



Hands-On Seven

Arabidopsis thaliana

We are going to locate the potential splice sites in an Arabidopsis gene (AtRad1):

- (I) By performing a pairwise sequence alignment of the Arabidopsis AtRAD1 gene cDNA and the Arabidopsis AtRad1 genomic DNA, and
- (II) By using a gene prediction Web site.

Part One:

A) Go to http://www.cs.sjsu.edu/~khuri/Rabat_2013 and obtain the following two sequences:

Arabidopsis Rad1 genomic DNA “arabidopsis_rad1_genomic.html” and its cDNA sequence: Arabidopsis Rad1 cDNA “arabidopsis_rad1_cDNA.html”.

B) Choose “LALIGN: DNA:DNA” under “Program” and “DNA (forward only)” under “Enter first (query) sequence” and align the Arabidopsis AtRAD1 gene cDNA and the Arabidopsis AtRad1 genomic DNA sequences using LALIGN at: http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=lalign. Or you can use CLUSTALW at <http://www.ebi.ac.uk/Tools/msa/clustalw2>.

C) Answer the following questions:

1) What do the gaps between the aligned genomic sequence and the cDNA represent?

2) Look for 5' and 3' splice junctions near the ends of the gapped regions. Go to http://arabidopsis.org/info/splice_site.jsp and choose “List of known Arabidopsis Splice Sites”. Consult “Arabidopsis Intron Splice-Site Tables”. Use the Arabidopsis table of consensus splice sites to deduce where they are on the genomic sequence and use this information to find the ends of the exons.

Note that the alignment program is unaware of the splice junctions and merely places gaps according to the dynamic programming algorithm.

So the alignment will often have to be modified slightly at the ends of the gaps in order to make the alignment reflect the best splice junctions.

D) Predict and indicate in the table below the positions of all exons in the genomic DNA sequence.

Exon Number	Exon Position	Intron Position
1		
2		
3		
4		
5		
6		
7		
8		
9		

Part Two:

Submit the genomic sequence to the Genscan server and compare the results with those obtained in part one.

The Genscan server is available at: <http://genes.mit.edu/GENSCAN.html>.

Do not forget to choose “Arabidopsis” under “Organism”.

How accurate are the results obtained by the Genscan server? What differences (if any) exist between the Genscan output and the gene locations you predicted using LALIGN and/or CLUSTALW?

Exon Number	Exon Position	Intron Position
1		
2		
3		
4		
5		
6		
7		
8		
9		