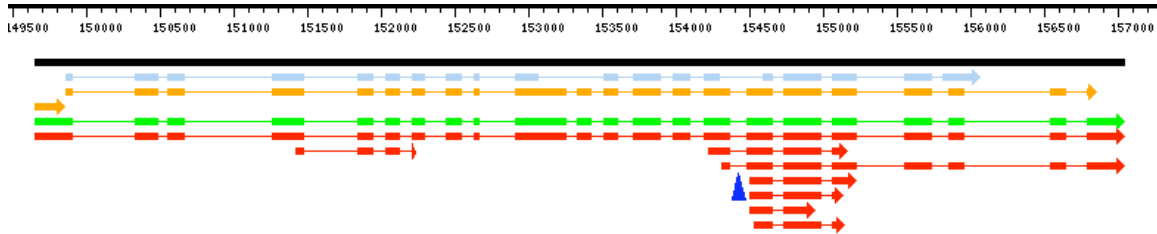


## Case Study #2: Identification of Non-Canonical Splice Sites

This example demonstrates GeneSeqer's ability to identify U12-type introns with AT-AC borders.



**Figure 2** - In this example 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. Using the GeneSeqer Web service at PlantGDB (Schlueter et al., 2003), 8 ESTs could be significantly aligned in this region. Among them, the two longest Predicted Gene Structures derived from two of the rice ESTs, 5 of the *Zea mays* (Corn) ESTs and 1 of the *Solanum tuberosum* (Potato) ESTs, all of which, in turn, predict a single gene structure with 22 exons (green). An open reading frame (orange) spans all the exons, and its translation identifies the gene as coding for a kinesin-related protein. Of particular interest is identification of the 15th intron as a U12-type intron with AT-AC borders (marked by the blue triangle). Once again, the spliced alignments generated by GeneSeqer indicate that the Genbank annotation is incorrect, which also is confirmed by the ESTs from other species. As illustrated here, GeneSeqer can use quite divergent ESTs to predict the correct gene structure.