

SplamiR – Predicting a new class of plant microRNAs

Christoph J. Thieme, Lydia Gramzow, and Günter Theißen

Christoph J. Thieme
AG Bioinformatics
Max Planck Institute for Molecular Plant Physiology
Potsdam-Golm, Germany
Email: thieme@mpimp-golm.mpg.de

Lydia Gramzow
Department of Genetics
Friedrich Schiller University Jena
Jena, Germany
Lydia.Gramzow@uni-jena.de

Günter Theißen
Department of Genetics
Friedrich Schiller University Jena
Jena, Germany
Gunter.Theissen@uni-jena.de

MicroRNAs have been recognized as major post-transcriptional regulatory elements. With about 21nt length they hybridize highly specific with mRNAs of coding genes. In cooperation with the RNA induced silencing complex (RISC), this allows microRNAs to accomplish the efficient down-regulation of these target mRNAs.

Recent discoveries in plants revealed a new mechanism of microRNA processing. Here, the primary unspliced transcript (pri^U-microRNA) spans large introns of several thousand nucleotides. Thus, the molecule cannot fold into the typical hairpin structure until the splicing events occurred. There are a number of *ab initio* microRNA prediction methods available. However, none of these methods are capable of detecting spliced microRNAs, since they focus on local structures and therefore cannot recognize a hairpin structure that is established post-splicing. Thus, the need for more generalized prediction methods is evident.

We developed SplamiR (Spliced plant microRNAs) to tackle this challenge. Our tool searches for putative spliced and unspliced microRNAs for a given mRNA sequence. It is applicable to any plant genome as it is not primarily based on finding sequence homologies.

SplamiR is a modular pipeline that consists of two main phases. First, we generate a database of all potential stems of the genomic input sequence. In phase two, stems with near-perfect complementarity to a given target gene are collected. Through linking splice site predictions and the evaluation of the resulting precursors in a classification step, SplamiR identifies a set of candidate microRNA precursors.

We applied SplamiR to the genomes of rice and maize. Our results show that the new class of spliced microRNAs can be successfully predicted and all known precursors of the MIR444 family are predictable despite their complex processing. Due to the classification step the number of false positives is quite low (less than two percent in simulation studies). Moreover, by using SplamiR we were able to find previously uncharacterized miRNA sequences in maize. The expression of these sequences is supported by an expressed sequence tag (EST).

Keywords: microRNA, splicing, prediction, RNA