

Tách dòng và xác định trình tự hai gene mã hóa methylketone synthase 2 (NtMK2-1 và NtMKS2-2) ở Thuốc lá (*Nicotiana tabacum*)

Trương Trần Diệu, Trần Thị Diễm Hương*,
Hồ Tiền Giang Em, Nguyễn Thị Hồng Thương

Khoa Sinh học-Công nghệ sinh học, Trường Đại học Khoa học Tự nhiên, ĐHQG TPHCM, Việt Nam

Nhận ngày 16 tháng 8 năm 2017

Chỉnh sửa ngày 09 tháng 9 năm 2017; Chấp nhận đăng ngày 10 tháng 10 năm 2017

Tóm tắt: 2-Methylketone được tích lũy ở một số loài thực vật là sản phẩm phụ của con đường sinh tổng hợp acid béo diễn ra ở lập thể. Methylketone synthase 2 (MKS2) là một thioesterase xúc tác bước phản ứng kế cuối trong con đường sinh tổng hợp methylketone, giúp chuyển hóa 3-ketoacyl-ACP thành 3-ketoacid, tiền chất trực tiếp để tổng hợp các 2-methylketone. Các nghiên cứu đã công bố cho thấy các loài thực vật khác nhau trong họ Solanaceae có thể có số lượng gene *MKS2* khác nhau và mỗi *MKS2* khác nhau có thể xúc tác tổng hợp các axit béo khác nhau về chiều dài chuỗi (C6-C18), mức độ bão hòa và trạng thái oxy hóa khử của khung carbon. Methylketone synthase 2 từ cà chua dại (*Solanum habrochaites*) (*ShMKS2*) là enzyme đầu tiên được xác định ở thực vật. Trong nghiên cứu này, với sự hỗ trợ của các công cụ tin sinh học, chúng tôi đã xác định được hai gene tương đồng với *ShMKS2* hiện diện trên hai contig AWOJ-SS748 và AWOJ-SS6425 trong cơ sở dữ liệu bộ gene của thuốc lá (*Nicotiana tabacum*), và đặt tên là *NtMKS2-1* và *NtMKS2-2*. Cả hai gene đều bao gồm 5 exon và 4 intron. Trình tự mã hóa (CDS) *NtMKS2-1* và *NtMKS2-2* hoàn chỉnh đã được phân lập thành công từ nguồn cDNA được chuẩn bị từ mô lá non của cây thuốc lá (*Nicotiana tabacum*), được giải trình tự và so sánh với trình tự dự đoán *in silico*. Protein *NtMKS2-1* và *NtMKS2-2* dự đoán bao gồm 211 amino acid, có tính kiềm với giá trị pI trong khoảng 9.37-9.61, trọng lượng phân tử khoảng 24 kDa và tương đồng với *ShMKS2* hơn 60%. Kết quả nghiên cứu này góp thêm dữ liệu để hỗ trợ nghiên cứu sự tiến hóa của hệ gene *MKS2* ở các loài thực vật thuộc họ *Solanaceae* và cung cấp nguồn gene cho các nghiên cứu cải biến con đường biến dưỡng ở thực vật và vi sinh vật nhằm mở rộng tiềm năng ứng dụng của các hợp chất 2-methylketone.

Từ khóa: *Nicotiana tabacum*, methylketone synthase 2 (MKS2), *NtMKS2-1*, *NtMKS2-2*.

1. Mở đầu

2-Methylketone là nhóm hợp chất hữu cơ, mang nhóm chức ketone ở vị trí nguyên tử carbon thứ hai, bao gồm một số hợp chất dễ bay

hoi có nguồn gốc từ acid béo như 2-heptanone, 2-nonanone, 2-undecanone, 2-tridecanone, 2-pentadecanone. Nhóm hợp chất này góp phần tạo nên mùi hương đặc trưng của nhiều tinh dầu thực vật. Vì vậy, methylketone từ lâu đã được sử dụng làm chất tạo hương trong công nghệ chế biến thực phẩm, nhất là những sản phẩm có nguồn gốc từ bơ, sữa. Tiềm năng ứng dụng của

* Tác giả liên hệ. ĐT.: 84-933698827.

Email: ttdhuong93@gmail.com

<https://doi.org/10.25073/2588-1140/vnunst.4624>

các hợp chất này khá đa dạng, phụ thuộc vào chiều dài chuỗi carbon, mức độ bão hòa và trạng thái oxy hóa - khử trong phân tử. Sự hiện diện của 2-methylketone ở thực vật đã được ghi nhận lần đầu tiên ở cây Cửu lý hương (*Ruta graveolens*) vào năm 1858, tuy nhiên cho đến gần đây gene mã hóa enzyme tham gia trong con đường sinh tổng hợp các hợp chất này mới được xác định. Cụ thể, hai gene mã hóa enzyme methylketone synthase 1 (ShMKS1) và methylketone synthase 2 (ShMKS2) đã được phân lập từ loài cà chua dại (*Solanum habrochaites*) vốn dĩ tổng hợp và tích lũy rất nhiều methylketone trong các lông tiết hiện diện trên bề mặt lá và thân của cây. ShMKS2 có hoạt tính thioesterase và có thể xúc tác thủy phân liên kết 3-ketoacyl-ACP, một hợp chất trung gian của quá trình sinh tổng hợp acid béo, thành các 3-ketoacid. ShMKS1 có hoạt tính decarboxylase xúc tác phản ứng loại nhóm carboxyl của các 3-ketoacid và sản phẩm sinh ra là các 2-methylketone có chiều dài khung carbon ngắn hơn một nguyên tử C so với các 3-ketoacid ban đầu [1].

Gene mã hóa protein tương đồng với ShMKS2 cũng đã được tìm thấy ở một số loài thực vật khác như *Solanum lycopersicum*, *Arabidopsis thaliana*... Các nghiên cứu tiếp theo đã chỉ ra rằng các loài thực vật khác nhau trong họ Solanaceae có thể có số lượng gene MKS2 khác nhau. Cà chua *Solanum lycopersicum* có ba gene [1], trong khi ớt *Capsicum annum* chỉ có một gene mã hóa protein tương đồng với ShMKS2 [dữ liệu nghiên cứu của nhóm]. Khi được biểu hiện tái tổ hợp trong *E. coli*, các gene mã hóa các protein enzyme MKS2 xúc tác tổng hợp các methylketone có chiều dài chuỗi carbon khác nhau. Thuốc lá (*Nicotiana tabacum*) cũng là một loài thực vật họ Solanaceae và được sử dụng phổ biến làm cây mô hình trong các thí nghiệm nghiên cứu trên thực vật bậc cao, tuy nhiên, thông tin về số lượng và chức năng của các gene MKS2 ở loài này vẫn chưa được tìm hiểu.

Hiện nay bản đồ bộ gene của *N. tabacum* đã được giải mã và công bố dưới dạng nhiều phân đoạn contig tách rời [2]. Việc tìm hiểu sự hiện

diện cùng chức năng của các gene MKS2 từ cây thuốc lá (*N. Tabacum*) sẽ cung cấp dữ liệu bổ sung, hỗ trợ cho nghiên cứu phân tích sự tiến hóa chức năng của hệ gene MKS2 ở thực vật thuộc họ Solanaceae. Trong đề tài này, chúng tôi tiến hành phân lập (tách dòng) gene MKS2 từ cây thuốc lá (*N. Tabacum*) với sự trợ giúp của các công cụ tin sinh học kết hợp kỹ thuật sinh học phân tử. Nguồn gene phân lập được có thể được sử dụng trong các nghiên cứu cải biến biểu dưỡng trên thực vật và vi sinh vật, nhằm mở rộng tiềm năng ứng dụng đa dạng của nhóm hợp chất 2-methylketone.

2. Vật liệu và phương pháp nghiên cứu

2.1. Vật liệu

Cây thuốc lá (*N. Tabacum*) dòng K326 được trồng từ hạt giống do công ty Profigen, Brazil cung cấp. Các cây này được trồng tại nhà màng Bộ môn sinh hóa, trường Đại học Khoa học Tự nhiên TP. Hồ Chí Minh.

2.2. Phương pháp

Dự đoán các trình tự gene MKS2 mới ở cây thuốc lá (N. Tabacum) bằng các công cụ tin sinh học

Các contig chứa trình tự mã hóa protein tương đồng cao với ShMKS2 được xác định thông qua tra cứu cơ sở dữ liệu bộ gene của *N. tabacum* bằng phần mềm TBLASTN với trình tự truy vấn là ShMKS2 (mã số Genbank GU987106). Cấu trúc của các gene *NtMKS2* mới được dự đoán bằng phần mềm FGENESH (www.softberry.com) và được hiệu chỉnh lại theo kết quả đối sánh trình tự genomic và với các trình tự CDS mã hóa MKS2 đã được công bố ở họ Cà (Solanaceae) [3].

Phân lập các trình tự CDS NtMKS2

RNA tổng số được tách chiết từ mô lá non cây thuốc lá *N. tabacum* bằng EZ-10 Spin Column Plant RNA Mini-Prep Kit (BioBasic). cDNA được tổng hợp từ RNA tổng số đã loại DNA bộ gene theo hướng dẫn của bộ kit

RevertAid First Strand cDNA Synthesis (Thermo Scientific). Mẫu cDNA này được sử dụng làm mạch khuôn để nhân bản trình tự mã hóa (CDS) MKS2 thông qua phản ứng PCR với các cặp mồi đặc hiệu được thiết kế dựa trên trình tự CDS dự đoán. Trình tự các cặp mồi đặc hiệu được thiết kế: NtMKS2-1-F (mồi xuôi 1): 5'-CTC GAG TAT GTC GCA AAC CCT AGT TTC C-3'; NtMKS2-1-R (mồi ngược 1): 5'-GGA TCC TTT TAG ATG CCT CCT GGC GA-3' và NtMKS2-2 (mồi xuôi 2): 5'-CTC GAG TAT GTC GCA AAC CCT AGT-3'; NtMKS2-2-R (mồi ngược 2): 5'-GGA TCC TTT TAG ATG CCT CCT GG-3'. Phản ứng PCR sử dụng Phusion DNA polymerase (Thermo Scientific) được thực hiện theo chu kỳ nhiệt: 98°C/30 giây; lặp lại 30 chu kỳ, mỗi chu kỳ: 98°C/10 giây -58°C/20 giây (cặp mồi NtMKS2-1-F/R) hoặc 52°C/20 giây (cặp mồi NtMKS2-2-F/R) -72°C/20 giây; 72°C trong 5 phút và giữ 4°C trong 15 phút.

Sản phẩm nhân bản được kiểm tra bằng điện di trên gel agarose 1%, sản phẩm mục tiêu được tinh sạch từ gel theo hướng dẫn của bộ kit GeneJET Gel Extraction và được chèn vào vector pJET1.2/blunt thông qua phản ứng nối. Hỗn hợp phản ứng nối được biến nạp vào tế bào khả nạp *E. coli* TOP10 bằng phương pháp hóa biến nạp. Các thể biến nạp *E. coli* TOP10/pJET-NtMKS2-1 được sàng lọc bằng phản ứng PCR khuôn lạc với mồi xuôi NtMKS2-1-F và mồi ngược pJET1.2-R (5'-AAG AAC ATC GAT TTT CCA TGG CAG-3'). Các thể biến nạp *E. coli* TOP10/pJET-NtMKS2-2 được sàng lọc bằng phản ứng PCR khuôn lạc với mồi xuôi pJET1.2-F (5'-CGA CTC ACT ATA GGG AGA GCG GC-3') và mồi ngược NtMKS2-2-R bằng *Taq* polymerase với chu kỳ nhiệt: 95°C/30 giây; lặp lại 30 chu kỳ, mỗi chu kỳ: 95°C/50 giây, 58°C/20 giây, 72°C/40 giây; 72°C trong 5 phút; 4°C trong 15 phút. Kết quả điện di sản phẩm PCR trên gel agarose 1% sẽ cho biết dòng tế bào *E. coli* có mang plasmid chứa gene mục tiêu hay không. Plasmid tái tổ hợp được tách chiết bằng bộ kit GeneJET Plasmid Miniprep,

được kiểm tra kích thước bằng phản ứng cắt với enzyme cắt giới hạn *XhoI* và *BamHI* và sau đó, được giải trình tự để đối sánh với trình tự gene đã dự đoán dựa trên phân tích tin-sinh học.

Xây dựng cây phát sinh loài

Sử dụng phần mềm sắp giống cột nhiều trình tự CLUSTAL W để so sánh trình tự protein của NtMKS2-1, NtMKS2-2 phân lập từ cây thuốc lá (*N. Tabacum*) với các trình tự MKS2 đã phân lập hoặc dự đoán từ một số loài thực vật trong họ Solanaceae như cà chua hoang dại (*S. Habrochaites*), cà chua thuần hóa (*S. Lycopersicum*), cà chua (*S. Pimpinellifolium*), cà tím *S. melongena* và ớt *Capsicum annuum* nhằm xác định mức độ tương đồng giữa các protein này và xây dựng cây phát sinh loài dựa trên kết quả so sánh trình tự protein bằng phần mềm MEGA6 [4].

3. Kết quả nghiên cứu

3.1. Kết quả dự đoán các trình tự contig chứa gene mã hóa chuỗi polypeptide có độ tương đồng cao với ShMKS2

Khi tiến hành TBLASTN cơ sở dữ liệu “*N. tabacum* Genome (Current version)” trên SNG với trình tự truy vấn là ShMKS2, chúng tôi tìm được hai contig có chứa gene mã hóa các protein có độ tương đồng cao với ShMKS2, đó là Ntab-K326 AWOJ-SS748 và Ntab-K326 AWOJ-SS6425 (Bảng 1).

Bảng 1. Các contig chứa gene mã hóa cho protein có trình tự tương đồng cao với ShMKS2

STT	Trình tự contig	Giá trị bit-score	Giá trị E
1	Ntab-K326	94.7	1e-19
	AWOJ-SS748		
2	Ntab-K326	90.9	2e-18
	AWOJ-SS6425		

3.2. Kết quả phân lập gene NtMKS2-1 và NtMKS2-2 in silico

Contig AWOJ-SS748, có kích thước 905,935 bp, chứa những đoạn nucleotide giống

cột cùng chiều với trình tự gene mã hóa cho protein ShMKS2. Từ kết quả phân tích dựa trên các phần mềm hỗ trợ dự đoán gene (FGENESH, FSPICE, CLUSTAL W) và kết quả hiệu chỉnh thủ công các vị trí cắt-nối (splicing sites) dựa trên thông tin sắp giống cột giữa trình tự contig trên và trình tự cDNA tương ứng của *ShMKS2*, chúng tôi dự đoán vị trí codon khởi đầu và kết thúc của gene mã hóa chuỗi polypeptide tương đồng với *ShMKS2* trên contig và đặt tên là *NtMKS2-1*.

Theo đó, gene *NtMKS2-1* gồm 12,160 bp, trong đó có 5 exon và 4 intron. Trình tự CDS mã hóa protein NtMKS2-1 hoàn chỉnh dài 636 bp và mã hóa protein chứa 211 amino acid, tương đồng 63% so với trình tự ShMKS2 hoàn chỉnh.

Tương tự, contig AWOJ-SS6425 (1,374,146 bp) cũng chứa một khung đọc mở (ORF) tương ứng với trình tự mã hóa cho một protein dài 211 amino acid và tương đồng 62% với ShMKS2 và tương đồng 91 % so với trình tự CDS *NtMKS2-1* dự đoán ở trên. Gene thứ hai được đặt tên là *NtMKS2-2* và có kích thước 9,217 bp với 5 exon và 4 intron.

3.3. Phân lập trình tự mã hóa *NtMKS2-1* và *NtMKS2-2* từ cây thuốc lá (*N. Tabacum*)

Kết quả điện di ở hình 1 cho thấy sản phẩm PCR nhân bản CDS *NtMKS2-1* và *NtMKS2-2* cho băng DNA có kích thước xấp xỉ 650 bp khi so sánh với thang DNA chuẩn, phù hợp với kích thước dự đoán của hai đoạn trình tự mã hóa *NtMKS2-1* và *NtMKS2-2* (Hình 1, giếng N1 và N2 tương ứng).

Sản phẩm PCR được tinh sạch từ gel agarose và được nối vào vector pJET1.2/blunt. Hỗn hợp phản ứng nối được biến nạp vào tế bào khả nạp *E. coli* TOP10. Kết quả sàng lọc các dòng tế bào *E. coli* TOP10 mang plasmid pJET-NtMKS2-1 hoặc pJET-NtMKS2-2 bằng phản ứng PCR khuẩn lạc với môi pJET1.2-F bắt cặp đặc hiệu với trình tự vector và môi ngược bắt cặp đặc hiệu với trình tự gene mục tiêu (Hình 2) cho các băng DNA có kích thước khoảng 700 bp, tương ứng với kích thước dự

đoán của sản phẩm nhân bản (Hình 2, giếng N1 và N2). Trong khi đó, đối chứng âm không xuất hiện vạch có cùng kích thước như trên (Hình 2, giếng 1). Các khuẩn lạc cho kết quả PCR dương tính được chọn nuôi nhân sinh khối để tách chiết plasmid. Plasmid tách chiết được sơ bộ kiểm tra bằng phản ứng cắt với tổ hợp enzyme cắt giới hạn *XhoI* và *BamHI* và được gửi giải trình tự.

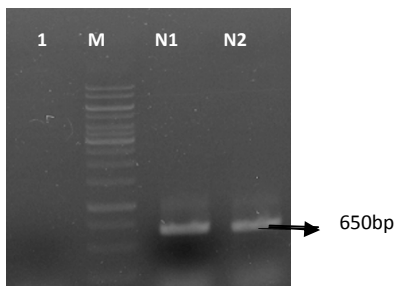
Kết quả sắp giống cột nhiều trình tự (multiple sequence alignment) cho thấy CDS *NtMKS2-1* và *NtMKS2-2* có trình tự giống 100% so với trình tự *NtMKS2-1* và *NtMKS2-2* đã dự đoán bằng phân tích tin sinh học.

3.4. Phân tích trình tự mã hóa *NtMKS2-1* và *NtMKS2-2* được phân lập từ *N. tabacum*

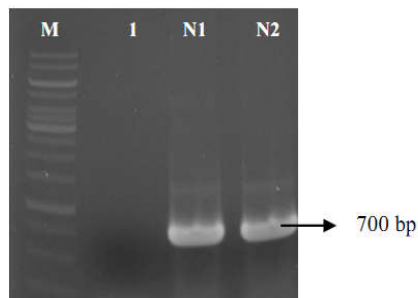
Kết quả sắp giống cột 11 trình tự protein MKS2 hiện diện ở 6 loài thực vật họ *Solanaceae* (bằng phần mềm CLUSTAL W) được mô tả ở hình 3. Theo đó, NtMKS2-1 và NtMKS2-2 đều chứa amino acid aspartate bảo tồn (được đóng khung - hình 3) cần thiết cho hoạt tính xúc tác của các enzyme MKS2 thuộc họ thioesterase có vùng gấp cuộn kiểu "hotdog". Cây phát sinh loài dựa trên kết quả so sánh các trình tự MKS2 được mô tả ở hình 4.

4. Kết luận

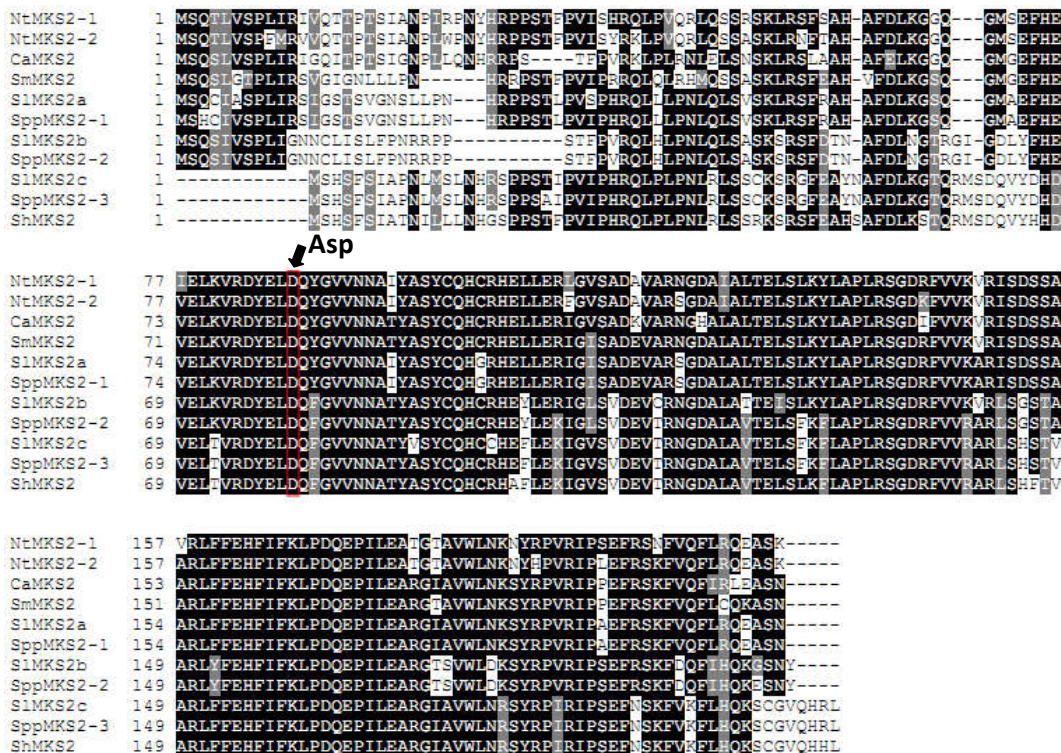
Với sự hỗ trợ của các công cụ tin sinh học và sinh học phân tử, chúng tôi đã phân lập được hai trình tự mã hóa protein NtMKS2-1 và NtMKS2-2 tương đồng 63% và 62% với ShMKS2. Cả hai protein NtMKS2-1 và NtMKS2-2 đều có chứa amino acid aspartate bảo tồn cần thiết cho hoạt tính xúc tác của các enzyme MKS2 thuộc họ thioesterase có vùng gấp cuộn kiểu "hotdog". Kết quả nghiên cứu cung cấp dữ liệu bổ sung hỗ trợ nghiên cứu sự tiến hóa của hệ gene MKS2 ở các loài thực vật thuộc họ *Solanaceae* và cung cấp nguồn gene cho các nghiên cứu cải biến con đường biến dưỡng ở thực vật và vi sinh vật nhằm mở rộng tiềm năng ứng dụng của các hợp chất 2-methylketone.



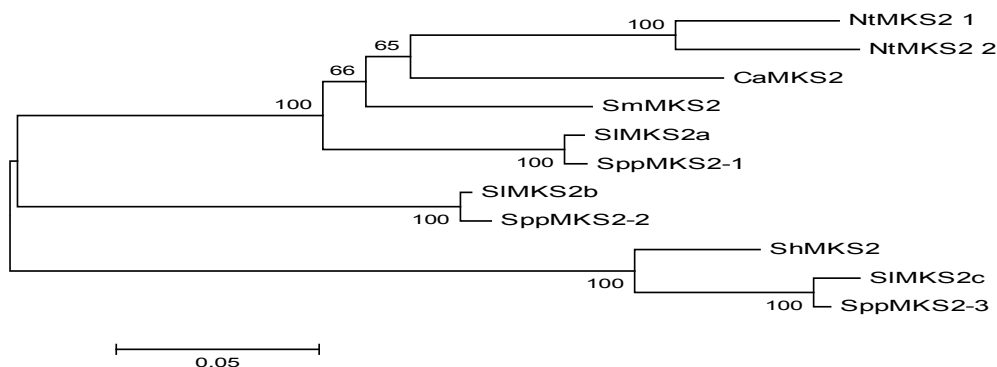
Hình 1. Kết quả điện di sản phẩm PCR nhân bản *NtMKS2-1* và *NtMKS2-2*.
Giếng M: Thang chuẩn DNA 1kb
Giếng 1: Đối chứng âm
Giếng N1, N2: Sản phẩm PCR nhân bản *NtMKS2-1* và *NtMKS2-2*



Hình 2. Kết quả sàng lọc dòng tế bào *E. coli* TOP10 mang plasmid tái tổ hợp bằng PCR khuẩn lạc.
Giếng M: Thang chuẩn DNA 1kb
Giếng 1: Đối chứng âm
Giếng N1, N2: Sản phẩm PCR khuẩn lạc tương ứng với dòng vi khuẩn mang plasmid tái tổ hợp pJET1.2/blunt-*NtMKS2-1* và pJET1.2/blunt-*NtMKS2-2*



Hình 3. So sánh trình tự protein của NtMKS2-1 và NtMKS2-2 từ cây thuốc lá *N. tabacum* với các trình tự tương đồng từ cà chua dại *S. habrochaites* (ShMKS2), cà chua thuần hóa *S. lycopersicum* (SIMKS2a, SIMKS2b, SIMKS2c), cà chua *S. pimpinellifolium* (SppMKS2-1, SppMKS2-2, SppMKS2-3), và ớt *Capsicum annuum* (CaMKS2).



Hình 4. Cây phát sinh loài dựa trên so sánh các trình tự protein của các gene *MKS2* từ *N. tabacum*, *S. habrochaites*, *S. lycopersicum*, *S. pimpinellifolium*, *S. melongenea* và *C. annuum*.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Quỹ phát triển khoa học và công nghệ quốc gia (NAFOSTED) trong đề tài mã số 106-NN.02-2013.35.

Tài liệu tham khảo

- [1] Yu G., Nguyen T. T. H., Guo Y., Schaubinhold I., Auldridge M. E., Bhuiyan N., Ben-Israel I., Iijima Y., Fridman E., Noel J., Pichersky E., Enzymatic functions of wild tomato methylketone synthase 1 and 2. *Plant. Physiol.*, 154 (2010) 67-77.
- [2] Sierro, N. et al, The tobacco genome sequence and its comparison with those of tomato and potato. *Nature Communications* 5 (2014).
- [3] Mai Huỳnh Hạnh Phúc, Đinh Minh Hiệp, Nguyễn Thị Hồng Thương, Sử dụng các công cụ tin sinh học để xác định các gen Methylketone Synthase 2 (MKS2) mới từ loài cà chua *Solanum pimpinellifolium*, *Tạp chí Sinh học* 36 (1se) (2014), 237-243.
- [4] Koichiro Tamura et al, MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, *Mol. Biol. Evol.* 30 (12), (2013) 2725-2729.

Cloning and Sequencing of Two Genes Encoding Methylketone Synthase 2 (MKS2) from the Tobacco (*Nicotiana tabacum*)

Truong Tran Dieu, Tran Thi Diem Huong,
Ho Tien Giang Em, Nguyen Thi Hong Thuong

Faculty of Biology and Biotechnology, VNUHCM - University of Science

Abstract: 2-Methylketones are accumulated in some plants as a byproduct of the fatty acid biosynthesis that takes place in plastids. Methylketone synthase 2 (MKS2) is a thioesterase that catalyzes the penultimate reaction in the methylketone biosynthetic pathway. It converts 3-ketoacyl-ACPs into 3-ketoacids which are precursors for the synthesis of 2-methylketones. Previous studies have shown that species in the Solanaceae family might have different number of MKS2 genes. Besides, each MKS2 could hydrolyze a specific group of endogenous fatty acyl-acyl carrier protein substrates that

varies in chain length (C6-C18), degree of saturation and oxidation state. In this study, we had identified two homologous genes of *ShMKS2*, designated as *NtMKS2-1* and *NtMKS2-2*, located at two contigs AWOJ-SS748 and AWOJ-SS6425 in the *Nicotiana tabacum* genome database. Both genes comprise of five exons and four introns. The full-length cDNA sequences encoding *NtMKS2-1* and *NtMKS2-2* have been successfully isolated from young leaf tissues of tobacco *N. tabacum*, sequenced and aligned with the corresponding sequences predicted by *in silico* analysis. Both of the deduced amino acid sequences encode proteins of 24 kDa that share more than 60% identity to *ShMKS2* and contain a conserved aspartate residue essential to the catalytic core of the hotdog-fold thioesterases. This study provided additional data to gain insights into the evolution of the *MKS2* genes in species from the Solanaceae family.

Keywords: *Nicotiana tabacum*, methylketone synthase 2 (MKS2), NtMKS2-1, NtMKS2-2.