

# Estudio *in silico* de factores transcripcionales básicos de zipper de leucina de *Carica papaya* L.

*In silico* study of basic leucine zipper transcription factors of *Carica papaya* L.

Domínguez Bucheli K.<sup>1</sup>, Endara Chiriboga E.<sup>1</sup>, Idrovo Espín F.<sup>2\*</sup>

<sup>1</sup>Universidad de Las Américas, Facultad de Ingenierías y Ciencias Aplicadas  
email: kmdominguez@udlanet.ec  
email: eendara@udlanet.ec

<sup>2</sup>Universidad Central del Ecuador. Facultad de Ciencias Químicas.  
email: fmidrovo@uce.edu.ec\*

Artículo Original

Recibido: octubre 2020

Publicado: abril 2021

DOI: <https://doi.org/10.29166/quimica.v7i1.2811>

**Resumen:** Los factores transcripcionales regulan la expresión de genes al interactuar directamente con el ADN. La clase de factores básicos de zipper de leucina o bZIP en *arabidopsis* posee 75 miembros con dominios de aminoácidos similares entre todos sus miembros. En este trabajo se estudió *in silico* el genoma de papaya con la finalidad de encontrar posibles genes ortólogos bZIP en papaya. Se encontró 37 secuencias posibles, se realizó el análisis bioinformático de las secuencias. Se agrupó las secuencias y nombró de forma similar a la clasificación de *arabidopsis*. Los resultados expuestos a continuación podrían ser utilizados para evaluar de forma experimental la función biológica posible de estos genes en papaya.

**Palabras clave:** bZIP, factores transcripcionales, papaya, caricáceas

**Abstract:** Transcription factors regulate gene expression by interacting directly with DNA. The class of basic leucine zipper factors or bZIP in *arabidopsis* has 75 members with similar amino acid domains among all its members. In this work, the papaya genome was studied *in silico* in order to find possible ortholog bZIP genes in papaya. 37 possible sequences were found, the bioinformatics analysis of the sequences was performed. The sequences were grouped and named similarly to the *arabidopsis* classification. The results presented below could be used to experimentally evaluate the possible biological function of these genes in papaya.

**Key Words:** bZIP, transcription factors, papaya, caricaceae

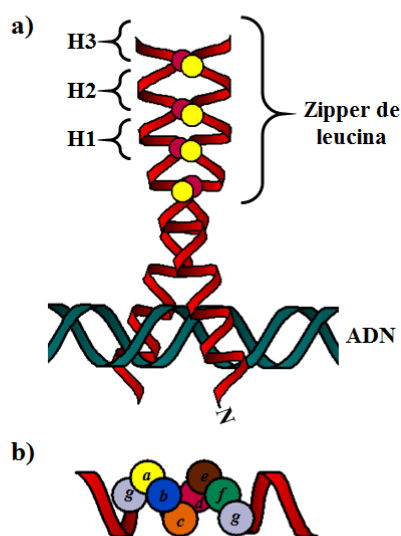
## 1. Introducción

Los factores transcripcionales (FT) son proteínas que se unen a secuencias específicas de ADN y regulan la transcripción (1) activándola o

reprimiéndola en respuesta a señales específicas intracelulares o ambientales (2).

La base de datos TRANSFAC 7.0 (3) agrupa a los factores transcripcionales de mamíferos en 5 superclases principales, cada una de estas se divide a su vez en clases y estas en familias. Los bZIP pertenecen a la superclase de dominios básicos y a la clase de factores básicos de zipper de leucina (del inglés Basic Leucine Zipper factors), el dominio está formado por una repetición periódica de leucinas por cada 7 residuos de aminoácidos (denominado heptámero) y la conformación resultante es una  $\alpha$ -hélice (4, 5) que permiten la formación del dímero de zipper de leucina (figura 1a). Cada posición de los residuos se designa con letras empezando por la posición g y seguidas de las posiciones a, b, c, d, e y f. Los residuos de leucina de forma general se ubican en cada posición d (6) de acuerdo como se ilustra en la figura 1b.

Adicionalmente, en el extremo N-terminal existe una región básica que facilita la unión del ADN (4).



**Figura 1.** Descripción gráfica de los factores transcripcionales bZIP a) dominio básico de zipper de leucina y forma de unión al ADN b) composición del heptámero en función de la posición de aa.

Para *Arabidopsis thaliana* Jakoby et al. (7) clasificaron a los bZIP dentro de una familia con 75 miembros, familia compuesta a su vez por 10

grupos que compartían similitud de secuencias proteicas.

El genoma secuenciado de *Carica papaya* se publicó el en el año 2008 (8) y desde entonces ha servido como referencia para el estudio de genes involucrados en diferentes procesos metabólicos de la papaya (9-12) la caracterización de secuencias bZIP del grupo D (13) así como también genes ortólogos de otras caricáceas (14, 15)

El presente trabajo es un estudio *in silico* a partir del genoma secuenciado de *C. papaya* que contempló la descripción de factores transcripcionales básicos de zipper de leucina en esta especie vegetal.

## 2. Parte Experimental

### 2.1. Materiales

Se utilizó las bases de datos bioinformáticos TAIR de arabisopsis (16) y PHYTOZOME (17) para papaya. Adicionalmente la plataforma *en línea* para predicción de secuencias proteicas FGESH (18) y finalmente los softwares, MEGA7 (19), PerfcBlast (20).

### 2.2. Métodos

#### 2.2.1 Búsqueda de secuencias

Se tomaron como referencia las 75 secuencias proteicas bZIP de arabisopsis (7) y se buscaron las secuencias de nucleótidos a partir de sus números de acceso en la base de datos TAIR (16). El genoma completo de papaya se descargó desde PHYTOZOME (17). Con las secuencias de arabisopsis se realizó un tBLASTx local (21) contra el genoma de papaya mediante PerfcBlast (20). Se seleccionó las secuencias codificantes con valores E menores que  $10^{-17}$  y porcentajes de

identidad, superiores a 25% dentro del genoma de *C. papaya*.

### 2.2.2 Análisis bioinformático

Se realizó la predicción proteica de las secuencias codificantes mediante FGENESH (18), el alineamiento (ClustalW), modelo y filogenia (Maximum Likelihood) de las predicciones proteicas se realizó mediante MEGA7 (19). Se utilizó la secuencia 26S (número de acceso AAF22522.1)

que codifica para una subunidad del proteasoma de *arabidopsis* como *outgroup*.

## 3. Resultados y Discusión

### 3.1. Número y función probable de genes *CpbZIP*

A partir de la búsqueda tBLASTx local se encontró 37 secuencias *CpbZIP* (tabla 1), se agrupó cada secuencia en siete grupos de acuerdo con la clasificación para *arabidopsis* propuesta por Jakoby et al. (7).

**Tabla 1.** Distribución de secuencias de nucleótidos de genes que codifican para bZIP de *A. thaliana* y posibles ortólogos en *C. papaya*.

Grupo/Nº <i>AtbZIP</i>	Función conocida en <i>A. thaliana</i>	<i>CpbZIP</i> (este estudio)
A/13	Respuesta a señalización por ABA, respuesta a estrés osmótico y sequía en tejidos vegetativos (22-26).	7
B/3	Modificación de las histonas durante estrés, respuesta a estrés salino en el retículo endoplasmático (27,28).	1
C/4	Desarrollo de semillas, respuestas mediadas por ABA a estrés abiótico (29,30).	2
D/10	Dependiendo del factor transcripcional, activador o represor de la expresión de genes de defensa (31).	6, reportados previamente en papaya (13)
E/2	Desarrollo del polen (32).	3
F/3	Adaptación a bajas concentraciones de zinc (33).	1
G/8	Germinación de semillas, elongación de hipocótilos (34)	3
H/2	Captación de nitrógeno, fotomorfogénesis (35).	1
I/13	Señalización osmosensora (36)	5
S/17	Respuesta a la luz (37).	8

Claramente se observó que las secuencias *CpbZIP* están presentes en un número menor que sus contrapartes de *arabidopsis*. Esta característica se repite en prácticamente todos los genes de las dos especies al compararlas entre sí, en promedio existe una reducción del 20% de genes en papaya. Presumiblemente esto se debe a las diferencias morfológicas entre *arabidopsis* y papaya (8).

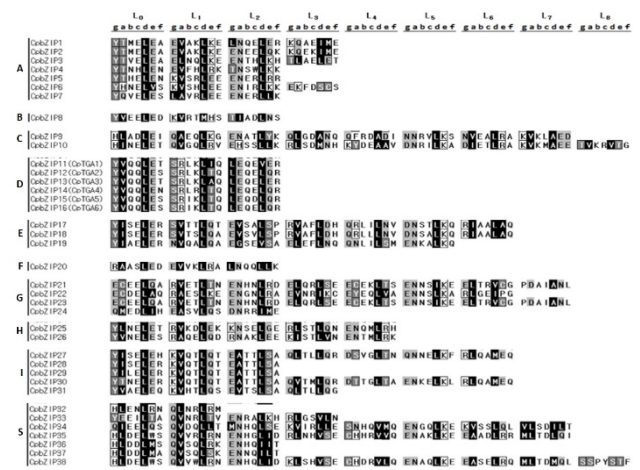
En los grupos H y C el número de genes de *arabidopsis* fue el doble que los de papaya, probablemente por efecto de las dos duplicaciones genómicas consecutivas (de sus siglas en inglés *whole genome duplications* WGD) que sufrió el genoma de *arabidopsis* en comparación con el de papaya que no tuvo duplicación alguna (38). Para el resto de los genes de *arabidopsis* cuyo número

no es doble de los genes de papaya, probablemente algunas copias de genes en arábidoopsis se perdieron por el proceso llamado de *fraccionamiento* que puede ocurrir luego de un evento de WGD (39).

Las funciones de los genes que codifican para bZIP en arábidoopsis son plenamente conocidas (tabla 1) y como una primera aproximación, es factible esperar respuestas biológicas parecidas entre genes ortólogos con secuencias similares, pero de diferentes especies. De hecho, algunos de los genes CpbZIP del grupo D respondieron a ácido salicílico (13) y podrían estar involucrados en la defensa en plantas en contra de patógenos de la misma forma como lo hacen en arábidoopsis. Sin embargo, para conocer las funciones de los demás genes de papaya encontrados en este trabajo, se debería demostrar experimentalmente. Las predicciones *in silico* son un excelente punto de partida, pero las diferencias morfológicas y fisiológicas entre arábidoopsis y papaya, se remontan de acuerdo con Fang et al. (40) hace 72 millones de años.

**Alineamiento y composición del dominio básico de zipper de leucina**

Las 37 secuencias proteicas CpbZIP compartieron dominios básicos de zipper de leucina similares, formados por los heptámeros de residuos de leucina y espaciados regularmente cada 7 aa (figura 2) lo que concuerda con lo expuesto previamente por Hakoshima (4).



**Fig. 2.** Alineamiento de secuencias predichas de aminoácidos CpbZIP en sus respectivos grupos, se detallaron también los heptámeros L0 a L8 dentro de los dominios bZIP y los tipos de aminoácidos alifáticos (cuadrados negros), básicos (cuadrados blancos), ácidos (cuadrados grises claros) e hidroxílicos (cuadrados grises oscuros).

Como criterio para seleccionar las 37 secuencias CpbZIP se consideró que cada heptámero debía contener uno de los aminoácidos L, M, I, V, C, A y S en la posición *d* ya que Moitra et al. (41) determinaron que la estabilización del dominio básico zipper de leucina depende del aminoácido en esa posición en particular y que el orden descendiente de estabilidad lo confería el aminoácido en la secuencia indicada. De forma general entre todas las CpbZIP predichas, la leucina es el aminoácido que se repite con mayor frecuencia en la posición *d* dentro de todos los heptámeros (L0 a L8).

Los heptámeros con aa diferentes no fueron considerados y las secuencias se cortaron hasta el heptámero anterior completo. Este es el motivo por el cual se encontró CpbZIP, con números de heptámeros y longitud variables (figura 2).

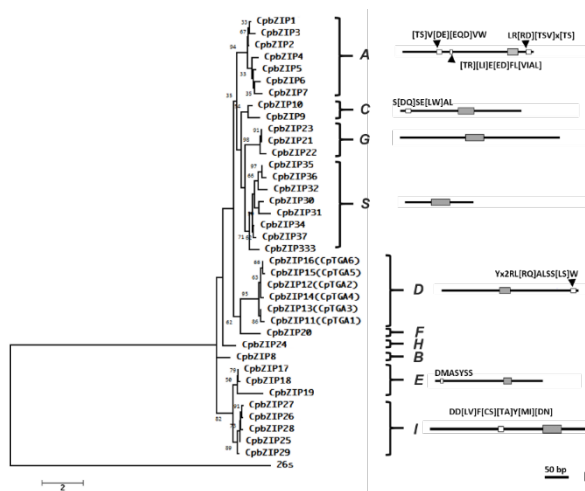
Los aa ubicados en posiciones *g* y *e* en cambio, son los responsables de la formación de dímeros y la interacción entre aa cargados de forma opuesta (positiva/negativa) promueven la dimerización



mientras que la interacción entre aa cargados de forma similar (positiva/positiva o negativa/negativa) inhiben la dimerización en FT bZIP de arabidopsis (42). Esto coincide con lo encontrado en las secuencias CpbZIP que mostraron la presencia de aa negativos y positivos en estas posiciones en particular que podrían interactuar entre sí de una forma específica, promoviendo la dimerización o inhibiéndola, según las características de los aa involucrados.

### 3.3 Árbol filogenético de secuencias CpbZIP

La filogenia de las 37 secuencias generó tres clados (Figura 3), el primero formado por los miembros de los grupos A (4 subclados), D (3 subclados), G (2 subclados), S (3 subclados), H, C y F y un clado más pequeño formado por los grupos I (3 subclados) y E (2 subclados) y el tercer clado formado únicamente por la a secuencia B.



**Fig. 3.** Árbol filogenético (Maximum Likelihood) de CpbZIP. El mejor método de sustitución de aa para el conjunto de datos fue calculado con MEGA 7 (19) el análisis involucró 38 secuencias de aa. Todas las posiciones que contenían gaps y datos faltantes fueron eliminadas. Se adjunta la firma característica de cada grupo, el bZIP se representa como un rectángulo gris.

Se observó que la clasificación de heptámeros propuesta en la figura 2 es consistente con este

árbol es decir las secuencias se agruparon en los clados correspondientes a su grupo en particular.

El árbol filogenético también mostró duplicación específica de linajes entre los subclados, lo que quiere decir que cada miembro podría tener funciones similares dentro de cada grupo que deberían demostrarse de forma experimental. Junto al árbol filogenético se incluyó las firmas de cada grupo que son representaciones gráficas que ilustran la posición del bZIP en cada grupo y otros dominios conservados que no necesariamente cumplen una función biológica determinada. Estas firmas fueron originalmente descritas también por Jakoby et al. (7). Las letras sin corchetes representan la secuencia de aa en esas posiciones en particular, las mismas que son constantes en todos los miembros del grupo, mientras que las letras entre corchetes representan aminoácidos variables que pueden ocupar esa posición en particular (solo una de las encerradas entre corchetes).

Mediante este trabajo se determinó la presencia de genes con características de FT bZIP en el genoma de papaya desde un enfoque *in silico* a partir bases de datos genómicos y predicción de genes, como estrategia para la notación de genes no descritos aún y cuya función biológica requiere ser confirmada de forma experimental. De acuerdo con Martínez (43), el advenimiento de tecnologías de secuenciación, el número creciente de genomas vegetales disponibles y las herramientas bioinformáticas han permitido la notación de genes con la finalidad de incrementar el conocimiento de las familias génicas y de la evolución.

### 4. Conclusiones

Se encontró *in silico* 37 secuencias que podrían ser ortólogos a factores transcripcionales básicos de

zipper de leucina (CpbZIP) en el genoma de papaya. En función de la similitud de las secuencias, estas se nombraron y agruparon en 10 grupos, de forma análoga a la clasificación de secuencias caracterizadas bZIP de *Arabidopsis*. Se elaboró el alineamiento y filogenia de las secuencias encontradas cuyas funciones biológicas podrían ser evaluadas de forma experimental a partir de esta información generada.

### Agradecimientos

Al Dr. Jorge Santamaría por su valioso apoyo y constante colaboración en todos los trabajos y proyectos científicos ejecutados.

### Referencias

1. T. R. Hughes in *A Handbook of Transcription Factors*, Ed.: T. R. Hughes, Springer Science+Business Media B.V., New York, 2011, pp. 1.
2. Salgado, H.; Martínez-Antonio, A.; Janga, S. *FEBS Lett.* 2007, 58, 3499–3506.
3. Matys, V.; et al. *Nuc. Ac. Res.* 2006, 34, D108–D110.
4. Hakoshima, T. *eLS*, 2014, 1–5.
5. Landschulz, W.H.; Johnson, P.F.; McKnight, S.L. 1988. *Science*. 1998, 240:1759–64.
6. O'Shea E.K.; Klemm, J.D.; Kim, P.S.; Alber, T. *Science*. 1991, 254, 539–44.
7. Jakoby, M.; Weisshaar, B.; Dröge-Laser, W.; Vicente-Carbajosa, J.; Tiedemann, J.; Kroj, T.; Parcy, F. *Trends Plant Sci.* 2002, 7, 3, 106–111.
8. Ming, R.; et al. *Nature*. 2008, 452, 24, 991–996.
9. Takata, N.; Saito, S.; Tanaka, C.; Uemura, M. *BMC Evol. Biol.* 2010, 10, 126.
10. Peraza-Echeverría, S.; Santamaría, J.M.; Fuentes, G.; Menéndez-Cerón, M.; Vallejo-Reyna, M.A.; Herrera-Valencia, V.A. *Genes Genom.* 2012, 34, 4, 379–390.
11. Estrella-Maldonado H.; et al. *Plant Cell Tiss. Organ Cult.* 2016, 126, 2, 187–204.
12. Estrella-Maldonado, H.; et al. *Plant Growth Regul.* 2018, 37, 2, 502–516.
13. Idrovo-Espín, F.M.; Peraza-Echeverría, S.; Fuentes, G.; Santamaría, J.M. *Plant Physiol. Bioch.* 2012, 54, 113–122.
14. Arizala-Quinto, E.; Viteri, G.; Idrovo-Espín, F. J. *Basic App. Gen.* 2019, 30, 1, 1–3.
15. Cevallos-Vilatuña T.; Garzón-Salazar K.; Idrovo-Espín F. *Rev. Amazónica Ciencia Tec.* 2019, 8, 1, 1–11.
16. Swarbreck, D.; et al. *Nuc. Ac. Res.* 2008, D1009–14.
17. Goodstein, D.M.; et al. *Nuc. Ac. Res.* 2012, D1178–86.
18. FGENESH. <http://www.softberry.com/berry.phtml?topic=fgenes&group=programs&subgroup=gfind> (accesado 6 de abril 2020).
19. Kumar, S.; Stecher, G.; Tamura, K. *Molec. Biol. Evol.* 2016, 33, 7, 1870–1974.
20. Santiago-Sotelo, P.; Ramirez-Prado, J. H. *BioTechn.* 2012, 53, 5, 299–300.
21. Altschul S. F.; Madden T. L.; Schäffer A. A.; Zhang J.; Zhang Z.; Miller W.; Lipman D. J.; *Nucl. Ac. Res.* 1997, 25, 3389–3402.
22. Uno, Y.; Furihata, T.; Abe, H.; Yoshida, R.; Shinozaki, K.; Yamaguchi-Shinozaki, K. *Proc. Natl. Acad. Sci. U. S. A.* 2002, 99, 11632–11637.
23. Choi, H.; Hong, J.; Ha, J.; Kang, J.; Kim, S. Y. *J. Biol. Chem.* 2000, 275, 1723–1730.
24. Kang, J.; Choi, H.; Im, M.; Kim, S. *Plan Cell.* 2002, 14, 343–357.
25. Yoshida, T.; et al. *Plant J.* 2010, 61, 672–685.
26. Yoshida, T.; et al. *Plant Cell Environ.* 2015, 38, 35–49.

27. Liu, J.; Srivastava, R.; Howell, S. *Plant Cell*. 2007, 19, 4111-4119.
28. Song, Z.T.; Sun, L.; Lu, S. J.; Tian, Y.; Ding, Y.; Liu, J. *Proc. Natl. Acad. Sci. U. S. A*. 2015, 3, 112, 9, 2900–2905.
29. Ehlert, A.; Weltmeier, F.; Wang, X.; Mayer, C.; Smeeckens, S.; Carbajosa, V.; Lase, W. *Plant J*. 2006, 46, 890–900.
30. Matioli, C.C.; et al. *Plant Physiol*. 2011, 157, 2, 692-705.
31. Kesarwani, M.; Yoo, J.; Dong, X. *Plant Physiol*. 2007, 144, 1, 336–346.
32. Gibalova, A.; Renák, D.; Matczuk, K.; Dupl'áková, N.; Cháb, D.; Twell, D.; Honys, D. *Plant Molec. Biol*. 2009, 70, 581–601.
33. Assunção, A.; et al. *Proc. Natl. Acad. Sci. U. S. A*. 2010, 107, 10296–10301.
34. Hsieh, W.; Hsieh, H.; Wu, S. *Plant Cell*. 2012, 24, 3997–4011.
35. Huang, L.; Zhang, H.; Zhang, H.; Deng, X.; Wei, N. *Plant Sci*. 2015, 238, 330–339.
36. Tsugama, D.; Liu, S.; Takano, T. *PLOS One*. 2014, 9, 8, e103930.
37. Rook, F.; Weisbeek, P.; Smeeckens, S. *Plant Molec. Biol*. 1998, 37, 171–178.
38. Bowers, J.E.; Chapman, B.A.; Rong, J.; Paterson, A.H. *Nature*. 2003, 422, 433–438.
39. Langham, R.J.; Walsh, J.; Dunn, M.; Ko, C.; Goff, S. A.; Freeling, M. *Genetics*. 2004, 166, 2, 935–945.
40. Fang, J.; et al. *Euphytica*, 2016, 209, 323–339
41. Moitra, J.; Szilák, L.; Krylov, D.; Vinson, C. *Biochem*. 1997, 36, 41, 12567–12573.
42. Deppmann, C., Acharya, A.; Rishi, V.; Wobbes, B.; Smeeckens, S.; Taparowsky, E.; Vinson, C. *Nuc. Ac. Res*. 2004, 32, 11, 3435–3445.
43. Martínez, M. *Comput. Struct. Biotechnol. J*. 2013, 8, 10, e201307001.