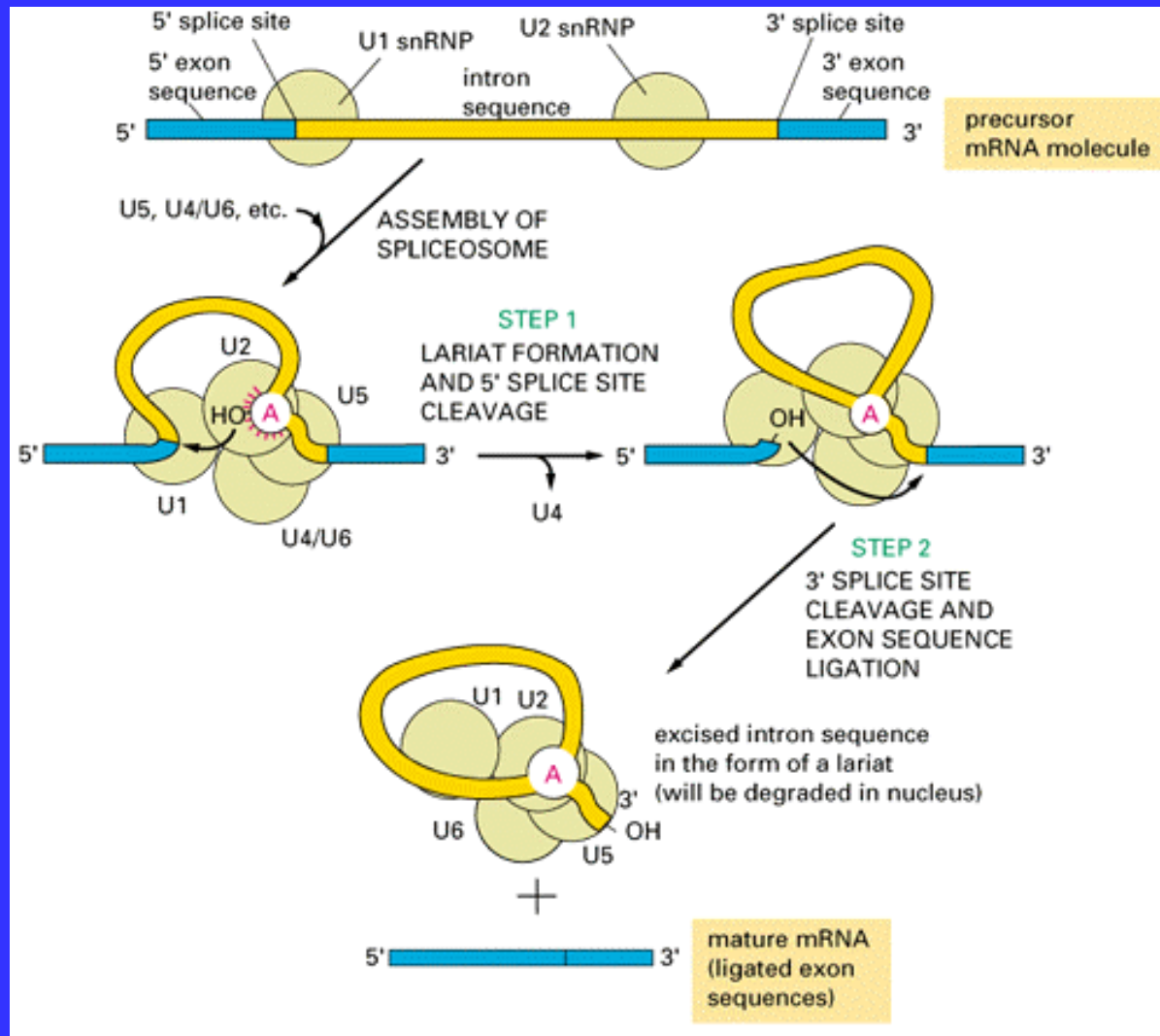


The role of pre-mRNA secondary structure in gene splicing

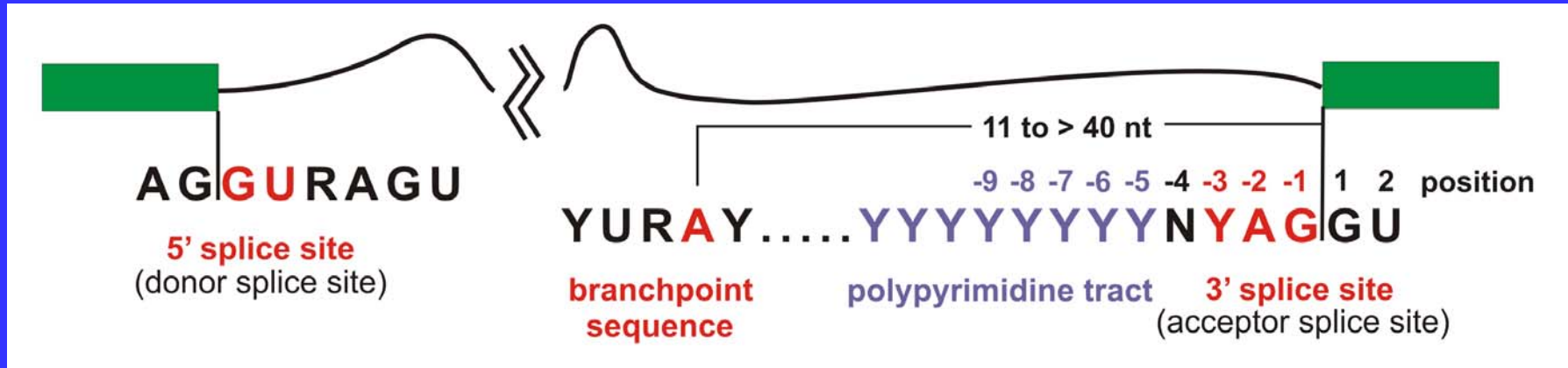
Sanja Rogic

Computer Science Department, UBC

Gene splicing



Splice sites consensus sequences



R = purines (A, G)

Y = pyrimidines (C, U)

- consensus sequences are necessary but not sufficient for splicing

Motivation

- splice site recognition by the spliceosome is still not well understood - **new biological insights**
- low accuracy of computational splice site prediction is a major reason for limited accuracy of gene-finding programs - **improving accuracy of gene-finding**
- pre-mRNA is not linear but has secondary structure
- biological studies suggest that pre-mRNA secondary structure can affect splicing

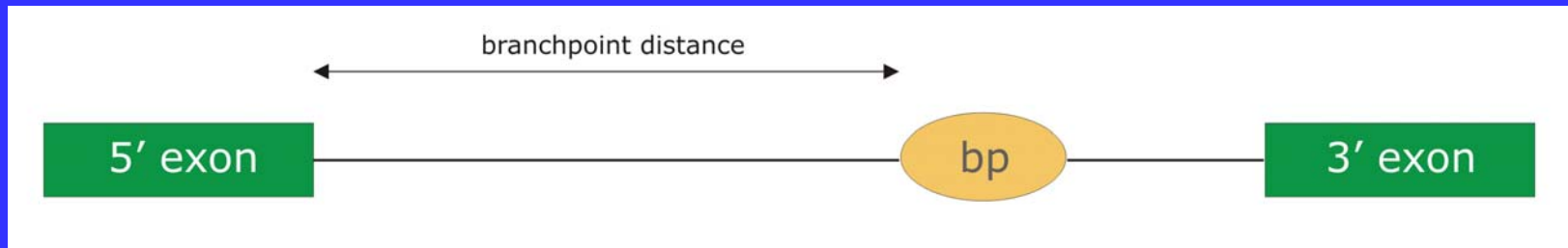
Hypothesis and goal

- secondary structure of pre-mRNA plays a role in splicing
- characterize secondary structure elements that are additional identifiers of intronic regions

Dataset – *S.cerevisiae* introns

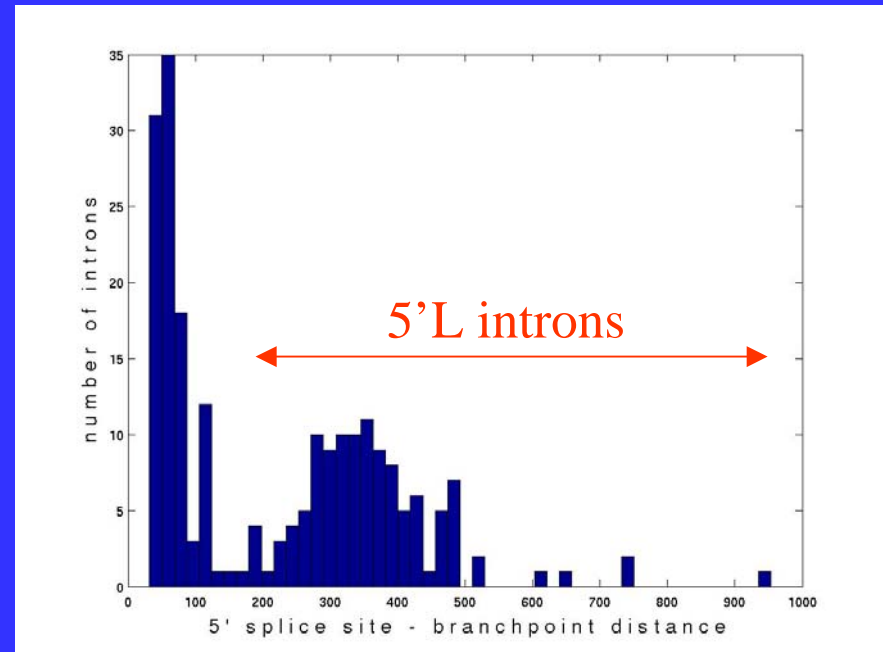
- 215 introns with consistent annotation between three databases: AYID, YIDB, and CYGD

Spatial architecture of yeast introns



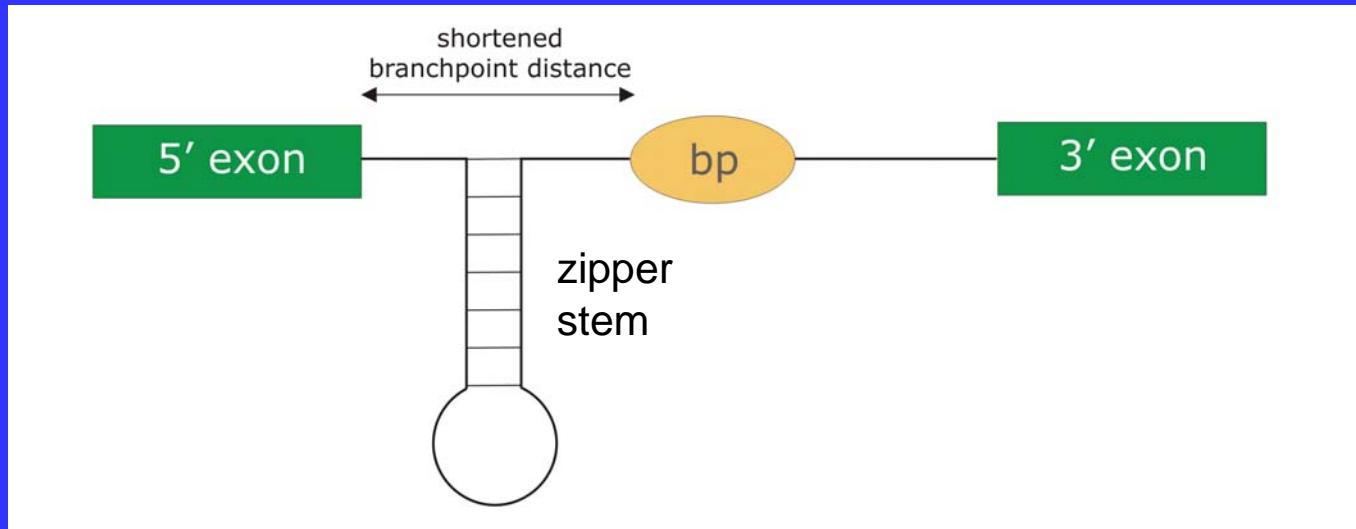
- *S.cerevisiae* introns have bimodal distribution of branchpoint distances

- 5' short (~ 40 nt) introns (5'S)
- 5' long (> 200 nt) introns (5'L)



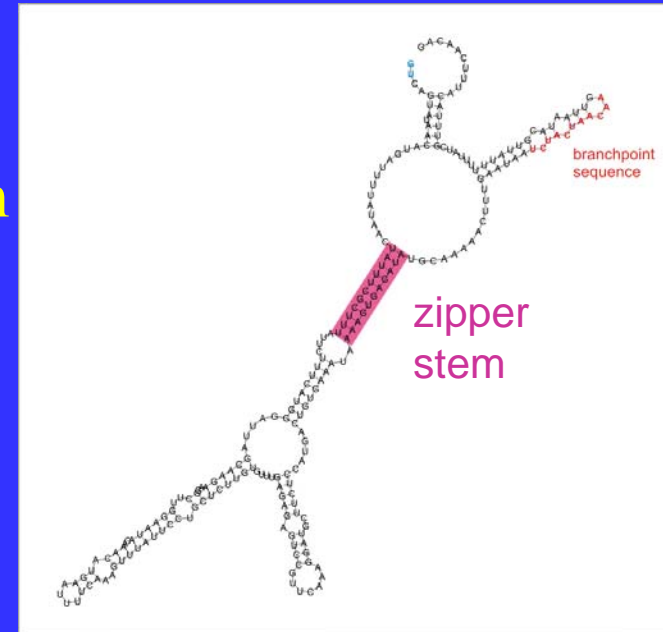
'Zipper' stem hypothesis

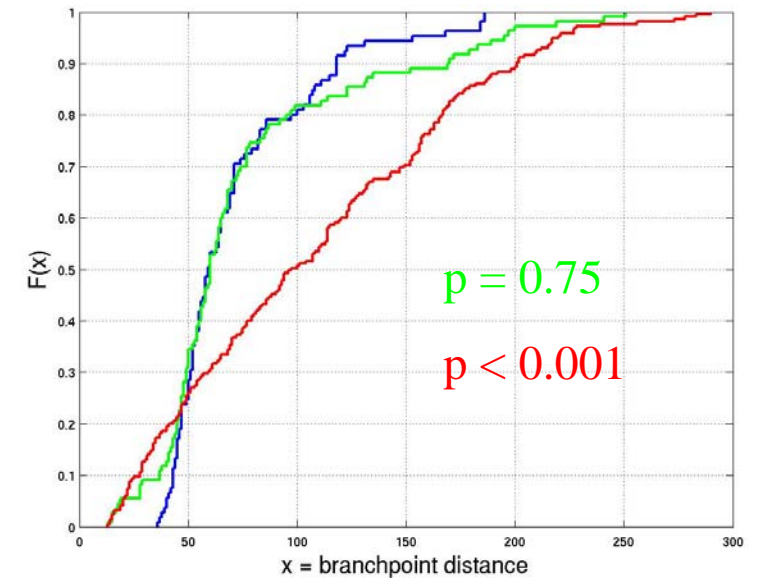
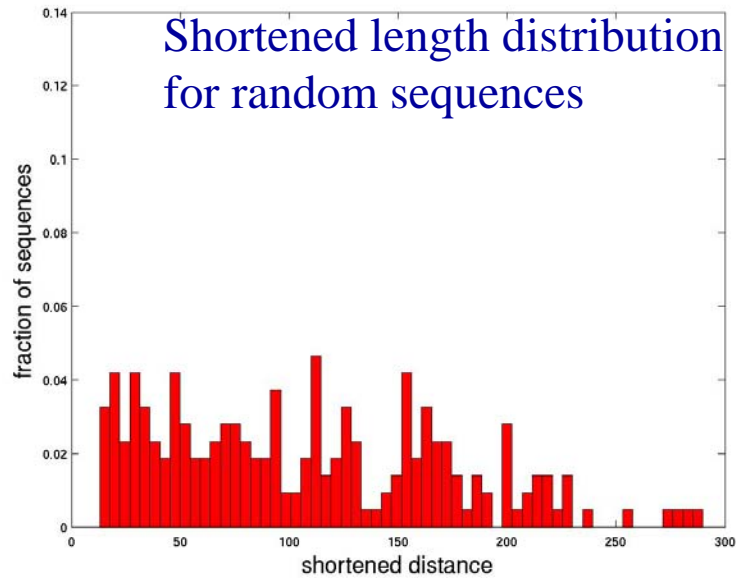
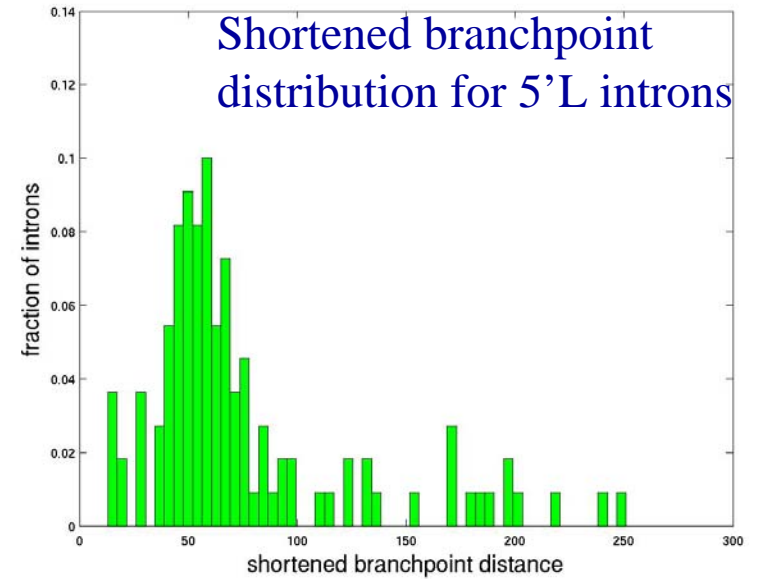
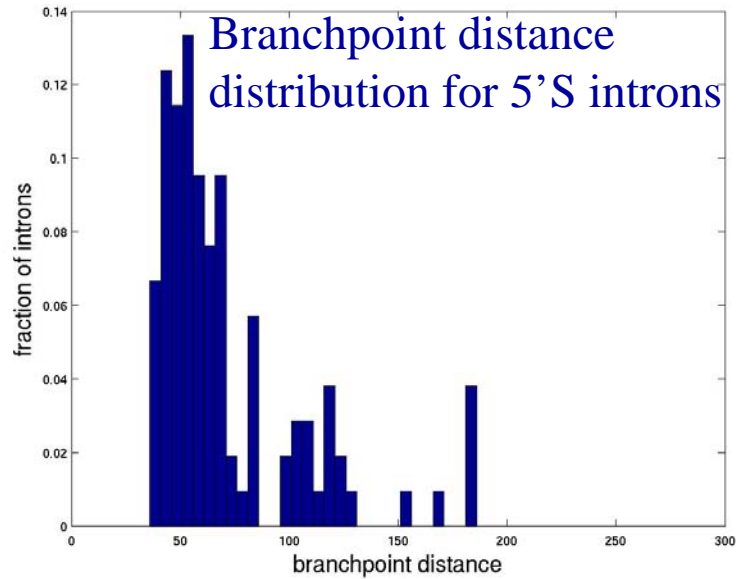
- 5'S introns have the optimal branchpoint distance for spliceosome assembly
- 5'L introns fold into secondary structure to shorten branchpoint distance to optimal
- stem/loop structures experimentally identified in several yeast introns - splicing inhibited when zipper stem disrupted



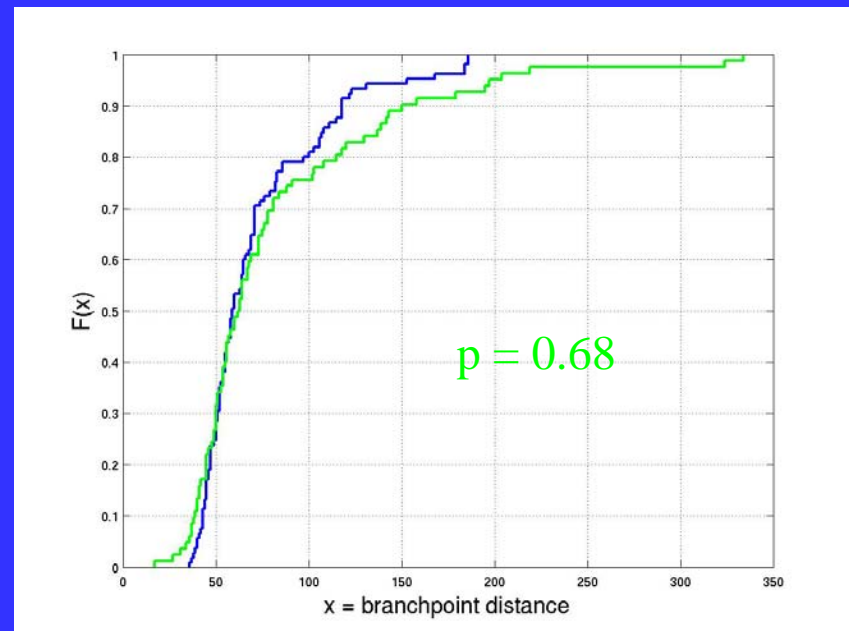
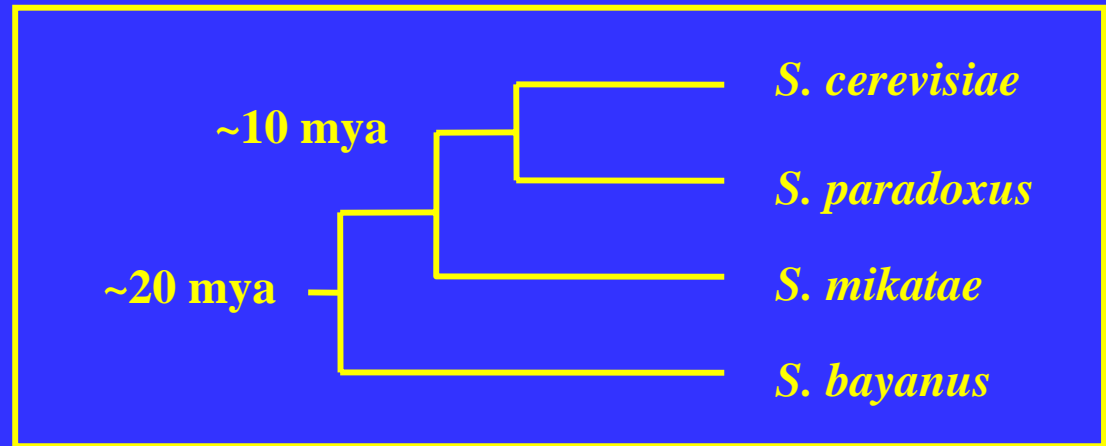
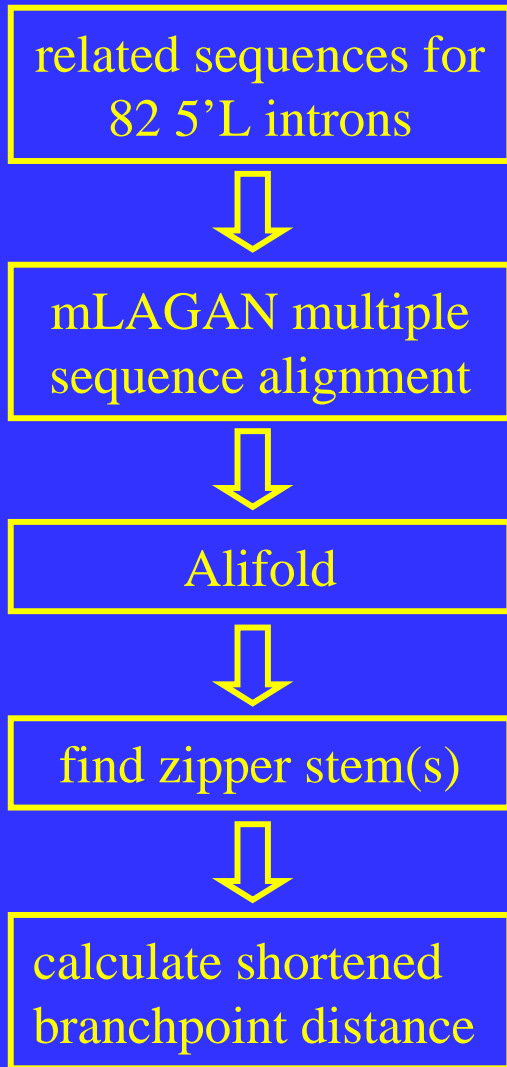
Identification of zipper stems

- computing secondary structure of introns:
 - using programs that predict RNA secondary structure based on minimum free energy (mfold, RNAfold)
 - using comparative (phylogenetic) secondary structure prediction
- secondary structures processed to identify one or more zipper stems (thermodynamical stability and loop size)
- shortened branchpoint distance calculated





Phylogenetic approach



Acknowledgements

- thesis supervisors:

Alan Mackworth, CS Department, UBC

Holger Hoos, CS Department, UBC

Francis Ouellette, UBiC

- colleagues in Beta lab

(Bioinformatics and Empirical & Theoretical Algorithmics)