Supplemental text:

Discussion

Conversion and GC content

It has been reported that gene conversion may increase GC content (Eyre-Walker and Hurst 2001), and hypothesized that conversion may be accompanied by DNA repair of nucleotide mismatches. If the repair process is biased towards G and C, an elevation in GC content will result (Galtier et al. 2001). However, this hypothesis has always been controversial (Eyre-Walker and Hurst 2001). Supporting evidence comes from analysis of the pseudoautosomal region (PAR) of the mouse Y chromosome (Perry and Ashworth 1999), showing that the GC3 of a gene recently translocated into the PAR region has increased from 0.50 to 0.70 in less than 1 million years. The analysis of multigene families and tandem genes from mammals, plants and yeast also related GC elevation to possible gene conversion (Benovoy et al. 2005; Galtier 2003; Kudla et al. 2004). However, the correlation between GC and sequence similarity, taken as an indirect sign of possible gene conversion, is low. Even if conversion is elevating GC content, it may account for only about 10-20% of the elevation (Benovoy et al. 2005). However, correlation does not mean causation. Here, we did not find evidence that conversion caused GC enrichment. Many grass genes are considerably enriched with G and C, with GC3 ~0.90, resulting in two distinct groups of genes (Carels and Bernardi 2000). Comparatively, our wholly converted gene set has only an averaged GC3 ~0.70, quite similar to that of the non-converted. This could not support the point that conversion has contributed sufficiently to the GC elevation to lead to two distinct groups of genes. If conversion increased GC content, we would find significant GC increase in the converted genes. We checked the GC content of the genes in the region at short arms' ends of Os11and Os12, where conversion may have recursively occurred and is possibly still on-going. However, we did not find GC elevation in those genes, which actually show lower GC content (0.65) than the average of all genes in colinearity (0.71). This strongly suggests that gene conversion may not contribute to GC elevation. Many other factors have been related to GC enrichment, such as transcription, translation, methylation and mutation bias (Eyre-Walker and Hurst 2001; Wong et al. 2002), making it still of much interest.

References

Benovoy, D., R.T. Morris, A. Morin, and G. Drouin. 2005. Ectopic gene conversions increase the G + C content of duplicated yeast and *Arabidopsis* genes. *Mol Biol Evol* **22:** 1865-1868.

Carels, N. and G. Bernardi. 2000. Two classes of genes in plants. *Genetics* **154:** 1819-1825.

Eyre-Walker, A. and L.D. Hurst. 2001. The evolution of isochores. *Nat Rev Genet* **2:** 549-555.

Galtier, N. 2003. Gene conversion drives GC content evolution in mammalian histones. *Trends Genet* **19:** 65-68.

Galtier, N., G. Piganeau, D. Mouchiroud, and L. Duret. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* **159:** 907-911.

Lin, Y.S., J.K. Byrnes, J.K. Hwang, and W.H. Li. 2006. Codon-usage bias versus gene conversion in the evolution of yeast duplicate genes. *Proc Natl Acad Sci U S A* **103:** 14412-14416.

Kudla, G., A. Helwak, and L. Lipinski. 2004. Gene conversion and GC-content evolution in mammalian Hsp70. *Mol Biol Evol* **21:** 1438-1444.

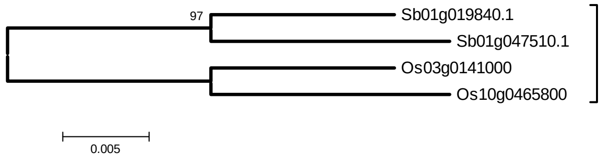
Perry, J. and A. Ashworth. 1999. Evolutionary rate of a gene affected by chromosomal position. *Curr Biol* **9:** 987-989.

Wong, G.K., J. Wang, L. Tao, J. Tan, J. Zhang, D.A. Passey, and J. Yu. 2002. Compositional gradients in Gramineae genes. *Genome Res* **12:** 851-856.

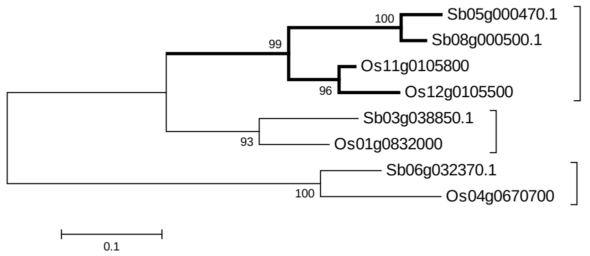
Supplementary figure 1 legend. Paralogous genes converted in both rice and sorghum. Genes in syntenic positions are marked out with brackets and those affected by conversion are noted with thick branch lines. (a) Ribosomal protein genes; (b) Phosphatidate cytidylyltransferase genes; (c) Esterase genes.

Supplementary figure 1.

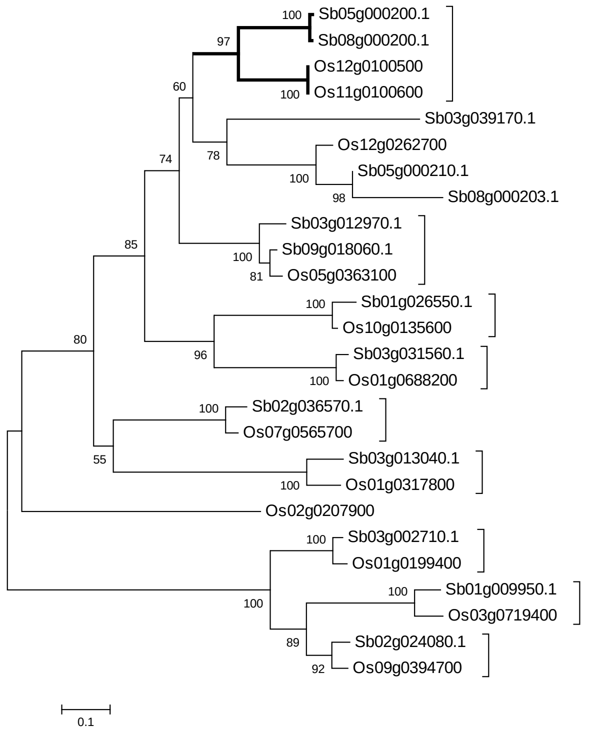
(a). Ribosomal protein genes



(b). Phosphatidate cytidylyltransferase genes

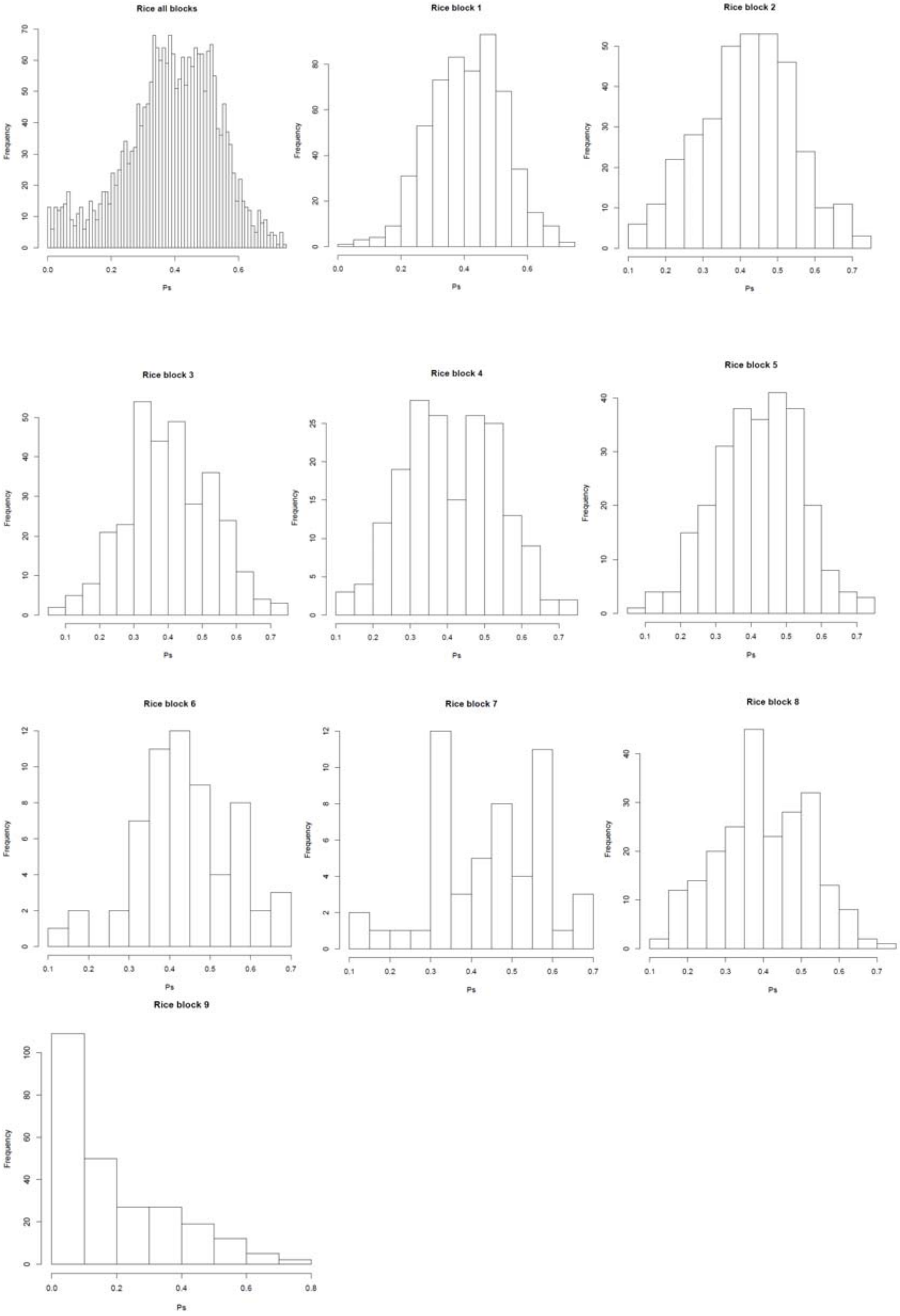


(c). Esterase genes



Supplementary figure 2 legend. Distribution of synonymous nucleotide substitution percentage (Ps) between syntenic paralogs in duplicated blocks in rice genome. Blocks correspond to those in Table 1 in order.

Supplementary figure 2.



Supplementary figure 3 legend. Distribution of synonymous nucleotide substitution percentage (Ps) between syntenic paralogs in duplicated blocks in sorghum genome. Blocks correspond to those in Table 1 in order.

Supplementary figure 3.