, it seems that subfunctionalization might be involved in *TFIIA $\gamma$*  evolution in grasses. The case of *TFIIA $\gamma$*  genes agree with previous notion that subfunctionalization would lead to functional specialization when one of the duplicate genes became better at performing the original function of the progenitor gene [22]. Nevertheless, the possibility that positive selection on some specific sites immediately after duplication of *TFIIA $\gamma$*  gene in ancestor of grasses cannot be excluded entirely given short length of the *TFIIA $\gamma$*  gene and the inference power of methods in our case [67].

One important point for the evolution of *TFIIA $\gamma$*  genes is the evidence that both *TFIIA $\gamma$ 1* and *TFIIA $\gamma$ 5* genes were effectively involved in response to biotic or abiotic factors. In addition to *xa5* mutation that lead to resistance to rice bacterial blight, a recent study documented that the expression of *TFIIA $\gamma$ 1* could express 400-fold greater than normal when infected by specific bacterial races (PXO99<sup>A</sup>) that cause blight disease [9]. Our EST database search also found the frequent presence of *TFIIA $\gamma$ 1* gene in drought-stressed cDNA library both in rice and sorghum, implying its inducibility by drought stress [Additional file 6]. As pointed out by previous authors, gene redundancy might create subtle fitness advantage that was only evident in particular stages of the life cycle or under particular environments [25,68,69]. Therefore, the fate of the duplicated *TFIIA $\gamma$*  genes can be alternatively explained by the Dykhuizen-Hartl effect [31,34], which predicts that one of duplicate genes evolves under relaxed purifying selection and the fixed mutations later convey a selective advantage in a novel environment or genetic background. It is noted that the V39E substitution in the  $\alpha$ -helix domain of *TFIIA $\gamma$ 5* was confined only to some varieties of *O. sativa*, suggestive of its recent emergence [7,8] [Additional file 5].

The involvement of the duplicated *TFIIA $\gamma$*  genes in adversity response could also be explained by the buffering hypothesis [27], which suggests that selection for a buffering effect was a mechanism for duplicate gene preservation after whole genome duplication. By exploring the footprints of selection associated with genome duplication in *Arabidopsis* ecotypes and rice subspecies, Chapman et al. (2006) found that functional buffering might be important against genetic turbulence after

genome duplication and could continue to act ~60 million years later. Retention of duplicate genes, particularly for complex genes and gene network, plays a critical role for genetic robustness of biological systems [22,25,27,70,71]. TFIIA is a complex consisting of three polypeptides and assumed recently to be tightly regulated with a particular role in differentiation and development [6]. Further biochemical and molecular investigations on the respective functions and the interactions between TFIIA and the other two components will be required to better understanding of the biology of the transcription factor TFIIA and to provide useful insights into the evolution of *TFIIA* and its counterparts.

## Conclusions

Based on phylogenetic reconstruction of the *TFIIA* genes from main lineages of angiosperms, we demonstrated that two *TFIIA* genes (*TFIIA1* and *TFIIA5*) arose from a whole genome duplication that happened in the common ancestor of grasses. Likelihood-based analyses with different models showed no evidence of positive selection but a signature of relaxed selective constraint after the *TFIIA* duplication. In particular, the nonsynonymous/synonymous rate ratio ( $\omega = d_N/d_S$ ) of the *TFIIA1* sequences was two times higher than that of *TFIIA5* sequences, indicating highly asymmetric rates of protein evolution in rice tribe and its relatives. Our expression data and EST database search further indicated that after whole genome duplication, the expression of *TFIIA1* gene was significantly reduced while *TFIIA5* remained constitutively expressed and maintained the ancestral role as a subunit of the TFIIA complex. These observations are not consistent with the neofunctionalization model that predicts that one of the duplicated genes acquires a new function and instead, implicate that subfunctionalization might be involved in *TFIIA* evolution in grasses. The fact that both *TFIIA1* and *TFIIA5* genes were effectively involved in response to biotic or abiotic factors might be explained by either Dykhuizen-Hartl effect or buffering hypothesis.

## Additional material

**Additional file 1** *TFIIA*-like sequences included in this study.

**Additional file 2** Gene structure and the location of primers. Universal forward (P1 and P3) and reverse (P2 and P4) primers are shown above the genes and the copy-specific internal sequencing primers (P7 and P8) are shown below the gene. Exons are shown in boxes and the shaded boxes are coding regions.

**Additional file 3** GC contents (%) and ENC of *TFIIA1* and *TFIIA5* in *Oryzae* species and its relative.

**Additional file 4** Maximum likelihood tree using GTR+I+G model of evolution. Bootstrap values > 50% are shown above branches.

**Additional file 5** Amino acid alignment of the *TFIIA* genes. 2D structure in the bottom is predicted by PredictProtein <http://www.predictprotein.org/> using *O. sativa* sequences as references. H represents the alpha helix and E the beta strand.

**Additional file 6** EST hits of grass *TFIIA* genes in GenBank EST database.

## Abbreviations

TBP: TATA-binding protein; ENC: effective number of codons; EST: expressed sequence tags; ML: maximum likelihood; BI: Bayesian inference; MCMC: Markov chain Monte Carlo.

## Authors' contributions

SG and HZS designed the research and outlined the manuscript together. HZS performed the research. HZS and SG analyzed and interpreted the data. SG and HZS wrote the paper. Both authors have read and approved the final manuscript.

## Acknowledgements

We thank Qihui Zhu, Xin-Hui Zou, Yan-Hua Yang and other members of Ge's group for their helps during the experiment and data analyses. We also thank Frank White and Bin Yang of Kansas State University of USA for providing useful information during early stage of this study. We are grateful to the International Rice Research Institute (Los Banos, Philippines) for providing leaf and seed samples. This study was supported by the National Basic Research Program of China (2007CB815704), National Natural Science Foundation of China (30990240 and 30430030), and the grants from the Chinese Academy of Sciences.

## Author Details

<sup>1</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China and <sup>2</sup>The Graduate University of the Chinese Academy of Sciences, Beijing 100039, China

Received: 22 October 2009 Accepted: 4 May 2010

Published: 4 May 2010

## References

1. Riechmann JL, Ratcliffe OJ: **A genomic perspective on plant transcription factors.** *Curr Opin Plant Biol* 2000, **3**(5):423-434.
2. Xiong Y, Liu T, Tian C, Sun S, Li J, Chen M: **Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots.** *Plant Mol Biol* 2005, **59**(1):191-203.
3. Doebley J, Lukens L: **Transcriptional regulators and the evolution of plant form.** *Plant Cell* 1998, **10**(7):1075-1082.
4. Orphanides G, Lagrange T, Reinberg D: **The general transcription factors of RNA polymerase II.** *Genes Dev* 1996, **10**(21):2657-2683.
5. Ranish JA, Hahn S: **The yeast general transcription factor TFIIA is composed of two polypeptide subunits.** *J Biol Chem* 1991, **266**(29):19320-19327.
6. Hoiby T, Zhou H, Mitsiou DJ, Stunnenberg HG: **A facelift for the general transcription factor TFIIA.** *Biochimica et biophysica acta* 2007, **1769**(7-8):429-436.
7. Jiang GH, Xia ZH, Zhou YL, Wan J, Li DY, Chen RS, Zhai WX, Zhu LH: **Testifying the rice bacterial blight resistance gene xa5 by genetic complementation and further analyzing xa5 (Xa5) in comparison with its homolog TFIAGamma1.** *Mol Genet Genomics* 2006, **275**(4):354-366.
8. Iyer AS, McCouch SR: **The rice bacterial blight resistance gene xa5 encodes a novel form of disease resistance.** *Mol Plant Microbe Interact* 2004, **17**(12):1348-1354.
9. Sugio A, Yang B, Zhu T, White FF: **Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes OsTFIIAGamma1 and OsTFX1 during bacterial blight of rice.** *Proc Natl Acad Sci USA* 2007, **104**(25):10720-10725.
10. Semon M, Wolfe KH: **Consequences of genome duplication.** *Curr Opin Genet Dev* 2007, **17**(6):505-512.
11. Ohno S: **Evolution by Gene Duplication.** New York: Springer-Verlag; 1970.
12. Zhang W, Xiang QY, Thomas DT, Wiegmann BM, Frohlich MW, Soltis DE: **Molecular evolution of PISTILLATA-like genes in the dogwood genus *Cornus* (Cornaceae).** *Mol Phylogenet Evol* 2008, **47**(1):175-195.
13. Streisfeld MA, Rausher MD: **Relaxed Constraint and Evolutionary Rate Variation between Basic Helix-Loop-Helix Floral Anthocyanin Regulators in *Ipomoea*.** *Mol Biol Evol* 2007, **24**(12):2816-2826.

14. Preston JC, Kellogg EA: **Reconstructing the evolutionary history of paralogous APETALA1/FRUITFULL-like genes in grasses (Poaceae).** *Genetics* 2006, **174**(1):421-437.
15. Baum DA, Yoon HS, Oldham RL: **Molecular evolution of the transcription factor LEAFY in Brassicaceae.** *Mol Phylogenet Evol* 2005, **37**(1):1-14.
16. Hileman LC, Baum DA: **Why do paralogs persist? Molecular evolution of CYCLOIDEA and related floral symmetry genes in Antirrhineae (Veronicaceae).** *Mol Biol Evol* 2003, **20**(4):591-600.
17. Moore RC, Purugganan MD: **The evolutionary dynamics of plant duplicate genes.** *Curr Opin Plant Biol* 2005, **8**(2):122-128.
18. Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T: **Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus.** *Proc Natl Acad Sci USA* 2003, **100**(Suppl 2):14587-14592.
19. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**(5494):1151-1155.
20. Nei M, Roychoudhury AK: **Probability of fixation and mean fixation time of an overdominant mutation.** *Genetics* 1973, **74**(2):371-380.
21. Hughes AL: **The evolution of functionally novel proteins after gene duplication.** *Proc Biol Sci* 1994, **256**(1346):119-124.
22. Zhang J: **Evolution by gene duplication: an update.** *Trends Ecol Evol* 2003, **18**:292-298.
23. Walsh B: **Population-genetic models of the fates of duplicate genes.** *Genetica* 2003, **118**(2-3):279-294.
24. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151**(4):1531-1545.
25. Nowak MA, Boerlijst MC, Cooke J, Smith JM: **Evolution of genetic redundancy.** *Nature* 1997, **388**(6638):167-171.
26. Scannell DR, Wolfe KH: **A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast.** *Genome Res* 2008, **18**(1):137-147.
27. Chapman BA, Bowers JE, Feltus FA, Paterson AH: **Buffering of crucial functions by paleologous duplicated genes may contribute cyclicity to angiosperm genome duplication.** *Proc Natl Acad Sci USA* 2006, **103**(8):2730-2735.
28. Bielawski JP, Yang Z: **Maximum likelihood methods for detecting adaptive evolution after gene duplication.** *J Struct Funct Genomics* 2003, **3**(1-4):201-212.
29. Ohta T: **Further Examples of Evolution by Gene Duplication Revealed through DNA-Sequence Comparisons.** *Genetics* 1994, **138**(4):1331-1337.
30. Shiu SH, Byrnes JK, Pan R, Zhang P, Li WH: **Role of positive selection in the retention of duplicate genes in mammalian genomes.** *Proc Natl Acad Sci USA* 2006, **103**(7):2232-2236.
31. Zhang J, Rosenberg HF, Nei M: **Positive Darwinian selection after gene duplication in primate ribonuclease genes.** *Proc Natl Acad Sci USA* 1998, **95**(7):3708-3713.
32. Yang Z, Wong WS, Nielsen R: **Bayes empirical bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**(4):1107-1118.
33. Yang Z, Bielawski JP: **Statistical methods for detecting molecular adaptation.** *Trends Ecol Evol* 2000, **15**(12):496-503.
34. Dykhuizen D, Hartl DL: **Selective neutrality of 6PGD allozymes in E. coli and the effects of genetic background.** *Genetics* 1980, **96**(4):801-817.
35. Guo YL, Ge S: **Molecular phylogeny of Oryzae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes.** *Am J Bot* 2005, **92**(9):1548-1558.
36. Vaughan DA: **The wild relative of rice: a genetic resources handbook.** International Rice Research Institute, Manila, Philippines; 1994.
37. GPWG: **Phylogeny and subfamilial classification of the grasses (Poaceae).** *Ann Missouri Bot Gard* 2001, **88**:373-457.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
39. Hagen-Larsen H, Laerdahl JK, Panitz F, Adzhubei A, Hoyheim B: **An EST-based approach for identifying genes expressed in the intestine and gills of pre-smolt Atlantic salmon (Salmo salar).** *BMC Genomics* 2005, **6**:171.
40. Martin SA, Caplice NC, Davey GC, Powell R: **EST-based identification of genes expressed in the liver of adult Atlantic salmon (Salmo salar).** *Biochem Biophys Res Commun* 2002, **293**(1):578-585.
41. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Series* 1999, **41**:95-98.
42. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DF: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucl Acids Res* 1997, **24**:4876-4882.
43. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5**(2):150-163.
44. Wright F: **The 'effective number of codons' used in a gene.** *Gene* 1990, **87**(1):23-29.
45. Tajima F: **Simple methods for testing the molecular evolutionary clock hypothesis.** *Genetics* 1993, **135**(2):599-607.
46. Comeron JM: **K-estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals.** *Bioinformatics* 1999, **15**(9):763-764.
47. Swofford DL: **PAUP\*: phylogenetic analysis using parsimony (\* and other methods), version 4.10b10.** Sunderland, MA, USA: Sinauer Associates; 2001.
48. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**(8):754-755.
49. Posada D, Buckley TR: **Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests.** *Syst Biol* 2004, **53**:793-808.
50. Bremer B, Bremer K, Chase MW, Reveal JL, Soltis DE, Soltis PS, Stevens PF, Anderberg AA, Fay MF, Goldblatt P, et al.: **An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II.** *Bot J Linn Soc* 2003, **141**(4):399-436.
51. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**(8):1586-1591.
52. Yang Z, Nielsen R: **Synonymous and nonsynonymous rate variation in nuclear genes of mammals.** *J Mol Evol* 1998, **46**(4):409-418.
53. Endo T, Ikeo K, Gojobori T: **Large-scale search for genes on which positive selection may operate.** *Mol Biol Evol* 1996, **13**(5):685-690.
54. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148**(3):929-936.
55. Yang Z, Nielsen R: **Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages.** *Mol Biol Evol* 2002, **19**(6):908-917.
56. Teshima KM, Innan H: **Neofunctionalization of duplicated genes under the pressure of gene conversion.** *Genetics* 2008, **178**:1385-1398.
57. Davis JC, Petrov DA: **Preferential duplication of conserved proteins in eukaryotic genomes.** *PLoS Biol* 2004, **2**(3):E55.
58. Gaut BS: **Evolutionary dynamics of grass genomes.** *New Phytol* 2002, **154**(1):15-28.
59. Kellogg EA: **Evolutionary history of the grasses.** *Plant Physiol* 2001, **125**(3):1198-1205.
60. Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, et al.: **The Genomes of Oryza sativa: a history of duplications.** *PLoS Biol* 2005, **3**(2):e38.
61. Wang X, Shi X, Hao B, Ge S, Luo J: **Duplication and DNA segmental loss in the rice genome: implications for diploidization.** *New Phytol* 2005, **165**(3):937-946.
62. Paterson AH, Bowers JE, Chapman BA: **Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics.** *Proc Natl Acad Sci USA* 2004, **101**(26):9903-9908.
63. Zhang P, Gu Z, Li WH: **Different evolutionary patterns between young duplicate genes in the human genome.** *Genome Biol* 2003, **4**(9):R56.
64. Conant GC, Wagner A: **Asymmetric sequence divergence of duplicate genes.** *Genome Res* 2003, **13**(9):2052-2058.
65. Peer Y Van de, Taylor JS, Braasch I, Meyer A: **The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes.** *J Mol Evol* 2001, **53**(4-5):436-446.
66. He X, Zhang J: **Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution.** *Genetics* 2005, **169**(2):1157-1164.
67. Nunney L, Schuenzel EL: **Detecting natural selection at the molecular level: A reexamination of some "Classic" examples of adaptive evolution.** *J Mol Evol* 2006, **62**:176-195.
68. Wendel JF: **Genome evolution in polyploids.** *Plant Mol Biol* 2000, **42**(1):225-249.

69. Hughes MK, Hughes AL: **Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*.** *Mol Biol Evol* 1993, **10**(6):1360-1369.
70. Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH: **Role of duplicate genes in genetic robustness against null mutations.** *Nature* 2003, **421**(6918):63-66.
71. Gu X: **Evolution of duplicate genes versus genetic robustness against null mutations.** *Trends Genet* 2003, **19**(7):354-356.

doi: 10.1186/1471-2148-10-128

**Cite this article as:** Sun and Ge, Molecular evolution of the duplicated TFIIA? genes in *Oryzaeae* and its relatives *BMC Evolutionary Biology* 2010, **10**:128

**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)

