Feltus_etal Supplemental Methods

Gene Identification

A detailed gene prediction analysis has been done for all 384 predicted amplicons. The predicted amplicon sequences were obtained by first BLAST (Altschul et al, 1990) aligning the primer sequences to the rice pseudomolecule (TIGRv1.0) and then extracting the predicted amplicon sequence using in-house Perl scripts.

The predicted amplicons were then BLAST aligned against 59,712 gene (TU) sequences (including exons, introns and upstream/downstream untranslated regions) from TIGR ($E \leq 1xe^{-10}$; ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_2.0/all_chrs/all.seq).

The BLAST output has been imported to Microsoft Access Database (in house protocol) for easy handling of the data. By using the database, the BLAST output was queried for higher percent identity matches where only those hits were considered which have shown a percent identity of 95- 100%. The database is also queried for redundant hits which then were screened and removed manually.

The predicted gene (TU) descriptions from BLAST report were lacking information about chromosomal location. This has raised question on the feasibility of predicted genes for a given locus. For this, database on all TU models, consisting chromosomal locations, were downloaded separately from TIGR (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_2.0/all_chrs/all.TU_model.brief_info) and indexed with data from TU models obtained by BLAST and also the chromosomal localization of each individual locus. This in turn has validated the feasibility or authenticity and localization of the predicted gene for a particular locus on a particular chromosome.

Out of 384 loci screened for gene prediction, predicted gene have been designated to only 379 loci whereas remaining 5 loci were found to have no BLAST hits.