

ANÁLISIS COMPARATIVO DE DIFERENTES MÉTODOS Y PUNTOS DE CORTE PARA LA DETERMINACIÓN DE EXPRESIÓN GÉNICA DIFERENCIAL EN BASES DE DATOS COMPLEJAS OBTENIDAS DE TRANSCRIPTOMAS DE MAÍZ

Peñas Ballesteros, A.^{1,2,3}; Baricalla, A.A.^{1,3}; Iglesias, J.^{2,3}

¹ CITNOBA (UNNOBA - UNSAdA - CONICET), Monteagudo 2772, Pergamino, Buenos Aires, Argentina

² Instituto Nacional de Tecnología Agropecuaria-Estación Experimental Pergamino (INTA-EEA Pergamino), Av. Frondizi (Ruta 32) Km 4,5, Pergamino, Buenos Aires, Argentina

³ Universidad Nacional del Noroeste de la provincia de Buenos Aires (UNNOBA), Monteagudo 2772, Pergamino, Buenos Aires, Argentina

INTRODUCCIÓN

- Existe una amplia variedad de datos transcriptómicos públicamente disponibles en NCBI, provenientes de estudios donde se ha evaluado la respuesta del cultivo de maíz frente a *Fusarium verticillioides*, *F. graminearum* y *Ustilago maydis*
- Dos de las herramientas más utilizadas para determinar los genes que se expresan diferencialmente son DESeq2 y edgeR.
- edgeR permite la utilización de dos métodos con enfoques diferentes para ajustar y contrastar los datos: la prueba de razón de verosimilitud (LRT) y la prueba F de cuasi verosimilitud (QL-F test).

OBJETIVO

Comparar diferentes herramientas, enfoques y puntos de corte utilizados para determinar expresión génica diferencial en un metaanálisis basado en datos provenientes de múltiples transcriptomas de maíz expuestos a diversos patógenos

MATERIALES Y MÉTODOS

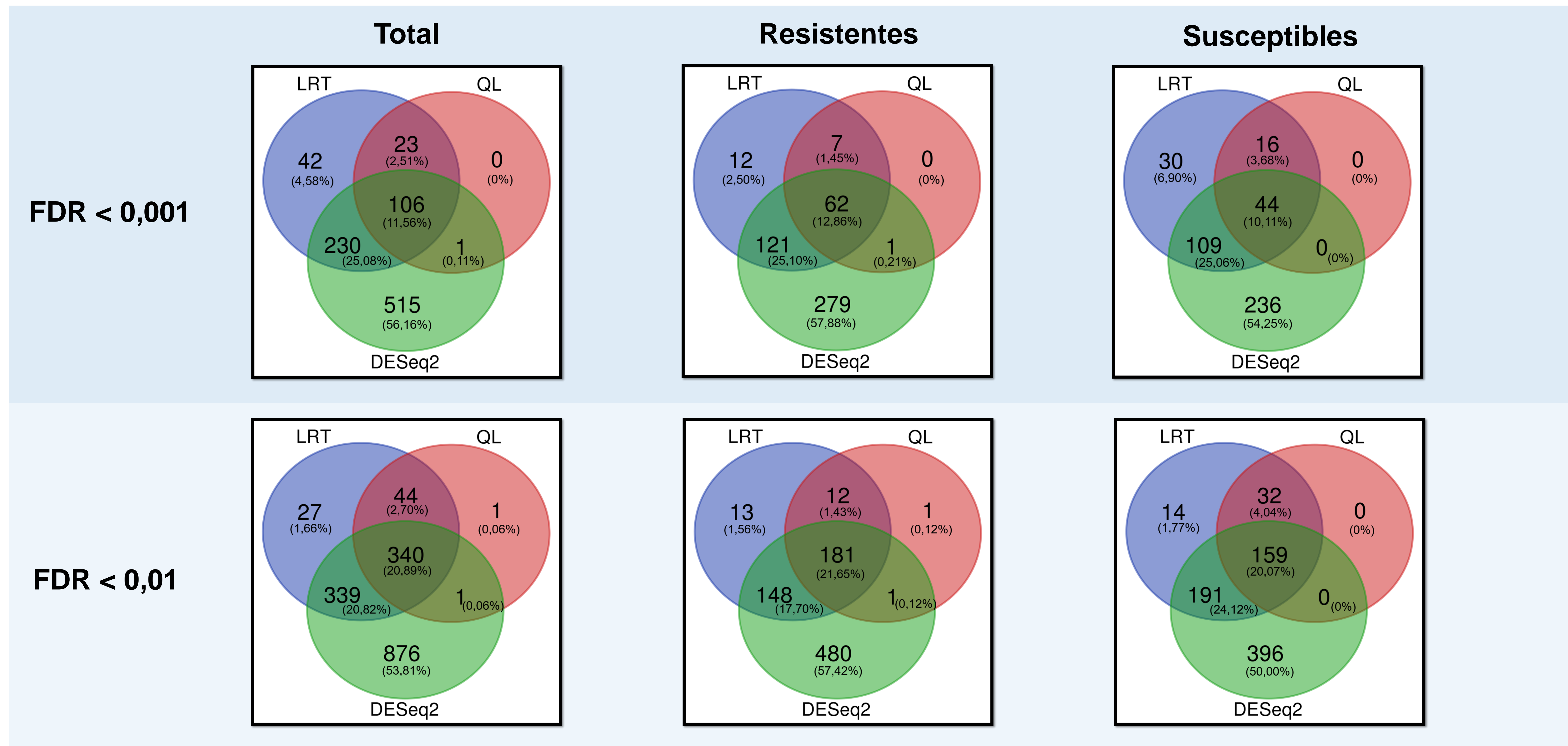
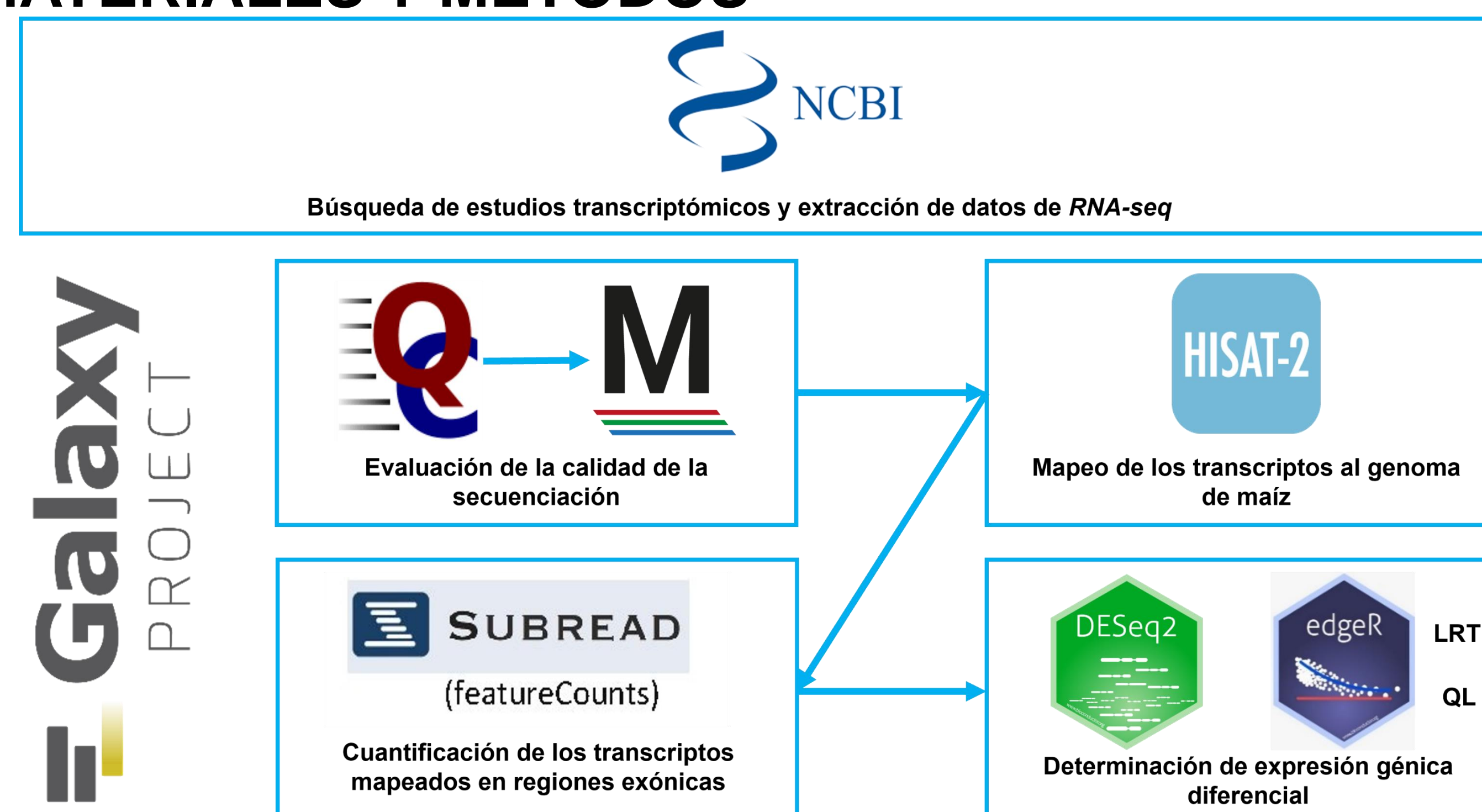


Figura 1. Diagramas de Venn de los genes diferencialmente sobreexpresados para ambos genotipos en conjunto (total), genotipos resistentes (resistentes) y genotipos susceptibles (susceptibles) en respuesta a *Fusarium verticillioides*, *F. graminearum* y *Ustilago maydis*, entre 48 y 72 horas después de la infección, determinados a partir de DESeq2 y edgeR (LRT y QL) utilizando 0,001 o 0,01 como puntos de corte. Se observa el porcentaje de genes sobre el total de genes únicos encontrados para cada categoría.

RESULTADOS

A partir del contraste entre genotipos resistentes y susceptibles, realizado en base a un diseño con dos factores: Genotipo y Proyecto (variable que permite la integración de los datos, metaanálisis), se observó que:

- La cantidad de genes diferencialmente expresados fue mayor con DESeq2 al tener en cuenta el mismo punto de corte, mientras que la menor cantidad de genes se obtuvo a partir del método QL de edgeR.
- Al comparar los genes diferencialmente expresados, las herramientas coincidieron en casi todos los genes encontrados a partir de QL.

CONCLUSIONES

- EdgeR podría ser una herramienta más restrictiva a la hora de encontrar genes diferencialmente expresados, sobre todo al utilizar el método de cuasi verosimilitud (QL).
- DESeq2 representaría una herramienta más permisiva, permitiendo encontrar una mayor cantidad de genes diferencialmente expresados.
- Para el set de datos estudiado, el método LRT de edgeR permitiría encontrar una cantidad óptima de genes diferencialmente expresados, dando un balance entre cantidad de genes, procesos biológicos en los que estos se encuentran involucrados y a través de los cuales se puede abordar una mejor comprensión del fenómeno biológico.

8vo Simposio Argentino de Jóvenes Investigadores en
Bioinformática
7SAJIB

[Ver pdf](#)