

IDENTIFICACIÓN DE LA SECUENCIA PROMOTORA Y TERMINADORA DE UN GEN DE *Pichia pastoris* (*Komagataella spp.*)

Yanelis Robainas-del-Pino, Karla Fernández-Cano, José María Viader-Salvadó, Martha Guerrero-Olazarán
 Universidad Autónoma de Nuevo León, UANL, Facultad de Ciencias Biológicas, Instituto de Biotecnología,
 San Nicolás de los Garza, N.L., México C.P. 66455. martha.guerrero01@uanl.edu.mx

Palabras clave: *Pichia pastoris*, regiones reguladoras, transcriptomas

Introducción. *Pichia pastoris* es un importante hospedero para la producción de proteínas recombinantes, debido a su fácil manipulación génica, requisitos mínimos de crecimiento, altos rendimientos de biomasa y modificaciones postraduccionales de tipo eucariota, unido a que secreta pocas proteínas endógenas al medio extracelular (1). A pesar de su utilidad demostrada, se trabaja en la mejora de este sistema de expresión buscando regiones reguladoras que puedan garantizar la sobreproducción de una proteína recombinante (2). Existen varios genomas reportados de este hospedero, sin embargo, las anotaciones son un poco divergentes o no están anotadas todas las regiones reguladoras. Recientemente en nuestro grupo de trabajo se identificaron ocho genes cuya expresión fue dependiente de la fase de crecimiento y la fuente de carbono en cepas recombinantes de *P. pastoris*, entre ellos el gen hipotético que denominamos P7.

El objetivo de este trabajo fue identificar dos posibles regiones reguladoras del gen P7 de *P. pastoris* mediante un análisis bioinformático.

Metodología. El gen P7 tiene orientación negativa en la cepa CBS7435 de *P. pastoris*, pero no tiene anotadas las regiones reguladoras. Su región promotora coincide con la región promotora del gen que está río arriba, en la secuencia comprendida entre las dos regiones codificantes (CDS) adyacentes, y la región terminadora coincide con la región promotora del gen que está río abajo, en la secuencia comprendida entre los dos CDS adyacentes. Se mapearon 6 transcriptomas de la cepa KM71 de *P. pastoris* a diferentes condiciones de cultivo, determinados por RNA-seq, contra la secuencia del genoma de referencia de la cepa CBS7435 utilizando la plataforma Galaxy y la herramienta bioinformática HISAT2 para localizar zonas sin transcritos en las regiones río arriba y río abajo del CDS del gen P7. Se visualizaron los niveles de transcritos en las dos regiones analizadas con el programa Integrative Genomics Viewer (IGV). Adicionalmente se utilizó el programa UTRscan de la plataforma ITBtools para identificar motivos de secuencia característicos de regiones 5'UTR y 3'UTR de eucariotas. Por último, se buscaron 5 secuencias reportadas como consenso en terminadores de levaduras: el elemento de eficiencia que determina la eficiencia de la formación del

extremo 3' terminal del ARNm maduro; elemento de posicionamiento que promueve el sitio corte; dos regiones de poli-T que flanquean el sitio de corte; y el sitio de corte y poliadenilación (3).

Resultados. En todos los transcriptomas analizados se encontraron regiones río arriba y río abajo del CDS del gen P7 donde no hubo la presencia de transcritos. Se determinaron las posiciones de inicio y final de las zonas no transcritas. La posición de inicio de la región promotora tuvo un coeficiente de variación de 0.003% y la posición final de 0.002%. La longitud media de esta región fue de 139 ± 12 nt. La posición de inicio de la región contigua río abajo de la región 3'UTR tuvo un coeficiente de variación de 0.004% y la posición final de la región terminadora de 0.003%. La longitud media de esta región fue de 158 ± 11 nt. Estos resultados demuestran que con el mapeo de transcriptomas es posible anotar una secuencia promotora y terminadora. Se identificaron motivos de secuencia característicos de regiones 5'UTR río arriba del CDS del gen P7 como son un marco de lectura abierto río arriba (uORF) y un sitio interno de entrada al ribosoma (IRES). También se pudieron identificar las regiones 5'UTR de los genes adyacentes al gen P7 por la presencia de un uORF, las cuales coincidieron con lo anotado para la cepa GS115 (4). Con el análisis de secuencias consenso se determinaron las 5 secuencias reportadas en terminadores de levaduras: el elemento de eficiencia, seguido a 30 nt el sitio de posicionamiento, 10 nt después el sitio de poli-T río arriba del sitio de corte, a 20 nt le sigue el sitio de corte (CAA) y 5 nt después está el sitio de poli-T final, por lo que se confirmó la ubicación de dicha región reguladora.

Conclusiones. Se identificaron las regiones reguladoras del gen P7 de *P. pastoris*.

Agradecimientos. Los autores agradecen el apoyo del CONACYT por las becas otorgadas (YRDP y KFC) y al proyecto CB-2016-01 (286093).

Bibliografía.

1. Lee JY et al. (2017) *Protein Expr. Purif.* 137: 7-12.
2. Love KR et al. (2018) *Curr. Opin. Biotechnol.* 53: 50-58.
3. Graber JH et al. (2003) *Trends Genet.* 19:473-476.
4. Love KR et al. (2016) *BMC Genomics* 17: 1-17.

