Module 3 final analysis

DATA
I am supplying you two sequence data sets that contain sequence reads for four Lycopersicon hirsutum accessions (LA 1223, LA1347, LA 1361, and LA1560) and LA 0490, a cultivated tomato (L. esculentum). The sequence data sets are for two genes, CD31 and Pto. Please use the sequence data to address the following questions:

QUESTIONS
Are there genetic differences between individuals within an accession (i.e. is there evidence for heterozygosity)?

Are there genetic differences between accessions?

Devise a simple assay to detect one of the genetic differences in each gene.

Can you identify intron and exon sequences in any of the data?

Hypothesis testing:
Do differences between “genes” indicate different evolutionary history of “wild” and “cultivated” accessions? Do the two genes tell us a consistent story about the evolutionary history of the accessions?

RELEVANT INFORMATION
In order to answer these questions you will need the following information: The sequence for CD31 in LA0490 is based on a cDNA clone while the sequences for the L. hirsutum accessions are based on PCR amplified genomic DNA. L. hirsutum is an outcrossing species while L. esculentum is inbreeding. I have edited the 5’ portion of the CD31 sequences, but not the 3’. In addition, this file still contains both Forward and Reverse-reverse compliment sequences. The Pto sequences are fully edited. Finally, the “base calling” algorithm used by the ABI uses the following codes for ambiguous bases:

R = A or G  K = G or T  H = A or C or T  D = A or G or T
Y = C or T  S = G or C  B = C or G or T  N = A or C or G or T
M = A or C  W = A or T  V = A or C or G

From the nomenclature committee of the International Union of Biochemistry [(1986) J. Biol. Chem. 261, 13]