

Fig. A. Isolation and sequencing of endosperm-expressed unigenes. A flow diagram is shown with the different steps to construct the various endosperm-specific cDNA libraries, to normalize and subtract abundant cDNAs, to filter and assemble the maize endosperm unigene set.

Fig. B. Length distribution of rice full-length cDNAs and maize endosperm cDNAs. The degree to which cDNAs represented their corresponding proteins was determined using analytical methods within the Sputnik application (see text). Paired EST sequences were compared against proteins in both the Swissprot database as well as a non-redundant protein database using the BLASTX algorithm (see text). Results were filtered using a threshold of 10^{-10} . Known information on the length of the top scoring match protein and the overlap coordinates with the paired EST reads from the cDNA were used to establish the degree to which the cDNA matched the protein. The 5' end of the sequences were considered in detail to investigate whether the respective cDNA represented a full length clone. This was achieved using two measures. The start position of the overlap was determined from the filtered BLASTX data and was placed in windows of 9 amino acids throughout the length of the match protein. A second equivalent plot was reproduced, but the classification of a sequence was dependent on at least 50 nucleotides of sequence within the EST upstream of the start of the sequence overlap. This is indicative of additional 5' UTR (untranslated region) sequence. The length of cDNAs based on sequence analysis has then been plotted against the number of clones present in each size class. The number of rice clones is much larger than the maize clones because the rice clones represent all tissues, while the maize clones represent only endosperm-expressed genes.

Fig. C. Chromosomal distribution of homoeologous rice genes. The 12 chromosomes of rice are presented as vertical lines relative to their estimated sizes in Megabases (Mb). Homoeologous genes detected by the maize

endosperm unigene set are marked by horizontal blue lines relative to their position in the chromosomes.