Primer Design for Automated Virus Diagnosis in Plants

C. MEDEIROS, D. TAVARES, K. ROCHA, M. MONTEIRO, L. GONÇALVES, P. M. LÚCIO

[1] Universidade Federal do Rio Grande do Norte CT-DCA-UFRN Cep. 59.072-970, Natal, RN, Brasil
{kliger,meika,douglas,lmarcos,cesar}@dca.ufrn.br,pmarinho@ufrnet.br

Abstract. We present current efforts toward development of a software platform to design specific and universal primers for pathogens in RT-PCR technique. By applying a selection filter and segregation of conserved and polymorphic sequences into plant virus genome it is possible to perform a data mining to search efficient primers to help in diagnostic works.

1 Introduction

One of the main issues to be observed in virus diagnosis of plants using RT-PCR is the proper primer design. Primers must be able to anneal to the target DNA in a predictable location and to extend by Taq DNA Polymerase. Programs are available to design primers, and they have a finite probability of producing errors. We describe the steps involved in this process and current efforts towards automating it. We show that designing primers by computers can produce excellent results. Basically, our proposal is to select a region for primer placement where the probability of diagnosis errors using RT-PCR is very low [3].

2 Methodology

In order to be efficient, a primer must anneal to the target DNA in a predictable location, and at the level of virus species. We have constructed a database containing virus's sequences, and a system to align and separate FASTAs sequences' domains. Our strategy performs a data bank searching, finding sequences and generating multiple alignments. Those sequences can share similarity with domains and differentiate between polymorphic domains. We've been working with polymorphic domains to design high specificity primers.

3 Partial Results

In the current version, we first perform a comparison algorithm in a minimum subsequence of size k. The query performs a cross-match to obtain the subsequence, finding one to be compared with the virus sequence. The following information list has to be stored based on relative similarity occurrences: ID compared sequence, subsequences initial position and size, and occurrence position. See Figure 1 for more details on the system. We applied the current algorithm to some selected virus sequences and the results show the primers to be unambiguous for each sequence.

4 Conclusions and Perspectives

The proposed program can be applied to other pathogens by just a simple adaptation. Polymorphism analysis within virus alignment sequences can be useful to identify molecular markers. Next step is to better study the algorithm performance and to make a comparison of data mining over another existing techniques. A Primer Data Bank for plant virus will also be built.

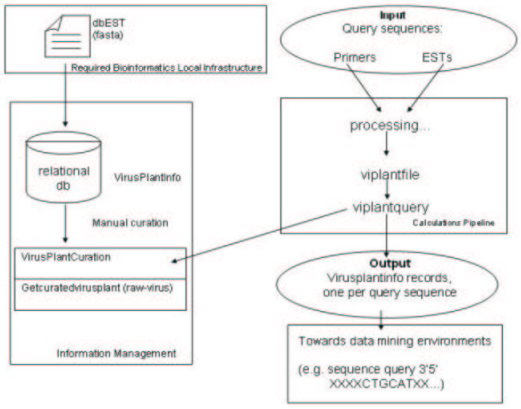

Figure 1: Infographic.

References

[1] Thomson K. at al., *Identification of Zucchini yellow mosaic potyvirus by RT- PCR and analysis of sequence variability?, Journal of Virological Methods*, 55, p. 83-96. ed.. Addison-Wesley (1995).

[2] Gitton F. at al., *A two-step multiplex RT-PCR method for simultaneous detection of soil-borne wheat mosaic virus and wheat spindle streak mosaic virus from France*, Plant Pathology, 48, p. 635-641. (1999).

[3] Antoniw, J., *A new method for designing PCR primers specific for groups of sequences and its application to plant viruses*, Molecular Biotechnology, 4, 111- 119. (1995).