

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ**

**BÁO CÁO TÓM TẮT  
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ CẤP BỘ**

**XÂY DỰNG CƠ SỞ DỮ LIỆU DNA MÃ VẠCH  
CHO CÁC GIỐNG CÂY ĂN TRÁI ĐẶC SẢN  
CỦA VIỆT NAM KHU VỰC NAM BỘ**

**Mã số: B2019-TCT-562-11**

**Chủ nhiệm đề tài: TS. Đỗ Tấn Khang**

**Cần Thơ, Tháng 01/2022**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ**

**BÁO CÁO TÓM TẮT  
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ CẤP BỘ**

**XÂY DỰNG CƠ SỞ DỮ LIỆU DNA MÃ VẠCH  
CHO CÁC GIỐNG CÂY ĂN TRÁI ĐẶC SẢN  
CỦA VIỆT NAM KHU VỰC NAM BỘ**

**Mã số: B2019-TCT-562-11**

**Xác nhận của cơ quan chủ trì đề tài**

**Chủ nhiệm đề tài**

**TS. Đỗ Tấn Khang**

**Cần Thơ, Tháng 01/2022**

---

## **DANH SÁCH CÁC THÀNH VIÊN THAM GIA ĐỀ TÀI VÀ ĐƠN VỊ PHỐI HỢP CHÍNH**

### **1/ Những thành viên tham gia thực hiện đề tài**

1. TS. Đỗ Tấn Khang
2. TS. Nguyễn Phạm Anh Thi
3. TS. Trần Thanh Mến
4. TS. Nguyễn Văn Ấy
5. PGS.TS. Trần Nhân Dũng
6. CN. Trần Văn Bé Năm
7. CN. Trần Gia Huy
8. CN. Đinh Đào Tấn Phát
9. CN. Huỳnh Ngọc Hơ

### **2/ Đơn vị phối hợp chính**

Viện Cây ăn quả Miền Nam

Tên cơ quan chủ quản: Viện Cây ăn quả Miền Nam

Điện thoại: 02733 893 129      Fax: 02733 893 122

Địa chỉ: Long Định – Châu Thành – Tiền Giang

Họ và tên thủ trưởng tổ chức: TS. Võ Hữu Thoại

---

## MỤC LỤC

<b>PHẦN I. MỞ ĐẦU .....</b>	<b>1</b>
1. Lý do chọn đề tài .....	1
2. Mục tiêu nghiên cứu .....	2
3. Cách tiếp cận.....	2
4. Đối tượng và phạm vi nghiên cứu.....	2
5. Nội dung nghiên cứu.....	2
6. Phương pháp nghiên cứu .....	2
6.1. Phương tiện.....	2
6.1.1. Hóa chất.....	2
6.1.2 Thiết bị.....	3
6.1.3 Dụng cụ.....	3
6.2 Phương pháp .....	3
6.2.1 Ly trích DNA thực vật.....	3
6.2.2 Điện di DNA.....	3
6.2.3 Khuếch đại các đoạn DNA mã vạch.....	3
6.2.4 Phân tích đa dạng di truyền bằng chỉ thị phân tử ISSR .....	3
6.2.5 Phân tích dữ liệu.....	3
<b>PHẦN II. KẾT QUẢ VÀ THẢO LUẬN .....</b>	<b>4</b>
<b>CHƯƠNG I. KHẢO SÁT CÂY ĐÀU DÒNG CỦA 10 LOẠI CÂY ĂN TRÁI ĐẶC SẢN.....</b>	<b>4</b>
1.1 Khảo sát cây đầu dòng của 10 loại cây ăn trái đặc sản .....	4
1.2 Đặc tính nông học cây đầu dòng.....	4
<b>CHƯƠNG 2. KẾT QUẢ PHÂN TÍCH TRÌNH TỰ DNA MÃ VẠCH VÀ DI TRUYỀN.....</b>	<b>4</b>
2.1 Cây bưởi Da Xanh.....	4
2.1.1 Kết quả phân tích trình tự DNA mã vạch và SNPs .....	4
2.1.2 Đa dạng di truyền dấu phân tử ISSR .....	4
2.2 Cây cam Mật.....	4
2.2.1 Kết quả phân tích trình tự DNA mã vạch và SNPs .....	4
2.2.2 Đa dạng di truyền các giống cam dựa vào dấu phân tử ISSR .....	5
2.3 Cây chôm chôm Nhãn.....	5
2.3.1 Kết quả phân tích vùng trình tự DNA mã vạch .....	5
2.3.2 Kết quả phân tích đa hình bằng dấu phân tử ISSR .....	5
2.4 Cây dâu hạ châu.....	6
2.4.1 Kết quả phân tích vùng trình tự DNA mã vạch .....	6
2.4.2 Đa dạng di truyền dựa vào dấu phân tử ISSR.....	6
2.5 Cây măng cầu xiêm.....	7
2.5.1 Kết quả phân tích trình tự DNA mã vạch và SNPs .....	7
2.5.2 Phân tích đa dạng di truyền bằng dấu phân tử ISSR .....	7
2.6 Cây quýt hồng Lai Vung.....	8
2.6.1 Kết quả phân tích trình tự DNA mã vạch và SNPs .....	8
2.6.2 Kết quả phân tích đa dạng di truyền dựa vào dấu phân tử ISSR .....	8
2.7 Cây sầu riêng Ri-6 .....	8
2.7.1 Kết quả phân tích trình tự DNA mã vạch và SNPs .....	8

---

2.7.2 Kết quả điện di sản phẩm PCR với dấu phân tử ISSR .....	9
2.8 Cây thanh long ruột đỏ.....	9
2.8.1 Kết quả khuếch đại các trình tự DNA mã vạch .....	10
2.8.2 Kết quả phân tích trình tự DNA mã vạch, xác định SNPs.....	10
2.9 Cây vú sữa lò rèn .....	10
2.9.1 Kết quả phân tích trình tự DNA mã vạch, xác định SNPs.....	10
2.9.2 Kết quả phân tích sự đa hình chỉ thị ISSR.....	11
2.10 Cây xoài cát hòa lộc.....	11
2.10.1 Kết quả phân tích trình tự DNA mã vạch và SNPs .....	11
2.10.2 Kết quả phân tích đa dạng dấu phân tử ISSR .....	11
<b>CHƯƠNG 3. CƠ SỞ DỮ LIỆU DNA MÃ VẠCH .....</b>	<b>12</b>
<b>PHẦN III.....</b>	<b>13</b>
<b>KẾT LUẬN VÀ ĐỀ NGHỊ .....</b>	<b>13</b>
3.1 Kết luận.....	13
3.2 Đề nghị.....	13

---

## DANH SÁCH CÁC TỪ VIẾT TẮT

<b>AFLP</b>	Amplified fragment length polymorphism
<b>BLAST</b>	Basic local alignment search tool
<b>COI</b>	Cytochrome c oxidase I gene
<b>CTAB</b>	Cetyl trimethylammonium bromide
<b>DNA</b>	Deoxyribonucleic acid
<b>ĐBSCL</b>	Đồng bằng sông Cửu Long
<b>EB</b>	Extraction buffer
<b>EDTA</b>	Ethylenediamine tetraacetic acid
<b>ISSR</b>	Inter Simple Sequence Repeat
<b>ITS</b>	Internal transcribed spacer
<b><i>matK</i></b>	Maturase K gene
<b>NCBI</b>	National Center for Biotechnology Information
<b>Nu</b>	Nucleotide
<b>NTSYSpc</b>	Numerical taxonomy system for personal computer
<b>PCR</b>	Polymerase chain reaction
<b>PVP</b>	Polyvinylpyrrolidone
<b>RAPD</b>	Randomly Amplified Polymorphic DNA
<b>rDNA</b>	Ribosomal DNA
<b>RFLP</b>	Restriction fragment length polymorphism
<b>SNP</b>	Single Nucleotide Polymorphism
<b>SSR</b>	Simple sequence repeat
<b>TAE</b>	Tris-acetate EDTA buffer
<b>TE</b>	Tris EDTA buffer
<b>UPGMA</b>	Unweighted pair group method with arithmetic mean

## THÔNG TIN KẾT QUẢ NGHIÊN CỨU

### 1. Thông tin chung

- Tên đề tài: Xây dựng cơ sở dữ liệu DNA mã vạch cho các giống cây ăn trái đặc sản của Việt Nam khu vực Nam Bộ
- Chủ nhiệm đề tài: TS. Đỗ Tấn Khang – Viện Nghiên cứu và Phát triển Công nghệ Sinh học
- Cơ quan chủ trì: Trường Đại học Cần Thơ
- Thời gian thực hiện: 01/2019 đến 6/2021

### 2. Mục tiêu đề tài

Mục tiêu chung: Xây dựng được cơ sở dữ liệu DNA mã vạch cho 10 loại trái cây đặc sản của Việt Nam khu vực Nam Bộ.

Mục tiêu cụ thể:

Khảo sát và tìm được các cây đầu dòng của 10 loại cây ăn trái đặc sản ở Nam Bộ.

Giải được trình tự DNA mã vạch và phân tích được đa dạng di truyền của từng loại DNA mã vạch của từng loại cây ăn trái đặc sản dựa trên các phần mềm tin sinh học và cơ sở dữ liệu DNA, từ đó xây dựng được cơ sở dữ liệu DNA mã vạch chuẩn có giá trị nhận dạng chính xác 10 loại cây ăn trái đặc sản Nam Bộ.

### 3. Tính mới và sáng tạo

DNA mã vạch là một trong những công cụ ứng dụng mới trong công nghệ sinh học nông nghiệp để nhận dạng giống cây trồng. Việc đưa công cụ này vào nhận diện cây giống và các sản phẩm từ cây ăn trái là một trong những bước tiến của nông nghiệp trong ứng dụng công nghệ 4.0.

Đề tài đã xác định được các vị trí chuyên biệt của từng giống cây ăn trái đặc sản trong nhóm cây ăn trái có nhiều đặc điểm tương đồng cả về hình thái và về gen.

### 4. Kết quả nghiên cứu

- Bưởi Da Xanh: Dựa trên mức độ sai khác về nucleotide, ITS là trình tự tiềm năng để nhận diện giống bưởi Da Xanh. với độ tương đồng dao động trong khoảng 0,66-0,95 với khoảng dao động là 0,07 được chia thành 3 nhóm chính: nhóm I với Nhóm 2 có hệ số tương đồng là 0,76, nhóm I, II với nhóm 3 có hệ số tương đồng thấp nhất là 0,66.

- Cam Mật: gen ycf1b cho thấy có 5 SNPs có thể phân biệt giữa các giống cam Mật với các giống khác. Giản đồ phả hệ dựa trên dấu ISSR cho thấy có sự khác biệt về mặt di truyền giữa các mẫu cam Mật không hạt và các mẫu cam khác trong nghiên cứu.

- Chôm chôm: Gene matK có thể được sử dụng làm DNA mã vạch để nhận diện giống chôm chôm rừng ruột vàng, chôm chôm Java và giống chôm chôm lai Tiến Cường của Việt Nam. Phương pháp phân tích đa hình dựa trên môi ISSR này đã chia tám giống – 14 mẫu chôm chôm thành bốn nhóm lớn. Với hệ số tương đồng giữa các giống dao động từ 72 đến 100%

- Dâu hạ châu: Trình tự atpF-H thấy được trình tự của dâu Hạ Châu khác biệt hoàn toàn với các mẫu dâu đối chứng ở 11 vị trí nucleotide. Dựa trên giản đồ, cho thấy sự đa dạng di truyền giữa 12 mẫu dâu thể hiện qua hệ số tương đồng là 74%. Các mẫu dâu Hạ Châu có hệ số tương đồng với nhau khá cao nằm trong khoảng từ 0,875 - 1 và nằm chung ở 1 nhóm trên giản đồ.

- Mãng cầu xiêm: Kết quả phân tích cho thấy gene rpoC1, kết quả nghiên cứu đã đưa ra 7 vị trí sai khác nucleotide. Trong đó có 2 vị trí đặc trưng cho loài bình bát, 4 vị trí có khả năng nhận diện loài măng cầu. Kết quả trên cho thấy ISSR phù hợp để đánh giá đặc tính của ba loài Annona và có sự khác biệt đáng kể về đa dạng di truyền giữa các loài

- Quýt hồng: Kết quả phân tích trình tự vùng ITS ở cho thấy có 14 vị trí khác nhau giữa quýt hồng và các giống quýt khác. Từ kết quả xây dựng giản đồ phả hệ ISSR cho thấy, mức độ tương đồng của các mẫu biến thiên từ 0,49 đến 1,00.
- Sầu riêng Ri-6: Phát hiện 6 SNPs giữa các trình tự của các mẫu sầu riêng. Trong đó có một SNP đặc trưng cho giống sầu riêng Ri-6 là vị trí 444, ở sầu riêng Ri-6 là G trong khi các giống còn lại là C. Kết quả phân tích dựa vào dấu phân tử ISSR đã phân chia 20 giống sầu riêng thu được ở bốn tỉnh của vùng ĐBSCL chia thành các 5 nhóm có hệ số tương đồng dao động khá xa trong khoảng 0,61 – 0,97.
- Thanh long ruột đỏ: Giống thanh long ruột đỏ Viện Cây ăn quả miền Nam được xác định bởi ba locus, *atpF-H* + *rbcL* + *matK*.
- Vú sữa lò rèn: Khi dựa trên trình tự DNA vùng *atpF-atpH* cho thấy có sự khác biệt giữa năm giống vú sữa Lò Rèn, vú sữa bơ hồng, vú sữa bơ tím, vú sữa tím và vú sữa Mica. Phân tích đa dạng di truyền bằng dấu ISSR cho thấy các giống vú sữa có độ tương đồng từ 0,68 đến 1,00 trong đó các giống vú sữa Lò Rèn được tách thành một nhánh riêng.
- Xoài cát Hòa Lộc: Qua kết quả phân tích vùng trình tự *matK* xuất hiện đa dạng vị trí biến đổi xảy ra ở hầu hết các vị trí ở 16 trình tự, trong đó có 27 vị trí có thể nhận diện xoài cát Hòa Lộc với các mẫu còn lại. Dựa trên dấu phân tử ISSR, các mẫu xoài được chia thành 2 nhóm với hệ số tương đồng trung bình là 0,67.

## 5. Sản phẩm

*Sản phẩm khoa học:*

- 05 bài báo quốc tế

+ 01 bài đăng trên Asian Journal of Plant Sciences 20(3) (2021) – Q3

+ 01 bài đăng trên Biodiversitas 2021 – Q3 (vượt 01 so với thuyết minh)

+ 01 bài đăng trên Asian Journal of Agriculture and Biology 2021(2) – Q4

+ 02 bài đăng trên International Journal of Agriculture and Biological Sciences 04(2021)

(vượt 02 so với đăng ký thuyết minh).

- 02 bài báo trong nước

+ 01 bài đăng trên Tạp chí Khoa học Trường Đại học Cần Thơ

+ 01 bài đăng trên Tạp chí Khoa học và Công nghệ Nông nghiệp Việt Nam

- 01 Sách chuyên khảo: “DNA mã vạch và đa dạng di truyền cây ăn trái”.

*Sản phẩm đào tạo:*

- Đào tạo thành công 06 thạc sĩ (vượt 04 so với đăng ký thuyết minh)

*Sản phẩm ứng dụng:*

- 01 Cơ sở dữ liệu về DNA mã vạch của 10 giống cây ăn trái đặc sản của Việt Nam (<https://dnafv.ctu.edu.vn/>)

- 01 Thư viện DNA mã vạch của 10 giống cây ăn trái đặc sản của Nam Bộ

## 6. Hiệu quả, phương thức chuyển giao kết quả nghiên cứu và khả năng ứng dụng vào thực tế

### 6.1. Hiệu quả

#### Đối với tổ chức chủ trì và các cơ sở ứng dụng kết quả nghiên cứu

Kết quả nghiên cứu này là một bước đột phá của vùng đồng bằng sông Cửu Long nói chung và của Trường đại học Cần Thơ nói riêng. Cơ sở dữ liệu DNA của các loại cây ăn trái đặc sản của Nam Bộ được lưu giữ tại trường làm cơ sở cho các tra cứu về gen cho các nghiên cứu liên quan. Đề tài còn hỗ trợ luận văn và luận án cho học viên cao học và nghiên cứu sinh của trường. Các luận văn và luận án này là nguồn tài liệu tham khảo quý cho các học viên tiếp theo cũng như cho các nhà khoa học cùng lĩnh vực nghiên cứu.

#### Đối với lĩnh vực giáo dục và đào tạo



---

Công nghệ sinh học của Việt Nam đang phát triển khá chậm so với thế giới, do đó những công trình về gene sẽ tăng tốc cho sự phát triển của khoa học nước ta. Đề tài cho ra các sản phẩm khoa học là các bài báo và sách về công nghệ DNA của thực vật. Đây là một trong những lĩnh vực quan trọng của công nghệ sinh học, có thể sử dụng như các tài liệu giảng dạy ở các cơ sở đào tạo ngành công nghệ sinh học và các ngành gần có liên quan.

**Đối với lĩnh vực khoa học và công nghệ có liên quan**

Xác định và lưu giữ cây đầu dòng cho việc sản xuất và nhân giống cây ăn trái ở khu vực Nam Bộ. Kết quả nghiên cứu tạo tiền đề quan trọng cho hướng nghiên cứu ứng dụng chỉ thị phân tử vào việc phân loại, giám định, đánh giá đa dạng di truyền, bảo tồn và quản lý thương mại nguồn tài nguyên sinh vật ở nước ta, cụ thể là các loại cây ăn trái đặc sản trên toàn lãnh thổ.

**Đối với phát triển kinh tế-xã hội**

Ngoài giá trị khoa học xác định giống cây ăn trái, một hệ thống quản lý giống dựa vào DNA mã vạch sẽ giúp chúng ta quản lý được tài nguyên di truyền để khai thác hiệu quả nhất và có đủ cơ sở khoa học để chứng minh được quyền sở hữu quốc gia đối với nguồn tài nguyên sinh học, khi Việt Nam gia nhập tổ chức thương mại quốc tế.

**6.2. Phương thức chuyển giao kết quả**

- Cơ sở dữ liệu DNA mã vạch của 10 loại cây ăn trái được công khai trên internet và được xem như chứng minh thư của các loại cây ăn trái này. Các tổ chức nghiên cứu trong nước và thậm chí nước ngoài có thể sử dụng làm trình tự tham khảo khi nghiên cứu về gen của các cây trồng tương tự.
- Có khả năng ứng dụng trong quản lý giống cây trồng ở các trung tâm giống và trung tâm ứng dụng tiến bộ khoa học công nghệ.

**Cơ quan chủ trì**

**Chủ nhiệm đề tài**

**TS. Đỗ Tấn Khang**

## INFORMATION ON RESEARCH RESULTS

### 1. General information

- Project title: Establishing DNA barcodes database of special fruits in Southern area, Vietnam
- Coordinator: PhD. Do Tan Khang
- Implementing institution: Can Tho University
- Duration: from 01/2019 to 6/2021

### 2. Objective(s)

General objective: Establishing DNA barcodes database of 10 special fruits in Southern of Vietnam.

Detailed objectives:

Evaluating the status of 10 approved special fruit trees in Southern of Vietnam.

Sequencing DNA barcodes and analysing genetic diversity of each DNA barcode of fruits based on bioinformatics softwares for establishing DNA barcodes database which can be applied in authenticating 10 special fruits in Southern of Vietnam.

### 3. Creativeness and innovativeness

DNA barcoding and metabarcoding have potential in the context of quality control of both well and poorly regulated supply systems. Standardisation of protocols for DNA barcoding and DNA sequence-based identification are necessary before DNA-based biological methods can be implemented as routine analytical approaches and approved by the competent authorities for use in regulated procedures.

### 4. Research results

- “Da xanh” pomelo: Based on the degree of nucleotide difference, ITS is a potential sequence to identify the green-skinned pomelo variety. The similarity ranging from 0.66-0.95 with a range of 0.07 divided into 3 main groups: group I with Group 2 with similarity coefficient of 0.76, group I, II with similarity coefficient of 0.76. group 3 has the lowest similarity coefficient of 0.66.
- “Mat” orange (*Citrus sinensis* L.): the *ycf1b* gene shows that there are 5 SNPs that can distinguish honey orange varieties from other varieties. The pedigree based on the ISSR marker showed that there were genetic differences between the seedless orange samples and the other orange samples in the study.
- “Nhan” rambutan: The *matK* gene can be used as a barcoded DNA to identify the yellow-fleshed forest rambutan, Java rambutan and Tien Cuong hybrid rambutan from Vietnam. This ISSR primer-based polymorphism analysis divided eight varieties – 14 rambutan samples into four large groups. With similarity coefficients between varieties ranging from 72 to 100%.
- “Ha Chau” Burmese grape: *AtpF-H* sequence showed that the sequence of “Ha Chau” Burmese grape was completely different from the control samples at 11 nucleotide positions. Based on the diagram, it shows that the genetic diversity between 12 strawberry samples is expressed through a similarity coefficient of 74%. “Ha Chau” mulberry samples have a relatively high similarity coefficient, ranging from 0.875 to 1 and are in the same group on the diagram.
- “Xiem” soursop (*Annona* spp.): The analysis results showed that the *rpoC1* gene, the research results gave 7 positions of nucleotide difference. In which, there are 2 positions specific to the species of bowl, 4 positions are capable of identifying the custard apple species. The above results show that the ISSR is suitable for characterization of three *Annona* species and there is a significant difference in genetic diversity between species.

- “Hong Lai Vung” mandarin: The results of sequencing analysis in the ITS region show that there are 14 different positions between “Hong Lai Vung” mandarin and other varieties of mandarin. From the results of ISSR marker, the similarity level of the samples varied from 0.49 to 1.00.
- “Ri-6” durian: Detecting 6 SNPs between sequences of durian samples. Among them, there is a specific SNP for durian variety Ri-6 at position 444, in durian Ri-6 it is G while the rest is C. Analysis results are based on molecular markers ISSR has divided 20 Durian varieties obtained in four provinces of the Mekong Delta were divided into five groups with similarity coefficients ranging from 0.61 to 0.97.
- “Ruot do” dragon fruit: Red flesh dragon fruit variety was identified by three loci, *atpF-H + rbcL + matK*.
- “Lo Ren” star apple: Based on the DNA sequence of *atpF-atpH* region, there are differences between five varieties of Lo Ren suckling, pink buttermilk, purple buttermilk, purple breast, and Mica. Genetic diversity analysis using ISSR markers showed that all varieties have similarity from 0.68 to 1.00 in which the Lo Ren suckling varieties were classified into a separated clade.
- “Hoa Loc” mango: Through the analysis of the *matK* sequence region, a variety of variable positions occurred in most of the 16 sequences, of which 27 positions could identify Hoa Loc mango. with the rest of the samples. Based on the ISSR molecular marker, the mango samples were divided into 2 groups with an average similarity coefficient of 0.67.

## 5. Products

### *Scientific products:*

#### - 05 international articles

- + 01 article published in Asian Journal of Plant Sciences 20(3) (2021) – Q3
- + 01 article published in Biodiversitas 2021 – Q3
- + 01 article published in Asian Journal of Agriculture and Biology 2021(2) (2021) – Q4
- + 02 article published in International Journal of Agriculture and Biological Sciences 04 (2021)

#### - 02 national articles

- + 01 article published in Scientific Journal of Can Tho University
- + 01 article published in Journal of Sciences and Agricultural Technology of Vietnam
- 01 book: “DNA barcode and genetic diversity of fruits”.

### *Educational products:*

- Supporting for 06 master theses.

### *Applicable products:*

- Library of DNA barcodes of 10 fruit trees.
- Database of DNA barcodes of 10 fruit varieties (<https://dnarf.ctu.edu.vn/>).

## 6. Effects, transfer alternatives of research results and applicability

### 6.1 Effects

- For organizations: storing and maintaining genetic resources of fruit trees in order to use for research and agricultural production.
- The applied institutions: the result brings scientific and practical benefits to researchers and farmers, and it helps farmers in selecting of varieties.
- Evaluating and concluding the feasible of DNA barcode application in agriculture.

### 6.2 Method of transferring results

DNA barcoding can be used for authenticating products based on single herbal ingredients and DNA metabarcoding for assessment of species diversity in processed products, and both methods should be used in combination with appropriate hyphenated chemical methods for quality control. Standardisation of protocols for DNA barcoding and DNA sequence-based identification are

---

necessary before DNA-based biological methods can be implemented as routine analytical approaches and approved by the competent authorities for use in regulated procedures.

## PHẦN I. MỞ ĐẦU

### 1. Lý do chọn đề tài

Cây ăn quả chủ lực trồng tập trung ở Nam Bộ được Bộ Nông nghiệp và Phát triển nông thôn phê duyệt, trong đó 12 loại cây ăn trái chủ lực được đề nghị bao gồm thanh long, xoài, chôm chôm, sầu riêng, vú sữa, bưởi, nhãn, chuối, dứa (khóm), cam, măng cầu và quýt. Tổng diện tích cây ăn quả chủ lực trồng tập trung là 257.000 ha, trong đó vùng Đồng bằng sông Cửu Long 185.100 ha, vùng Đông Nam Bộ 71.900 ha. Trong đó, diện tích trồng thanh long là 24.800 ha, xoài 45.900 ha, chôm chôm 18.300 ha, sầu riêng 15.000 ha, vú sữa 5.000 ha, bưởi 27.900 ha, nhãn 29.800 ha, chuối 28.900 ha, dứa 21.000 ha, cam 26.250 ha, măng cầu 8.300 ha và quýt 5.850 ha. Quy hoạch đặt mục tiêu xây dựng ngành hàng trái cây chủ lực trồng tập trung ở Nam Bộ theo hướng sản xuất hàng hóa lớn trên cơ sở phát huy lợi thế so sánh, nâng cao sức cạnh tranh của sản phẩm trên thị trường trong nước và nước ngoài. Ngoài ra, trái cây Việt Nam được người tiêu dùng thế giới đánh giá cao về chất lượng và mang tính đặc trưng của vùng. Trong thời gian qua, nhiều loại trái cây Việt Nam đã được thâm nhập và mở rộng xuất khẩu vào các thị trường như Hoa Kỳ, Úc, Malaysia, Châu Âu (Pháp, Đức, Anh, Hà Lan), Nhật Bản, Hàn Quốc, Singapore...

DNA mã vạch, một khái niệm được đưa ra bởi Heber (2003), là trình tự nucleotide của một chuỗi DNA ngắn, có cùng nguồn gốc tổ tiên (orthologous), trong đó có vùng ít bị thay đổi (rất ổn định – bảo thủ) và có vùng dễ thay đổi trong quá trình tiến hóa. Dựa vào mức độ thay đổi trong trình tự DNA này để đánh giá sự sai khác di truyền giữa các sinh vật. Như vậy, DNA mã vạch là một phương pháp định danh mới sử dụng một hoặc nhiều đoạn DNA chuẩn ngắn nằm trong hệ genome của sinh vật đang nghiên cứu, nhằm xác định sinh vật đó thuộc về loài nào. Sau khoảng 10 năm nghiên cứu và phát triển DNA mã vạch, đến nay các nhà khoa học đã công bố trên hàng nghìn công trình khoa học trên các tạp chí khoa học chuyên ngành, với hơn 3.483.696 trình tự mã vạch DNA ở 215.513 loài sinh vật, trong đó động vật có 144.402 loài, thực vật có 54.478 loài, nấm và các dạng sinh vật khác có 16.633 loài. Nhiều kết quả nghiên cứu đã cho thấy, có nhiều đoạn DNA đặc trưng được sử dụng làm DNA mã vạch, các đoạn DNA mã vạch có thể là những đoạn DNA nằm ở trong nhân như: 18S, 5,6S, 26S, 5S và vùng ITS; nằm ở ty thể như: *Cytb* và vùng kiểm soát (control region); nằm ở lục lạp như: *matK*, *rcbL*, *atpB*, *ndnF* (Cuenoud, 2002; Kress *et al.*, 2008; Aron *et al.*, 2008; Spooner *et al.*, 2009). Có khoảng 8 locus gen đã được sử dụng làm mã vạch DNA ở các loài thực vật, bao gồm cả hệ gen nhân và hệ gen lục lạp (vùng xen *atpF-atpH*, gen *matK*, gen *rcbL*, gen *rpoC1*, vùng xen *psbK-psbI*, vùng xen *trnH-psbA* và vùng gen nhân ITS).

Hướng nghiên cứu xây dựng cơ sở dữ liệu DNA mã vạch đang được nhiều quốc gia, nhiều nhà khoa học trên thế giới rất quan tâm phát triển, đặc biệt trong những năm gần đây và sẽ là một xu thế nghiên cứu trong thời gian tới. DNA mã vạch được xem là một công cụ mới, hỗ trợ có hiệu quả trong nghiên cứu về phân loại, phát hiện loài mới, giám định loài và các mẫu có nguồn gốc từ sinh vật sống hoặc đã chết thậm chí đã qua chế biến, vì vậy mã vạch DNA có rất nhiều ứng dụng trong nghiên cứu cũng như thực tiễn (Nimis, 2010; Bruni, 2010; Bell, 2011; Hebert, 2003; Lahaye, 2008; Liu, 2010). Ưu điểm của công nghệ này là định danh loài nhanh chóng. Nếu xây dựng mẫu chuẩn thì trong vòng 4 giờ đã cho kết quả rất chính xác. Có thể phát triển và áp dụng nó ở các tổ chức hoạt động trong lĩnh vực kiểm định.

Để đáp ứng nhu cầu hội nhập và giữ vững thị phần trên thị trường quốc tế, trái cây của Việt Nam cần được đảm bảo về an toàn, chất lượng lẫn hình thức, trong đó việc truy xuất nguồn gốc là vấn đề hết sức quan trọng. Hơn nữa, việc chứng minh chất lượng của trái cây được sản xuất từ giống gốc hoặc cây đầu dòng vô cùng phức tạp và gặp nhiều khó khăn, do đó cần phải có một phương pháp đáng tin cậy để giải quyết vấn đề này. Chính vì thế, việc xây dựng cơ sở dữ liệu

DNA mã vạch cho các loại cây ăn trái đặc sản của Việt Nam là một nhiệm vụ vô cùng quan trọng và cấp thiết hiện nay.

## **2. Mục tiêu nghiên cứu**

Mục tiêu chung: Xây dựng được cơ sở dữ liệu DNA mã vạch cho 10 loại trái cây đặc sản của Việt Nam khu vực Nam Bộ.

Mục tiêu cụ thể:

- Khảo sát và tìm được các cây đầu dòng của 10 loại cây ăn trái đặc sản ở Nam Bộ.
- Giải được trình tự DNA mã vạch và phân tích được đa dạng di truyền của từng loại DNA mã vạch của từng loại cây ăn trái đặc sản dựa trên các phần mềm tin sinh học và cơ sở dữ liệu DNA, từ đó xây dựng được cơ sở dữ liệu DNA mã vạch chuẩn có giá trị nhận dạng chính xác 10 loại cây ăn trái đặc sản Nam Bộ.

## **3. Cách tiếp cận**

Dựa trên nhu cầu về đăng ký sở hữu trí tuệ các loại cây ăn trái đặc sản và tính cấp thiết về xây dựng cơ sở dữ liệu DNA cho quốc gia, đề tài được thực hiện bằng việc khảo sát tìm cây đầu dòng có nhiều đặc tính tốt, đặc biệt là phẩm chất trái để thu mẫu DNA. Việc phân tích sẽ tìm ra các trình tự hoặc vùng gen đặc trưng cho loại cây ăn trái để đăng ký và lưu giữ trên cơ sở dữ liệu.

## **4. Đối tượng và phạm vi nghiên cứu**

### **Đối tượng nghiên cứu**

DNA của các loại cây ăn trái gồm thanh long ruột đỏ, xoài cát Hòa Lộc, sầu riêng Ri-6, vú sữa lò rèn, bưởi Da Xanh ruột hồng, dâu hạ châu, chôm chôm đường, cam Mật, quýt hồng và măng cầu Xiêm.

### **Phạm vi nghiên cứu**

Các loại cây ăn trái được trồng ở vùng đồng bằng sông Cửu Long và một số tỉnh vùng Tây Nam Bộ. Cụ thể:

Thanh long ruột đỏ (Tiền Giang, Bến Tre, Long An, Đồng Nai)

Xoài cát Hòa Lộc (Đồng Tháp, Tiền Giang, Cần Thơ, Vĩnh Long)

Sầu riêng Ri-6 (Bến Tre, Tiền Giang, Vĩnh Long)

Vú sữa lò rèn (Tiền Giang, Bến Tre, Vĩnh Long, Cần Thơ)

Bưởi Da Xanh ruột hồng (Bến Tre, Tiền Giang, Vĩnh Long)

Chôm chôm đường (Bến Tre, Tiền Giang, Vĩnh Long)

Cam Mật (Bến Tre, Tiền Giang, Vĩnh Long)

Quýt hồng (Đồng Tháp, Vĩnh Long, Tiền Giang)

Măng cầu Xiêm (Tiền Giang, Vĩnh Long, Cần Thơ, Bến Tre)

Dâu hạ châu (Cần Thơ, Hậu Giang)

## **5. Nội dung nghiên cứu**

### **Nội dung 1: Khảo sát cây đầu dòng của 10 loại cây ăn trái đặc sản**

### **Nội dung 2: Giải trình tự DNA mã vạch của các loài cây ăn trái để có thể nhận dạng giống cây ăn trái.**

Ly trích DNA

Phân tích dấu phân tử ISSR

Giải trình tự DNA mã vạch

### **Nội dung 3: Phân tích dữ liệu và xây dựng cơ sở dữ liệu**

Phân tích đa dạng di truyền của từng loại DNA mã vạch của từng loại cây ăn trái

Xây dựng cơ sở dữ liệu DNA mã vạch cây ăn trái

## **6. Phương pháp nghiên cứu**

### **6.1. Phương tiện**

#### **6.1.1. Hóa chất**

- Ly trích DNA: Dung dịch trích DNA thực vật EB (Extraction Buffer), SDS 10%, isopropanol, TE 1X, CTAB buffer 2%, ethanol tuyệt đối, ethanol 70%, TE 0,1X.
- Thành phần PCR: Nước khử ion, Mytaq master mix 2X (Bioline, Anh), oligonucleotide (IDT, Mỹ)
- Điện di: Tris Acetic EDTA 50X, dung dịch tải mẫu Loading buffer, thang DNA chuẩn 100 bp (Bioline, Anh).

#### 6.1.2 Thiết bị

- Máy nghiền mẫu, bể ủ nhiệt, máy ly tâm (Eppendorf, Đức), máy ly tâm chân không, hệ thống nước khử ion, máy luân nhiệt (Bio-rad C1000, Mỹ), máy vortex, bể điện di (Embitech, Mỹ), hệ thống chụp hình gel (Bio-rad Gel doc XR, Mỹ)

#### 6.1.3 Dụng cụ

- Bộ micropipette, chai nắp xanh, ống đông thủy tinh, đầu cone xanh, đầu cone vàng, tube Eppendorf 2,0 mL và 1,5 mL, tube PCR 200  $\mu$ L.

### 6.2 Phương pháp

#### 6.2.1 Ly trích DNA thực vật

Quy trình được thực hiện theo mô tả của Roger và Bendich (1988) với một số cải tiến.

#### 6.2.2 Điện di DNA

#### 6.2.3 Khuếch đại các đoạn DNA mã vạch

##### a) Phương pháp tiến hành

- Khuếch đại 8 trình tự DNA mã vạch bằng kỹ thuật PCR bao gồm: atpF-atpH, rbcL, pbA-trnH, matK và rpoC1.
- Sản phẩm PCR được kiểm tra bằng kỹ thuật điện di trên gel agarose 2% ở hiệu điện thế 50V trong 30 phút, sau đó quan sát các băng DNA bằng máy chụp hình gel (Bio-Rad Gel Doc XRTM imaging system, Mỹ).
- Phổ điện di các sản phẩm PCR được quan sát trên phần mềm Quantity One (Bio-Rad) phiên bản 4.6.1.
- Sản phẩm PCR được tinh sạch và giải trình tự theo phương pháp Sanger tại công ty Nextgen

#### 6.2.4 Phân tích đa dạng di truyền bằng chỉ thị phân tử ISSR

Thành phần phản ứng PCR gồm 20  $\mu$ L với các thành phần: BiH<sub>2</sub>O; Mix (gồm Buffer KCl, MgCl<sub>2</sub>, dNTPs, Taq polymerase); mỗi ISSR; DNA đã được kiểm tra độ tinh sạch.

Tám mỗi ISSR dùng trong nghiên cứu được lấy từ bộ primer của phòng thí nghiệm Công nghệ Sinh học phân tử.

- Sản phẩm PCR được điện di trên gel agarose 3% với điện thế 25V trong 120 phút và quan sát dưới tia UV.

#### 6.2.5 Phân tích dữ liệu

##### a) DNA mã vạch

- Trình tự DNA được hiệu chỉnh bằng phần mềm Bioedit phiên bản 7.0, các trình tự nhiễu ở 2 đầu được loại bỏ, chất lượng trình tự được kiểm tra qua độ cao của peak nucleotide.
- Các trình tự DNA được dàn hàng (Alignment) bằng thuật toán Clustal W trên phần mềm Bioedit.
- Sử dụng chương trình MEGA X để xác định các vị trí bảo tồn và biến đổi, số SNPs, đột biến Indels.

##### b) Chỉ thị phân tử ISSR

- Các phổ điện di của các sản phẩm PCR được mã hóa theo hệ nhị phân bằng chương trình Excell 2013. Giản đồ phân nhánh được xây dựng theo phương pháp UPGMA để so sánh sự liên quan di truyền của các dòng khác nhau nhằm phát hiện sự đồng dạng.

## PHẦN II. KẾT QUẢ VÀ THẢO LUẬN

### CHƯƠNG I. KHẢO SÁT CÂY ĐÀU DÒNG CỦA 10 LOẠI CÂY ĂN TRÁI ĐẶC SẢN

#### 1.1 Khảo sát cây đầu dòng của 10 loại cây ăn trái đặc sản

Trong những năm gần đây, Viện Cây ăn quả miền Nam đã phối hợp cùng Sở Nông nghiệp & PTNT, Chi cục TT & BVTV, Trung tâm giống, TT ứng dụng Công nghệ cao các tỉnh tại ĐBSCL và ĐNB đã bình tuyển và được Sở NN & PTNT công nhận Cây đầu dòng trên nhiều chủng loại cây ăn quả. Các cây đầu dòng của 10 loại cây ăn trái trong nghiên cứu đã được xác định vị trí.

#### 1.2 Đặc tính nông học cây đầu dòng

Các đặc tính nông học đã được khảo sát và đã được mô tả dưới dạng bảng.

### CHƯƠNG 2. KẾT QUẢ PHÂN TÍCH TRÌNH TỰ DNA MÃ VẠCH VÀ DI TRUYỀN

#### 2.1 Cây bưởi Da Xanh

##### 2.1.1 Kết quả phân tích trình tự DNA mã vạch và SNPs

Kết quả phân tích vùng trình tự *ycf1b* của các giống bưởi trong nghiên cứu cho thấy không có sự khác biệt quá lớn trong cùng một giống cũng như giữa các giống bưởi khác nhau của các đối tượng được nghiên cứu. Điểm sai khác duy nhất trong trình tự của các giống nghiên cứu xuất hiện ở trình tự của bưởi Ruby khi alignment có một sự sai khác xuất hiện tại vị trí 754. Tuy nhiên qua kiểm tra lại peak vị trí này nằm ở peak thứ 754 trên trình tự bưởi Ruby vị trí này nằm trong 30 nucleotit cuối của trình tự nên độ tin cậy cho sự sai khác này là không cao. Từ kết quả thu được có thể suy luận rằng vùng trình tự *ycf1b* có tính bảo tồn cao giữa các giống bưởi thu được trong khu vực ĐBSCL.

##### 2.1.2 Đa dạng di truyền dấu phân tử ISSR

Theo kết quả với độ tương đồng dao động trong khoảng 0,66-0,95 với khoảng dao động là 0,07 được chia thành 3 nhóm chính: nhóm I với Nhóm 2 có hệ số tương đồng là 0,76, nhóm I, II với nhóm 3 có hệ số tương đồng thấp nhất là 0,66.

Từ giản đồ phá hệ cho thấy bưởi Da Xanh tập trung ở nhóm I và Nhóm II tách biệt hoàn toàn với nhóm III với hệ số tương đồng là 0,66 nên có thể sử dụng dấu phân tử ISSR để phân biệt bưởi Da Xanh Với 2 giống bưởi của nhóm III là bưởi Ruby và bưởi 5 Roi. Vì mẫu BDTG được thu từ Viện Cây Ăn Quả miền Nam còn 4 mẫu bưởi Da Xanh còn lại được thu trực tiếp tại vườn đây là nhóm bưởi Da Xanh phục vụ cho sản xuất. Nguyên nhân dẫn đến sự khác biệt của bưởi Ruby so với bưởi năm roi cũng như bưởi Da Xanh là do bưởi Ruby là giống cây có nguồn gốc từ thái lan được đem cây giống về Việt Nam không lâu nên dẫn đến việc có khác biệt di truyền so với các giống bưởi bản địa như bưởi Da Xanh và bưởi năm roi.

#### 2.2 Cây cam Mật

Các mẫu trái cam, lá cam Mật, cam xoàn, cam sành, cam dây, cam cara được thu tại Cần Thơ, Hậu Giang, Bến Tre, Tiền Giang.

##### 2.2.1 Kết quả phân tích trình tự DNA mã vạch và SNPs

Dựa vào kết quả phân tích vùng gen *ycf1b* cho thấy có 8 vị trí SNPs có thể phân biệt giữa các giống cam. Trong đó có 5 vị trí (577, 594, 597, 637 và 638) cho thấy cam Mật khác với các giống cam khác. Ở vị trí 577 và 594 nucleotide của cam Mật lần lượt là T và C, trong khi các giống cam khác là G. Ở vị trí 637 và 638 các giống cam khác mất 2 nucleotide G so với cam Mật. Đây là các vị trí SNPs quan trọng có thể dùng để phân biệt cam Mật và các giống cam khác trong nghiên cứu.

Các trình tự ITS, *rpoC1* không tìm thấy sự khác biệt giữa các giống cam nghiên cứu.

Ở vùng trình tự *atpF-atpH* có 26 vị trí sai khác của nucleotide giữa các trình tự. Các vị trí sai khác lần lượt là 420; 446; 449; 452; 457; 458; 460; 465; 469; 470; 471; 496; 503; 506; 512; 523; 531; 536; 537; 540; 549; 550; 553; 572; 574 và 575.



Ở vùng trình tự *trnH-psbA*, có 12 vị trí sai khác của nucleotide giữa các trình tự. Các vị trí sai khác lần lượt là 338; 340; 353; 358; 369; 372; 380; 381, 382; 383; 384 và 389.

Ở một số nghiên cứu khác trên vùng gen *psbA-trnH*, Vũ Huyền Trang và cs. (2013) đã nghiên cứu xây dựng mã vạch DNA cho Sâm Ngọc linh trên cơ sở 5 chỉ thị DNA mã vạch *psbA-trnH*, *matK*, *trnL*, *rbcL* và ITS. Nhóm tác giả đã chứng minh trong 5 chỉ thị mã vạch nghiên cứu, *psbA-trnH* là chỉ thị có tiềm năng nhất, cho phép phân biệt Sâm Ngọc linh với các loài sâm khác trên thế giới với độ chính xác cao.

Vùng *psbK-psbI* có 18 vị trí sai khác của nucleotide giữa các trình tự, từ vị trí 206 đến 213 các giống cam nghiên cứu đều bị đột biến mất nucleotide, chỉ có giống cam dây không bị mất. Ngược lại từ vị trí 277 đến 284; 358 và 359 các giống cam đều xuất hiện nucleotide, chỉ giống cam dây bị mất đột biến mất nucleotide.

#### 2.2.2 Đa dạng di truyền các giống cam dựa vào dấu phân tử ISSR

Dấu phân tử ISSR được Capparelli *et al.* (2004) sử dụng để phân tích mối quan hệ di truyền của các mẫu chanh thu thập tại Italy, góp phần vào việc phân biệt và chọn các giống khác nhau về các đặc điểm hình thái nông học cũng như góp phần vào nguồn thông tin cho việc tiến hóa và nguồn gốc phát sinh loài.

Kết quả phân tích dấu ISSR trên các mẫu cam cho thấy ISSR03, ISSRK2 và ISSRK3 đều xuất hiện các băng đa hình cho thấy sự đa dạng di truyền giữa các giống và đủ điều kiện để xây dựng bản đồ phả hệ.

Mức độ tương đồng của các mẫu biến thiên từ 0,89 đến 1. Tại mức tương đồng 0,89 có sự phân chia nhóm thành hai nhóm chính và nhiều nhóm phụ.

### 2.3 Cây chôm chôm Nhãn

Thí nghiệm được thực hiện trên mẫu lá tươi của các giống chôm chôm bao gồm: Chôm chôm Nhãn, chôm chôm Java, chôm chôm Thái, chôm chôm Tta, chôm chôm Indo, chôm chôm rừng ruột vàng và chôm chôm vỏ vàng. Thu mẫu lá cây của các giống trên tại các vườn chôm chôm đầu dòng ở các vùng trọng điểm trồng chôm chôm của cả nước, bao gồm: huyện Chợ Lách - tỉnh Bến Tre, cù lao An Bình - tỉnh Vĩnh Long, huyện Phong Điền - thành phố Cần Thơ và huyện Krông Năng - tỉnh Đắk Lắk.

#### 2.3.1 Kết quả phân tích vùng trình tự DNA mã vạch

Kết quả phân tích cho thấy, bằng việc sử dụng cặp mồi *matK-4600/trnK-2R*, có chín vị trí SNP đáng tin cậy được ghi nhận. Các vị trí SNP khác được thể hiện trong phần phụ lục 4, tuy nhiên khi kiểm tra song song trình tự tại các vị trí này, nhận thấy các peak đều bị nhiễu hoặc không rõ, cũng như không có độ tin cậy cao do trình tự giống lặp lại không trùng khớp. Vì những lý do trên, chỉ có 9 SNP được sử dụng để phân tích khả năng nhận diện của cặp mồi trên các giống loài chôm chôm trong nghiên cứu.

Trong đó, có một số SNP chỉ hiện diện ở một giống duy nhất, thể hiện tính đặc trưng cho giống, có thể kể đến là: Vị trí 454 sau khi alignment, trong khi các giống loài chôm chôm khác đều là nu. A, thì nu. G lại xuất hiện ở duy nhất giống chôm chôm rừng. Vị trí 637, chỉ có giống chôm chôm Java xuất hiện nu. A. Tại các vị trí 669, 670 và 718 chỉ có giống chôm chôm lai Tiến Cường xuất hiện SNP đặc trưng lần lượt là C, A và A.

Từ kết quả này, một kết luận được đưa ra: Cặp mồi *matK-4600/trnK-2R* khuếch đại gen *matK*, có thể được sử dụng làm DNA mã vạch để nhận diện giống chôm chôm rừng ruột vàng, chôm chôm Java và giống chôm chôm lai Tiến Cường của Việt Nam. Ngoài ra, những dấu SNP này còn là cơ sở để vẽ cây quan hệ di truyền, từ đó khảo sát được sự đa dạng về kiểu gen của các giống chôm chôm được sử dụng trong đề tài.

#### 2.3.2 Kết quả phân tích đa hình bằng dấu phân tử ISSR

Để khuếch đại DNA của các mẫu chôm chôm Việt Nam nhằm phân tích đa hình, 2 môi ISSR đã được sử dụng theo lược khảo từ nghiên cứu của Madihah *et al.* (2018). Nghiên cứu của Madihah đã chỉ ra 6 môi ISSR cho kết quả biểu hiện thành công sự đa hình giữa các giống chôm chôm, trong đó 2 môi ISSR 10 và ISSR 23 là cho tỉ lệ băng đa hình cao nhất, lần lượt là 92% và 100% băng đa hình trên tổng số băng. Dựa trên kết quả này, đề tài cũng khảo sát đa hình các giống chôm chôm tại Việt Nam bằng 2 môi ISSR 10 và ISSR 23.

Phương pháp phân tích đa hình dựa trên môi ISSR này đã chia tám giống – 14 mẫu chôm chôm thành bốn nhóm lớn. Với hệ số tương đồng giữa các giống dao động từ 72 đến 100%. Ở mức tương đồng 84,8%, có hai giống chôm chôm thuộc nhóm 1, nhóm 2 có bốn giống, nhóm 3 chỉ có giống chôm chôm rừng và nhóm 4 chỉ gồm duy nhất giống lai Tiên Cường.

#### **2.4 Cây dâu hạ châu**

Tổng cộng đã thu được 12 mẫu dâu (dâu Hạ Châu, dâu đỏ, dâu xanh, dâu vàng, dâu Xiêm) từ 2 tỉnh Hậu Giang, Bến Tre và Thành phố Cần Thơ.

##### **2.4.1 Kết quả phân tích vùng trình tự DNA mã vạch**

Ở các vị trí nucleotide số 393, 716, 723, 736, 742, 743, 745, 747, 762, 767 và 768 của dâu Hạ Châu khác biệt hoàn toàn với các mẫu dâu còn lại.

Dựa vào kết quả trên, thấy được trình tự của dâu Hạ Châu khác biệt hoàn toàn với các mẫu dâu đối chứng ở 11 vị trí nucleotide. Vùng trình tự *atpF-atpH* cũng đã xác thực thành công một số cây thuốc trong nghiên cứu công bố trước đây (Tehen *et al.*, 2014). Qua đó có thể nói vùng gen *atpF-atpH* có khả năng được sử dụng để phân biệt dâu Hạ Châu với các giống dâu khác. Ở vùng trình tự *psbK-psbI* cho thấy được sự khác biệt hoàn toàn giữa trình tự của dâu Hạ Châu với các mẫu dâu đối chứng tại vị trí nucleotide số 94, 95, 113, 121, 215 và 431.

Kết quả phân tích trình tự gen *rbcL* ở các mẫu dâu cho thấy, chỉ có vị trí nu 35 có thể phân biệt được dâu hạ châu (nu A) và các dâu khác (nu C).

##### **2.4.2 Đa dạng di truyền dựa vào dấu phân tử ISSR**

Kết quả điện di PCR của 8 môi ISSR trên 12 mẫu dâu, cho thấy có 7 môi cho kết quả đa hình, mỗi ISSR 27 cho ra kết quả đơn hình. Tổng số băng được khuếch đại là 91 băng trong đó có 55 băng cho kết quả đa hình (chiếm tỉ lệ 60,44%), trung bình 6,88 băng đa hình cho mỗi môi. Kết quả này là khá tốt khi ở nghiên cứu của Hoàng Đăng Hiếu *et al.* (2016) khi sử dụng 10 môi ISSR trong việc đánh giá di truyền ở quần thể Ba Kích nhưng chỉ có 6 môi cho ra kết quả đa hình.

Trong 8 môi, mỗi ISSR 03 thể hiện sự đa hình giữa các mẫu cao. Tất cả các băng được khuếch đại đều là băng đa hình (chiếm tỉ lệ 100%) và có chỉ số PIC = 0,52 cao nhất trong các mẫu. Trọng lượng phân tử lớn nhất ở kích thước khoảng 850 bp và nhỏ nhất khoảng 220 bp. Hầu hết ở các mẫu dâu đều có sự khác biệt về số lượng băng và vị trí kích thước trên gel. Do đó, có thể sử dụng môi này cho các nghiên cứu đa dạng trên các giống dâu sau này.

Ngoài ra, các môi ISSR 12, ISSR 22, ISSR 31, ISSR K1, ISSR K2 cũng cho thấy sự đa hình rõ nét, băng đa hình chiếm tỉ lệ khá cao lần lượt là 54,55%, 63,64%, 69,23%, 68,75% và 55,56%. Đặc biệt ở hai môi ISSR 22 và ISSR K2 có sự khác biệt so với các môi còn lại là chỉ 11/12 mẫu dâu được khuếch đại, mẫu dâu Xiêm (Xi2) không cho băng trên gel, chứng tỏ cả hai môi ISSR 22 và ISSR K2 đều không bắt cặp với DNA khuôn từ mẫu dâu này hay nói cách khác trên bộ gen của mẫu này không có trình tự bổ sung bởi hai môi ISSR này.

Còn lại môi ISSR K3 cho ra kết quả đa hình nhưng không cao (chiếm tỉ lệ 22,22%) và có chỉ số PIC ở mức độ thấp (0,14), hầu hết các mẫu đều khuếch đại cho số băng giống nhau, chỉ có vài mẫu có sự khác biệt nhỏ. Và môi ISSR 27 cho ra kết quả đơn hình ở tất cả các băng được khuếch đại.

Dựa trên giản đồ, cho thấy sự đa dạng di truyền giữa 12 mẫu dâu thể hiện qua hệ số tương đồng là 74% hay nói cách khác 12 mẫu dâu này có khoảng cách di truyền là 26%.

Qua các kết quả trên, ta có thể thấy một vài mẫu dâu có độ tương đồng rất cao lên đến 100%, tuy nhiên giữa một số mẫu khác có độ khác biệt tương đối cao ở cấp độ dưới loài. Chứng tỏ sự đa dạng di truyền của các mẫu dâu nghiên cứu khá cao.

## 2.5 Cây măng cầu xiêm

Có 5 giống cây được sử dụng trong thí nghiệm: măng cầu ta, măng cầu Xiêm, na, na Đài Loan và bình bát. Các mẫu lá được thu ở các vườn trái cây, trại cây giống tại TP Cần Thơ, tỉnh Hậu Giang và tỉnh Vĩnh Long.

### 2.5.1 Kết quả phân tích trình tự DNA mã vạch và SNPs

Kết quả cho thấy sự đa dạng di truyền được thể hiện rõ giữa các loài khác nhau, gồm có 9 vị trí nucleotide khác biệt (SNPs) 6, 105, 111, 135, 202, 218, 219, 438, 441. Trong đó, vị trí nucleotide 6, 105, 111 ở trình tự loài bình bát tương ứng là T, T, C có sự sai khác với loài măng cầu và na là nucleotide C, A, T. Trong các trình tự loài măng cầu, có 6 vị trí SNPs so với loài na là các vị trí 135, 202, 218, 219, 438, 441. Nucleotide ở các vị trí tương ứng của loài măng cầu là C, C, C, T, G, T khác biệt so với loài na là T, A, T, C, A, C. Riêng mẫu na Đài Loan – Ngã bảy không có sự sai khác ở bất kỳ vị trí nucleotide nào so với loài măng cầu. Kết quả trên cho thấy khả năng phân biệt loài dựa trên vùng gen *rbcL* còn thiếu hiệu quả ở một vài loài, điều này phù hợp với kết luận của Larranaga và Hormaza (2015) về mã vạch *rbcL* dùng cho chi *Annona* có hiệu quả phân biệt loài kém hơn so với vùng gen *matK*.

Kết quả thu được sau khi so sánh các trình tự vùng gen *rpoC1* của các loài thuộc chi *Annona* nghiên cứu là 7 vị trí SNPs lần lượt là 23, 136, 154, 184, 303, 399, 435. Trong đó, bình bát có 2 vị trí 23 và 136 là nucleotide G, C khác biệt so với 2 loài còn lại là nucleotide A, T. Các mẫu măng cầu có 4 vị trí SNPs (154, 184, 303, 435) là các nucleotide C, G, T, G tương ứng với các 2 loài khác là các nucleotide T, A, C, A. Trong các mẫu na, ở vị trí 399, mẫu na là nucleotide A và na Đài Loan là nucleotide G.

Kết quả nghiên cứu đã chỉ ra được 9 vị trí nucleotide khác biệt của vùng gen *rbcL* ở các loài thuộc chi *Annona*. Trong đó có 3 vị trí có khả năng nhận diện loài bình bát, 2 vị trí đặc trưng cho loài măng cầu, 4 vị trí có tiềm năng nhận diện loài na. Đối với vùng gen *rpoC1*, kết quả nghiên cứu đã đưa ra 7 vị trí sai khác nucleotide. Trong đó có 2 vị trí đặc trưng cho loài bình bát, 4 vị trí có khả năng nhận diện loài măng cầu.

### 2.5.2 Phân tích đa dạng di truyền bằng dấu phân tử ISSR

Sản phẩm khuếch đại của mỗi ISSR 3 và ISSR 31 được điện di trên gel agarose 3%, đa số các mẫu cho phổ băng rõ, vị trí các băng có sự khác biệt giữa các loài. Khi so sánh với thang chuẩn cho thấy kích thước băng trải dài từ 1000 – 100 bp.

Đối với kết quả điện di với mỗi ISSR 3, đoạn mỗi khuếch đại được tổng cộng 15 băng, trong đó có 8 băng đa hình, chiếm tỉ lệ 53,34%. Măng cầu xuất hiện 4 băng đa hình ở vị trí lần lượt là 700 bp, 450 bp, 400 bp, 200 bp. Các băng ở vị trí 500 bp, 350 bp là 2 băng đa hình đặc trưng cho bình bát. Na có 2 băng đa hình ở vị trí lần lượt là 425 bp và 300 bp.

Đối với kết quả điện di với mỗi ISSR 31, đoạn mỗi khuếch đại được tổng số 18 băng, trong đó có 12 băng đa hình, chiếm tỷ lệ 66,67%. Các băng ở vị trí 1100 bp, 1000 bp, 800 bp, 425 bp, 350 bp là 4 băng đa hình đặc trưng cho bình bát. Măng cầu xuất hiện 6 băng đa hình ở vị trí lần lượt là 900 bp, 800 bp, 725 bp, 525 bp, 375 bp, 275 bp. Na có 2 băng đa hình ở vị trí lần lượt là 300 bp và 225 bp.

Giản đồ phân nhánh được xây dựng dựa trên hệ số tương đồng, trong đó 7 mẫu phân ra làm 2 nhánh chính với hệ số tương đồng nằm trong khoảng từ 0,48 đến 0,77. Giản đồ phân ra thành 2 nhóm lớn (I và II). Nhóm I phân thành 2 nhóm nhỏ: I<sub>a</sub> có 2 mẫu (bình bát – Cần Thơ và

binh bát – Long Mỹ) với hệ số tương đồng 66,50% và I<sub>b</sub> có 2 mẫu (na – Cần Thơ và na Đài Loan – Ngã Bảy) với hệ số tương đồng là 59,50%. Nhóm II cũng được chia thành 2 nhóm nhỏ: II<sub>a</sub> gồm 2 mẫu (mãng cầu ta – Cần Thơ và măng cầu Xiêm – Ngã Bảy) với hệ số tương đồng cao nhất là 77,00% và II<sub>b</sub> gồm 1 mẫu măng cầu Xiêm – Vĩnh Long. Mãng cầu Xiêm – Vĩnh Long so với 2 mẫu măng cầu có hệ số tương đồng là 69,50%.

## 2.6 Cây quýt hồng Lai Vung

Tổng cộng đã thu được 11 mẫu quýt (quýt hồng và quýt đường) từ 3 tỉnh Đồng Tháp, Tiền Giang và TP. Cần Thơ.

### 2.6.1 Kết quả phân tích trình tự DNA mã vạch và SNPs

Dựa vào kết quả phân tích trình tự vùng ITS cho thấy có 14 vị trí khác nhau giữa quýt hồng và các giống quýt khác. Các vị trí SNPs có thể dùng nhận diện quýt hồng Lai Vung với các quýt khác gồm 571, 573, 610, 616, 618, còn các vị trí còn lại chỉ phân biệt được quýt hồng và quýt đường. Trong nghiên cứu của Gonzalez *et al* (2009), trong số 285 loài được chọn để khảo sát ở rừng Amazonian, ở vùng gen ITS phân biệt được 41%. Tương tự, vùng gen ITS cũng phân biệt được 50% các loài lan hải Ấn Độ (Parveen *et al.*, 2012). Trong nghiên cứu của Newmaster *et al* (2006), ITS đã được khuyến nghị sử dụng làm DNA mã vạch cho thực vật vì sự nhận diện cao ở mức độ loài.

Cho thấy sự đa dạng di truyền thể hiện ở các giống khác nhau, bao gồm 13 vị trí có khác biệt. Đặt biệt, ở giống quýt hồng có thể nhận diện nhanh bởi sự khác biệt ở nucleotide vị trí 108 (quýt hồng là nu G, các giống còn lại bị đột biến mất điểm), vị trí 834 (quýt hồng là nu G, các giống còn lại bị đột biến mất điểm), vị trí 841 (quýt hồng là nu C, các giống còn lại là nu A). Nguyễn Bá Phú (2011), đã khuếch đại thành công quýt đường có hạt và không hạt Lai Vung ở vùng gen *matK*, tuy nhiên không phân biệt được 2 giống này ở vùng gen *matK*. Lahaye *et al* (2008), đã khảo sát hơn 1036 loài thực vật ở Mesoamerican (Trung bộ Châu Mỹ), kết quả là *matK* có khả năng nhận diện được tương đối cao (80,56%). Tuy nhiên khi kết hợp *matK* và *trnH-psbA* thì khả năng nhận diện hiệu quả hơn (trên 90%).

### 2.6.2 Kết quả phân tích đa dạng di truyền dựa vào dấu phân tử ISSR

Kết quả kiểm tra PCR với mỗi ISSR cho thấy các giếng có xuất hiện các băng đa hình thể hiện được sự đa dạng di truyền giữa các giống và kết quả này có thể sử dụng để xây dựng sơ đồ phả hệ.

Kết quả kiểm tra ISSR cho thấy có 15 băng được tạo ra, trong đó có 11 băng đa hình chiếm tỷ lệ 73,33%. Số băng mỗi đoạn mỗi ISSR đạt từ 4 đến 6 băng, trung bình mỗi mỗi đạt 5 băng. Từ kết quả điện di cho thấy, ISSR K1 có 5 băng DNA được khuếch đại, trong đó chỉ có 1 băng đa hình (chiếm tỷ lệ 20%), còn lại 4 băng đồng hình (chiếm tỷ lệ 80%). Đối với mỗi ISSR K2, có 6 băng DNA được khuếch đại, trong đó tất cả các băng đều là băng đa hình (chiếm tỷ lệ 100%), thể hiện tính đa hình cao với các mẫu nghiên cứu. Đối với mỗi ISSR K3, có 4 băng DNA được khuếch đại, trong đó tất cả các băng đều là băng đa hình (chiếm tỷ lệ 100%), thể hiện tính đa hình cao với các mẫu nghiên cứu.

## 2.7 Cây sầu riêng Ri-6

Mẫu lá sầu riêng được thu từ các tỉnh Tiền Giang, Bến Tre, Vĩnh Long và Cần Thơ có ký hiệu và nơi thu nhận. Riêng mẫu lá đầu dòng sầu riêng Ri-6 được thu ở Viện Cây ăn quả miền Nam.

### 2.7.1 Kết quả phân tích trình tự DNA mã vạch và SNPs

Qua kết quả phân tích và so sánh 14 trình tự dựa trên vùng ITS, có độ dài dao động từ 600 - 700 bp sau khi được hiệu chỉnh. Phát hiện 6 SNPs giữa các trình tự của các mẫu sầu riêng. Trong đó có một SNP đặc trưng cho giống sầu riêng Ri-6 là vị trí 444, ở sầu riêng Ri-6 là G trong khi các giống còn lại là C.

Kết quả cho thấy có 9 SNPs, trong các trình tự của giống Ri-6, có 6 vị trí khác biệt ở giống này. Vị trí 662 của trình tự Ri-6, sau khi kiểm tra peak cho thấy nu G có tín hiệu thấp. Tuy nhiên vẫn thể hiện được sự khác biệt với các vị trí trình tự còn lại. Vị trí 739 và 740 của trình tự Ri-6, sau khi xem tín hiệu peak ở vị trí 760, 761 trên Chromatogram cho thấy chỉ xuất hiện đỉnh cho một nu A nên hai vị trí này có thể do đột biến. Vị trí 759 cũng ở trình tự Ri-6 là nu G các trình tự khác là nu T, sau khi kiểm tra cho thấy tín hiệu nu G chưa tốt và bị nhiễu. Vị trí 780 xuất hiện nu C trên trình tự Ri-6 các trình tự còn lại là nu G, kiểm tra peak cho thấy tín hiệu nu C thấp nhưng không có tín hiệu nhiễu bên dưới. Trong trình tự giống Chín Hóa, cho thấy 2 điểm không xuất hiện nu là 750 và 763, sau khi kiểm tra peak (đoạn trình tự 765 – 805) tín hiệu khá tốt các peak có đỉnh cao và phân biệt được các nu, nên đây có thể là hai vị trí đột biến cho thấy được khác biệt của giống Chín Hóa với các giống còn lại. Trình tự các giống Sầu Hạt Lép, cho thấy có 2 vị trí khác biệt. Vị trí 798 xuất hiện sai khác giữa nu G và T, nên trình tự Sầu hạt lép nên được thay thế bằng ký tự K. Trình tự sầu riêng 6 Hữu xuất hiện sai khác tại vị trí 793 là nu G các trình tự còn lại là nu C, kiểm tra cho thấy tín hiệu nu G thấp nhưng có tín hiệu của nu A bên dưới mà không phải là nu C.

Qua kết quả so sánh và kiểm tra đối chứng giữa 17 trình tự vùng gen *matK* có độ dài khoảng 800 bp. Tìm được một số vị trí sai khác giữa các trình tự trong cùng giống Ri-6 (R6\_CT, R6\_VCAQMN), giống Sầu Hạt Lép (SHL\_BT) và cả hai thể của giống Chín Hóa (9H-BT) và giống Sáu Hữu (6H\_TG). Trong các điểm sai khác chưa tìm được vị trí đặc trưng cho giống sầu riêng Ri-6 nên vẫn chưa thể nhận diện được giống sầu riêng Ri-6 với các giống sầu riêng được sử dụng trong nghiên cứu. Tuy nhiên, có thể nhận diện được hai cá thể Chín Hóa\_Bến Tre và Sáu Hữu\_Tiền Giang. Qua kết quả phân tích cho thấy sự đa dạng trong trình tự vùng gen *matK* giữa các cá thể sầu riêng với nhau.

#### 2.7.2 Kết quả điện di sản phẩm PCR với dấu phân tử ISSR

Có tổng số 57 phân đoạn được nhân lên với trung bình 11,4 băng/đoạn môi. Trong đó có 40 phân đoạn đa hình chiếm tỉ lệ 72%. Số lượng băng đa hình dao động từ 6 (ISSR22) đến 20 (ISSR13) với trung bình đạt 8, kích thước các băng dao động trong khoảng 200 - 2500 bp. Trong số 5 chỉ thị của nghiên cứu này, ISSR13 cho nhiều phân đoạn nhất với 20 phân đoạn, trong đó có 11 phân đoạn đa hình chiếm 65%, kể đến là ISSR22 cho 15 phân đoạn với 11 phân đoạn đa hình chiếm 73,3%. Còn lại ISSR22, ISSR3 và ISSR03 cho số phân đoạn thấp hơn lần lượt là 6,7 và 9. Kết quả này cao hơn trong một nghiên cứu tương tự trên sầu riêng của Vanijajiva (2012), với 5 chỉ thị ISSR đã khuếch đại được 50 phân đoạn, có kích thước từ 50-2000bp và có 19 phân đoạn đa hình chiếm 38%. So sánh với quần thể 55 giống sầu riêng được nghiên cứu ở Indonesia của Angeliena *et al.* (2019), đã sử dụng 11 đoạn môi ISSR cho 83 alen đa hình (93,25%).

Từ phổ điện di các phân đoạn được mã hóa dưới dạng hệ số nhị phân 0:1. Sau đó được xử lý bằng phần mềm NTSYSpc2.10 thu được hệ số tương đồng của 20 giống sầu riêng. Sự khác biệt của 20 giống sầu riêng được ghi nhận dựa trên sự đa hình về kiểu gen được khuếch đại bởi 5 môi ISSR có hệ số tương đồng dao động trong khoảng 0,61-0,97. Cây phân loại được xây dựng dựa trên hệ số tương đồng với phương pháp UPGMA. Dựa vào giản đồ hình cây, 20 mẫu sầu riêng được chia thành 5 nhóm chính.

Kết quả phân tích dựa vào dấu phân tử ISSR đã phân chia 20 giống sầu riêng thu được ở bốn tỉnh của vùng ĐBSCL chia thành các 5 nhóm có hệ số tương đồng dao động khá xa trong khoảng 0,61 – 0,97.

#### 2.8 Cây thanh long ruột đỏ

Dựa trên kết quả bình tuyển cây đầu dòng của Viện cây ăn quả miền Nam và khảo sát các tỉnh trồng thanh long trọng điểm, đề tài đã tiến hành thu 10 mẫu thanh long đại diện cho 4 giống phổ

biến ở Nam Bộ hiện nay là thanh long ruột đỏ, ruột tím hồng, ruột trắng và vỏ vàng ruột trắng. Về đặc điểm hình thái, các giống được phân biệt chủ yếu bởi màu sắc vỏ, ruột và hình dạng trái.

#### 2.8.1 Kết quả khuếch đại các trình tự DNA mã vạch

Các trình tự DNA mã vạch bao gồm 4 gene mã hóa protein là *ycf1b*, *rbcL*, *rpoC1* và *matK*, nhóm không mã hóa gồm 3 trình tự *atpF-H*, *psbA-trnH* *psbK-I* và vùng ITS trong bộ gene nhân đã được khuếch đại. Thông qua độ sáng của sản phẩm PCR trên gel agarose, cho thấy gen *rpoC1*, *rbcL* và 2 vùng *atpF-H*, *psbA-trnH* đạt hiệu quả khuếch đại cao. Băng điện di của gen *matK* và *ycf1b* có độ sáng kém hơn nhưng vẫn có thể quan sát rõ trên gel. Các trình tự *rpoC1*, *rbcL*, *atpF-H*, *psbA-trnH* và *matK* chỉ xuất hiện 1 băng duy nhất phản ánh tính đặc hiệu của mỗi trong quá trình PCR. Mẫu đối chứng âm không xuất hiện băng đảm bảo không xảy qua ngoại nhiễm trong quá trình PCR. Do đó, 5 vùng trên được sử dụng để giải trình tự và phân tích hiệu quả nhận diện loài trong nội dung nghiên cứu tiếp theo.

#### 2.8.2 Kết quả phân tích trình tự DNA mã vạch, xác định SNPs

Dựa trên chromatogram, chất lượng của trình tự DNA đã được khẳng định, vùng *psbA-trnH* với nhiều ký tự không xác định trong khi *atpF-H*, *matK*, *rpoC1* và *rbcL* thể hiện tín hiệu nucleotide rõ ràng. Các vị trí bảo tồn và thay đổi trong số các trình tự này đã được kiểm tra do kết quả liên kết. Nhìn chung, có sự bảo thủ cao giữa các loài *Hylocereus* đối với bốn trình tự mã vạch DNA. Tuy nhiên, sự xuất hiện của SNPs và đột biến indel cho thấy sự đa dạng nucleotide của các loài *Hylocereus* như vậy. Vùng trình tự *atpF-H* thay đổi cao nhất với 6 vị trí thay đổi. Ba gen bao gồm *matK*, *rpoC1* và *rbcL* có một lượng nhỏ các vị trí thay đổi so với *atpF-H*. Về đột biến indel, 9 đột biến đã được phát hiện trên trình tự này trong khi tần số của đột biến này khá thấp ở ba locus khác. Do đó, có thể thấy rằng trình tự không mã hóa biến đổi nhiều hơn trình tự mã hóa và được đề xuất như một trình tự mã vạch tiềm năng để phân biệt loài.

Một phát hiện mới từ trình tự *atpF-H* là sự chèn một đoạn DNA ngắn gồm 9 nucleotide đột biến ATTAGGTAC ở *Hylocereus* sp. DF4 (ruột đỏ và quả tròn) so với 9 giống thanh long còn lại. Các đột biến đã làm cho giống này trở nên đặc biệt so với những giống khác. Kết quả này khẳng định rằng vùng đệm *atpF-H* có tính đa hình cao và có thể xác định được giống thanh long này. Hơn nữa, năm đột biến thay thế đã được phát hiện và phân loại thành ba SNPs như T / C, C / A và G / A.

Phân tích trình tự từ gen *rbcL* đã chứng minh rằng gen có nguồn gốc từ lạp thể này có giá trị trong việc phân định loài. Hai giống *H. polyrhizus* DF1 (ruột đỏ) và *Hylocereus* sp. DF5 (ruột tím hồng) từ Viện cây ăn quả Miền Nam đã được phân biệt với các giống khác khác bởi 4 SNPs, bao gồm G / C, A / C, G / T, T / C và 2 đột biến thêm nucleotide A và G.

### 2.9 Cây vú sữa lò rèn

Tổng cộng có 19 mẫu vú sữa gồm các giống vú sữa Lò Rèn, vú sữa bơ hồng, vú sữa bơ tím và vú sữa tím được thu mẫu ở Viện cây ăn quả Miền Nam, thu nhà vườn và trại giống ở các tỉnh Tiền Giang, Bến Tre, Vĩnh Long, Cần Thơ.

#### 2.9.1 Kết quả phân tích trình tự DNA mã vạch, xác định SNPs

Sau khi giải trình tự sản phẩm PCR tám mỗi ITS, *matK*, *rpoC1*, *atpF-atpH*, *ycf1b*, *rpoC1*, *trnH-psbA* và *psbI-psbK* của sáu giống vú sữa: vú sữa Lò Rèn, vú sữa bơ hồng, vú sữa bơ tím, vú sữa tím, vú sữa Mica và vú sữa cà na thu được tổng cộng 113 trình tự DNA, trong đó số trình tự đã giải và sau kiểm tra được sử dụng có số lượng lần lượt là ITS 19, *matK* 18, *rbcL* 15, *atpF-atpH* 16, *rpoC1* 12, *psbI-psbK* 18, *trnH-psbA* 15.

Sản phẩm sau khi giải trình tự được kiểm tra chỉ có mỗi *atpF-atpH* cho kết quả khác nhau ở các giống.

Phân tích trình tự DNA của vùng gen *atpF-atpH* sau khi được hiệu chỉnh có kích thước 504 bp, thực hiện phân tích bằng phần mềm MEGA X cho kết quả vùng bảo tồn là 489/504 vị trí,

vùng biến đổi là 15/504 trong đó 8 vị trí khác biệt trên trình tự DNA của 1 giống so với các giống còn lại và 3 vị trí có sự khác biệt giữa 2 giống so với các giống còn lại.

Khi dựa trên trình tự DNA vùng *atpF-atpH* cho thấy có sự khác biệt giữa năm giống vú sữa Lò Rèn, vú sữa bơ hồng, vú sữa bơ tím, vú sữa tím và vú sữa Mica. Điều này cũng cho thấy đặc điểm hình thái có thể tương tự nhau nhưng khi xét về di truyền phân tử thì có sự khác nhau. Kết quả này chứng tỏ có thể sử dụng vùng gen *atpF-atpH* để đánh giá mối quan hệ di truyền và xác định một số cá thể của các giống vú sữa.

#### **2.9.2 Kết quả phân tích sự đa hình chỉ thị ISSR**

Sau khi điện di sản phẩm PCR mỗi ISSRK1 và mỗi ISSRK2 đều cho kết quả băng sáng, rõ, các băng gel đều xuất hiện băng đơn hình và đa hình, cho thấy sự đa dạng di truyền giữa các giống và đủ điều kiện xây dựng sơ đồ phả hệ.

Khi xem xét các băng điện di mỗi ISSRK1 và ISSRK2 cho thấy ở vị trí có kích thước 300 bp và 370 bp không xuất hiện băng ở ba mẫu vú sữa bơ hồng (Bến Tre), vú sữa Mica (Bến Tre) và vú sữa cà na (Cần Thơ), khi các mẫu vú sữa còn lại đều xuất hiện băng. Như vậy, tại kích thước băng ở vị trí trên có thể phân biệt được nhóm các cá thể gồm vú sữa bơ hồng (Bến Tre), vú sữa Mica (Bến Tre) và vú sữa cà na (Cần Thơ). Ở kích thước băng 680 bp của mỗi ISSRK2 chỉ có một mẫu vú sữa bơ hồng (Cần Thơ) không xuất hiện băng nên có thể nhận biết được riêng cá thể này trong tất cả các mẫu. Tương tự, mỗi ISSRK2 ở vị trí băng có kích thước 710 bp không xuất hiện băng ở hai mẫu gồm vú sữa bơ hồng (Cái Bè) và vú sữa bơ hồng (Bến Tre) nên có thể nhận diện được nhóm hai cá thể này trong các mẫu được nghiên cứu.

Sự giống nhau về mặt di truyền của các giống vú sữa được ghi nhận dựa trên sự đa hình về kiểu gen được khuếch đại bởi hai môi và có hệ số tương đồng dao động trong khoảng 0,68 – 1,00.

### **2.10 Cây xoài cát hòa lộc**

Các giống xoài được thu ở các tỉnh Tiền Giang (huyện Cái Bè và Viện Cây ăn quả miền Nam), Bến Tre, Vĩnh Long, Hậu Giang, Sóc Trăng, Cần Thơ (Cờ Đỏ, Ninh Kiều và Bình Thủy) và An Giang. Mẫu cây đầu dòng xoài cát Hòa Lộc được cung cấp từ Viện Cây ăn quả miền Nam.

#### **2.10.1 Kết quả phân tích trình tự DNA mã vạch và SNPs**

Kết quả cho thấy các giống xoài khác nhau ở 52 vị trí biến đổi đối với trình tự ITS. Đặc biệt là ở giống Hòa Lộc, có thể nhận diện nhanh bởi sự khác biệt ở nucleotide vị trí 624 (Hòa Lộc là nu A, tất cả trình tự còn lại là nu G, riêng trình tự X.UC và X.CHAU-TG, X.NGOCVAN-VCAQMN là C. Ngoài ra, tại vị trí 559, 589, 633 (tương ứng AAC, tương ứng) khác với vị trí nucleotide Hòa Lộc (GGG, tương ứng) khi so sánh trình tự HOALOC với các trình tự của X.DUDU-AG, X.UC, X.CHAU-TG, X.NV-VCAQMN, X.CHAU-TG.

Có 5 vị trí khác biệt trong trình tự gen *atpF-atpH* tương ứng với các vị trí peak là 34, 35, 51, 784, 812. Trong đó, X-HOALOC có thể nhận diện được sự khác biệt với các giống khác ở 3 vị trí 40, 787 và 816. Tại vị trí 40, X-HOALOC xuất hiện nu A, ở các giống còn lại không biểu hiện nu, riêng trình tự xoài X-NGOCVAN biểu hiện nu G. Ngoài ra, tại vị trí 787 và 816, trình tự X-HOALOC là nu A, các giống còn lại không biểu hiện nu.

#### **2.10.2 Kết quả phân tích đa dạng dấu phân tử ISSR**

Dựa vào dấu phân tử ISSR đã chia các giống xoài thành 3 nhóm chính. Tỷ lệ phần trăm đa hình cao 76, 47% xấp xỉ với tỉ lệ nghiên cứu trên xoài sử dụng 100 mỗi ISSR thu được tỉ lệ 213 băng đơn hình trên tổng số 335 băng thu được, chiếm tỉ lệ 63,58% (Damodaran *et al.*, 2012) qua đó thể hiện được sự đa dạng của các giống xoài về mặt di truyền.

### **CHƯƠNG 3. CƠ SỞ DỮ LIỆU DNA MÃ VẠCH**

Cơ sở dữ liệu được viết dưới dạng trang web trong server của trường Đại học Cần Thơ bằng ứng dụng Joomla. Cơ sở dữ liệu được viết bằng cả tiếng Việt và tiếng Anh. Cấu trúc của cơ sở dữ liệu như sau :

Trang chủ bao gồm phần giới thiệu chung về cơ sở dữ liệu.

Phần cơ sở dữ liệu gồm trình tự DNA mã vạch của 6 vùng trình tự thuộc 10 cây ăn trái.

Danh sách các môi DNA mã vạch và chu trình nhiệt cũng được đưa vào cơ sở dữ liệu.

Để sử dụng cơ sở dữ liệu, người dùng truy cập vào trang web: <http://www.dnavf.ctu.edu.vn>.

Các tính năng của cơ sở dữ liệu bao gồm: tải trình tự về dưới định dạng fasta và hiển thị vùng trình tự đặc trưng của mỗi trình tự.



### **PHẦN III**

#### **KẾT LUẬN VÀ ĐỀ NGHỊ**

##### **3.1 Kết luận**

Các trình tự DNA mã vạch của 10 loại cây ăn trái đặc trưng của Việt Nam vùng Nam Bộ đã được giải và phân tích. Mỗi loại cây ăn trái đã xác định trình tự DNA đặc trưng. Sự đa dạng di truyền của cây ăn trái cũng đã được xác định.

Các trình tự đã được xây dựng thành cơ sở dữ liệu có thể sử dụng trong tra cứu và phát triển công cụ nhận dạng cây ăn trái.

##### **3.2 Đề nghị**

- Khai thác cơ sở dữ liệu để phát triển công cụ nhận diện cây ăn trái.
- Tiếp tục bổ sung cơ sở dữ liệu DNA mã vạch với các loại cây ăn trái khác.