

XÁC ĐỊNH VÀ PHÂN TÍCH CÁC GENE MÃ HÓA PROTEIN LIÊN KẾT STRESS (SAP) Ở CÂY ĐU ĐỦ (*Carica papaya* L.) BẰNG PHƯƠNG PHÁP *IN SILICO*

Lê Thị Mận¹, Trần Thị Thanh Huyền^{2*}, Lê Chí Toàn³, Trần Thị Thanh Loan^{1,4} và Cao Phi Bằng¹

¹Khoa Khoa học Tự nhiên, Trường Đại học Hùng Vương

²Khoa Sinh học, Trường Đại học Sư phạm Hà Nội

³Khoa Sinh - Kỹ thuật Nông nghiệp, Trường Đại học Sư phạm Hà Nội 2

⁴Trường Trung học cơ sở Tiên Kiên, Lâm Thao, Phú Thọ

Tóm tắt. Nhờ sử dụng các phương pháp nghiên cứu *in silico*, chúng tôi đã xác định được 7 gene mã hóa các protein liên quan stress (SAP) ở trong hệ gene của cây Đu đủ (*Carica papaya* L.). Các gene SAP của cây Đu đủ có kích thước từ 416 tới 813 nucleotide, không có hoặc chỉ có một intron. Các protein suy diễn có kích thước từ 114 tới 270 amino acid, khối lượng phân tử nằm trong khoảng 13,10 kDa tới 29,63 kDa. Các protein suy diễn có tính kiềm với pI dao động từ 8,82 đến 9,95. Căn cứ vào kết quả phân tích cấu trúc và cây phả hệ, các SAP của cây Đu đủ được phân chia thành hai nhóm nhóm I (năm gene) và phân nhóm II (hai gene). Các SAP này của cây Đu đủ có mức độ bảo tồn cao về cấu trúc với hai vùng bảo tồn A20-AN1 (nhóm I) hoặc vùng AN1 (nhóm II). Phân tích dữ liệu RNA-seq được xây dựng từ lá đu đủ trong điều kiện thường và điều kiện đông lạnh cho thấy tất cả 7 gene SAP của cây Đu đủ đều biểu hiện và đều được cảm ứng biểu hiện mạnh hơn trong điều kiện đông lạnh so với điều kiện thường. Gene *CpSAP2* có mức độ biểu hiện tương đối mạnh nhất trong các gene SAP của cây Đu đủ. Kết quả nghiên cứu này là cơ sở cho các nghiên cứu về tách dòng gene, phân tích chức năng của các gene trong họ SAP và chọn giống ở cây Đu đủ trong đáp ứng với các điều kiện stress vô sinh của môi trường.

Từ khóa: biểu hiện gene, cây Đu đủ (*Carica papaya* L.), cây phả hệ, *in silico*, protein liên kết stress (SAP).

1. Mở đầu

Các protein liên kết stress (stress-associated protein, SAP) thuộc về nhóm protein kẹp kềm, có chứa vùng A20/AN1, được coi như một thành tố quan trọng trong sự chống chịu stress của thực vật [1]. Mức độ phiên mã của các gene SAP tăng dưới tác động của nhiều stress vô sinh. Các protein SAP cũng được phát hiện ở thực vật chịu tác động của các stress như nóng, lạnh, hạn, mặn, kim loại nặng [1, 2]. Các SAP còn được chứng minh hoạt động như là ubiquitin ligase, cảm ứng oxi hóa khử, tác nhân điều hòa biểu hiện gene đáp ứng stress ở thực vật [2].

Ở mức độ hệ gene hoàn chỉnh, các gene mã hóa SAP đã được xác định trong hệ gene của các cây lúa (18 gene), *A. thaliana* (14 gene) [3], cà chua (13 gene) [4], bông (37 gene) [5], đậu tương (27 gene) [6]. Mặc dù những nghiên cứu đã bổ sung các họ gene SAP ở thực vật nhưng những nghiên cứu tương tự vẫn cần được thực hiện ở các đối tượng cây trồng khác nhằm làm sáng tỏ vai trò của các protein này trong tính chống chịu của thực vật với các stress môi trường.

Ngày nhận bài: 24/2/2021. Ngày sửa bài: 19/3/2021. Ngày nhận đăng: 26/3/2021.

Tác giả liên hệ: Trần Thị Thanh Huyền. Địa chỉ e-mail: tranthanhhuyen@hnue.edu.vn

Cây Đu đủ (*Carica papaya* L.) là cây ăn quả quan trọng được trồng ở vùng nhiệt đới và cận nhiệt đới. Quả đu đủ chín mềm và ngọt, giàu dinh dưỡng, chứa lượng lớn tiền vitamin A, vitamin C và các chất chống oxy hóa [7]. Cây Đu đủ chịu tác động của nhiều nhân tố sinh thái như ánh sáng, nhiệt độ, chế độ nước, gió cũng như nhiều tác nhân stress sinh học [8].

Nghiên cứu này hướng tới mục tiêu xác định và phân tích đặc điểm của các gene *SAP* ở cây đu đủ, loại cây ăn quả nhiệt đới quan trọng, bằng các phương pháp tin sinh học, phương pháp nghiên cứu hiện đại, phổ biến trong công nghệ sinh học hiện đại [9].

2. Nội dung nghiên cứu

2.1. Cơ sở dữ liệu và phương pháp nghiên cứu

* Cơ sở dữ liệu hệ gene của cây Đu đủ

Trình tự hệ gene của cây Đu đủ (*Carica papaya* L.) được khai thác từ phytozome V12 (<http://www.phytozome.net/>) [10].

* Xác định các gene *SAP* ở cây Đu đủ

Mười bốn protein *SAP* của cây *A. thaliana* [3] được sử dụng làm trình tự khuôn dò để tìm kiếm các gene tương đồng trên toàn hệ gene của cây Đu đủ bằng chương trình TBLASTN [11].

* Xây dựng cây phả hệ

Các protein suy diễn *SAP* của cây Đu đủ được sắp dãy bằng MAFFT [12] cùng các protein *SAP* của cây *A. thaliana* và cây lúa. Cây phả hệ được xây dựng nhờ phần mềm MEGA X [13].

* Phân tích *in silico* các gene *SAP* của cây Đu đủ

Các đặc điểm vật lý, hóa học của các gene/protein *SAP* được phân tích bằng công cụ ProtParam trên server ExPASy (Expert Protein Analysis System) [14]. Vị trí khu trú dưới tế bào được phân tích nhờ ProtComp 9.0 [15]. Chức năng của các gene *CpSAP* được phân tích nhờ công cụ agriGO v2.0 [16].

* Phân tích sự biểu hiện gene

Sự biểu hiện của các gene được xác định qua phân tích dữ liệu RNA-seq của cây Đu đủ trong ngân hàng Gene Expression Omnibus GSE130188 [17], gồm có dữ liệu RNA-seq từ mẫu lá cây Đu đủ trong điều kiện thường (26°C) (GSM3733617) và trong điều kiện đông lạnh ở -60 °C sau quá trình giảm mỗi 0,5 °C/ngày trong thời gian 180 ngày (GSM3733618). Mức độ biểu hiện tương đối của các gene *CpSAP* được tính bằng cách lấy giá trị biểu hiện của gene quan tâm (FPKM) ở điều kiện đông lạnh chia cho giá trị biểu hiện của gene đó (FPKM) ở điều kiện thường. Giá trị thu được được chuẩn hóa (chia cho) tỉ lệ giữa giá trị biểu hiện của gene *eIF4E* (FPKM) ở điều kiện đông lạnh và giá trị biểu hiện của gene *eIF4E* (FPKM) ở điều kiện thường. Gene *eIF4E* được chọn làm gene chuẩn vì gene này đã được chứng minh là có mức độ biểu hiện ổn định trong các điều kiện khác nhau ở cây Đu đủ và phù hợp để chuẩn hóa trong các nghiên cứu biểu hiện gene ở các điều kiện khác nhau của cây Đu đủ [18].

2.2. Kết quả nghiên cứu và thảo luận

2.2.1. Xác định họ gene *SAP* ở cây Đu đủ

Tổng số 7 gene có thể mã hóa cho các *SAP* được xác định trong hệ gene của cây Đu đủ (Bảng 1). Kiểm tra Pfam [19] các protein suy diễn chỉ ra tất cả các *CpSAP* có chứa các vùng bảo tồn AN1-like Zinc finger (PF01428) trong khi chỉ 5 trong 7 *CpSAP* có chứa vùng bảo tồn A20-like zinc finger (PF01754), ngoại trừ *CpSAP2* và *CpSAP5*. Với bảy gene, đu đủ có số lượng gene *SAP* ít hơn so với ở các loài đã biết như lúa (18 gene), *A. thaliana* (14 gene) [3], cà chua (13 gene) [4], bông (37 gene) [5], đậu tương (27 gene) [6].

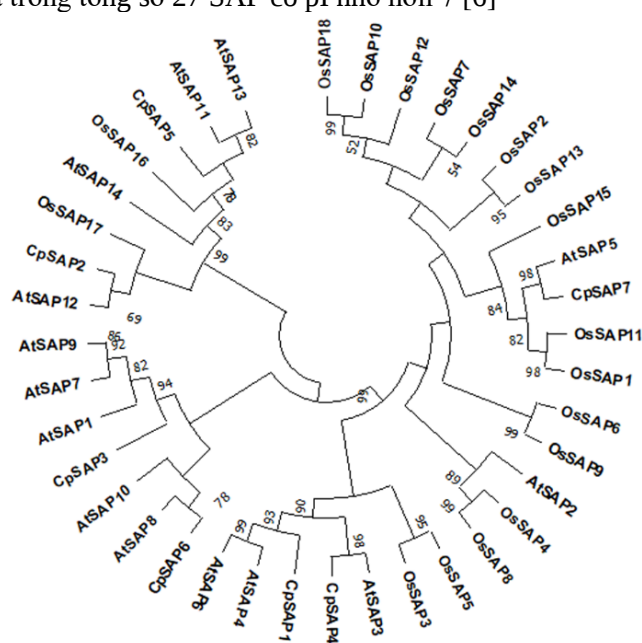
Bảng 1. Các trình tự SAP ở cây Đu đủ

Gene	Phân nhóm	Tên locus	GS (bp)	PL (aa)	MW (kD)	pI	GRAVY	In	SCL
<i>CpSAP1</i>	A20-AN1	evm.TU.supercontig_5.42	522	173	18,78	8,90	-0,381	0	Chlo, Mito
<i>CpSAP2</i>	AN1-AN1	evm.TU.supercontig_14.84	653	187	20,79	8,90	-0,566	1	Nucl
<i>CpSAP3</i>	A20-AN1	evm.TU.supercontig_29.64	507	168	17,58	8,82	-0,348	0	Chlo
<i>CpSAP4</i>	A20-AN1	evm.TU.supercontig_32.102	534	177	19,04	8,85	-0,530	0	Chlo, Nucl
<i>CpSAP5</i>	AN1-AN1-CH2-CH2	evm.TU.supercontig_92.40	813	270	29,63	8,91	-0,566	0	Nucl
<i>CpSAP6</i>	A20-AN1	evm.TU.supercontig_107.19	416	114	13,10	9,95	-0,595	1	Nucl
<i>CpSAP7</i>	A20-AN1	evm.TU.supercontig_2168.1	561	186	19,80	9,05	-0,511	0	Chlo

Chú thích: GS = Kích thước gen, PL = Chiều dài phân tử protein, MW = Khối lượng phân tử protein, pI = điểm đẳng điện, In = Số lượng intron, SCL = Khu trú dưới tế bào, Chlo: lục lạp, Mito: ti thể, Nucl: nhân tế bào

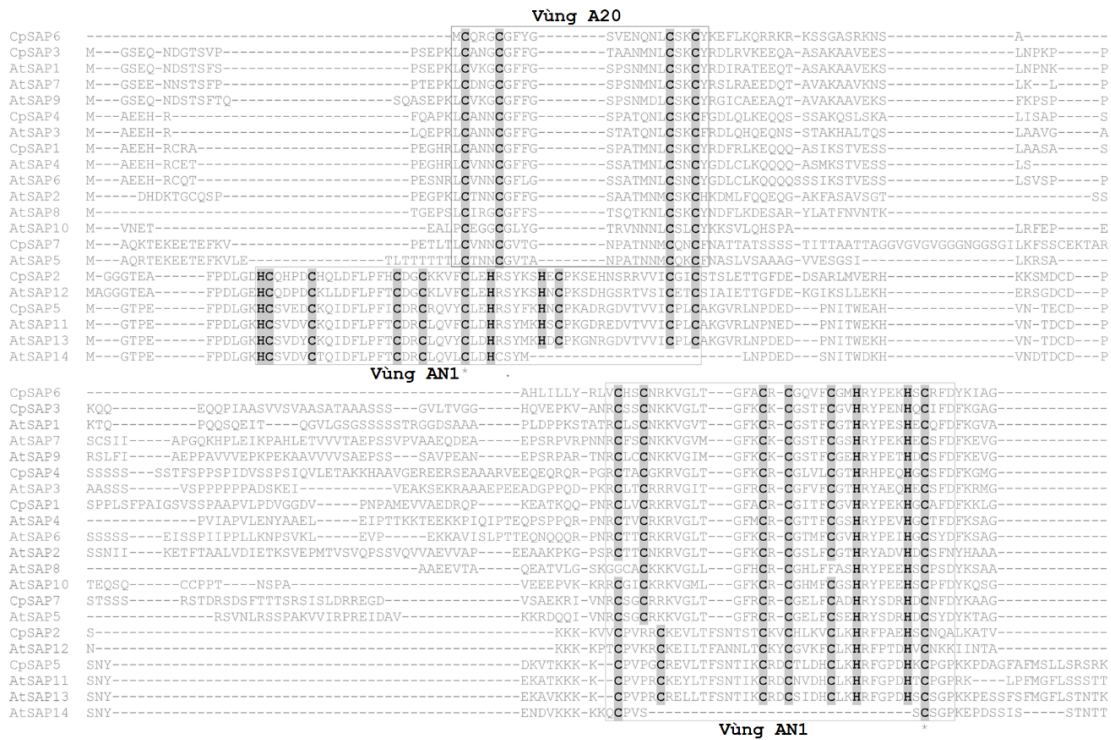
2.2.2. Phân tích đặc điểm và tiến hóa các SAP ở cây Đu đủ

Các gene SAP của cây Đu đủ có độ dài trình tự nucleotide khác nhau, từ 416 tới 813 nucleotide. Chỉ hai gene, *CpSAP2* và *CpSAP6* có một intron, các gene còn lại mã hóa liên tục. Đặc điểm này tương đồng với phần lớn các SAP đã biết như ở cây bông [5]. Ở cây *A. thaliana*, chỉ có hai gene có số intron lớn hơn 1 là *AtSAP2* (2 intron) và *AtSAP14* (3 intron). Tương tự, ở cây lúa chỉ duy nhất gene *OsSAP8* có hai intron [3]. Các protein suy diễn có độ dài từ 114 tới 270 amino acid, khối lượng phân tử nằm trong khoảng 13,10 kDa tới 29,63 kDa. Các protein suy diễn có tính kiềm, giá trị điểm đẳng điện (pI) lý thuyết của các SAP ở cây Đu đủ khá tương đồng, dao động từ 8,85 tới 9,95 (Bảng 1). Các đặc điểm này khá giống với đặc điểm của các SAP của cây Lúa và *A. thaliana* [3], Cà chua [4], Bông [5]. Riêng ở cây Đậu tương có ba trong tổng số 27 SAP có pI nhỏ hơn 7 [6]



Hình 1. Cây phả hệ Maximum-Likelihood được thiết lập từ các SAP ở các loài đu đủ (*Cp*), *A. thaliana* (*At*) và lúa (*Os*) nhờ phần mềm MEGAX [13] với 1000 bootstrap lặp lại
Giá trị bootstraps được thể hiện ở gốc của nhánh

Cây phả hệ (Hình 1) cho thấy các SAP ở cây Đu đủ được xếp vào hai nhóm chính, nhóm I gồm phần lớn các gene (5 trong tổng số 7 gene), trong khi nhóm II chỉ gồm có hai gene, *CpSAP2* và *CpSAP5*. Các SAP nhóm I có hai vùng bảo tồn A20-AN1, nhóm II chỉ có vùng bảo tồn AN1 (2 vùng AN1) mà không có vùng A20 (hình 2). Đặc điểm này giống với ở cây đậu tương [6]. Tuy nhiên, ở cây Đu đủ không có thành viên nào mang một vùng bảo thủ AN1 như đã được tìm thấy ở các loài *A. thaliana* (AtSAP14), lúa (OsSAP13-OsSAP15), Cà chua (SlSAP10), Bông (GhSAP15A, GhSAP15B và GhSAP19), đậu tương (GmSAP15 và GmSAP17). Ở Lúa và Đậu tương còn tìm thấy gene chỉ chứa vùng bảo tồn A20 (OsSAP18 và GmSAP23) [3, 5, 6, 20]. Cây phả hệ cũng cho thấy rằng các SAP ở cây Đu đủ đã có mặt từ cây tổ tiên chung giữa cây một lá mầm và cây hai lá mầm. Hơn nữa, ở cây Đu đủ không phát hiện ra bất cứ hiện tượng nhân gene nào sau sự phát sinh loài. Hiện tượng tiến hóa này ở cây Đu đủ khác hoàn toàn so với ở các cây lúa, *Arabidopsis*, cà chua, bông và đậu tương.

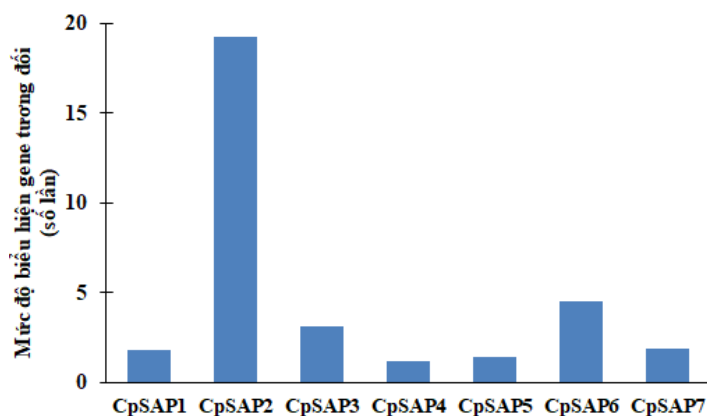


Hình 2. Các trình tự peptide của cây Đu đủ và *A. thaliana* được sắp dãy bởi MAFFT [12] thể hiện rõ các vùng bảo tồn A20 và AN1

Các amino acid Cystein và Histidine bảo tồn được đánh dấu bằng nền xám



Hình 3. So sánh trình tự peptide của *CpSAP7* với *AtSAP5* và *OsSAP1*
Các amino acid Cystein và Histidine bảo tồn được đánh dấu bằng nền xám



Hình 5. Mức độ biểu hiện tương đối của các gene SAP của cây đu đủ

3. Kết luận

Chúng tôi đã xác định được 7 gene trong họ gene SAP ở cây Đu đủ và phân tích bằng phương pháp *in silico*. Cấu trúc gene và các đặc điểm lí - hóa của các protein SAP của cây Đu đủ cũng được phân tích. Các SAP của cây Đu đủ được xếp vào hai nhóm, nhóm I (có hai vùng bảo tồn A20-AN1) và II (chỉ có vùng bảo tồn AN1). Sự biểu hiện gene của các gene SAP ở lá cây Đu đủ đã được khảo sát trong điều kiện thường và điều kiện đông lạnh thông qua ngân hàng RNA-Seq. Tất cả các gene SAP của cây Đu đủ đều biểu hiện, và mức độ biểu hiện tương đối cho thấy các gene này đáp ứng với điều kiện đông lạnh (nhiệt độ thấp), mạnh nhất là gene *CpSAP2*. Kết quả này có ý nghĩa lớn, mở đường cho việc tách dòng gene và phân tích chức năng của các gene trong họ SAP ở cây Đu đủ trong đáp ứng với các điều kiện stress vô sinh khác.

Lời cảm ơn. Công trình này được hoàn thành với sự hỗ trợ kinh phí từ chương trình nghiên cứu khoa học cơ bản của Trường Đại học Hùng Vương (Đề tài mã số: 24/2020/HĐKH.HV20-24).

TÀI LIỆU THAM KHẢO

- [1] J. Giri, P. K. Dansana, K. S. Kothari, G. Sharma, S. Vij, and A. K. Tyagi, 2013. SAPs as novel regulators of abiotic stress response in plants. *Bioessays*, Vol. 35, No. 7, pp. 639-48, doi: 10.1002/bies.201200181.
- [2] R. R. Kumar S. K. Sharma, S. Goswami, G. P. Singh, R. Singh, K. Singh, H. Pathak and R. D. Rai, 2013. Characterization of differentially expressed stress-associated proteins in starch granule development under heat stress in wheat (*Triticum aestivum* L.). *Indian J. Biochem Biophys*, Vol. 50, No. 2, pp. 126-38.
- [3] S. Vij and A. K. Tyagi, 2006. Genome-wide analysis of the stress associated protein (SAP) gene family containing A20/AN1 zinc-finger(s) in rice and their phylogenetic relationship with Arabidopsis. *Mol Genet Genomics*, Vol. 276, No. 6, pp. 565-575.
- [4] A. U. Solanke, M. K. Sharma, A. K. Tyagi, and A. K. Sharma, 2009. Characterization and phylogenetic analysis of environmental stress-responsive SAP gene family encoding A20/AN1 zinc finger proteins in tomato. *Mol Genet Genomics*, Vol. 282, No. 2, pp. 153-64, doi: 10.1007/s00438-009-0455-5.
- [5] W. Gao, L. Long, X. Tian, J. Jin, H. Liu, H. Zhang, F. Xu and C. Song, 2016. Genome-wide identification and expression analysis of stress-associated proteins (SAPs) containing A20/AN1 zinc finger in cotton. *Mol Genet Genomics*, Vol. 291, No. 6, pp. 2199-2213, doi: 10.1007/s00438-016-1252-6.

- [6] X.-Z. Zhang, W.-J. Zheng, X.-Y. Cao, X.-Y. Cui, S.-P. Zhao, T.-F. Yu, J. Chen, Y.-B. Zhou, M. Chen, S.-C. Chai, Z.-S. Xu, and Y.-Z. Ma, 2019. Genomic analysis of stress associated proteins in Soybean and the role of GmSAP16 in abiotic stress responses in *Arabidopsis* and soybean, (in English). *Front Plant Sci*, Vol. 10, No. 1453, doi: 10.3389/fpls.2019.01453.
- [7] R. Ming, Q. Yu, P. H. Moore, R. E. Paull, N. J. Chen, M.-L. Wang, Y. J. Zhu, M. A. Schuler, J. Jiang and A. H. Paterson., 2012. Genome of papaya, a fast growing tropical fruit tree. *Tree Genetics & Genomes*, Vol. 8, No. 3, pp. 445-462, doi: 10.1007/s11295-012-0490-y.
- [8] E. Campostrini and D. M. Glenn, 2007. Ecophysiology of papaya: a review. *Brazilian Journal of Plant Physiology*, Vol. 19, No. 4, pp. 413-424.
- [9] Cao Phi Bằng, 2015, Xác định và phân tích *in silico* các gen *DREB2* ở cây quýt (*Citrus clementina*). *Tạp chí Khoa học Trường Đại học Sư phạm Hà Nội*, Vol. 60, No. 4, tr. 127-131, doi: 10.18173/2354-1059.2015-00018.
- [10] R. Ming, S. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte, J. H. Saw, P. Senin, W. Wang, B. V. Ly, K. L. T. Lewis, S. L. Salzberg, L. Feng, M. R. Jones, R. L. Skelton, J. E. Murray, C. Chen, W. Qian, J. Shen, P. Du, M. Eustice, E. Tong, H. Tang, E. Lyons, R. E. Paull, T. P. Michael, K. Wall, D. W. Rice, H. Albert, M.-L. Wang, Y. J. Zhu, M. Schatz, N. Nagarajan, R. A. Acob, P. Guan, A. Blas, C. M. Wai, C. M. Ackerman, Y. Ren, C. Liu, J. Wang, J. Wang, J.-K. Na, E. V. Shakhov, B. Haas, J. Thimmapuram, D. Nelson, X. Wang, J. E. Bowers, A. R. Gschwend, A. L. Delcher, R. Singh, J. Y. Suzuki, S. Tripathi, K. Neupane, H. Wei, B. Irikura, M. Paidi, N. Jiang, W. Zhang, G. Presting, A. Windsor, R. Navajas-Pérez, M. J. Torres, F. A. Feltus, B. Porter, Y. Li, A. M. Burroughs, M.-C. Luo, L. Liu, D. A. Christopher, S. M. Mount, P. H. Moore, T. Sugimura, J. Jiang, M. A. Schuler, V. Friedman, T. Mitchell-Olds, D. E. Shippen, C. W. dePamphilis, J. D. Palmer, M. Freeling, A. H. Paterson, D. Gonsalves, L. Wang and M. Alam, 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature*, Vol. 452, pp. 991-996, doi: 10.1038/nature06856.
- [11] E. M. Gertz, Y. K. Yu, R. Agarwala, A. A. Schaffer and S. F. Altschul, 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol*, Vol. 4, pp. 41, doi: 10.1186/1741-7007-4-41.
- [12] K. Katoh and D. M. Standley, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, Vol. 30, No. 4, pp. 772-80, doi: 10.1093/molbev/mst010.
- [13] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, Vol. 35, No. 6, pp. 1547-1549, doi: 10.1093/molbev/msy096.
- [14] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, and A. Bairoch, 2005. *Protein identification and analysis tools on the ExPASy server*, in "The proteomics protocols handbook", Springer, p. 571-607.
- [15] <http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=help&subgroup=proloc>
- [16] T. Tian, Y. Liu, H. Yan, Q. You, X. Yi, Z. Du, W. Xu and Z. Su, 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res*, Vol. 45, No. W1, pp. W122-w129, doi: 10.1093/nar/gkx382.
- [17] <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130188>
- [18] X. Zhu, X. Li, W. Chen, J. Chen, W. Lu, L. Chen and D. Fu, 2012. Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. *PLoS One*, Vol. 7, No. 8, p. e44405, doi: 10.1371/journal.pone.0044405.

- [19] R. D. Finn, A. Bateman, J. Clements, P. Coghill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta, 2014. Pfam: the protein families database. *Nucleic Acids Res*, Vol. 42, No. Database issue, pp. D222-30, doi: 10.1093/nar/gkt1223.
- [20] A. U. Solanke, M. K. Sharma, A. K. Tyagi, and A. K. Sharma, 2009. Characterization and phylogenetic analysis of environmental stress-responsive SAP gene family encoding A20/AN1 zinc finger proteins in tomato. *Mol Genet Genomics*, Vol. 282, No. 2, pp. 153-164, doi: 10.1007/s00438-009-0455-5.

ABSTRACT

Identification and analysis of stress-associated protein (SAP) encoding genes in papaya (*Carica papaya* L.) by using *in silico* methods

Le Thi Man^{1,2}, Tran Thi Thanh Huyen^{3,*}, Le Chi Toan⁴, Tran Thi Thanh Loan^{1,5}, Cao Phi Bang^{1,2}

¹Faculty of Natural Sciences, Hung Vuong University

²Biotechnology Research Group, Hung Vuong University

³Faculty of Biology, Hanoi National University of Education

⁴Faculty of Biology-Agriculture, Hanoi Pedagogical University No. 2

⁵Tien Kien Secondary School, Lam Thao, Phu Tho

By using the *in silico* methods, total of seven stress-associated protein encoding genes have identified in the genome of papaya (*Carica papaya* L.). The full-length genomic sequence of papaya SAP genes were ranging from 416 to 813 nucleotides. These genes had single intron or were intronless. The predicted protein sequences included from 114 to 270 amino acids, according to the molecular weight ranged from 13.10 to 29.63 kDa. These proteins were alkaline with a pI value ranging from 8.82 to 9.95. Based on the protein structure and the phylogenetic analysis, the papaya SAP were divided into two groups, I (five members) and II (two members). The papaya SAP had a highly conserved level of structure including two conserved regions, A20-AN21 (group I), or AN21 domains only (group II). RNA-seq analysis showed that all of seven SAP genes expressed in papaya leaves and were induced by freeze thaw awakening treatment (in comparison with control treatment). *CpSAP2* showed the highest relative expression level in compared to other SAP genes of papaya. The results of this work have an important significance and are base of the further research on gene cloning, functional analysis of SAP genes and breeding of papaya in response to environmental abiotic stresses.

Keywords: gene expression, *in silico*, papaya (*Carica papaya* L.), phylogenetic tree, stress-associated protein (SAP).