SRB Genome Sequence Gene Prediction Decision Tree

1. Are there GENES present in this contig?
   - Yes: Gene Structure Prediction Programs
     - FGENESH
     - GENEMARKHMM
     - GENSCAN
     - Predicted gene(s)
       - Yes: Is the SAME gene predicted by at least TWO programs?
         - No: Go to PART 2
         - Yes: BLASTN to the NCBI ESTDB [Non-human, non-mouse ESTs (est_others)]
       - No: Go to PART 2
   - No: Go to PART 2

2. Are there cDNAs/ESTs corresponding to the contig?
   - Yes: Matching to cDNA/EST
     - No: Does the MATCHING REGION correspond to the region of the PREDICTED gene?
       - Yes: The PREDICTED GENE is REAL
       - No: Is the Matching region within 100-200 bp from the initial/terminal region of the PREDICTED gene?
         - Yes: The matching region may represent an UNTRANSLATED region of a message
         - No: Study the matched cDNA/EST (whether it contains a coding sequence)
   - No: BLASTP to the NCBI NR PROTEIN database

3. Are there REPEAT sequences present?
   - Yes: Repeat Detection Programs
     - PLANT REPEAT DATABASE
     - REPEAT MASKER
     - Repeat(s)
       - No: Examine the alignment of sequences and mark the region(s) on the CONTIG as REPEAT(S)
   - No: Go to PART 2

BLASTN to the NCBI ESTDB [Expressed Sequence Tags (EST)]