

SỬ DỤNG CÁC CÔNG CỤ TIN-SINH HỌC ĐỂ XÁC ĐỊNH CÁC GEN METHYLKETONE SYNTHASE 2 (MKS2) MỚI TỪ LOÀI CÀ CHUA *Solanum pimpinellifolium*

Mai Huỳnh Hạnh Phúc¹, Đinh Minh Hiệp², Nguyễn Thị Hồng Thương^{1*}

¹Trường Đại học Khoa học tự nhiên, ĐHQG tp. Hồ Chí Minh, *thuongnth@yahoo.com

²Ban quản lý Khu nông nghiệp công nghệ cao tp. Hồ Chí Minh

TÓM TẮT: 2-Methylketone là nhóm hợp chất hữu cơ mang nhóm chức ketone ở vị trí carbon thứ hai, nhóm hợp chất này có nhiều ứng dụng rộng rãi trong bảo vệ thực vật, trong công nghiệp tạo hương và trong sản xuất nhiên liệu sinh học. Gần đây, hai cDNA mã hóa cho hai protein tham gia trong sự sinh tổng hợp methylketone ở cà chua hoang dã *Solanum habrochaites* subsp. *glabratum* đã được xác định và được ký hiệu là methylketone synthase 1 (ShMKS1) và methylketone synthase 2 (ShMKS2). Để xác định các MKS2 mới có khả năng sử dụng cơ chất 3-ketoacyl-acyl carrier protein (3-ketoacyl-ACP, chất trung gian của quá trình sinh tổng hợp acid béo xảy ra trong lục lạp) khác nhau về độ dài và mức độ không bão hòa của khung carbon, chúng tôi sử dụng công cụ tìm kiếm TBLASTN với ShMKS2 là trình tự mẫu để truy vấn cơ sở dữ liệu bộ gen của loài cà chua *Solanum pimpinellifolium*. Với sự hỗ trợ của các công cụ dự đoán gen khác, chúng tôi đã xác định được ba gen tương đồng với ShMKS2 trên bốn contig (contig 3697822, 6568413, 6704221 và 6708991) hiện diện trong cơ sở dữ liệu bộ gen của *S. pimpinellifolium*. Ba gen này mã hóa cho các protein có trình tự tương đồng với trình tự protein ShMKS2 hơn 65% và tương đồng với trình tự của các protein SIMKS2 hơn 98%; chúng tôi ký hiệu ba gen mới này là *SppMKS2-1*, *SppMKS2-2* và *SppMKS2-3*. Cả ba gen MKS2 này của *S. pimpinellifolium* đều có năm exon và bốn intron (các vị trí của chúng được bảo tồn khi so sánh với các vị trí exon và intron trong các gen tương đồng ở *S. lycopersicum*). Chúng tôi tiến hành so sánh trình tự protein của các MKS2 tương đồng hiện diện ở ba loài thuộc chi *Solanum* này và dựa vào đó xây dựng cây phát sinh loài.

Từ khóa: cà chua, gen mã hóa protein, gen tương đồng, trình tự protein.

MỞ ĐẦU

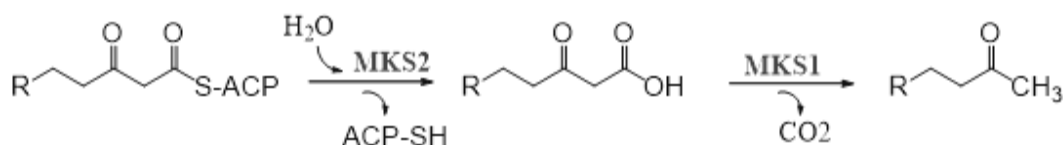
2-Methylketone (methylketone) là nhóm hợp chất hữu cơ mang nhóm chức năng ketone ở nguyên tử carbon thứ hai. Methylketone từ lâu đã được biết đến là nhóm hợp chất tạo hương quan trọng trong công nghiệp sản xuất phô mai và các sản phẩm từ sữa [5]. Antonious et al. (2003, 2004) [1, 2] cho thấy, methylketone không gây độc trên người và gia súc, và là nhóm hợp chất có tính kháng sâu. Gần đây, theo nghiên cứu của Goh et al. (2012) [4], methylketone bắt đầu thu hút sự chú ý của các nhà nghiên cứu năng lượng sinh học vì chúng có trị số kích nổ cetan cao, hứa hẹn sẽ là lựa chọn mới trong sản xuất nguồn năng lượng có thể tái sinh. Mới đây, Yu et al. (2010) [7] đã tìm thấy ở cây cà chua hoang dại *S. habrochaites* hai enzyme mới gồm methylketone synthase 2 (ShMKS2) và methylketone synthase 1 (ShMKS1) tham gia trong chuyển hóa chất trung gian của con đường sinh tổng hợp acid

béo là 3-ketoacyl-ACP (còn gọi là β -ketoacyl-ACP) thành methylketone (hình 1). ShMKS2 hoạt động như một enzyme thioesterase, xúc tác sự thủy phân liên kết thioester của 3-ketoacyl-ACP và hoạt động hiệu quả nhất trên cơ chất 3-ketomyristoyl-ACP (14C) và 3-ketolauroyl-ACP (12C), tạo thành hai 3-ketoacid tương ứng là 3-ketomyristic acid (14C) và 3-ketolauric acid (12C), sau đó ShMKS1 xúc tác sự decarboxyl hóa các 3-ketoacid vừa được phóng thích, tạo ra sản phẩm methylketone. Các gen mã hóa cho protein có độ tương đồng cao với ShMKS2 hiện diện trong nhiều loài thực vật [3]. Loài cà chua thuần hóa *S. lycopersicum* có ba gen mã hóa cho các protein tương đồng với ShMKS2, được ký hiệu là *SIMKS2a*, *SIMKS2b* và *SIMKS2c*. Khi được biểu hiện tái tổ hợp trong vi khuẩn *E. coli*, ShMKS2 chủ yếu tổng hợp 2-tridecanone (C13) trong khi *SIMKS2a* chủ yếu tổng hợp 2-undecanone (C11) trong môi trường nuôi cấy [3, 7].

Để khai thác tiềm năng ứng dụng rộng rãi của nhóm hợp chất methylketone như đã giới thiệu ở trên, việc xây dựng bộ sưu tập các gen mã hóa cho enzyme tham gia trong sự tổng hợp các methylketone từ nhiều loài thực vật khác nhau (trong đó mỗi enzyme có khả năng sử dụng hiệu quả nhất một cơ chất 3-ketoacyl-ACP khác nhau về độ dài và mức độ không bão hòa của khung carbon) là bước đầu tiên cần thực hiện. Trình tự bộ gen của loài cà chua *S. pimpinellifolium* đã được giải mã và được

công bố dưới dạng các phân đoạn contig tách rời.

Dựa vào các trình tự protein MKS2 đã biết từ *S. habrochaites* và *S. lycopersicum*, kết hợp với các công cụ hỗ trợ trong tin-sinh học, chúng tôi tiến hành xác định trình tự các gen mã hóa cho protein MKS2 mới từ loài cà chua *S. pimpinellifolium*, so sánh trình tự protein của các MKS2 tương đồng hiện diện ở ba loài thuộc chi *Solanum* này và dựa vào đó xây dựng cây phát sinh loài.



Hình 1. Sự tổng hợp methylketone [7]

VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Tìm kiếm các trình tự gen mới bằng các công cụ tin-sinh học

Sử dụng trình tự protein ShMKS2 đã được Yu et al. (2010) [7] công bố trước đây làm trình tự mẫu và sử dụng công cụ tìm kiếm TBLASTN để tìm trong cơ sở dữ liệu bộ gen của loài cà chua *Solanum pimpinellifolium* những phân đoạn contig mang gen mã hóa cho protein có trình tự tương đồng cao với protein ShMKS2.

Sử dụng công cụ dự đoán cấu trúc gen FGENESH (www.softberry.com) để dự đoán sơ bộ cấu trúc của gen *SppMKS2* hiện diện trong mỗi contig được tìm thấy. Cấu trúc gen *SppMKS2* hiện diện trong mỗi contig được kiểm tra lại một cách thủ công bằng cách đối chiếu với các trình tự gen và cDNA mã hóa cho các gen *MKS2* đã biết ở cà chua hoang dại

S. habrochaites và cà chua thuần hóa *S. lycopersicum*.

Xây dựng cây phát sinh loài

Sử dụng phần mềm sắp giống cột nhiều trình tự CLUSTAL 2.1 để so sánh trình tự protein của các *SppMKS2* mới từ loài *S. pimpinellifolium* với các trình tự *MKS2* đã biết từ cà chua hoang dại *S. habrochaites* và cà chua thuần hóa *S. lycopersicum* nhằm xác định mức độ tương đồng giữa các protein này và xây dựng cây phát sinh loài dựa trên so sánh trình tự các protein *MKS2* của các loài thuộc chi *Solanum*.

KẾT QUẢ VÀ THẢO LUẬN

Các contig chứa gen mã hóa cho protein có trình tự tương đồng cao với ShMKS2 được tìm thấy trong cơ sở dữ liệu bộ gen cà chua *Solanum pimpinellifolium*

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

Các trình tự cho kết quả giống cột ý nghĩa	Giá trị bit-score	Giá trị E
contig:unspecified:6704221:1:1720:1 contig 6704221	69,3	1e-24
contig:unspecified:3697822:1:12874:1 contig 3697822	70,1	5e-22
contig:unspecified:6708991:1:1385:1 contig 6708991	100	1e-20
contig:unspecified:6568413:1:11515:1 contig 6568413	95,9	3e-19

Chúng tôi sử dụng trình tự ShMKS2 đã công bố làm trình tự mẫu và sử dụng công cụ tìm kiếm TBLASTN để tìm trong cơ sở dữ liệu bộ gen của loài cà chua *Solanum pimpinellifolium* (http://solgenomics.net/tools/blast/index.pl?db_id=114) những phân đoạn contig chứa gen mã hóa cho protein có trình tự tương đồng cao với ShMKS2. Kết quả tìm kiếm cho ra 4 contig được trình bày như trong bảng 1.

Kết quả phân lập gen *SppMKS2-1* trên contig 3697822

Contig 3697822 chứa những đoạn nucleotide giống cột (align) ngược chiều với trình tự nucleotide mã hóa cho ShMKS2 nên được chuyển đổi sang trình tự bổ sung bằng công cụ COMPLEMENTARY SEQUENCE để việc xác định cấu trúc gen dễ dàng hơn (http://www.bioinformatics.org/sms/rev_comp.h

tml). Kết hợp công cụ dự đoán cấu trúc gen FGENESH, công cụ xác định các vị trí nối FSPLICE, phần mềm sắp giống cột nhiều trình tự CLUSTAL 2.1 và sự điều chỉnh thủ công, chúng tôi tìm thấy trên contig 3697822 một gen chứa một khung đọc mở (ORF) mã hóa cho một protein dài 208 amino acid có trình tự tương đồng với trình tự protein ShMKS2 (bao gồm cả trình tự chuyển tiếp) 68,97% và tương đồng với trình tự của protein SIMKS2a 98,56%. Gen này được ký hiệu là *SppMKS2-1* (hình 2). Sự điều chỉnh thủ công chủ yếu bao gồm sự điều chỉnh lại một số vị trí nối đã được dự đoán chưa chính xác bởi công cụ FSPLICE, dựa trên so sánh đối chiếu các trình tự exon và intron tương ứng của *SppMKS2-1* với các trình tự bộ gen và cDNA mã hóa cho các gen MKS2 đã biết ở cà chua *S. lycopersicum*.

ATGTCTCACTGCATCGTTTCCCGTTGATTTCGCAGCATTGGATCCACTTCAGTCGGTAACTCACTGTTGCCGAATCATCCG
 CCACCGTCTACATTACCGGTCACTCCCTACCGGCAGCTCCCTGCTTCCAAATTTACAGTTATCCGTCAGTAAATTGAGGAGT
 TTTTCGAGCTCATGCTTTTGGATCTCAAAGGTAGCCAAGGGTATGTTTATATATATATATCTTTTACTCCATCAATCCCATT
 TCTGAAGTATTTGATTAGGCGCGGAGTTTATGGATAAAAAGGAAGACCTTTAAAAATTTGTGATATAAAACAACCGATATATA
 TATGTGTATGTATGGATATATTTGATTTGTTATAAATCATCTAATGAAATGAAAAAGTAAAAAGTGTATTAATAATAGAAA
 TGTGATATGGTTAAGTAAGAAGTTTGAAGTCAAAGTGTACCCTGTGATAAAGGTGTCTCAGGTGATCTTTGTAAGTGGAA
 AAATTGAAGTTAAATTTGTTATGGAATATAGAAAGGTGTCTCAAGGTGATCTCGTAAAAATGGGAAGTTGAAATCAAGTTGT
 TATCGAATATAGAAAGGTTTCTCAGGGTGTCTCGTAAAAATTTGGAATTTTGAAGTCAAATTTGTTACTGAATATAGGAAGGT
 GTCATGGGTAGTAACTTACAGTTCCATTCAAAAATTCATCCTGTATGACAAAACATAGTCCGGATCATGCTTTGGATGACCG
 ATGAGGGTTGTCTAGGTTGTCAATGAGGGTAAAGTAAAGTCAATTATGATCAGATACTCTTTAAGTATTTGATTCATTGGC
 TTGTGTCCACTTGATTTCAACTGAATGGCAGAGGAGTTATGTAGTTTGTGTAAGTACTAGTTGGGCTTTAGATATAGTTGA
 TTGATTGGTTTTGCTGTAGCTTCTGTAGGTTTGAAGTTGATTAGAAGTATGTTTCTCCATCTGAATGAAGGGCTATGC
 ATTTTCAATTTCTACAATTTGGTGGAACTGATTGATTGAATAATGTTTTTTTTTTATCAGAATTTGGAAAAGGTTTTTTTT
 TTGGGAAAGAAAAATGAAAAACCTTTTATTCTTTTGTGTGCGAGCGTTTTATAGGCTTCCCTTTCTGTAGTTTCATTTT
 AAGTTTCAGCAAGAATTTGGTATTTTGTAGTTTGTCTCATTGACATAGTCTATTTTTTCTTATTTATAGGAGCTTACCTTTTGC
 TCTTGTCTTGCAGAAATGGCTGAGTTCCATGAAGTTGAAGTCAAAGTCCGGGACTATGAATTGGATCAGTATGGTGTGTAA
 ACAATGCTATTTATGCAAGTTATTGCCAACATGGTAAGGTTATGGTTTCGATCTGTACTTCAGTTTACAACCTACCATATT
 ATACATGTGCTTTCATTCATCAAAAAGCATATAAATACTGCGCTTTTCCCTTTTAAATGAAAAAGGATTTACTCAAGGGAGAA
 ATTTTTTCTGGCAACTGTTATGAGTAGAAAGCTAGAAAATTACTTTTTTTTTTTTTAAAAAACTGAAGTAAACTAGAAATTAC
 TGAAAAAGGATCTTTTGTATCTGTTCACATCTTTGTAAACCTATAGTTAGATCATCTGTACCCGTTTATGGAATGT
 GTTTCTCTCTCAATAACTTGAGATGATGCCACCAAAAATGGATGATGAATATGATTTCTTTGTCTGCTTATTACTAGAA
 ACATGTTGAATCCCAAGTTTGAAGGGATCTGATGTGGTCAATGACTGTTTGAATCTTGCATTTACACATGCTAACGATAAA
 GCCAATATCCACTTTGTATGTGAACATAATTGATTGCCAAATAGTTGTTTTGCCAGAAGCTCAGAAGTTGCTCAGTTATAAAT
 CAATAAATTTAAGTTAATAATATGTCTATCCTAATGAAAAAGAAAGTTAATAAATTTGTCTATTCAAAAATGTTGTTAAGTAAT
 TGGCACGGTTTACCTGATTACCCGTGATATGGAATCAAGGATATCAAAAATCAAGTCTTCCCAACGTAATAAGATCTG
 TTACATGTGAGTGACCTATGTATACAAGTTGAGTTTTTTTAATAAACCAATAAAAAGTTTCTGTTTAAATTTCTATAAAT
 TATATCAGATCTTTCTAGTTCTCGACTATTATTGAAGTATACTGACAAGATGATTACTTTAAAGGATTTAAATTAACCTCT
 TTATCTTTGTCAAGATCAATACTTTGAGGGATTTGAAGTCTGCTTGTAAAAAAAGGAATTAACCTAACAGCTGCAAGGTTT
 CTTACGCTAAATTTCAAAAATGGGGCCAGTATACTACTCTTATTACAAAATTTTGGCGTATGAGTTCTACCTATAATAGACA
 AAGTTACTGGTATCTGTAGGTGAAAAAAAAGATCTCTTCTTAAAAAGCTTAGAGTAATGAGAATTTACTTGTTCATAATG
 CTATTATATGATCAGACAATCTGGTGGATTATTTGGAAAGAGAGAACTAGAGATTTTCAAGGCAAAAAGGAGAAATGTAAC
 CAGTCTTGAGAAATGTACTTCTTTGCTTTCTTTTCGGAGCAATGTGGCAAAATGAACATGATTAATGATGTGGAGGCCAT
 GGTTCACCTTATTTAGTCACTACACAGTTAGTGATGACTACCTTTGATGTGTTCTCTTCTTATAACAACCTGATGTTTG
 TTACATTTATAAAAATTTACCTTATCAAAAATAAATAAATTTAGAAATGATGATCAGGACTTTTGCATGAAAAGAACAGTAAA
 AAGAAAAATAATAACAGTTACCCATCCAGTTAAATAGAACTAATTAGAGATAACCCAGTGTCATTTTTCTAGAGGCAAA

CAATAATATTTAGATAAAGCAAGCAAGATTTGGAACTCCAAAAGGTGATAGTTTCTTTAGTTGATTACTTCTGTGTAGA
 TAGAGTTCGAGAAAAGTTTACTTCCGTGTAGTTTCTTTACTGATATTTTCAATTTTTCAATAAGTACCCTTCCAAC
 TCAATTAAGTGAATTTTGTGATGGCATTAGTGTAAAGCAACTATGTCAGCTTTATAGTATTTAAGTGGAAAGTGTAGCA
 AAAGGTGGAGCTAGGTTTAACTTGCATGACTTGAACCCAAATGCGCAGAAAGGTCTGTCTTTTTTCATGATATAGTAA
 AACAAATGATGAGTATAGAGAAAAGAGATTTTTGAAAATAAGCTGACATTTTTCTGATAATCTAGGTTGTTACCTCAAG
 GAAGGTTGTCTTTTTCGATAAGTAGCTAATTTTTATTGCTTCAAAAAACAGCCTAAGCTTGTATTGCATTTGCATGTGTAC
 ATGCTACATAGTGCATTAATACCTCTGCTTCTCAGTACTATCTACTGAAAACTAAGCAATTTATCATATTTCCCTA
 TATCATATACATCATGTCTACAGTAAGAAGAGAAATAAATCATAAATGTAACCTCGTAAATGCTTTCTGATTTGCTCTAAA
 AATTCTTCAATCTTTCTGTCCAAAACACCGACTAGATGCTAACTGGCACTGTGTACATATTTCTGTCTATGCAATCTC
 CTTTGTCTTTCAATGCTGTGTAGGTTCTTCACTATTTTTGGTATAGTCTATTAATAAATCAGTCTGGTGCCTAAAGCTCTT
 GCTATGCACGGGGTCTAGGGAAGGCTGGACCAGAAGGGTCTATTGTATGCGGTCTTACCCTGCATTTTTGGAAGAGGCTGT
 TTCAATGGCTTGTAACTGTGACCTCCAGGTCACATGGCAGTAACTTTCTAGTTATGACAAGGCTCCCTTCTTTGGTA
 TTGGTATAGAATTTTAGTATAGTCTGTGTCATATTAATAATGCTTAGGAGGAACCTCCATAGCTGTGAAGCCATTGAGAAG
 TGTACAACTAGAAAACAGATAATTTGCATCCTTCTCTCTCTTGCAGAGATAATATCTCCAGAAAACATCAATCCCCT
 TCTCTGAAATTTGTGCAAGTTAGGCTAGAAGCATGTGCAATATCCAGATTAACACTTTCTGTGCTTTGGCTTTGTATAAA
 TCTCTCTTTAGCCAAAAGGATTGTGATGACTTCAACCTAAGTTCACTGTGTAGGGTGGTGTCCAAGTTAGAGAATCT
 GGTTCATTTGATTGTTGTTGTCCTGTTCTCGTAACATTTGAGTCACTTTTCCAGCTCCTCATTTACGAGAGGGAAA
 ACAGTCACTAGTTACAACCTGATCAAGAAAAAAGTAGCAGTAGTTGTCATTAATGAAGTGAAGTCTTTTCTCCATATTTT
 TCCCTTTCCCTAAGGAGAAGTTCTATGTTGAATCTTTTGTATTCTGGGATTTTGTCTAGCCTCCTTCTGTACAAGGAC
 GTTACCTTGTGTATATTATCATATACTGGATATGACATTTGTCCATATCAAAAACCTTTCAATGACGACAATTTAACTAAT
 CTTGTAGTTATGACTTATTTTTAATAAATGAACAGGTCGTCATGAGCTTCTAGAAAAGGATTGGTATAAGTGTGATGAAG
TGGCAGCAGTGGTGCAGCCTAGCCTAACAGAGCTGCTAAAGTATCTAGCACCTTAAGGATGACCCCTCATATC
TAAACATCTTAAAGAACCAAGAAATATGCAACCAGAACTTTAGACCTTGGTTAAGTGTCTATTCAATTTGAATTTGTT
 TCACAAAACCTTTGCATTTGAATATGAAGTTTAGATCTTGGGATACATAGAAATGAAGAATAAATGTTAATTGCAAGTGT
 GAGAAGTTTGGATTAGCATAAATAGGAAAGTTAATGTCAATGGATAATGGTTCCGCTAAATGAAGCTTTTTACAGCTGAT
 TATAAATATGTGACACTGCCTTCTTTCAAATTTACTTGGGACACTGTCTTGTATTCTATAAATTACTTGTCTTTTCTCT
 CAGTAAGTATAAGAACTTTACTTTACCATGAATTTGGAGAACTACAACAAAATAAAGATTAGTCTACATTTCCGTTAATCT
 TTATTGACTTGTCTTCAATTTGATTATGCTACAATTAATAAAGCTATTTTATTTAGATATCATCTGGCTCTAAGTTAACA
 ATTTGTTCAACAAACCTTGTGTTCTGTACTATCAGACTCAGTCACTTACTTGGGACGTGAGCTTCTTTCTTCTGAACAGG
 ACTGGTGTGATCTTTATAACTTCAAACCTTGAATGAACTGCTGAAATTTATGTTATCCTGCTGTTCTCATTACTTTTCACT
 CATTGGTTCAGAGTGGAGATAGATTTGTCGTGAAGGCAGAAATATCTGATTCTTCAAGCTGCTGCTTTGTTTTTCGAACT
TCATCTTCAAGCTTCCAGATCAAGAGGTCAGTTACCCTATTTACCGCTTTTTTTTTTTTTTTTGGAAACAAAACCACTTCAT
ATCTCAATGATTCTGTTACTACTTTTTTCCAGCCATCTTGGAGGCAAGAGGAATAGCAGTGTGGCTCAATAAAAGTTAC
CGTCTGTCCGAATCCGGCAGAGTTGATCAAAATTTGTTTCAAGTCTTCCGCAAGGAGGATCCAACTAA

Hình 2. Cấu trúc gen *SppMKS2-1* (các exon được gạch dưới)

Kết quả phân lập gen *SppMKS2-2* trên contig 6704221 và 6708991

Contig 6704221 chứa những đoạn nucleotide giống cột cùng chiều với trình tự gen mã hóa cho protein ShMKS2. Trong khi đó contig 6708991 chứa những đoạn nucleotide giống cột ngược chiều với trình tự gen mã hóa cho ShMKS2 nên được chuyển đổi sang trình tự bổ sung bằng công cụ COMPLEMENTARY

SEQUENCE. Kết hợp cả hai contig, sử dụng các công cụ tin-sinh học (FGENESH, FSPICE, CLUSTAL 2.1) và điều chỉnh lại các vị trí nối, chúng tôi tìm thấy một gen chứa một khung đọc mở (ORF) mã hóa cho một protein dài 204 amino acid và có trình tự tương đồng với trình tự protein ShMKS2 65,20% và tương đồng với trình tự protein SIMKS2b 99,02%. Gen này được ký hiệu là *SppMKS2-2* (hình 3).

ATGTCACAATCCATAGTTTCCCTTTGATTGGCAACAATTTGCCTTATCTCACTGTTTCCGAATCGTCCGCTCCACCATCTACA
TTTCCGGTCAGGCAACTCCATCTTCCAAATTTACAGTTATCAGCCAGTAAATCGCGGAGTTTGGACACTAATGCATTTGAT
CTCAATGGTACAGGAGGATGATATATATATATCTATTACATCCTCTGTCCCAATTCAGATCGCGCAAATATGACAATT
 TTGAAGTCAAATTTACTGAATATAGAAACGTGTCATTAATTTGCTCGTTGACATAGTCGATTATTTATTTGTAACCTTTG
 CAGAATAGGTGACCTATATTTCCATGAAGTTGAACTCAAAGTCAAGGACTATGAATGGATCAATTTGGTGTGTAAACAA
 TGCTACTTATGCAAGTTATTGTCAACATTTGAAGGTTTACTGTTTTGATAATCGATCGTACACAAATTACAATATTTCAA
 TAAATGAACAGGCGTGCATGAATATCTAGAAAAAATTTGGCCTAAGTGTGATGAAGTATGTCGCAATGGTGTGATGATG
CAACAACAGAAAATTTCACTCAAGTATCTAGCACCTCTAAGGGTATGTCGAATTTCACTCCTGTTTATGCTTTCATGATTTTGT
 TATATATACTACTTGTAGGTTTTATTTGTCTTAAATTTCTTATTAGAAAAAGGTTTTGGATTGACTATTTCTTTTCTA
 GTAGCAAAAGGTTTTAGACTCTATAAATAGAGACATGTTCTTCTAACTTAATCANNNNNNNNNNNNNNNNNTCTTAAAGGC
 TTTGAGAGTTTTGGTTAGAGGGAGAAATTTGTGGTCAACAAGCATGATACCTTATCACTTGTGTGAACCTCCCATGATTTCT

GAATGAATGGTTGAGGTTGTTTCTCTCTGTATTTTGTACTATTTATAGTGGATTGCTCATCTCCTTTGTGGACGTAGGTC
 ACGTTAAATCTTTGTGCTTTTGGTATATTTCTCGTTGTCTTCTTACTCGTGATCTTTCGAGGTTTGTCTTGGCTAGCTTCC
 GCGTTTACACCTGCTTATTTTCGGTCCCTAACACTACTTGGCATGTACTTCAAGTCGAATTTGGAGTATTTAAATTTCTGG
 AGATACACAGAGGTGACTTTATTAGTCATATGGGAAAACAGAAGTGTAGTCTTTTATGGCTACAAATGTGAATACAAC
 TACTTAAATTCAGCTATGTTATCATTTCTTTGATCATTGGTTTAGAGTGGAGATAGATTTCGTCGTGAAGGTGAGATTAT
 CCGGCTCTACAGCTGCTCGTTTGTATTTTCGAGCATTTCATCTTCAAGCTTCCAGATCAAGAGGTGAGTTACGTACATCTAA
 TTATCATTTCAATTAAGAGGATAACTTTATAACTAGTGAATCTTAATGATTTTTCTTGAATTTACATACCCTATCT
 TGAAGCAAGAGGAACATCAGTGTGGCTTGATAAAAGCTACCGTCTGTTCGAATTCGGTCAGAGTTTTCAGATCAAAATTTG
 ATCAGTTTATTCATCAGAAGGAATCTAATTACTAA

Hình 3. Cấu trúc gen *SppMKS2-2* (các exon được gạch dưới)

Kết quả phân lập gen *SppMKS2-3* trên contig 6568413

Contig 6568413 (dài 11,515 nucleotide) chứa những đoạn trình tự nucleotide giống cột cùng chiều với trình tự gen mã hóa cho protein ShMKS2. Tương tự, sử dụng công cụ tin-sinh học như trên để phân tích trình tự contig này,

chúng tôi tìm thấy một gen chứa một khung đọc mở (ORF) mã hóa cho một protein dài 208 amino acid và có trình tự tương đồng với trình tự protein ShMKS2 (bao gồm cả trình tự chuyển tiếp) 91,83% và tương đồng với trình tự SIMKS2c 98,56%. Gen này được ký hiệu là *SppMKS2-3* (hình 4).

ATGTCTCATTGTTGAGCATTGCACCCAACTAATGTGCTGAATCATCGGTACCCGCCGCTGCAATTCCGGTCATCCCT
 CACCGGCACTCCCGCTCCCAAATTTACGATTATCGTCTGTAATCGAGGGGTTTGAAGCTTATAATGCGTTGCATCTC
 AAAGGTACCCAACGGTACGTGTGTGTATATATATATATATATTACTCTCTCTGTTTAGTGGCGGTACACAGAATTTTT
 GTTACCTTTTAAAAAAGTAAACAATAAATAAACAATGTAAACATAATATAAATAAAGAAACAAAATCTCTTGTAAATTT
 ATTTTTTTTTTCTATTGGTATGTGATTTTGCAGAAATGAGTGATCAGGTCTATGACCATGACGTTGAATCACAGTCAGGGA
 CTATGAGTTGGATCAGTTTGGTGTGTAATAATGCTACGTATGCAAGTTATTGTCAACATGTAAGTTTACTGTTTCGA
 TAATTGATCGTACACAAATACAATATTTGACTTATTTTTCAATAAATGAAATAGTTCGTCATGAGTTTCTAGAAAAAAT
 GGTGTTAGTGTGATGAAGTAACCGGAAATGGTGACGCATTAGCAGTAACAGAGCTCTCATTTAAGTTTCTTGACCACTA
 AGGGTATGATGACTTTTCGTCGCCGTTTATGTTTCATGTATTTGTTAAGTTCTGTTATACCTTAGTCGAATTTGGAGATTTA
 AAAAATTTGGAGATCCAACCTCAAATGCCTGATATAATATGTTTTGTTTTCAGAGTGGAGATAGATTCTGTTGGTGGGCGCG
 ATTATCCCACTCTACAGTAGCTCGATTGTTTTTCGAGCATTTCATCTTCAAGCTTCCAGATCAAGAGGTTAGTTACCTCTA
 TTATCATACAAATTAAGAGTCACTTTATACTTGTCAAATCTTACTGTATTTTTCTTAAATTTTCACAGCCTATATTGGAG
 GCAAGAGGAATAGCAGTGTGGCTCAATAGAAGTTACCGTCTTATTTCGAATTCGGTCAGAGTTCAATTCAAAATTTGTTAAG
 TTCCTTACCAGAAGAGTTGCGGTGTACAACATCGTCTCTAG

Hình 4. Cấu trúc gen *SppMKS2-3* (các exon được gạch dưới)

S1MKS2c	1	-----M	S	H	S	F	S	I	A	P	N	M	S	L	N	H	R	S	P	P	S	T	P	V	I	P	H	R	Q	L	L	P	N	L	R	L	S	S	C	K	S	R	G	F	E	A	Y	N	A									
SppMKS2-3	1	-----M	S	H	S	F	S	I	A	P	N	M	S	L	N	H	R	S	P	P	S	A	T	P	V	I	P	H	R	Q	L	L	P	N	L	R	L	S	S	C	K	S	R	G	F	E	A	Y	N	A								
ShMKS2	1	-----M	S	H	S	F	S	I	A	T	N	L	L	L	N	H	G	S	P	P	S	T	P	V	I	P	H	R	Q	L	L	P	N	L	R	L	S	S	R	K	S	R	S	F	E	A	H	S	A									
S1MKS2a	1	M	S	Q	C	I	A	S	P	L	I	R	S	I	G	S	T	S	V	G	N	S	L	L	P	N	H	R	P	P	S	T	L	P	V	S	P	H	R	Q	L	L	P	N	L	Q	L	S	V	S	K	L	R	S	F	R	A	H
SppMKS2-1	1	M	S	H	C	I	V	S	P	L	I	R	S	I	G	S	T	S	V	G	N	S	L	L	P	N	H	R	P	P	S	T	L	P	V	I	P	H	R	Q	L	L	P	N	L	Q	L	S	V	S	K	L	R	S	F	R	A	H
S1MKS2b	1	M	S	Q	S	I	V	S	P	L	I	G	N	---	C	L	I	S	L	F	P	N	R	R	P	P	S	T	F	P	V	R	---	Q	L	H	L	P	N	L	Q	L	S	A	S	K	S	R	S	F	D	T	N					
SppMKS2-2	1	M	S	Q	S	I	V	S	P	L	I	G	N	---	C	L	I	S	L	F	P	N	R	R	P	P	S	T	F	P	V	R	---	Q	L	H	L	P	N	L	Q	L	S	A	S	K	S	R	S	F	D	T	N					

S1MKS2c	52	F	D	L	K	G	T	Q	R	M	S	D	Q	Y	D	H	V	E	L	T	V	R	D	Y	E	L	D	Q	F	G	V	V	N	N	A	T	Y	V	S	Y	C	Q	H	C	H	E	F	L	E	K	I	G	V	S	V	D	E		
SppMKS2-3	52	F	D	L	K	G	T	Q	R	M	S	D	Q	Y	D	H	V	E	L	T	V	R	D	Y	E	L	D	Q	F	G	V	V	N	N	A	T	Y	A	S	Y	C	Q	H	C	R	H	E	F	L	E	K	I	G	V	S	V	D	E	
ShMKS2	52	F	D	L	K	S	T	Q	R	M	S	D	Q	Y	H	H	V	E	L	T	V	R	D	Y	E	L	D	Q	F	G	V	V	N	N	A	T	Y	A	S	Y	C	Q	H	C	R	H	A	F	L	E	K	I	G	V	S	V	D	E	
S1MKS2a	60	F	D	L	K	G	S	Q	---	G	M	A	E	F	H	E	V	E	L	K	V	R	D	Y	E	L	D	Q	Y	G	V	V	N	N	A	T	Y	A	S	Y	C	Q	H	R	H	E	L	L	E	R	I	G	T	S	A	D	E		
SppMKS2-1	60	F	D	L	K	G	S	Q	---	G	M	A	E	F	H	E	V	E	L	K	V	R	D	Y	E	L	D	Q	Y	G	V	V	N	N	A	T	Y	A	S	Y	C	Q	H	R	H	E	L	L	E	R	I	G	T	S	A	D	E		
S1MKS2b	53	F	D	L	N	G	T	R	G	I	---	G	D	L	F	H	E	V	E	L	K	V	R	D	Y	E	L	D	Q	F	G	V	V	N	N	A	T	Y	A	S	Y	C	Q	H	C	R	H	E	Y	L	E	R	I	G	L	S	V	D	E
SppMKS2-2	53	F	D	L	N	G	T	R	G	I	---	G	D	L	F	H	E	V	E	L	K	V	R	D	Y	E	L	D	Q	F	G	V	V	N	N	A	T	Y	A	S	Y	C	Q	H	C	R	H	E	Y	L	E	K	I	G	L	S	V	D	E

S1MKS2c	112	T	R	N	G	D	A	L	A	V	T	E	L	S	F	K	F	L	A	P	L	R	S	G	D	R	F	V	V	R	A	R	L	S	H	S	T	V	A	R	L	F	F	E	H	F	I	F	K	L	P	D	Q	E	P	I	L	E	A	R
SppMKS2-3	112	T	R	N	G	D	A	L	A	V	T	E	L	S	F	K	F	L	A	P	L	R	S	G	D	R	F	V	V	R	A	R	L	S	H	S	T	V	A	R	L	F	F	E	H	F	I	F	K	L	P	D	Q	E	P	I	L	E	A	R
ShMKS2	112	T	R	N	G	D	A	L	A	V	T	E	L	S	L	K	F	L	A	P	L	R	S	G	D	R	F	V	V	R	A	R	L	S	H	F	T	V	A	R	L	F	F	E	H	F	I	F	K	L	P	D	Q	E	P	I	L	E	A	R
S1MKS2a	117	A	R	S	G	D	A	L	A	L	T	E	L	S	L	K	Y	L	A	P	L	R	S	G	D	R	F	V	V	K	A	R	I	S	D	S	S	A	A	R	L	F	F	E	H	F	I	F	K	L	P	D	Q	E	P	I	L	E	A	R
SppMKS2-1	117	A	R	S	G	D	A	L	A	L	T	E	L	S	L	K	Y	L	A	P	L	R	S	G	D	R	F	V	V	K	A	R	I	S	D	S	S	A	A	R	L	F	F	E	H	F	I	F	K	L	P	D	Q	E	P	I	L	E	A	R
S1MKS2b	112	C	R	N	G	D	A	L	A	T	T	E	I	S	L	K	Y	L	A	P	L	R	S	G	D	R	F	V	V	K	V	R	L	S	G	S	T	A	A	R	L	F	F	E	H	F	I	F	K	L	P	D	Q	E	P	I	L	E	A	R

- production of methylketones in tomato trichomes. *Plant Physiol.* 151:1952-1964.
4. Goh E. B., Baidoo E. E. K., Keasling J. D., and Beller H. R., 2012. Engineering of bacterial methylketone synthesis for biofuels. *Appl. Environ. Microbiol.*, 78(1): 70-80.
 5. Kinsella J. E., Hwang D., 1976. Biosynthesis of flavors by *Penicillium roqueforti*. *Biotechnol. Bioeng.*, 18: 927-938.
 6. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol. Biol. Evol.*, 28(10): 2731-2739.
 7. Yu G., Nguyen T. T. H., Guo Y., Schauvinhold I., Auldridge M. E., Bhuiyan N., Ben-Israel I., Iijima Y., Fridman E., Noel J., Pichersky E., 2010. Enzymatic functions of wild tomato methylketone synthase 1 and 2. *Plant. Physiol.*, 154: 67-77.

**USING BIOINFORMATIC TOOLS
TO IDENTIFY NEW METHYLKETONE SYNTHASE 2 (MKS2) GENES
IN THE CURRANT TOMATO *Solanum pimpinellifolium***

Mai Huynh Hanh Phuc¹, Dinh Minh Hiep², Nguyen Thi Hong Thuong¹

¹VNU HCM-University of Science

²Management Board of Agricultural Hi-Tech Park, HCMC

SUMMARY

2-Methylketone is the organic compound containing a ketone functionality on the second carbon. This group of compounds has been known for their widespread applications in plant protection, fragrance and flavor industry, and biofuel production. Previously, two cDNAs encoding two proteins necessary for methylketone biosynthesis in the wild tomato species *Solanum habrochaites* subsp. *glabratum* were identified and designated methylketone synthase 1 (ShMKS1) and methylketone synthase 2 (ShMKS2). In searching of new MKS2 variants that are capable of using substrates of different chain lengths or degree of unsaturation, we used ShMKS2 as a query sequence for TBLASTN searches against the currant tomato (*Solanum pimpinellifolium*) genome. With the support of other bioinformatics tools for gene prediction, we identified three homologous genes of ShMKS2 on four contigs (contig 3697822, 6568413, 6704221, and 6708991) found in the genomic database of *Solanum pimpinellifolium*. Three genes encoded proteins with more than 65% identity to previously reported ShMKS2 and more than 98% identity to SIMKS2s; we named these genes as *SppMKS2-1*, *SppMKS2-2*, and *SppMKS2-3*. All three of these *Solanum pimpinellifolium* MKS2 genes had five exons and four introns (whose positions are conserved in comparison with the exon and intron positions in the homologous *S. lycopersicum* genes). The protein sequences of the MKS2 homologs from three *Solanum* species were compared and a phylogenetic tree was constructed.

Keywords: *Solanum pimpinellifolium*, DNA barcoding, methylketone synthase 2 (MKS2).

Ngày nhận bài: 15-7-2013