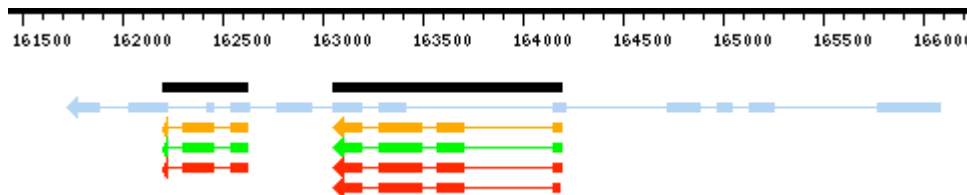


## Case Study #1: Example of Rice Genome Annotation using Maize ESTs

The rice ESTs in our collection proved to be insufficient in doing the annotation. Only when maize ESTs were added was it possible to locate this particular gene.



**Figure 1** - In this example a collection of 94 ESTs were used, consisting of 7 *Arabidopsis thaliana* ESTs, 28 *Glycine max* ESTs, 15 *Medicago truncatula* ESTs, 5 *Oryza sativa* ESTs, 1 *Secale cereale* EST, 4 *Solanum tuberosum* ESTs and 35 *Zea mays* ESTs. We see in the above figure three Predicted Gene Structures that are aligned with three of the maize ESTs (red) resulting in two disjoint Alternative Gene Structures (green) in the region from 161Kb to 166Kb along the rice contig AP003768. These Predicted Gene Structures do not show up using the rice ESTs alone. The two Alternative Gene Structures have similarity scores of about 0.76, and both of these disagree with the GenBank annotation which is in error (light blue).

The greater accuracy afforded by the dynamic programming approach adopted in GeneSeqer is obtained at the expense of greater computational efforts. Practical implementation of the algorithm requires efficient selection of restricted genomic DNA regions and matching ESTs from a typically large EST collection in order to minimize or eliminate the computer time spent on deriving locally optimal, but insignificant alignments. GeneSeqer incorporates a string matching scheme based on pre-processing of the input EST data set that allows fast target selection for detailed analysis by the dynamic programming algorithm. This procedure will need to be modified and optimized for applications to vertebrate gene structures with long introns. An unpublished current version of GeneSeqer uses the vmatch engine to chain significant matches (presumably exons) over long distances and limits the dynamic programming to potential exon/intron boundaries, resulting in significant compute space and time savings (Brendel, Gremme, and Kurtz, unpublished).