

ĐÁNH GIÁ ĐA DẠNG DI TRUYỀN VÀ CẤU TRÚC QUẦN THỂ NGUỒN GEN XOÀI (*Manifera* spp.) BẰNG CHỈ THỊ GBS - SNP

Nguyễn Thị Lan Hoa^{1,2,*}, Trần Đức Trung¹, Bùi Thị Thu Giang², Hà Minh Loan², Lê Thị Hoa², Nguyễn Hoàng¹, Roland Schafleitner³, Jean Lin³, Nguyễn Thị Thanh Thủy⁴

TÓM TẮT

Xoài (*Mangifera indica* L.) là cây ăn quả có giá trị kinh tế cao tại Việt Nam. Mục tiêu của nghiên cứu này nhằm đánh giá mức độ đa dạng di truyền của 14 mẫu giống xoài có nguồn gốc từ các tỉnh thành trong cả nước bằng chỉ thị đa hình đơn nucleotide (SNP). Kỹ thuật GBS (Genotyping by sequencing) được ứng dụng để phát hiện và sàng lọc 1.860 chỉ thị SNP chất lượng cao và sử dụng trong phân tích cấu trúc quần thể (bằng phần mềm STRUCTURE), phân nhóm di truyền (bằng công cụ PowerMarker) cũng như các chỉ số di truyền của các mẫu giống xoài. Kết quả cho thấy, bộ mẫu giống nghiên cứu có mức độ đa dạng di truyền tương đối cao, dựa trên chỉ số đa dạng gen *GD* (0,259) và hệ số thông tin đa hình *PIC* (0,227) của các chỉ thị SNP. Bên cạnh phân nhóm của mẫu giống đối chứng nhóm ngoại Thanh Trà khác loài, 2 phân nhóm mẫu giống xoài đã được xác định - một phân nhóm bao gồm các mẫu giống xoài phía Nam có nền di truyền đồng nhất và một phân nhóm của các mẫu giống xoài còn lại có nền di truyền pha trộn. Kết quả phản ánh lịch sử lai giống, chọn tạo khép kín giữa các mẫu giống xoài phía Nam đã góp phần xây dựng cơ sở di truyền cho các nghiên cứu lưu giữ và khai thác nguồn gen xoài tại Việt Nam.

Từ khóa: Xoài, *Mangifera* spp., SNP, GBS, cấu trúc quần thể.

1. ĐẶT VẤN ĐỀ

Xoài (*Mangifera indica* L.) là cây ăn quả có giá trị kinh tế cao và được trồng phổ biến ở các vùng nhiệt đới và cận nhiệt đới trên thế giới [1]. Không chỉ được sử dụng làm thực phẩm (ăn tươi, sấy khô, nước ép, chất điều vị...), các bộ phận của cây xoài như: Rễ, thân, lá, vỏ quả, hạt... còn được sử dụng làm vị thuốc trong y học cổ truyền chữa trị một số chứng bệnh về đường tiêu hóa, thiếu máu, hen suyễn... Tính đến năm 2020, diện tích canh tác xoài trên toàn thế giới đạt khoảng 6 triệu hecta với tổng sản lượng ước tính 51 triệu tấn, trong đó Ấn Độ và Trung Quốc là 2 quốc gia sản xuất xoài hàng đầu [1]. Việt Nam là quốc gia sản xuất xoài đứng thứ 13 thế giới về quy mô với diện

tích canh tác hơn 114 nghìn ha, tổng sản lượng hơn 940 nghìn tấn [2]. Đây đồng thời là mặt hàng trái cây xuất khẩu thế mạnh của nước ta trong những năm vừa qua.

Trong quá trình sinh trưởng và phát triển, cây xoài chịu ảnh hưởng của nhiều yếu tố sinh học (bệnh héo nhanh, bệnh loét cây, ruồi đục quả...) cũng như yếu tố phi sinh (nhiệt độ cực đoan, mưa bất thường, nhiễm mặn...). Thêm vào đó biến đổi khí hậu, mối đe dọa chính cho sản xuất nông nghiệp, cũng gây ra nhiều tác động tiêu cực ảnh hưởng trực tiếp đến năng suất và chất lượng quả xoài [3]. Chính vì vậy, nghiên cứu đánh giá đa dạng di truyền nhằm làm sáng tỏ quan hệ di truyền giữa các mẫu nguồn gen và tìm kiếm, xác định những biến dị di truyền liên quan đến tính trạng có lợi là hoạt động quan trọng, quyết định đến sự thành công của công tác chọn tạo, cải thiện giống xoài. Tương tự các loài cây trồng khác, đa dạng di truyền giữa các nguồn gen xoài có thể được đánh giá thông qua đặc điểm hình thái hoặc kiểu gen ADN hoặc kết hợp cả hai. Đánh giá các

¹ Trung tâm Tài nguyên Dược liệu, Viện Dược liệu

² Trung tâm Tài nguyên Thực vật, Viện Khoa học Nông nghiệp Việt Nam

³ The World Vegetable Center, Taiwan

⁴ Vụ Khoa học, Công nghệ và Môi trường, Bộ Nông nghiệp và Phát triển nông thôn

* Email: lanhoanimm@gmail.com

chỉ tiêu hình thái là phương thức đơn giản và nhanh chóng nhất để xác định các mẫu nguồn gen tiềm năng. Trong khi đó, chỉ thị phân tử với ưu điểm phản ánh chính xác bản chất di truyền mà không chịu tác động của điều kiện môi trường là cách tiếp cận phổ biến trong các nghiên cứu đánh giá đa dạng di truyền.

Các nghiên cứu đánh giá đa dạng di truyền nguồn gen xoài đã được thực hiện sử dụng hệ chỉ thị ngẫu nhiên không cần thông tin trình tự hệ gen đối tượng nghiên cứu như: RAPD [4], ISSR [5], ScoT [6, 7] và chỉ thị đặc hiệu phát triển dựa trên thông tin hệ gen như SSR [8, 9], InDel, SNP [10]. So với các loại chỉ thị phân tử khác, chỉ thị SNP được xác định có nhiều ưu điểm vượt trội như: Chỉ thị đồng trội, có số lượng lớn và phân bố đồng đều trên toàn hệ gen. Tuy nhiên, việc phát triển và ứng dụng chỉ thị SNP trên quy mô lớn bị hạn chế bởi các yêu cầu về kỹ thuật cũng như trang thiết bị nghiên cứu hiện đại và đắt tiền hơn so với các loại chỉ thị phân tử khác. Bên cạnh đó, các loài thực vật thường có hệ gen lớn và phức tạp với tỷ lệ cao trình tự lặp, trong đó nhiều loài chưa được giải mã hệ gen hoàn chỉnh cũng là trở ngại chính trong việc ứng dụng chỉ thị SNP.

Cùng với sự phát triển của công nghệ giải trình tự thế hệ mới, kỹ thuật xác định kiểu gen

thông qua giải trình tự (GBS-Genotyping-By-Sequencing) đã trở thành lựa chọn tối ưu cho các nghiên cứu đánh giá đa dạng di truyền nguồn gen thực vật ở quy mô toàn hệ gen [11]. Gần đây hệ gen xoài được giải mã đầy đủ [12, 13] đã cung cấp cơ sở dữ liệu trình tự rất hữu ích, tạo điều kiện cho việc ứng dụng kỹ thuật GBS phục vụ đánh giá đa dạng di truyền và xa hơn là công tác chọn tạo giống xoài bằng chỉ thị phân tử.

Trong nghiên cứu này, sự đa dạng di truyền giữa bộ mẫu nguồn gen xoài thu thập từ một số địa phương trong cả nước lần đầu tiên được đánh giá bằng kỹ thuật GBS. Kết quả nghiên cứu này cung cấp cơ sở khoa học cho việc lựa chọn nguồn gen xoài phù hợp cho công tác bảo tồn và chọn tạo giống xoài mới trong thời gian tới tại Việt Nam.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

Mẫu lá non của 21 mẫu giống xoài đại diện ở các địa phương trên cả nước và mẫu Thanh trà sử dụng làm đối chứng nhóm ngoại (outgroup) được thu thập từ Vườn Bảo tồn của Viện Khoa học Kỹ thuật Nông lâm nghiệp miền núi phía Bắc, Viện Nghiên cứu Rau quả và Viện Cây ăn quả miền Nam. Ký hiệu, tên mẫu giống và thông tin liên quan được thể hiện ở bảng 1.

Bảng 1. Danh sách mẫu giống xoài sử dụng trong nghiên cứu này

STT	Tên giống	Ký hiệu	Nguồn gốc	Tên khoa học
1	Xoài cát Yên Châu	M1	Sơn La*	<i>Mangifera indica</i>
2	Xoài Yên Minh vỏ hồng	M2	Hà Giang*	<i>Mangifera indica</i>
3	Xoài Yên Minh quả dài	M3	Hà Giang*	<i>Mangifera indica</i>
4	Xoài Vân Du	M4	Phú Thọ**	<i>Mangifera indica</i>
5	Xoài Kim Hoàng	M5	Phú Thọ**	<i>Mangifera indica</i>
6	Xoài tượng	M6	Bến Tre***	<i>Mangifera indica</i>
7	Xoài Thanh Ca	M7	Tiền Giang***	<i>Mangifera indica</i>

8	Xoài Tứ quý Vĩnh Long	M8	Vĩnh Long***	<i>Mangifera indica</i>
9	Xoài Hòn phấn Cần Thơ	M9	Cần Thơ***	<i>Mangifera indica</i>
10	Xoài Cát trắng Tiền Giang	M10	Tiền Giang***	<i>Mangifera indica</i>
11	Xoài bưởi xanh	M11	Tiền Giang***	<i>Mangifera indica</i>
12	Xoài cơm	M12	Khánh Hòa***	<i>Mangifera indica</i>
13	Xoài Tân Quang	M13	Tiền Giang***	<i>Mangifera indica</i>
14	Xoài nghệ Tiền Giang	M14	Tiền Giang***	<i>Mangifera indica</i>
15	Xoài cát mật	M15	Khánh Hòa***	<i>Mangifera indica</i>
16	Xoài ngọt Bến Tre	M16	Bến Tre***	<i>Mangifera indica</i>
17	Xoài xiêm giòn	M17	Bến Tre***	<i>Mangifera indica</i>
18	Xoài Lữ Phụng Tiên	M18	Bến Tre***	<i>Mangifera indica</i>
19	Thanh Trà	M19	NA***	<i>Bouea oppositifolia</i>
20	Xoài cát nghệ	M20	NA*	<i>Mangifera indica</i>
21	Xoài cát chu	M21	NA**	<i>Mangifera indica</i>
22	Xoài cát Hòa Lộc	M22	Tiền Giang***	<i>Mangifera indica</i>

Ghi chú: NA - Không có thông tin; * Thu tại Viện Nghiên cứu Rau quả; ** Thu tại Viện Khoa học Kỹ thuật Nông lâm nghiệp miền núi phía Bắc; *** Thu tại Viện Cây ăn quả miền Nam.

2.2. Phương pháp nghiên cứu

Tách chiết ADN tổng số: 100 mg lá của mỗi mẫu nguồn gen xoài được sử dụng để tách chiết ADN tổng số bằng DNeasy Plant Mini Kit (Qiagen) theo hướng dẫn của nhà cung cấp. Tinh nguyên vẹn, độ tinh sạch và nồng độ các mẫu ADN tổng số được đánh giá thông qua điện di trên gel agarose 1% và đo độ hấp thụ quang bằng thiết bị NanoDrop™ 2000.

Chuẩn bị thư viện GBS và giải trình tự: ADN tổng số của các mẫu nguồn gen xoài được định

lượng đồng nhất về nồng độ 25 ng/μL trước khi phân cắt bởi enzyme giới hạn *ApeKI* (New England Biolabs) tại 75°C trong 3 giờ. Thư viện GBS được thiết lập theo quy trình của Elshire và cs (2011) [11] và được gửi đi giải trình tự một chiều (single-end read, với chiều dài 150 bp) trên hệ thống Illumina HighSeq 2500 tại công ty Yourgene Health (Đài Loan).

Xử lý trình tự và phân tích xác định kiểu gen SNP: Trình tự thô của từng mẫu nguồn gen xoài được kiểm tra xử lý loại bỏ adapter và các vùng trình tự chất lượng kém bằng các công cụ FastQC

và Cutadapt với các thông số theo quy trình của Elshire và cs (2011) [11]. Các đoạn trình tự sau đó được ánh xạ (mapping) lên hệ gen tham chiếu của giống xoài Alphonso (mã Genbank: GCF_011075055.1) [10] bằng công cụ Bowtie2 để phát hiện và sàng lọc biến dị trình tự (multi - variant detection) giữa các mẫu nguồn gen xoài trên nền tảng gói công cụ NGSEP [14]. Dựa trên tiêu chí chất lượng giải trình tự tại vị trí SNP (Q-score) ≥ 20 , tần số allele thứ yếu (MAF – minor allele frequency) $> 0,01$ và tỷ lệ khuyết thiếu dữ liệu tại mỗi locus SNP $< 10\%$, dữ liệu kiểu gen SNP đạt yêu cầu giữa các mẫu nguồn gen xoài được trích xuất cho các phân tích tiếp theo.

Phân tích cấu trúc quần thể, xác định các chỉ số đa dạng di truyền và xây dựng mô hình phân nhóm: Cấu trúc quần thể của bộ mẫu giống xoài được xác định bằng công cụ STRUCTURE với 10 lần kiểm chứng độc lập ở mỗi giá trị K từ 1 - 10 (số quần thể tính toán dựa trên dữ liệu kiểu gen), 25.000 burnin và 50.000 MCMC theo mô hình hỗn tạp. Kết quả phân tích sau đó được xử lý bằng công cụ STRUCTURESELECTOR và xác định giá trị số quần thể K tối ưu theo phương pháp của Evanno [15]. Giá trị hệ số thành viên $q > 0,8$ được áp dụng để phân định các mẫu giống xoài vào các quần thể, trong khi đó mẫu giống với giá trị $q \leq 0,8$ được xem là có nền di truyền lẫn tạp. Các tham số di truyền, phân tích thành phần chính Principal Coordinates Analysis - PcoA, phân tích khoảng cách di truyền dựa trên tham số Nei 1983 và phân tích AMOVA xác định phương sai di truyền ở 2 cấp độ: (i) Giữa các quần thể và (ii) Trong các quần thể được phân nhóm dựa trên giá trị K tối ưu được thực hiện trên phần mềm PowerMarker V3.25 [16]. Phân nhóm di truyền Neighbor-Joining được mô hình hóa bằng phần mềm MEGA 11.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

Công tác thu thập, lưu giữ và đánh giá nguồn gen xoài, đặc biệt là các nguồn gen địa phương và mẫu giống từ các loài hoang dại có ý nghĩa vô cùng quan trọng, mang tính quyết định đến sự thành công của công tác chọn, tạo và phát triển giống xoài mới. Đặc biệt trong bối cảnh phát triển sản xuất hàng hóa, sự xói mòn đa dạng di truyền

nguồn gen xoài địa phương mang những đặc tính quý sẽ là những mất mát khó bù đắp cho nghiên cứu nâng cao chất lượng và khả năng chống chịu các điều kiện bất thuận vô sinh và hữu sinh của cây xoài [17].

3.1. Phân tích GBS và các chỉ số đa dạng di truyền dựa trên chỉ thị SNP

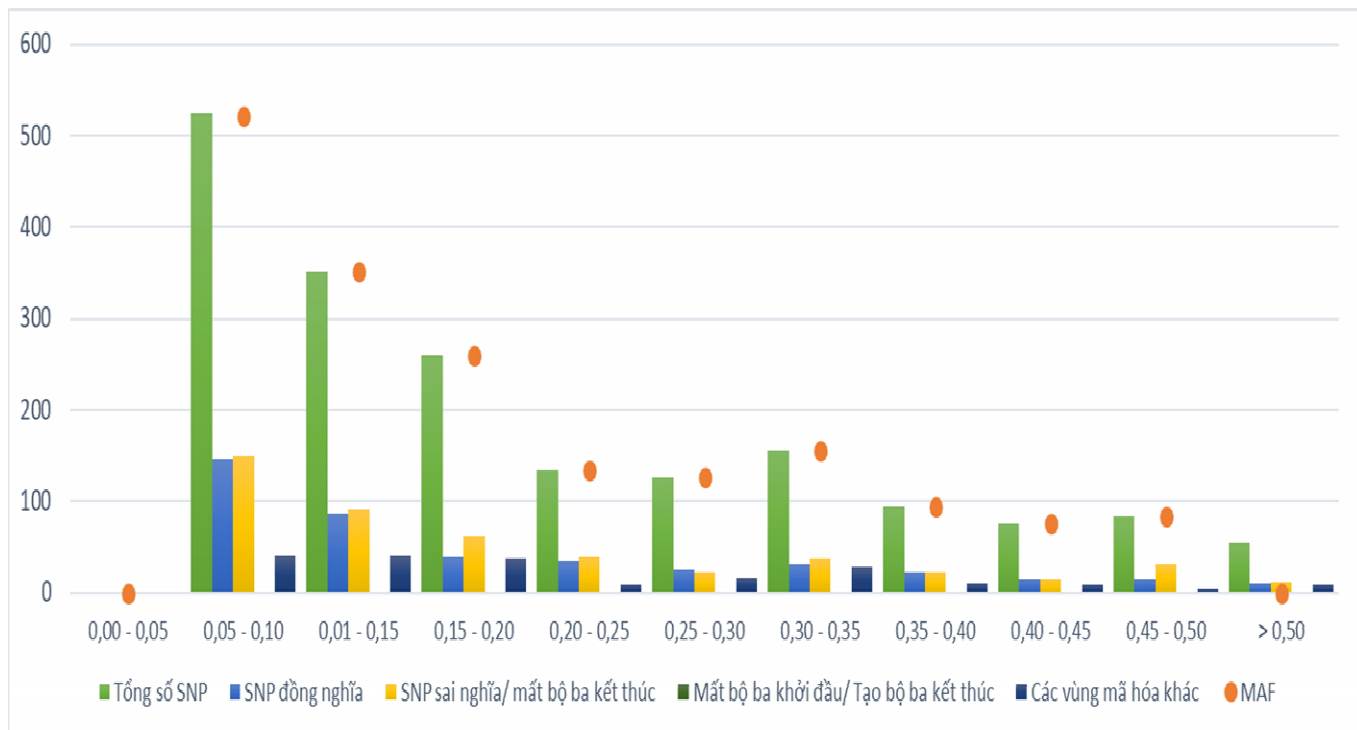
Cùng với sự phát triển của công nghệ giải trình tự, các chỉ thị phân tử dựa trên trình tự ADN (SNP, InDel) với ưu điểm số lượng và tần suất trên hệ gen lớn, mức độ đa hình cao, khả năng tái lập và độ tin cậy cao đã trở nên phổ biến và dần thay thế các thể hệ chỉ thị phân tử dựa trên phản ứng khuếch đại PCR (SSR, ISSR, RAPD...). Với sự tiến bộ không ngừng của công nghệ giải trình tự thông lượng lớn, kỹ thuật GBS có thể đánh giá kiểu gen của một số lượng lớn các đa hình trình tự SNP ở quy mô trên toàn hệ gen với hiệu suất cao và chi phí thấp [18]. Trong nghiên cứu này, tổng số 22 thư viện GBS đã được thiết lập cho bộ mẫu giống xoài nghiên cứu. Tuy nhiên, qua đánh giá chất lượng thư viện GBS, 7 thư viện của các mẫu: M05, M08, M09, M11, M13, M18 và M22 không đạt yêu cầu cho giải trình tự. Vì vậy bước giải trình tự cho các thư viện GBS được tiến hành cho 14 mẫu giống xoài và 01 mẫu thanh trà M19 và tạo ra 12.815.809 đoạn đọc với chất lượng cao. Số lượng đoạn đọc thu được từ mỗi mẫu giống biến thiên từ 304.093 (M10 - Xoài Cát trắng Tiền Giang) đến 1.444.961 (M04 - Xoài Vân Du), trung bình là 854.387,27 đoạn đọc cho mỗi mẫu giống. Dữ liệu trình tự GBS sau đó được ánh xạ lên hệ gen tham chiếu của giống xoài “Alphonso” với trung bình 417.792,87 đoạn đọc/thư viện (48,9%) ánh xạ thành công. Trong đó, mẫu M02 (Xoài Yên Minh vô hồng) có tỷ lệ ánh xạ thành công cao nhất (62,28%) và M19 (Thanh Trà) có tỷ lệ ánh xạ thấp nhất (20,07%). Do mẫu giống M19 (Thanh Trà) không thuộc chi *Mangifera*, kết quả này phản ánh đúng khả năng ánh xạ dữ liệu GBS từ các mẫu khác loài lên hệ gen tham chiếu của xoài.

Trình tự GBS của 15 mẫu ánh xạ thành công trên hệ gen tham chiếu bằng gói công cụ NGSEP đã xác định được 25.632 biến dị trình tự thuộc các nhóm khác nhau (SNP, InDel, STR và các loại biến dị đa allele khác). Trong đó, có 1.860 SNP hai allele

(bi-allelic SNP) có tỷ lệ khuyết dữ liệu <10% đã được lựa chọn cho các phân tích tiếp theo (mật độ trung bình 5,19 Mbp/1 SNP). Phân loại các SNP theo vị trí trên vùng gen được thống kê tại bảng 2 và hình 1.

Bảng 2. Thống kê phân loại các SNP xác định từ dữ liệu GBS của 15 mẫu giống xoài

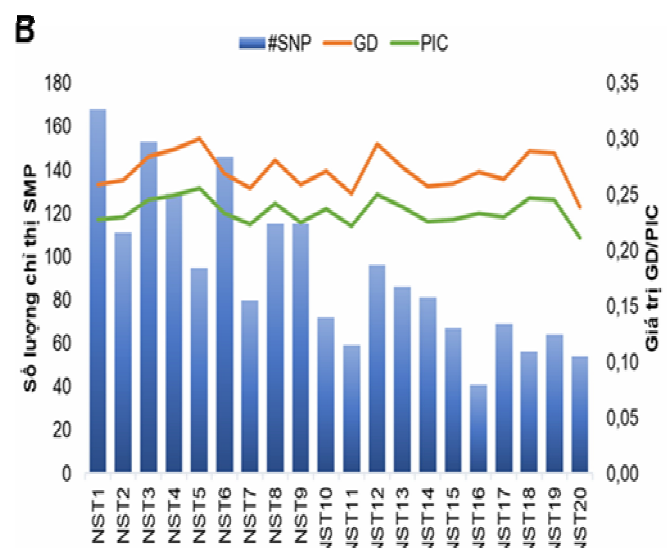
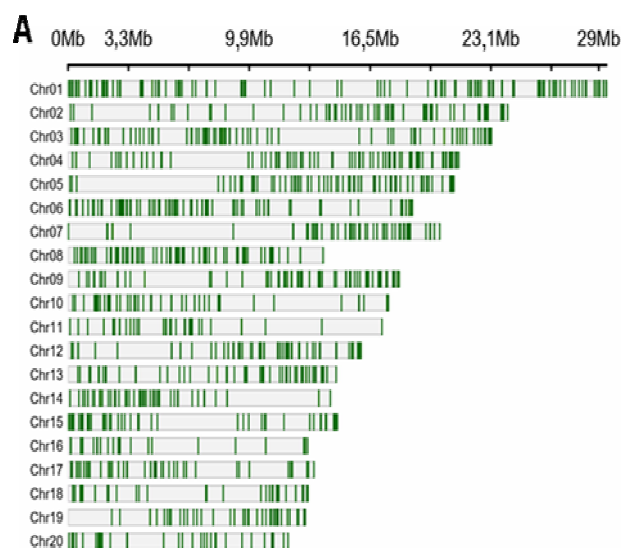
Phân loại	Số lượng SNP
Tổng số	1.860
SNP trên vùng mã hóa (coding)	1.115
Biến dị đồng nghĩa (synonymous variant)	426
Biến dị sai nghĩa (missense variant)	480
Mất bộ ba kết thúc (stop lost)	1
Tạo bộ ba kết thúc (stop gained)	4
Mất bộ ba khởi đầu (start lost)	0
Biến dị vùng bắt đầu cắt nối (đầu 5' intron) (splice donor variant)	0
Biến dị vùng tiếp nhận cắt nối (đầu 3' intron) (splice acceptor variant)	0
Biến dị vùng cắt nối trong exon (exonic splice region variant)	3
Biến dị vùng cắt nối (splice region variant)	26
Biến dị vùng 5' UTR (5' UTR variant)	74
Biến dị vùng 3' UTR (3' UTR variant)	101
SNP trên các vùng mã hóa khác (Other coding)	0
Biến dị trên các exon không mã hóa (non coding transcript exon variant)	0
Biến dị vùng phía trước phiên mã (upstream transcript variant)	88
Biến dị vùng phía sau phiên mã (downstream transcript variant)	40
Biến dị vùng intron (intron variant)	205
Biến dị vùng liên gen (intergenic variant)	412



Hình 1. Biểu đồ các nhóm biến dị di truyền phân bố theo tần số allel phụ (MAF) của chỉ thị SNP xác định được từ bộ 15 mẫu giống xoài bằng kỹ thuật GBS

Ngoài 3 SNP trên các contig/scaffold chưa được lắp ráp vào bộ nhiễm sắc thể (NST) của giống xoài Alphonso, 1.857 SNP phân bố trên từng NST với số lượng biến động từ 41 (NST số 16) đến 168 (NST số 01) (Hình 2A-B). So với nghiên cứu của Wang và cs (2020) [12] sử dụng kỹ thuật giải trình tự lại hệ gen (whole genome re-sequencing) xác định được hơn 19 triệu SNP từ bộ 52 mẫu

giống xoài hay phân tích hệ phiên mã của Mahato và cs (2016) [19] xác định được hơn 22 nghìn SNP, nghiên cứu này áp dụng kỹ thuật GBS có độ che phủ thấp (low-coverage GBS) [20] nên số lượng SNP thu được qua phân tích GBS là tương đối thấp. Mặc dù vậy, số lượng 1.857 chỉ thị SNP chất lượng cao xác định từ 15 mẫu giống hoàn toàn đáp ứng được mục tiêu đánh giá đa dạng di truyền.



Hình 2. Mô hình (A) và biểu đồ (B) phân bố của 1.857 SNP cùng các hệ số di truyền trên 20 nhiễm sắc thể hệ gen tham chiếu của giống xoài “Alphonso”

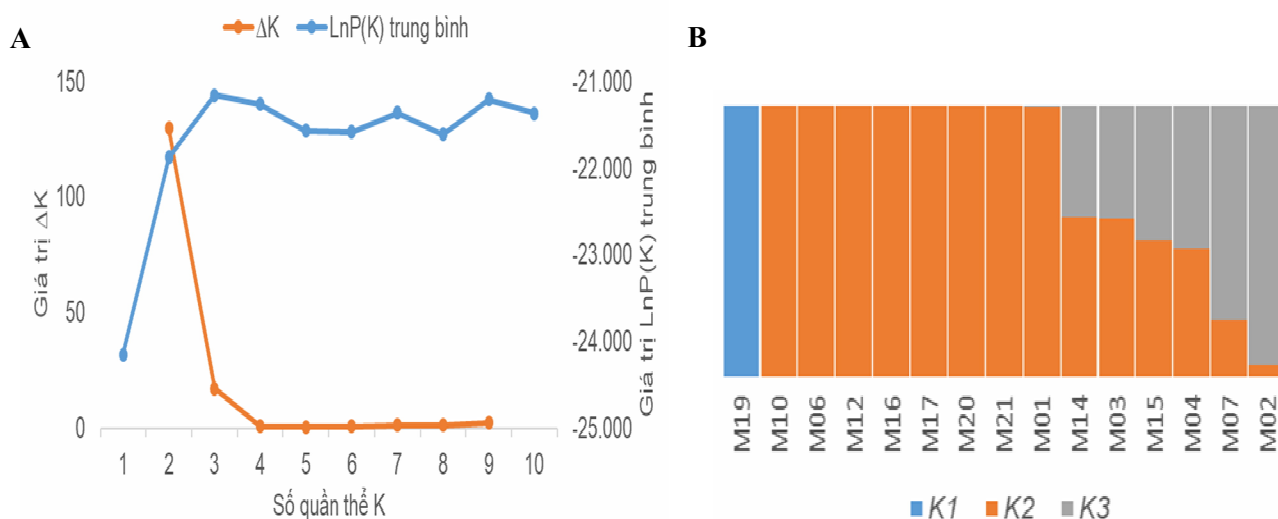
Dựa trên kiểu gen SNP, các thông số di truyền đã được xác định cho bộ mẫu giống xoài nghiên cứu. Theo đó, hệ số thông tin đa hình PIC của các locus SNP biến thiên trong khoảng 0,117 - 0,375 với giá trị trung bình theo từng NST thấp nhất và cao nhất ghi nhận lần lượt ở NST 20 (0,211) và NST 5 (0,256). Tương tự, chỉ số đa dạng gen (Gene diversity – GD) của các locus SNP nghiên cứu biến thiên trong khoảng 0,116 - 0,500, giá trị trung bình theo từng NST thấp nhất và cao nhất ghi nhận lần lượt ở NST 20 (0,238) và NST 5 (0,300) (Hình 2B).

Hệ số thông tin đa hình PIC là chỉ số giúp đánh giá hiệu quả của chỉ thị phân tử trong nghiên cứu di truyền một quần thể nhất định và luôn có giá trị tối đa $\leq 0,5$ đối với các chỉ thị SNP 2 alen (khi có 2 alen cùng xuất hiện trong quần thể với tần suất bằng nhau) [21]. Trong nghiên cứu này, giá trị trung bình của hệ số PIC đạt 0,227 cho thấy, bộ chỉ thị SNP xác định bằng kỹ thuật GBS có hiệu quả trong đánh giá đa dạng di truyền bộ mẫu giống nghiên cứu. Trong khi đó, chỉ số đa dạng gen GD - biểu hiện cho mức độ biến dị di truyền trên cơ sở tỷ lệ các alen của các locus được phân tích trong quần thể - đạt trung bình 0,259 cho thấy, mức biến dị di truyền tương đối cao của bộ mẫu giống xoài nghiên cứu [22]. Kết quả này phản ánh đúng nguồn gốc thu thập từ nhiều tỉnh, thành trong cả nước của bộ mẫu

giống xoài nghiên cứu.

3.2. Cấu trúc quần thể của bộ mẫu giống xoài

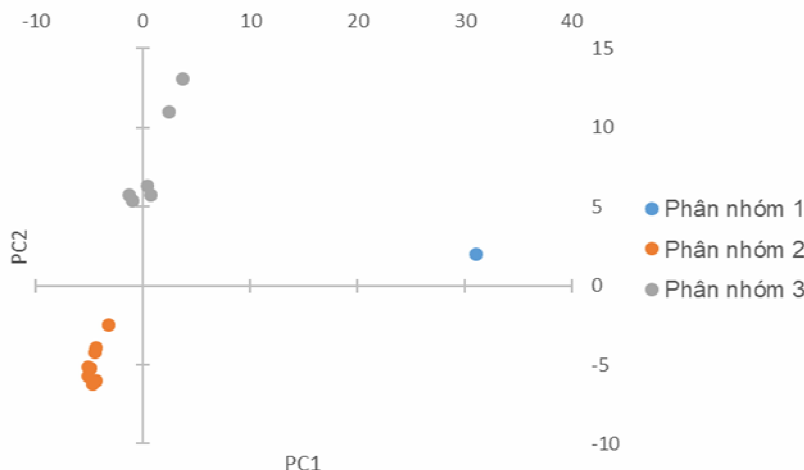
Nhằm xác định mô hình cấu trúc di truyền của bộ 15 mẫu giống xoài nghiên cứu, phân tích phân nhóm bằng phần mềm STRUCTURE đã được thực hiện với bộ dữ liệu 1.857 SNP được xác định trên 20 NST tham chiếu của giống xoài “Alphonso”. Kết quả phân tích STRUCTURE dự đoán số quần thể K tối ưu trong bộ mẫu nghiên cứu là 2 (dựa trên chỉ số ΔK) hoặc 3 (dựa trên giá trị trung bình $\text{LnP}(K)$) (Hình 3A). Kết hợp với thông tin phân loài của mẫu đối chứng nhóm ngoại M19 (Thanh Trà *Bouea oppositifolia*, không thuộc chi *Mangifera*), số quần thể tối ưu là 3 được lựa chọn. Theo đó, M19 là mẫu giống duy nhất có nền di truyền thuộc quần thể $K1$ (hệ số thành viên $q_{K1} = 1$), hoàn toàn khác biệt so với các mẫu giống xoài còn lại. Tám mẫu giống xoài (M01, M06, M10, M12, M16, M17, M20 và M21) có nền di truyền đồng nhất cao thuộc quần thể $K2$ ($q_{K2} = 0,99 \sim 1$). Trong khi đó, nền di truyền của sáu mẫu giống còn lại là pha trộn giữa quần thể $K2$ và $K3$ với tỷ lệ khác nhau. Năm mẫu giống (M03, M04, M07, M14 và M15) có nền di truyền pha trộn (giá trị $q < 0,8$) giữa quần thể $K2$ và $K3$ ($q_{K2} = 0,21 \sim 0,50$ và $q_{K3} = 0,40 \sim 0,78$). Riêng mẫu giống M02 có nền di truyền đồng nhất cao thuộc quần thể $K3$ ($q_{K3} = 0,95$) (Hình 3B).



Hình 3. Kết quả phân tích cấu trúc quần thể bộ mẫu giống xoài bằng chỉ thị SNP. (A) Biểu đồ xác định số quần thể K tối ưu. (B) Cấu trúc quần thể của bộ 15 mẫu giống xoài với giá trị số quần thể tối ưu $K = 3$.

Phân tích thành phần chính PCoA cho kết quả phân nhóm bộ mẫu giống xoài tương đồng với kết quả cấu trúc quần thể STRUCTURE. Phân nhóm 2 (gồm các mẫu giống ứng với quần thể *K2*) và phân nhóm 3 (gồm các mẫu giống có nền di truyền pha

trộn giữa quần thể *K2* và *K3*) tách biệt rõ ràng với phân nhóm 1 (mẫu giống M19 ứng với quần thể *K1*) trên biểu đồ phân bố dựa trên hai thành phần chính đầu tiên PC1 và PC2 đại diện 43,24% biến dị tích lũy.



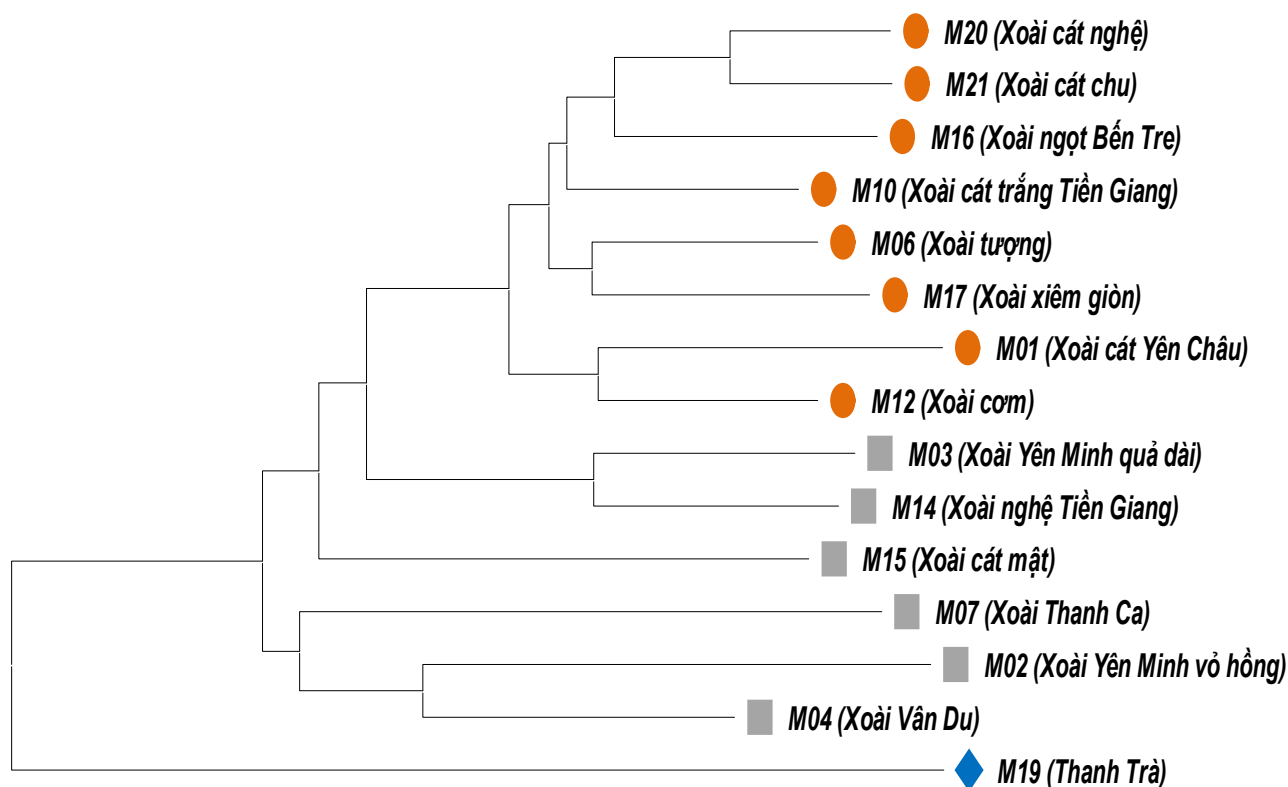
Hình 4. Phân nhóm thành phần chính PCoA của bộ 15 mẫu giống nghiên cứu

3.3. Phân nhóm di truyền bộ mẫu giống xoài

Dựa trên kiểu gen SNP của bộ mẫu giống nghiên cứu, mức độ khác biệt di truyền của các mẫu giống đã được tính toán dựa trên công thức Nei1983 bằng phần mềm PowerMarker. Giá trị khoảng cách di truyền trung bình của các cặp mẫu giống là 0,222; dao động từ 0,010 (M20 và M21) đến 0,497 (M10 và M19). Cây phân nhóm Neighbor-Joining dựa trên ma trận khoảng cách di truyền Nei1983 phân chia 15 mẫu giống thành 3 phân nhóm. Tương đồng với kết quả phân tích cấu trúc quần thể và phân tích thành phần chính PCoA, phân nhóm I bao gồm mẫu giống đối chứng nhóm ngoại (M19 - Thanh Trà) thuộc loài *Bouea oppositifolia* tách biệt với 2 phân nhóm còn lại. Phân nhóm II bao gồm 8 mẫu giống xoài, phần lớn là các mẫu có nguồn gốc thu thập từ các tỉnh phía Nam và có nền di truyền đồng nhất (thuộc quần thể *K2*). Mẫu giống M01 (Xoài cát Yên Châu - thu thập từ phía Bắc) thuộc phân nhóm II gợi ý về nguồn gốc du nhập từ các tỉnh phía Nam của mẫu giống này. Phân nhóm III bao gồm 6 mẫu giống, trong đó 3 mẫu giống M03 (Xoài Yên Minh quả dài), M14 (Xoài nghệ Tiền Giang) và M15 (Xoài cát mật) có vị trí phân nhóm tương đối gần với phân nhóm II, với nền di truyền pha trộn cân bằng giữa quần thể *K2* và *K3* đã được xác định bởi phân

tích cấu trúc quần thể (Hình 3B). Trong khi đó, các mẫu giống còn lại của phân nhóm III là M02 (Xoài Yên Minh vỏ hồng), M04 (Xoài Vân Du) và M07 (Xoài Thanh Ca) có quan hệ di truyền tương đối gần nhau và phân biệt khác rõ với phân nhóm II với nền di truyền đồng nhất cao hơn với quần thể *K3* (Hình 5).

Ngoài phân nhóm I do chỉ có một mẫu giống M19, các thông số di truyền trung bình đã được xác định cho các phân nhóm II và III. Theo đó, khoảng cách di truyền Nei1983 giữa 2 phân nhóm đạt 0,667 và các chỉ số *H*, *GD* và *PIC* của phân nhóm III (lần lượt là 0,251, 0,342 và 0,287) cao hơn của phân nhóm II (lần lượt là 0,167, 0,233 và 0,199). Bên cạnh đó, phân tích phương sai di truyền AMOVA cho thấy, chỉ có 1% biến dị di truyền xảy ra giữa phân nhóm II và III, trong khi đó 99% biến dị ghi nhận được trong các phân nhóm. Giá trị PhiPT giữa hai phân nhóm chỉ đạt 0,014 và không có ý nghĩa thống kê ($p = 0,322$). Kết quả này cho thấy, các mẫu giống xoài thuộc phân nhóm III có tính đa dạng di truyền cao hơn so với phân nhóm II và phù hợp với kết quả phân nhóm di truyền. Tương quan di truyền cũng gợi ý khả năng kết hợp các mẫu giống thuộc hai phân nhóm này trong các tổ hợp lai nhằm thu được biến dị có lợi về ưu thế lai cao nhất.



Hình 5. Cây phân nhóm di truyền của các mẫu giống xoài dựa trên 1.567 chỉ thị SNP xác định từ dữ liệu GBS.

Ký hiệu 3 phân nhóm tương ứng với kết quả xác định cấu trúc quần thể của bộ mẫu giống nghiên cứu

Tại Việt Nam các chỉ thị phân tử dựa trên kỹ thuật PCR đã được sử dụng trong đánh giá đa dạng di truyền các mẫu giống xoài. Nguyễn Thị Lang và cs (2020) [9] đã sử dụng 31 chỉ thị SSR để phân nhóm 30 mẫu đại diện cho 16 giống xoài thành 4 nhóm chính, trong đó có 7 giống nhập nội, thu thập ở đồng bằng sông Cửu Long. Tương ứng với mức độ tương đồng di truyền dao động trong khoảng 47 - 100%, kết quả phân nhóm phản ánh đúng nguồn gốc của nhóm các mẫu giống xoài nhập nội và xoài Việt Nam đồng thời cho thấy, sự khác biệt kiểu gen không đáng kể giữa các mẫu giống xoài Việt Nam. Trong khi đó, Nguyễn Thị Thanh Nhã và cs (2021) [7] đã đánh giá kiểu gen 30 mẫu đại diện cho 13 giống xoài thu thập ở các tỉnh miền Trung và Nam bộ bằng 22 chỉ thị SCoT. Các mẫu giống xoài có mức độ tương đồng di truyền tương đối cao (trong khoảng 83 - 97%) và được phân thành 2 nhóm chính. Hai nghiên cứu trên sử dụng các chỉ thị phân tử đa alen với số lượng alen trung bình đạt 4,1 (chỉ thị SSR) và 15,2 (chỉ thị SCoT), do đó có chỉ số *PIC* cao hơn so với

ghi nhận dựa trên chỉ thị SNP trong nghiên cứu này. Kết quả tương tự cũng đã được ghi nhận ở nghiên cứu phân tích di truyền sử dụng chỉ thị GBS-SNP của Pavan và cs (2018) [18], Cericola và cs (2018) [20].

Các công bố trước đây cũng cho thấy, một số lượng chỉ thị SNP giới hạn có thể đảm bảo phù hợp cho nghiên cứu đa dạng di truyền và phân tích cấu trúc quần thể ở xoài. Thông qua giải trình tự hệ phiên mã từ mẫu lá của một số giống xoài Ấn Độ, Singh và cs (2016) [23] đã phát triển và lựa chọn 1.159 chỉ thị SNP chất lượng cao để đánh giá phân chia 84 giống xoài tại Ấn Độ thành 2 phân nhóm có nền di truyền tương quan với nguồn gốc thu thập, trong đó có một phân nhóm có nền di truyền pha trộn rõ rệt. Trên quy mô lớn hơn với bộ 106 mẫu giống xoài thu thập trên toàn thế giới, Warschefsky và cs (2014) [24] đã xác định được 612 chỉ thị SNP chất lượng cao từ dữ liệu RADseq. Kết quả phân tích phân nhóm và cấu trúc di truyền đã cho thấy, nhóm các mẫu giống xoài khu vực Đông Nam Á có nền di truyền đồng nguồn với chỉ

một số ít mẫu giống có nền di truyền pha trộn từ các mẫu giống Ấn Độ, châu Phi và châu Mỹ. Tương tự, chỉ với 272 chỉ thị SNP, nền di truyền của bộ 208 mẫu giống xoài ở Australia đã được xác định thuộc 4 phân nhóm, trong đó phân nhóm các mẫu giống xoài Đông Nam Á có sự khác biệt di truyền rõ ràng với các phân nhóm còn lại [25]. Các kết quả trên thống nhất với nghiên cứu này, mặc dù số lượng mẫu giống được phân tích không lớn nhưng việc sử dụng bộ chỉ thị SNP phân bố trên toàn hệ gen đã giúp phân định nguồn gốc di truyền và tương quan với vùng địa lý của các mẫu giống xoài Việt Nam. Bên cạnh đó, nền di truyền đồng nhất được ghi nhận ở 2/3 số mẫu giống nghiên cứu phù hợp với các nghiên cứu đã công bố trước đây [7, 9] và có thể xem là đặc điểm di truyền đặc trưng của nguồn gen xoài Việt Nam, mặc dù trải qua lịch sử lai tạo, chọn lọc tương đối dài kể từ khi được du nhập từ Ấn Độ [23].

4. KẾT LUẬN

Trong nghiên cứu này, lần đầu tiên kỹ thuật GBS được áp dụng nhằm xác định kiểu gen SNP phục vụ đánh giá đa dạng di truyền và phân tích quần thể của bộ mẫu giống xoài Việt Nam. Thông qua 1.860 chỉ thị SNP chất lượng cao phân bố trên toàn hệ gen, các mẫu giống xoài nghiên cứu được xác định có mức độ đa dạng di truyền trung bình. Trong khi đó, phân tích cấu trúc quần thể cho thấy, 2 nhóm với nền di truyền khác biệt: Nhóm có nền di truyền đồng nhất (các mẫu giống xoài phía Nam) và nhóm có nền di truyền pha trộn (các mẫu giống xoài phía Bắc và phía Nam). Kết quả này không chỉ góp phần tạo lập cơ sở di truyền cho công tác lưu giữ và phát triển nguồn gen xoài mà còn minh chứng cho hiệu quả và tính khả thi của việc triển khai kỹ thuật GBS tại Việt Nam.

LỜI CẢM ƠN

Nghiên cứu này được thực hiện với sự tài trợ của Chương trình Công nghệ Sinh học Nông nghiệp - Thủy sản - Bộ Nông nghiệp và Phát triển nông thôn và Trung tâm Rau Thế giới - The World Vegetable Center.

TÀI LIỆU THAM KHẢO

1. Zhang, C., D. Xie, T. Bai, X. Luo, F. Zhang, Z. Ni and Y. Chen (2020). Diversity of a large collection of natural populations of mango

(*Mangifera indica* Linn.) revealed by agro-morphological and quality traits. *Diversity*, 12 (1): p. 27.

2. Bộ Nông nghiệp và PTNT (2021). Báo cáo tổng kết năm 2021, nhiệm vụ trọng tâm năm 2022 lĩnh vực trồng trọt.

3. Normand, F., P. -E. Lauri, and J. -M. Legave (2015). Climate change and its probable effects on mango production and cultivation. *Acta Horticulturae*, 1075: p. 21 - 32.

4. Souza, I., S. Valente, F. Britto, V. A. B. Souza and P. S. Lima (2011). RAPD analysis of the genetic diversity of mango (*Mangifera indica*) germplasm in Brazil. *Genetics and molecular research*, 10: p. 3080 - 3089.

5. Rocha, A., L. Salomão, T. Salomão, C. Cruz and D. Siqueira (2011). Genetic diversity of 'Ubá' mango tree using ISSR markers. *Molecular biotechnology*, 50: p. 108-113.

6. Zhou, L., *et al.* (2020). Evaluation of the genetic diversity of mango (*Mangifera indica* L.) seedling germplasm resources and their potential parents with start codon targeted (SCoT) markers. *Genetic Resources and Crop Evolution*, 67: p. 41 - 58.

7. Nguyễn Thị Thanh Nhã, Lê Thị Thanh Nga, Vũ Thị Huyền Trang (2021). Đánh giá đa dạng di truyền các mẫu giống xoài tại Trung bộ và Nam bộ Việt Nam bằng chỉ thị SCoT. *Tạp chí Nông nghiệp và Phát triển nông thôn*, 9 (2): p. 113 - 120.

8. Dillon, N., I. Bally, L. Hucks, C. Wright, D. J. Innes and R. G. Dietzgen (2013). Implementation of SSR markers in mango breeding in Australia. *Acta Horticulturae*, 992: p. 259 - 267.

9. Nguyễn Thị Lang, Nguyễn Thị Khánh Trân, Bùi Chí Công, Bùi Chí Hiếu, Bùi Chí Bửu (2020). Phân tích sự đa hình cây xoài (*Mangifera indica* L.) bằng kỹ thuật SSR tại đồng bằng sông Cửu Long. *Tạp chí Nông nghiệp và Phát triển nông thôn*, 12 (2): p. 18 - 27.

10. Cortaga, C. Q., J. A. P. Lachica, D. V. Lantican and E. T. M. Ocampo (2022). Genome - wide SNP and InDel analysis of three Philippine mango species inferred from whole - genome

sequencing. *Journal of Genetic Engineering and Biotechnology*, 20 (1): p. 46.

11. Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler and S. E. Mitchell (2011). A robust, simple genotyping - by - sequencing (GBS) approach for high diversity species. *PLoS One*, 6 (5): p. e19379.

12. Wang, P., Y. Luo, J. Huang, S. Gao, G. Zhu, Z. Dang, J. Gai, M. Yang, M. Zhu, H. Zhang, X. Ye, A. Gao, X. Tan, S. Wang, S. Wu, E. B. Cahoon, B. Bai, Z. Zhao, Q. Li, J. Wei, H. Chen, R. Luo, D. Gong, K. Tang, B. Zhang, Z. Ni, G. Huang, S. Hu and Y. Chen (2020). The genome evolution and domestication of tropical fruit mango. *Genome Biology*, 21 (1): p. 60.

13. Bally, I. S. E., A. Bombarely, A. H. Chambers, Y. Cohen, N. L. Dillon, D. J. Innes, M. A. Islas-Osuna, D. N. Kuhn, L. A. Mueller, R. Ophir, A. Rambani, A. Sherman, H. Yan and Consortium of Mango Genome (2021). The 'Tommy Atkins' mango genome reveals candidate genes for fruit quality. *BMC Plant Biology*, 21 (1): p. 108.

14. Perea, C., J. F. De La Hoz, D. F. Cruz, J. D. Lobaton, P. Izquierdo, J. C. Quintero, B. Raatz and J. Duitama (2016). Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. *BMC Genomics*, 17 (5): p. 498.

15. Li, Y. L. and J. X. Liu. (2018). StructureSelector: A web - based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, 18 (1): p. 176 - 177.

16. Liu, K. and S. Muse. (2005). PowerMaker: An integrated analysis environment for genetic maker analysis. *Bioinformatics*, 21: p. 2128 - 2129.

17. Dwivedi, S. L., S. Ceccarelli, M. W. Blair, H. D. Upadhyaya, A. K. Are and R. Ortiz (2016). Landrace germplasm for improving yield and abiotic stress adaptation. *Trends Plant Science*, 21 (1): p. 31 - 42.

18. Pavan, S., P. L. Curci, D. L. Zuluaga, E. Blanco and G. Sonnante (2018). Genotyping-by-

sequencing highlights patterns of genetic structure and domestication in artichoke and cardoon. *PLoS One*, 13 (10): p. e0205988.

19. Mahato, A. K., N. Sharma, A. Singh, M. Srivastav, Jaiprakash, S. K. Singh, A. K. Singh, T. R. Sharma and N. K. Singh (2016). Leaf transcriptome sequencing for identifying genic-SSR markers and SNP heterozygosity in crossbred mango variety 'Amrapali' (*Mangifera indica* L.). *PLoS One*, 11 (10): p. e0164325.

20. Cericola, F., I. Lenk, D. Fè, S. Byrne, C. S. Jensen, M. G. Pedersen, T. Asp, J. Jensen and L. Janss (2018). Optimized use of low-depth genotyping-by-sequencing for genomic prediction among multi-parental family pools and single plants in perennial ryegrass (*Lolium perenne* L.). *Frontiers in Plant Science*, 9: p. 369.

21. Eltaher, S., A. Sallam, V. Belamkar, H. A. Emara, A. A. Nower, K. F. M. Salem, J. Poland and P. S. Baenziger (2018). Genetic diversity and population structure of F(3:6) Nebraska winter wheat genotypes using genotyping-by-sequencing. *Frontiers in Genetics*, 9: p. 76.

22. Harris, A. M. and M. DeGiorgio. (2017). An unbiased estimator of gene diversity with improved variance for samples containing related and inbred individuals of any ploidy. *G3 (Bethesda)*, 7 (2): p. 671 - 691.

23. Singh, N. K., A. K. Mahato, P. K. Jayaswal, A. Singh, S. Singh, N. Singh, V. Rai, A. S. Mithra, K. Gaikwad, N. Sharma, S. K. Lal, M. L. Srivastava, J. Prakash, U. Kalidindi, S. K. Singh, A. K. Singh, K. Khan, R. K. Mishra, A. Bajpai, B. S. Sandhya, P. Nischita, K. V. Ravishankar, M. R. Dinesh, N. Kumar, S. Jaiswal, D. Kumar, A. Rai and T. R. Sharma (2016). Origin, diversity and genome sequence of mango (*Mangifera indica* L.). *Indian journal of history of science*, 51.

24. Warschewsky, E. J. and E. J. B. von Wettberg. (2019). Population genomic analysis of mango (*Mangifera indica*) suggests a complex history of domestication. *New Phytologist*, 222 (4): p. 2023 - 2037.

25. Wilkinson, M. J., R. Yamashita, M. E. James, I. S. E. Bally, N. L. Dillon, A. Ali, C. M.

Hardner and D. Ortiz-Barrientos (2022). The *indica*) gene pool. *Scientific Reports*, 12 (1): p. influence of genetic structure on phenotypic 20614.
diversity in the Australian mango (*Mangifera*

ASSESSMENT OF GENETIC DIVERSITY AND POPULATION STRUCTURE OF MANGO (*Mangifera* spp.) GERMPLASMS USING GBS - BASED SNP MARKERS

**Nguyen Thi Lan Hoa^{1,2}, Tran Duc Trung², Bui Thi Thu Giang¹, Ha Minh Loan¹,
Nguyen Hoang², Roland Schafleitner³, Jean Lin³, Nguyen Thi Thanh Thuy⁴**

¹*National Institute of Medicinal Materials*

²*Plant Resources Center*

³*The World Vegetable Center, Taiwan*

⁴*Ministry of Agricultural and Rural Development*

Summary

Mango (*Mangifera indica* L.) is an economically significant fruit tree in Vietnam. This study aims to evaluate the genetic diversity of 15 mango accessions collected nationwide using single nucleotide polymorphism (SNP) markers. The GBS (Genotyping by sequencing) technique was applied to explore and screen 1,860 high - quality SNP markers, which were then utilized in population structure analysis (using the STRUCTURE v.4.3 software) and genetic clustering (using the PowerMarker v.3.25 software) as well as determining genetic indices among collected mango samples. The results showed that the studied mango accessions harbor a relatively high level of genetic diversity, as demonstrated by the genetic diversity index (*GD*) and the polymorphic information coefficient (*PIC*) determined based on the SNP markers. In addition to the subgroup of the out-group, interspecific Thanh Tra accessions, two subgroups of the mango accessions were identified - one consisting of southern - originated accessions with a homogeneous genetic background and one of the remaining accessions having an admixture genetic background. This result reflected the history of closed breeding and selection among mango varieties in the southern region. It might contribute to an essential genetic foundation for research in the conservation and exploitation of mango genetic resources in Vietnam.

Keywords: *Mango, Mangifera spp., SNP, GBS, population structure.*

Người phản biện: TS. Lê Thị Tuyết Châm

Ngày nhận bài: 6/11/2023

Ngày thông qua phản biện: 01/12/2023

Ngày duyệt đăng: 26/01/2024