

## **Modeling of mRNA poly(A) sites in Arabidopsis**

Q. Quinn Li<sup>1\*</sup>, Jianting Li<sup>2</sup>, Han Rong<sup>2</sup>, Yingjia Shen<sup>1</sup>, and Guoli Ji<sup>2</sup>,  
<sup>1</sup>Dept. of Botany, Miami University, Oxford, OH 45056. <sup>2</sup>Dept. of Automation, Xiamen University, Fujian, P. R. China 361005.

One of the essential processing events during pre-mRNA maturation is the post-transcriptional addition of a poly(A) tail. The poly(A) track serves as an identification of the integrity of the messenger, is recognized by the mRNA export and translation machinery, and protects mRNA. The positions of poly(A) sites are predetermined by signals that reside in the pre-mRNA sequence that recognized by a complex of protein factors. However, there are no highly consensus signal elements that can be readily used to predict the poly(A) site of a given gene. Currently, the only way to recognize a poly(A) site is through identification of the actual poly(A) track at the end of a messenger.

We have recently compiled the potential polyadenylation signals of the model plant Arabidopsis, and revealed a special nucleotide profile in the pre-3'-UTR sequences. These signals, though with very low consensus, including three main elements: far upstream elements (FUE), near upstream elements (NUE), and cleavage elements (CE). However the distinct profile of the four nucleotides in the 3'-UTR offers some clues of the features of the poly(A) sites. We have devised a Hidden Markov Model (HMM) based algorithm by using available characteristics of known poly(A) sites in an authenticated poly(A) site database with high sensitivity and specificity. The accuracy of the program has been demonstrated by the prediction of many known poly(A) sites. The program also predicts the change of poly(A) site selections in the poly(A) signal mutants. The utility of the algorithm will be discussed.