

LatinR, Conferencia Latinoamericana sobre el Uso de R en Investigación + Desarrollo

Análisis de datos de expresión con R como herramienta para el estudio de familias génicas.

Chiriotto Tai S¹* y Saura-Sánchez Maite¹*

¹IFEVA-Conicet. Instituto de investigaciones fisiológicas y ecológicas vinculadas a la agricultura. Facultad de Agronomía - Universidad de Buenos Aires Av. San Martín 4453 (C1417DSE) Ciudad de Buenos Aires - ARGENTINA.

tchiriotto@agro.uba.ar; saurasanchez@agro.uba.ar

* Los autores tienen la misma contribución en el trabajo.

Palabras clave: proteínas BBX, microarreglos, RNAseq, software R.

Las proteínas B-box de *Arabidopsis thaliana* constituyen una familia de 32 miembros (BBX1 a BBX32) que están implicadas en múltiples procesos a lo largo de toda la vida de la planta. Participan en procesos de desetiología, floración y germinación entre otros. También, actúan en respuestas ambientales como en las respuestas a la percepción de plantas vecinas o a la salinidad. Los datos proporcionados por microarreglos y RNAseq hechos en plantas bajo diversas condiciones experimentales pueden suministrar información importante y robusta acerca de la expresión de los genes de interés y pueden servir para plantear nuevas hipótesis de trabajo. En la actualidad, la base de datos existente es extensa y variada. Los datos publicados pueden estar crudos, procesados y también de ambas formas. Como consecuencia, realizar análisis con herramientas convencionales como tablas Excel es complejo y lento, pudiendo cometer múltiples errores en los numerosos pasos del proceso. El objetivo del siguiente trabajo es analizar la familia de proteínas B-box mediante el software R en microarreglos y RNAseq de plantas en distintas condiciones experimentales.

Se seleccionaron y trabajaron con más de diez bases de datos de microarreglos y RNAseq publicados en Geo DataSet (<https://www.ncbi.nlm.nih.gov/gds>) y ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>). El análisis de estos datos se llevó a cabo con distintos paquetes de Bioconductor en R. Se utilizó el paquete ggplot2 para representar gráficamente los genes B-box según los cambios de expresión a lo largo de las distintas condiciones experimentales estudiadas. Este análisis permitió agrupar los genes B-box en sendos *clusters* según su patrón de expresión. Asimismo, se analizó la co-expresión con otros genes.

El uso de software R permitió de manera simple y relativamente rápida obtener datos homogenizados de distintas bases de datos de microarreglos y RNAseq. Se pudo crear diversos parámetros de búsqueda lo que permitió obtener información sin necesidad de nuevos análisis. Además, se pudo programar el formato de los gráficos y diseñarlos acorde a las necesidades requeridas.

Gracias a este tipo de análisis podemos estudiar grandes familias génicas con el fin de encontrar miembros que puedan compartir funciones biológicas o inferir nuevos procesos en los que participan estos genes.