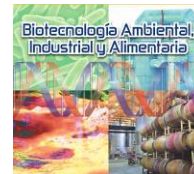

Talk

Searching for gene clusters related to virulence by coding sequence conservation



Álvaro Centrón Broco, Carlos Sánchez Casemiro-Soriguer, Ramón Ramos Barrales and Antonio J. Pérez Pulido (1)

1. Área de Genética, Centro Andaluz de Biología del Desarrollo, Ctra. de Utrera, km. 1, 41013, Sevilla.

Keywords: AnABlast; *Ustilago maydis*; Principal component analysis

ABSTRACT

Motivation: Due to the increasing world population, the need to improve food production is growing. This can be helped by fighting the pathogens which affect the main crops as maize, wheat, barley and sugar cane. Among those, biotrophic parasites such as smut fungi can be found. To study how those microorganisms infect their host, the model system *Ustilago maydis* can be used.

U. maydis secretes protein effectors to infect its host, and at least 25% of them are known to be grouped in 13 different gene clusters. In addition to these characterized clusters, 7 new clusters have been described in the bibliography but not experimentally tested. The aim of this work is to find out new clusters with features similar to the known ones (controls), mainly low conservation, which can affect the infection process.

Methods: To achieve this goal, candidate gene clusters were initially discovered based on coding sequence conservation via the computational tool AnABlast [1], which highlighted genomic coding region with conservation signal similar to the initial controls. Then, the candidates were functionally annotated using the tool Sma3s_v2 [2]. To select the best candidates, a principal component analysis (PCA) was done using the following factors, which were trained with the controls: sequence conservation obtained by a similarity search by Blast against close organisms (Ensembl fungi phylogeny), expression data during infection, and signal peptide presence (SignalP and TargetP), usually present in effectors.

Currently, a laboratory experiment has been began to elucidate if the chosen candidates affect the pathogenity, deleting them by homologous recombination.

Results: We have been able to identify 49 new clusters by comparing their coding signal with those already known. After the subsequent analysis three of them, and one from the bibliography have been chosen to be tested in laboratory to elucidate their virulence phenotype (swelling and tumors).

In the PCA our best candidate is located among the clusters previously described as pathogenic, showing genes being secreted with high levels of expression

Conclusions: In brief, we propose that putative cluster of virulence sequences could be found by the presented strategy. So, it could constitute a new silico approach to find out specific genes involved in different biological processes such as infection.

REFERENCES

1. Jimenez, J. et al. (2015) AnABlast: a new in silico strategy for the genome-wide search of novel genes and fossil regions *DNA Res.* 22:439-449.
2. Carlos S. Casimiro-Soriguer, Antonio Muñoz-Mérida, Antonio J. Pérez-Pulido. Sma3s: a universal tool for easy functional annotation of proteomes and transcriptomes. *Proteomics* 2017, in press.