

ĐA DẠNG NUCLEOTIDE VÙNG *ITS* GEN NHÂN VÀ CÁC GEN LỤC LẠP (*matK*, *rbcl*, *rpoC1*) LOÀI TRÁM ĐEN (*CANARIUM NIGRUM*) Ở MỘT SỐ TỈNH PHÍA BẮC, VIỆT NAM

Đinh Thị Phòng^{1,3}, Trần Thị Liễu^{1,✉}, Vũ Thị Thu Hiền¹, Hoàng Thanh Lộc²

¹Bảo tàng Thiên nhiên Việt Nam, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Viện Cải thiện giống và Phát triển lâm sản

³Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

✉ Người chịu trách nhiệm liên lạc. E-mail: tranthilieu@gmail.com

Ngày nhận bài: 05.10.2017

Ngày nhận đăng: 05.7.2018

TÓM TẮT

Trong nghiên cứu này, 03 vùng gen lục lạp (*matK*, *rbcl* và *rpoC1*) và 1 vùng gen nhân (*ITS*) đã được sử dụng để đánh giá mức độ đa dạng nucleotide cho 9 cá thể đại diện loài Trám đen (*Canarium nigrum*) ở tỉnh Bắc Giang, Hòa Bình và Phú Thọ (mỗi tỉnh 3 cá thể). Trình tự nucleotide của 4 vùng gen (*ITS*, *matK*, *rbcl* và *rpoC1*) đã được xác định với kích thước tương ứng là 696 bp, 798 bp, 702 bp và 522 bp. Kết quả so sánh trình tự nucleotide của 4 vùng gen giữa các mẫu nghiên cứu cho thấy có sự tương đồng 100%. Kết quả so sánh với trình tự các loài trong chi *Canarium* trên GenBank đã chỉ ra mức độ đa dạng nucleotide (π) cao nhất ở vùng gen *ITS* (0,02), thứ hai là vùng gen *matK* (0,007) và thấp nhất là vùng gen *rbcl* (0,003). Cây phát sinh chủng loại của loài Trám đen với các loài trong chi *Canarium* đã chỉ ra sự phân nhánh giữa các loài rõ nhất ở vùng gen *rpoC1*, thứ hai là vùng gen *rbcl*, thứ ba là vùng gen *matK* và thấp nhất là vùng gen *ITS* với giá trị bootstrap tại các điểm nút chính tương ứng dao động từ 98 đến 99%, 35 đến 67%, 65 đến 98% và 98 đến 99%. Loài Trám đen có mức độ tương đồng nucleotide gần nhất với loài *C. tramdenum* (KP093200) và *C. album* (KP093198) khi phân tích vùng gen *ITS* và loài *C. subulatum* (KR530509), *C. acutifolium* (KR530512) khi phân tích vùng gen *matK*. Các kết quả thu được cho thấy có thể sử dụng vùng gen *rpoC1* cho việc nhận dạng loài thuộc chi *Canarium*.

Từ khóa: *Canarium nigrum*, đa dạng nucleotide, *ITS*, *matK*, *rbcl*, *rpoC1*

MỞ ĐẦU

Trám đen (*Canarium pimela* K. D. Koenig; tên đồng nghĩa: *Canarium nigrum* (Lour.) Engl.) là loài thuộc chi Trám (*Canarium*) trong họ Burseraceae. Đây là loài có giá trị kinh tế cao nên được trồng và khai thác ở nhiều nơi như Bắc Giang, Phú Thọ, Hòa Bình, Tuyên Quang, Thái Nguyên, Lạng Sơn, Quảng Ninh, Thanh Hóa, Nghệ An, Quảng Bình, Quảng Nam, Đắk Lắk và Khánh Hòa. Quả Trám đen chứa nhiều chất dinh dưỡng được dùng làm thực phẩm, vị thuốc. Hạt Trám đen có nhiều dầu béo, vị bùi, có thể sử dụng ăn sống hoặc ép lấy dầu. Ngoài ra, các bộ phận khác của cây như lá, rễ cũng được dùng làm thuốc, nhựa dùng để thắp sáng, gỗ được sử dụng làm nhà, đóng đồ dùng, làm bột giấy... Trám đen là loài cây bản địa đa mục đích được trồng trong nhiều chương trình và dự án trồng rừng khác nhau ở các

tỉnh trung du miền núi phía Bắc và miền Trung. Ở phía Bắc, loài Trám đen được phân bố nhiều nhất ở ba tỉnh Bắc Giang, Hòa Bình và Phú Thọ. Đặc biệt, Trám đen ở Hoàng Vân, Hiệp Hòa, Bắc Giang thơm ngon nổi tiếng trong cả nước (http://thongtinkhcn.com.vn/vn/tin-tuc/detail.php?ELEMENT_ID=2883). Tuy nhiên, trong thực tế các mô hình trồng Trám đen tập trung theo hướng lấy quả vẫn thiếu thông tin về di truyền. Vì vậy, nghiên cứu đa dạng di truyền vùng gen nhân và lục lạp loài Trám đen sẽ cung cấp thêm dữ liệu về trình tự nucleotide các vùng gen, làm cơ sở cho công tác phát triển nhân rộng và bảo tồn loài Trám đen ở ba tỉnh nói trên.

Hai nhóm gen nhân và gen lục lạp thường được sử dụng trong các nghiên cứu về mối quan hệ chủng loại và nhận dạng loài ở nhiều đối tượng sinh vật (Yang *et al.*, 2007; Schoch *et al.*, 2012; Huang *et al.*,

2015). So với gen nhân thì gen lục lạp có mức độ bảo thủ hơn bởi việc thay thế chỉ một vài nucleotide (Sang *et al.*, 1997; Sun *et al.*, 2005). Đối với các loài Trám, cũng đã có một vài công bố về mối quan hệ chủng loại và nhận dạng loài (Clarkson *et al.*, 2002; Weeks, Simpson, 2004; 2007; Weeks *et al.*, 2005; Weeks, 2009; Liu *et al.*, 2014). Chẳng hạn, Weeks (2009) đã nghiên cứu tiên hóa của chi *Canarium* dựa trên phân tích trình tự 2 vùng gen nhân (ETS, NIA-i3) và 5 vùng gen lục lạp (*rbcL*, *rps16*, *psbA-trnH*, *trnL* và *trnL-trnF*). Kết quả đã chỉ ra *Canarium* có ít nhất hai dòng dõi tiên hóa liên quan đến các đặc tính quả và hạt. Hiện nay trên GenBank (2016) đang lưu giữ trên 500 trình tự nucleotide (<http://www.ncbi.nlm.nih.gov/nucleotide>) cho các loài thuộc chi *Canarium*, trong đó vùng gen nhân (*ITS*) (8 trình tự), *rbcL* (96 trình tự), *matK* (26 trình tự), *trnL-trnF* (32 trình tự), *psbA-trnH* (15 trình tự), *rpoC1* (1 trình tự)... Đây chính là nguồn cơ sở dữ liệu có thể khai thác ứng dụng trong xác định trình tự nucleotide cho quần thể Trám đen của Việt Nam.

Xuất phát từ những cơ sở khoa học trên, chúng

tôi đã tiến hành đánh giá mức độ đa dạng nucleotide vùng gen nhân (*ITS*) và 3 vùng gen lục lạp (*matK*, *rbcL* và *rpoC1*) của loài Trám đen ở ba tỉnh Bắc Giang, Hòa Bình và Phú Thọ.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Vật liệu

Chín mẫu lá lựa chọn từ 35 cây trội loài Trám đen do Viện Cải thiện giống và Phát triển lâm sản cung cấp (mỗi mẫu là một cá thể cây trội trưởng thành, cả 35 mẫu cây mang tính đại diện của mỗi quần thể trên cơ sở phân tích một số đặc điểm nông học, sản lượng và chất lượng quả) thu tại 03 địa điểm Bắc Giang, Hòa Bình và Phú Thọ. Mỗi địa điểm lựa chọn đại diện 3 mẫu ngẫu nhiên. Các mẫu được bảo quản trong silicagel tới khi sử dụng. Danh sách các mẫu nghiên cứu có ký hiệu và nơi thu thập được thể hiện trong bảng 1.

Trình tự nucleotide và kích thước sản phẩm PCR của 04 cặp mồi sử dụng trong nghiên cứu được thể hiện trong bảng 2.

Bảng 1. Nguồn gốc và ký hiệu của các mẫu Trám đen dùng trong nghiên cứu.

TT	Ký hiệu mẫu thu	Ký hiệu mẫu phân tích	Tọa độ thu mẫu		Độ cao (m) **	Địa điểm thu mẫu
			Vĩ độ (°N)	Kinh độ (°E)		
1	BGQT2	VNMN000863	21.23.58,5	105.57.22,6	31	Hoàng Văn,
2	BGQT3	VNMN000864	21.23.56,4	105.56.57,3	20	Hiệp Hoà, Bắc Giang
3	BGQT5	VNMN000865	21.23.04,6	105.56.53,2	16	
4	HBQT2	VNMN000866	20.32.05,4	105.15.52,2	98	Lỗ Sơn, Tân Lạc, Hòa Bình
5	HBQT4	VNMN000867	20.33.53,5	105.18.56	112	Thanh Hối, Tân Lạc, Hòa Bình
6	HBQT5	VNMN000868	20.34.10,9	105.18.47,2	97	
7	PTQT2	VNMN000869	21.26.56,5	105.13.51,2	37	
8	PTQT4	VNMN000870	21.26.28,3	105.14.30,4	32	Hà Lộc, TX. Phú Thọ
9	PTQT5	VNMN000871	21.27.12,3	105.14.10,8	43	Thọ, Phú Thọ

Ghi chú: * Mã hiệu mẫu lưu giữ tại Bảo tàng Thiên nhiên Việt Nam (VNMN); ** Độ cao so mực nước biển.

Bảng 2. Trình tự nucleotide và kích thước vùng gen đích theo lý thuyết của 04 cặp mồi sử dụng trong nghiên cứu.

Vùng gen	Ký hiệu mồi	Trình tự nucleotide (5'– 3')	Kích thước sản phẩm PCR (bp)	Tài liệu tham khảo
<i>ITS</i>	<i>ITS1/ ITS4</i>	TCCTCCGCTTATTGATATGC TCCGTAGGTGAACCTGCGG	750	White <i>et al.</i> , 1990
<i>rbcL</i>	<i>rbcL1F/ rbcL724R</i>	ATGTCACCACAAACAGAGACTAAAGC TCGCATGTACCTGCAGTAGC	750	Fay <i>et al.</i> , 1997
<i>rpoC1</i>	<i>rpoC1F/ rpoC1R</i>	GTGGATACACTTCTTGATAATGG TGAGAAAACATAAGTAAACGGGC	600	http://www.kew.org/barcoding/protocols.html
<i>matK</i>	<i>Kim3/ Kim1R</i>	CGTACAGTACTTTTGTGTTTACGAG ACCCAGTCCATCTGGAAATCTTGTTTC	950	Kim <i>et al.</i> , unpublished

Phương pháp

Tách chiết DNA tổng số

DNA tổng số được tách chiết từ mẫu lá theo phương pháp CTAB của Doyle và Doyle (1990), sau đó được tinh sạch sử dụng Genomic DNA purification kit (#KO512) của hãng Fermentas (Mỹ).

PCR nhân bản vùng gen đích

PCR được tiến hành trên máy PCR Systems 9700, với thể tích 25 µL, gồm các thành phần và chu trình nhiệt theo công bố của Vũ Thị Thu Hiền *et al.*, (2012).

Xác định trình tự nucleotide

Sản phẩm PCR sau khi tinh sạch được sử dụng làm DNA khuôn cho phản ứng xác định trình tự trên máy ABI PRISM® 3100 Avant Genetic Analyzer (Applied Biosystems) tại Phòng Thí nghiệm trọng điểm Công nghệ gen, Viện Công nghệ Sinh học.

Phản ứng đọc trình tự được thực hiện theo cả hai chiều xuôi và ngược.

Phân tích số liệu

Các trình tự DNA của loài Trám đen được phân tích, so sánh với trình tự đại diện của các loài thuộc chi *Canarium* trên GenBank (Bảng 3) để tìm ra vị trí sai khác, mức độ đa dạng nucleotide bằng phần mềm BioEdit (Hall, 1999) và Mega 4.0 (Tamura, 2007). Cây phát sinh chủng loại của loài Trám đen với các loài trong chi *Canarium* được xây dựng theo phương pháp NJ (Neighbor Joining) bằng phần mềm Mega 4.0. Trong đó có sử dụng trình tự nucleotide của một số loài thuộc chi khác trong cùng họ Burseraceae làm nhóm ngoài khi lập cây phát sinh chủng loại bao gồm trình tự của loài *Protium krukoffi* (mã số GenBank AY375511), loài *Protium divaricatum* (AY594475), loài *Protium pallidum* (JQ625811) và loài *Protium sagotinaum* (FJ038446).

Bảng 3. Danh sách trình tự các loài trong chi *Canarium* trên GenBank sử dụng trong nghiên cứu.

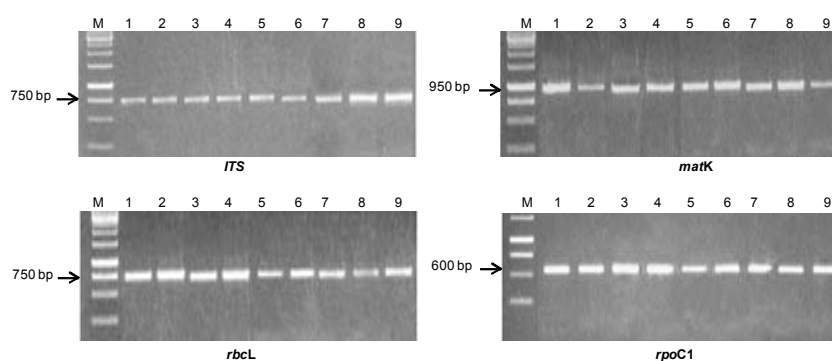
Tên loài	Mã số GenBank			Tên loài	Mã số GenBank	
	ITS	matK	rbcl		rbcl	rpoC1
<i>C. album</i>	KP093198			<i>C. multiflorum</i>	KT698497	
	KR531870				GU24602	
	DQ517524	KR530503	KJ440021	<i>C. muelleri</i>	9	
	KP093199	HQ415266	FJ466626	<i>C. ovatum</i>	FJ466636	
	JF421482			<i>C. obtusifolium</i>	KT698501	
<i>C. tramdenum</i>	KP093200	KP094017		<i>C. odontophyllum</i>	KT698499	
	KP093201	HQ415267	FJ466639	<i>C. pilosum</i>	FJ466637	
		KR530504		<i>C. littorale</i>	FJ466633	
<i>C. subulatum</i>		KR530509		<i>C. decumanum</i>	FJ466629	
	KR531871	KR530511		<i>C. latistipulatum</i>	KT698538	
<i>C. acutifolium</i>		AB924844		<i>C. rufum</i>	KT698508	
<i>C. zeylanicum</i>		KR530512	FJ976124	<i>C. pulchrebracteatum</i>	KT698504	
<i>C. pilosum</i>		KF521891		<i>C. patentinervium</i>	KT698503	
<i>C. littorale</i>		KJ708856		<i>C. whitei</i>	FJ466641	
<i>C. australasicum</i>		KJ708855		<i>C. zeylanicum</i>	FJ466642	
		JN564129		<i>C. vulgare</i>	FJ466640	
<i>C. madagascariense</i>		JN564128		<i>C. strictum</i>	FJ466638	
		KX146374	FJ466634			
<i>Canarium</i> sp.		KR530506	KT698509	<i>C. oleosum</i>		GQ248897
			KT698507			

KẾT QUẢ

Kết quả PCR nhân bản vùng gen đích

Bốn cặp mồi đặc hiệu (*ITS1/ITS4*, *rbcL1F/rbcL724R*, *rpoC1F/rpoC1R*, *Kim3F/Kim1R*) đã được sử dụng để nhân bản đoạn gen đích cho 9 cá thể nghiên cứu. Kết quả phân tích sản phẩm PCR được thể hiện ở hình 1 cho thấy, đã nhân bản thành công tất cả 4 đoạn gen đích với kích thước như lý thuyết dự đoán, cụ thể: ~ 750 bp đối với vùng gen *ITS* và vùng gen *rbcL*, ~ 950 bp đối với vùng gen

matK và ~ 600 bp đối với vùng gen *rpoC1*. Tất cả 36 trình tự DNA đích đều được giải mã thành công với độ tin cậy cao và không có sự thay đổi trình tự giữa các mẫu ở cả bốn vùng gen nghiên cứu. Vì vậy, các nghiên cứu tiếp theo chỉ lấy đại diện một mẫu VNMN000863 *C. nigrum*. Sau khi phân tích số liệu, loại bỏ trình tự mồi, các vị trí trống và các trình tự so le ở hai đầu, chúng tôi đã thu được trình tự vùng gen *ITS*, *matK*, *rbcL* và *rpoC1* của loài Trám đen với kích thước tương ứng là 696 bp, 798 bp, 702 bp và 522 bp. Ba mươi sáu trình tự đã được đăng ký trên GenBank với mã số từ MF166582 đến MF166617.



Hình 1. Sản phẩm PCR sau khi tinh sạch của 9 mẫu Trám đen phân tích với 4 vùng gen trên gel agarose 1,5% (M: marker phân tử 1 kb, giếng từ 1 - 9 là thứ tự của các mẫu Trám đen trong Bảng 1).

Đa dạng nucleotide 4 vùng gen của loài Trám đen với trình tự của các loài đã công bố trên GenBank

Hiện nay trên GenBank mới chỉ có 8 trình tự vùng gen *ITS*, 26 trình tự vùng gen *matK*, 96 trình tự vùng gen *rbcL* và 1 trình tự vùng gen *rpoC1* của 35 loài trong chi *Cannarium*. Vì số lượng trình tự tương đối nhiều nên trong nghiên cứu này chỉ một số trình tự nucleotide đã công bố trên GenBank được lựa chọn cho phân tích (chi tiết trong bảng 3). Đây là các trình tự đại diện cho những trình tự giống nhau của một loài hoặc những trình tự khác nhau ở dạng thêm vào, mất đi hay thay thế nucleotide. Kết quả cho thấy tại vùng gen *ITS*, trình tự VNMN000863 *C. nigrum* giống hoàn toàn (100%) với 2 trình tự KP093198 *C. album* và KP093200 *C. tramdenum* (cả hai đều có nguồn gốc ở Trung Quốc), nhưng lại có 15 vị trí nucleotide sai khác với 6 trình tự còn lại trên GenBank, chẳng hạn tại vị trí nucleotide thứ 49, 135, 173, 188, 232, 257 và 258 (GAGCGTC, tương ứng) được thay thế bằng (ACAAAGT, tương ứng) khi so sánh trình tự VNMN000863 *C. nigrum* với các trình

tự KR531870 *C. album*, DQ517524 *C. album*, KP093201 *C. tramdenum*, KP093199 *C. album*, JF421482 *C. album*, và KR531871 *C. subulatum* (Bảng 4).

Tại vùng gen *matK*, trình tự VNMN000863 *C. nigrum* giống hoàn toàn với 2 trình tự KR530509 *C. subulatum* và KR530512 *C. acutifolium* (đều có nguồn gốc ở Trung Quốc), và có 25 vị trí đột biến chèn vào hay thay thế nucleotide khi so sánh với 14 trình tự còn lại. Cụ thể, tại vị trí nucleotide thứ 42 (C) được thay thế bằng (T) khi so sánh trình tự loài VNMN000863 *C. nigrum* với trình tự KF521891 *C. zeylanicum* (có nguồn gốc ở Sri Lanka) và KX146374 *C. madagascariense* (có nguồn gốc ở Cộng hòa Madagascar). Hay tại vị trí nucleotide thứ 62, 200 và 202 (CGT, tương ứng) được thay bằng (AAC, tương ứng) khi so sánh trình tự VNMN000863 *C. nigrum* với các trình tự KR530503 *C. album*, HQ415266 *C. album*, KP094017 *C. tramdenum*, HQ415267 *C. tramdenum*, KR530504 *C. tramdenum* và KR530511 *C. subulatum* (Bảng 5).

Bảng 4. Các vị trí biến đổi nucleotide giữa loài Trám đen (VNMN000863 *C. nigrum*) với trình tự nucleotide của một số loài trên GenBank khi phân tích với vùng gen *ITS*.

Vị trí nucleotide	49	100	135	173	188	232	257	258	266	267	270	277	286	302	307
VNMN000863 <i>C. nigrum</i>	G	-	A	G	C	G	T	C	T	G	C	C	G	C	C
KP093198 <i>C. album</i>	.	-
KP093200 <i>C. tramdenum</i>	.	-
KR531870 <i>C. album</i>	A	C	C	A	A	A	G	T	C	T	T	T	T	T	T
DQ517524 <i>C. album</i>	A	C	C	A	A	A	G	T	C	T	T	T	T	T	T
KP093201 <i>C. tramdenum</i>	A	C	C	A	A	A	G	T	C	T	T	T	T	T	T
KP093199 <i>C. album</i>	A	C	C	A	A	A	G	T	C	T	T	T	T	T	T
JF421482 <i>C. album</i>	A	C	C	A	A	A	G	T	C	T	T	T	T	T	T
KR531871 <i>C. subulatum</i>	A	C	C	A	A	A	G	T	C	T	T	T	T	T	T

Bảng 5. Các vị trí biến đổi nucleotide giữa loài Trám đen (VNMN000863 *C. nigrum*) với trình tự nucleotide của một số loài trên GenBank khi phân tích với vùng gen *matK*.

Vị trí nu	42	62	81	138	176	200	202	207	211	224	235	236	283	309	321	333	341	366	399	452	479	540	549	607	632	
1	C	C	T	C	G	G	T	A	G	A	G	A	A	G	T	C	G	A	A	C	C	-	A	G	T	
2	-	.	.	.
3	.	A	.	.	.	A	C	A	-	.	.	.	
4	.	A	.	.	.	A	C	A	-	.	.	.	
5	.	A	.	.	.	A	C	A	-	.	.	.	
6	.	A	.	.	.	A	C	A	A	.	.	.	
7	.	A	.	.	.	A	C	A	-	.	.	.	
8	-	.	.	.	
9	.	A	.	.	.	A	C	A	-	.	.	.	
10	C	G	.	C	A	-	G	.	G	
11	T	.	.	.	C	.	C	T	.	.	.	A	-	.	A	.	
12	C	.	T	A	-	.	A	.	
13	C	G	A	-	.	A	.	
14	.	.	C	G	.	.	C	.	.	G	.	.	.	T	.	A	A	-	.	A	.	
15	.	.	C	G	.	.	C	.	.	G	.	.	.	T	.	A	.	G	.	.	A	-	.	A	.	
16	T	C	.	.	.	T	T	A	A	-	.	A	
17	C	C	.	A	-	.	.	.	

Ghi chú: Các trình tự nucleotide trên GenBank được sử dụng để so sánh: 1: VNMN000863 *C. nigrum*, 2: KR530512 *C. acutifolium*, 3: KR530503 *C. album*, 4: HQ415266 *C. album*, 5: KP094017 *C. tramdenum*, 6: HQ415267 *C. tramdenum*, 7: KR530504 *C. tramdenum*, 8: KR530509 *C. subulatum*, 9: KR530511 *C. subulatum*, 10: AB924844 *C. subulatum*, 11: KF521891 *C. zeylanicum*, 12: KJ708856 *C. pilosum*, 13: KJ708855 *C. littorale*, 14: JN564129 *C. australasicum*, 15: JN564128 *C. australasicum*, 16: KX146374 *C. madagascariense*, 17: KR530506 *Canarium* sp.

Đặc biệt, khi so sánh trên 2 vùng gen *rbcL* và *rpoC1*, trình tự VNMN000863 *C. nigrum* không giống hoàn toàn với bất kỳ trình tự nào trên GenBank. Có tất cả 23 vị trí đột biến mất đi hay thay thế nucleotide khi so sánh giữa trình tự loài VNMN000863 *C. nigrum* với 26 trình tự trên

GenBank đối với vùng gen *rbcL*. Chẳng hạn, tại vị trí nucleotide 547 duy nhất trình tự VNMN000863 *C. nigrum* xuất hiện nucleotide (T) khi so sánh với 26 trình tự còn lại. Hay tại vị trí nucleotide thứ 126 và 462 (TT, tương ứng) được thay thế bằng (CC, tương ứng) khi so sánh trình tự VNMN000863 *C.*

nigrum với trình tự FJ466629 *C. decumanum* (có nguồn gốc ở New Caledonia). Tại vị trí nucleotide thứ 24, 288, 462 và 492 (CGTT, tương ứng) được thay thế bằng (CTCC, tương ứng) khi so sánh trình tự VNMN000863 *C. nigrum* với trình tự FJ466634 *C. madagascariense* (có nguồn gốc ở Cộng hòa

Madagascar) (Bảng 6). Còn tại vùng gen *rpoC1* chỉ có 2 vị trí sai khác nucleotide là vị trí thứ 401 (G) được thay bằng (T) và vị trí thứ 405 xuất hiện thêm nucleotide mới (T) khi so sánh trình tự VNMN000863 *C. nigrum* với trình tự GQ248897 *C. oleosum*.

Bảng 6. Các vị trí biến đổi nucleotide giữa loài Trám đen (VNMN000863 *C. nigrum*) với trình tự nucleotide của một số loài trên GenBank khi phân tích với vùng gen *rbcl*.

Vị trí nu.	3	1 2	1 8	2 4	4 2	6 9	1 1	1 2	1 4	2 0	2 8	3 3	3 4	3 6	4 6	4 9	5 1	5 4	5 5	5 8	5 8	5 8	6 2
1	A	T	T	C	T	G	A	T	T	T	G	C	A	G	T	T	A	T	T	T	A	T	T
2	C
3	T	.	.	.	C
4	G
5	C	T	A	C	.	C	-	C	G	.	G	.	.
6	.	G
7	G
8	.	.	.	C	T	.	.	.	C	C
9	T	.	.	.	C	C
10	C	A	T	.	C	C	.	.
11	.	.	T	A	.	C
12	A	.	C
13	T	.	.	.	C
14	G	C	G
15
16	C
17	C	C
18	G
19	C	A	.	.	C	.	.	.	T	C
20	G	C	.	C	-	G
21	G	C	G
22	T	.	.	.	C	C
23	G	C	G
24
25	A	.	.	C
26	C
27	.	.	.	G

Ghi chú: Các trình tự nucleotide trên GenBank được sử dụng để so sánh: 1: VNMN000863 *C. nigrum*, 2: KT698509 *Canarium* sp., 3: KT698507 *Canarium* sp., 4: KJ440021 *Canarium album*, 5: FJ466626 *Canarium album*, 6: FJ976124 *Canarium acutifolium*, 7: FJ466639 *Canarium tramdenum*, 8: FJ466634 *Canarium madagascariense*, 9: KT698497 *Canarium multiflorum*, 10: GU246029 *Canarium muelleri*, 11: KT698491 *Canarium indicum*, 12: FJ466636 *Canarium ovatum*, 13: KT698501 *Canarium obtusifolium*, 14: KT698499 *Canarium odontophyllum*, 15: FJ466637 *Canarium pilosum*, 16: FJ466633 *Canarium littorale*, 17: FJ466629 *Canarium decumanum*, 18: FJ466628 *Canarium bengalense*, 19: FJ466627 *Canarium balansae*, 20: KT698538 *Canarium latistipulatum*, 21: KT698508 *Canarium rufum*, 22: KT698504 *Canarium pulchrebracteatum*, 23: KT698503 *Canarium patentinervium*, 24: FJ466641 *Canarium whitei*, 25: FJ466642 *Canarium zeylanicum*, 26: FJ466640 *Canarium vulgare*, 27: FJ466638 *Canarium strictum*.

Mức độ đa dạng nucleotide khi so sánh giữa trình tự loài Trám đen nghiên cứu với các trình tự các loài trong chi *Canarium* trên GenBank dao động từ 0,000 (so với KP093198 *C. album* và KP093200 *C. tramdenum*) đến 0,0424 (so với KR531870 *C. album*, DQ517524 *C. album*, KP093201 *C. tramdenum*, KP093199 *C. album*, JF421482 *C. album* và KR531871 *C. subulatum*) đối với vùng gen *ITS*; từ 0,000 (so với KR530512 *C. acutifolium* và KR530509 *C. subulatum*) đến 0,014 (so với JN564128 *C. australasicum*) đối với vùng gen *matK*; từ 0,000 (so với FJ466637 *C. pilosum* và FJ466641 *C. whitei*) đến 0,006 (so với FJ466626 *C. album* và FJ466634 *C. madagascariense*) đối với vùng gen *rbcL*; và 0,002 (so với GQ248897 *C. oleosum*) đối với vùng gen *rpoC1* (số liệu không chỉ ra ở đây). Sở dĩ có mức độ đa dạng nucleotide từ 0,000 đối với vùng gen *rbcL* là vì khi phân tích xử lý trình tự nucleotide trong phần mềm MEGA 4.0 sự xuất hiện hay không xuất hiện các nucleotide đã bị bỏ qua.

Trên cơ sở so sánh các trình tự nucleotide 4 vùng gen nghiên cứu của loài Trám đen với trình tự tương ứng của các loài trên GenBank, chúng tôi đã tìm ra các vị trí bảo thủ (C), vị trí biến đổi (V) và vị

trí mang thông tin tiến hóa (Pi) giữa các loài là 306, 14 và 14, tương ứng khi phân tích với vùng gen *ITS*; 638, 24 và 11, tương ứng khi phân tích với vùng gen *matK*; 620, 22 và 12, tương ứng khi phân tích với vùng gen *rbcL*; 409, 01 và 0, tương ứng khi phân tích với vùng gen *rpoC1*. Mức độ đa dạng nucleotide trung bình (π) giữa loài Trám đen với các loài trong chi *Canarium* thể hiện cao nhất khi phân tích với vùng gen *ITS* (0,02), thứ hai là vùng gen *matK* (0,007), thứ ba là *rbcL* (0,003) (Bảng 7) (riêng vùng *rpoC1* chỉ có 2 trình tự nên không so sánh được giá trị π). Kết quả nhận được còn cho thấy, nhìn chung các vùng gen trong hệ gen lục lạp (*matK*, *rbcL*) có tính bảo thủ cao hơn vùng gen nhân (*ITS*) khi phân tích trên loài Trám đen. Kết quả này cũng tương tự với các báo cáo trước đây của Dinh Thi Phong *et al.*, (2014) khi nghiên cứu mức độ đa dạng nucleotide vùng gen nhân (*ITS*) và 4 vùng gen lục lạp (*trnL*, *matK* và *psbA-trnH*) trên 8 loài *Dalbergia* của Việt Nam và nhóm Vũ Thị Thu Hiền *et al.*, (2014) khi đánh giá mức độ đa dạng nucleotide vùng gen nhân (*PIF*) và 3 vùng gen lục lạp (*trnL-trnF*, *psbA-trnH*, *matK*) trên 8 loài *Bambusa* của Việt Nam.

Bảng 7. Thống kê mức độ biến đổi và mức độ đa dạng nucleotide khi so sánh giữa loài Trám đen với các loài trên GenBank khi phân tích với 4 vùng gen.

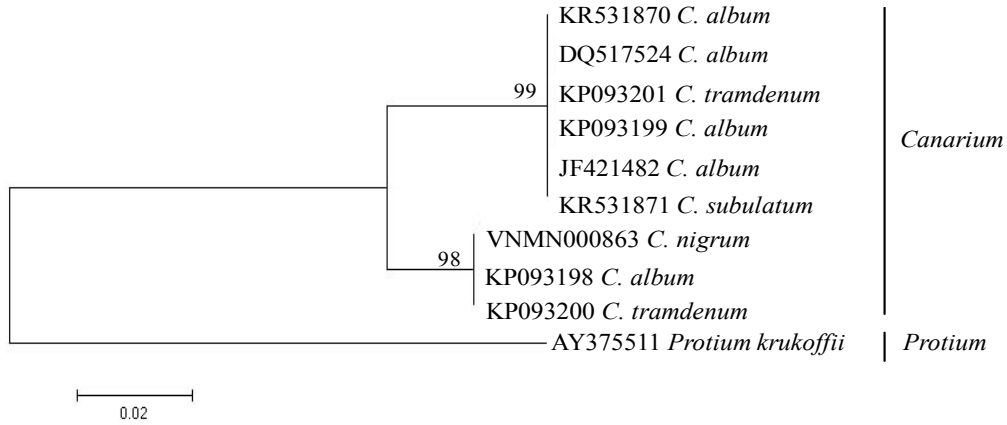
Vùng gen nghiên cứu	m	C	V	Pi	S	π
<i>ITS</i>	9	306	14	14	0	0,020
<i>matK</i>	19	638	24	11	13	0,007
<i>rbcL</i>	27	620	22	12	10	0,003
<i>rpoC1</i>	2	409	1	0	0	-

Ghi chú: m: số loài; C: vị trí bảo thủ; V: vị trí biến đổi; Pi: vị trí mang thông tin tiến hóa; S: vị trí singleton; π : mức độ đa dạng nucleotide; “-” không tính được.

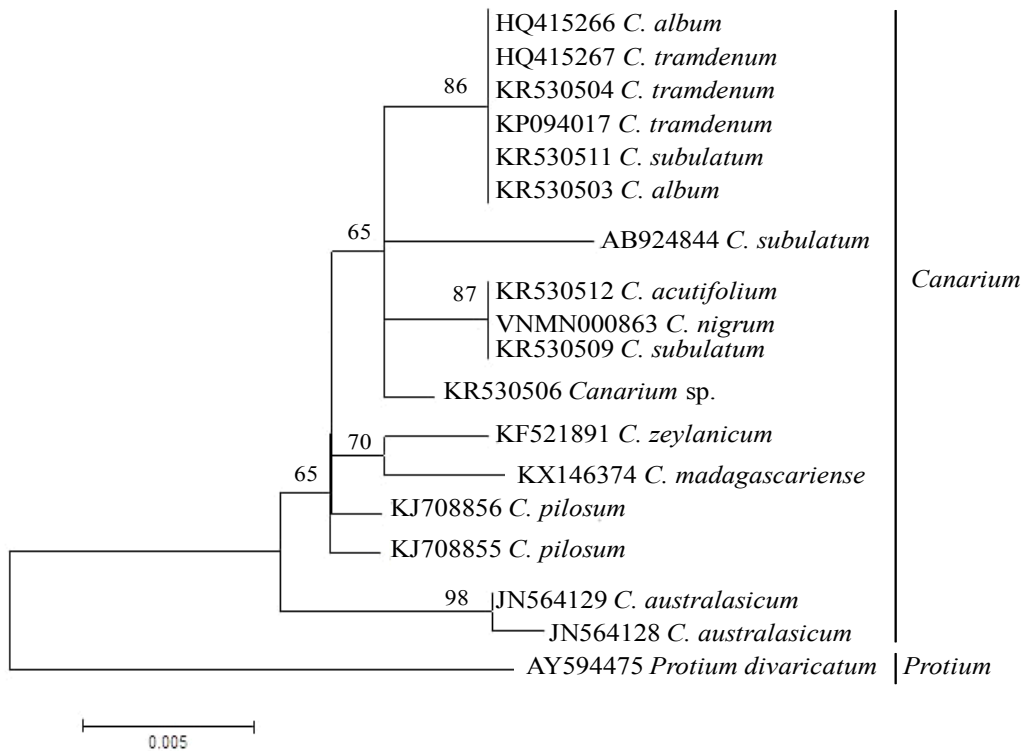
Vị trí phân loại của loài Trám đen trên cơ sở phân tích 4 vùng gen

Cây phát sinh chủng loại của loài Trám đen với các trình tự loài *Canarium* đã có trên GenBank đối với 4 vùng gen nghiên cứu (8 trình tự vùng gen *ITS*, 16 trình tự vùng gen *matK*, 26 trình tự vùng gen *rbcL* và 1 trình tự vùng gen *rpoC1*) được thể hiện trong Hình 2, 3, 4 và 5. Kết quả phân tích cho thấy, tất cả các loài trong cùng chi *Canarium* đều hình thành một nhánh tiến hóa riêng và liên quan mật thiết với nhau với các giá trị bootstrap tại các điểm nút tạo nhánh dao động từ 98 đến 99% đối với vùng gen *ITS* (Hình 2), từ 65 đến 98% đối với vùng gen *matK* (Hình 3), từ 35 đến

67% đối với vùng gen *rbcL* (Hình 4) và 98 đến 99% đối với vùng gen *rpoC1* (Hình 5). Dẫn liệu cũng chỉ ra những loài có mức độ tương đồng nucleotide cao đều nằm co cụm trong cùng một nhánh tiến hóa. Chẳng hạn đối với vùng gen *ITS*, trình tự loài Trám đen (VNMN000863 *C. nigrum*) và 2 trình tự loài Trám của Trung Quốc là KP093200 *C. tramdenum* (Trám đen) và KP093198 *C. album* (Trám trắng) đã lập thành một nhánh riêng biệt với giá trị bootstrap là 98% (Hình 2). Hay như tại vùng gen *matK* trình tự VNMN000863 *C. nigrum* và KR530509 *C. subulatum*, KR530512 *C. acutifolium* (đều có nguồn gốc ở Trung Quốc) hình thành một nhánh tiến hóa với với giá trị bootstrap là 87% (Hình 3).



Hình 2. Cây phát sinh chủng loại của loài Trám đen (VNMN000863 *C. nigrum*) với các loài trong chi *Canarium* trên GenBank bằng phần mềm MEGA 4.0 trên cơ sở phân tích vùng gen ITS.



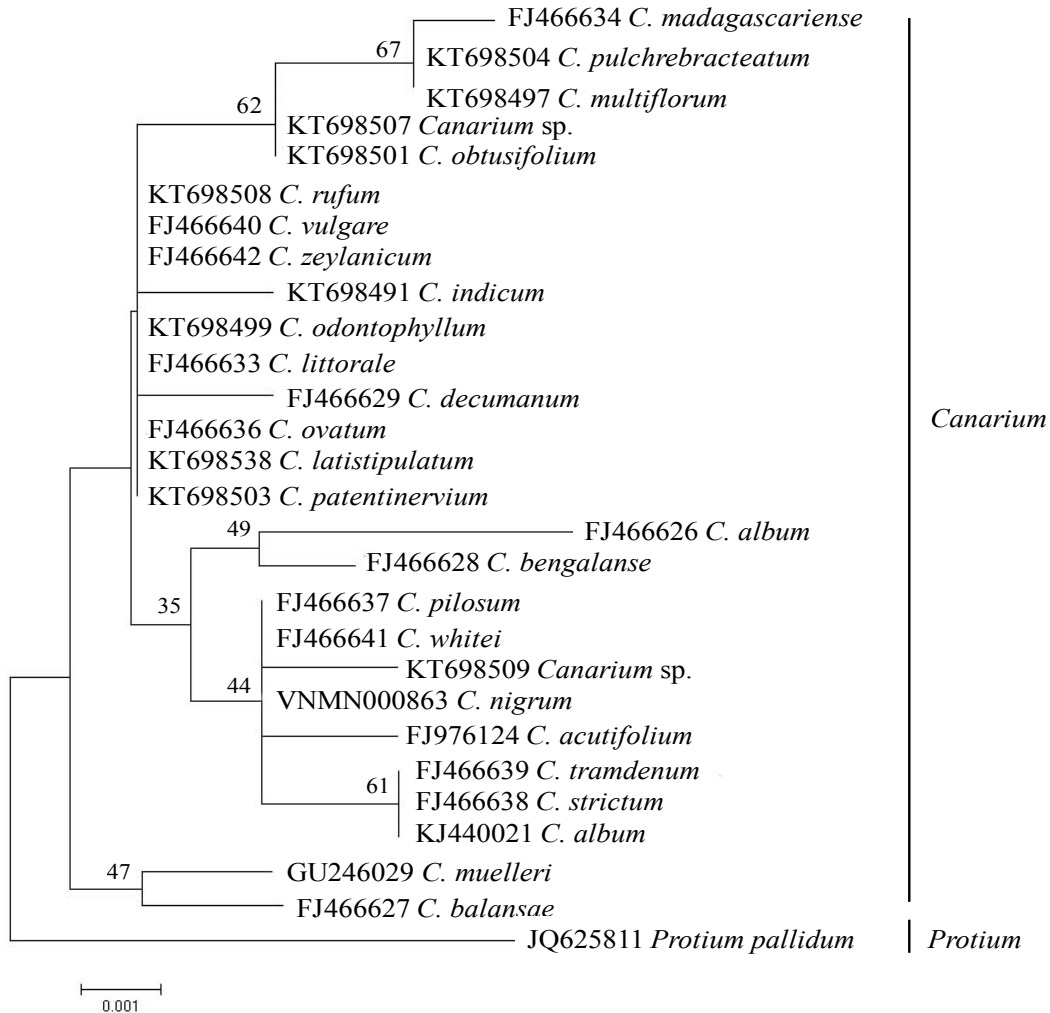
Hình 3. Cây phát sinh chủng loại của loài Trám đen (VNMN000863 *C. nigrum*) với các loài trong chi *Canarium* trên GenBank bằng phần mềm MEGA 4.0 trên cơ sở phân tích vùng gen *matK*.

Tại vùng gen *rbcL*, mặc dù có sự sai khác nucleotide đặc trưng giữa trình tự VNMN000863 *C. nigrum* với trình tự FJ466637 *C. pilosum* (nguồn gốc

ở Malaysia) và FJ466641 *C. whitei* (nguồn gốc ở New Caledonia) nhưng mức độ dạng nucleotide giữa chúng là 0,0% (do phần mềm MEGA 4.0 đã bỏ qua

sự khác biệt thêm hoặc bớt nucleotide khi so sánh) nên trong cây phát sinh chủng loại (Hình 4) trình tự

VNMN000863 *C. nigrum* và 3 trình tự này đã co cụm lại thành nhánh tiến hóa riêng biệt.



Hình 4. Cây phát sinh chủng loại của loài Trám đen (VNMN000863 *C. nigrum*) với các loài trong chi *Canarium* trên GenBank bằng phần mềm MEGA 4.0 trên cơ sở phân tích vùng gen *rbcL*.

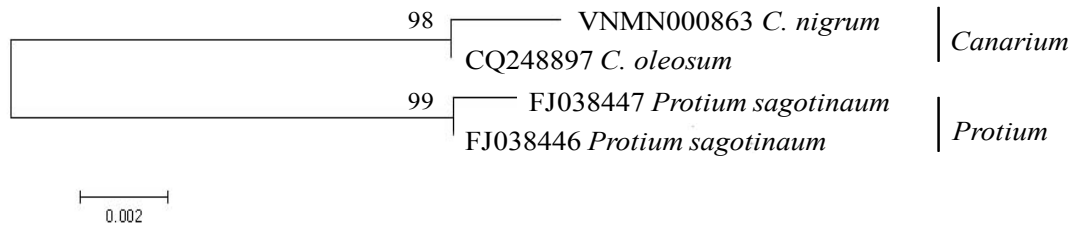
Còn tại vùng gen *rpoC1*, trình tự VNMN000863 *C. nigrum* có 2 vị trí nucleotide sai khác với trình tự GQ248897 *C. oleosum* nên trong cây phát sinh chủng loại có sự phân nhánh giữa 2 trình tự này (Hình 5).

Bốn vùng gen *ITS*, *matK*, *rbcL* và *rpoC1* đã được đề xuất như mã vạch DNA ở thực vật (Newmaster *et al.*, 2009; Liu *et al.*, 2010). Trong nghiên cứu này, cả 4 vùng gen đều nhân bản thành công các đoạn gen đích cho loài Trám đen và đã chỉ ra sự phân nhánh giữa các loài trong chi.

Trong đó vùng gen *rpoC1* có mức độ phân nhánh giữa các loài trong chi *Canarium* là rõ nhất, với giá trị bootstrap tại các điểm nút dao động từ 98 đến 99%. Một nghiên cứu khác của Weeks (2009) sử dụng vùng gen nhân (ETS, NIA-i3) và gen lục lạp (*rbcL*, *rps16*, *psbA-trnH*, *trnL*, *trnL-F*) để đánh giá sự tiến hóa của 16 loài *Canarium* trên thế giới cũng chỉ ra vùng gen *rbcL* có sự phân nhánh giữa các loài ở mức tương đối (Weeks, 2009). Trái với kết quả nghiên cứu của chúng tôi, vùng gen *ITS* lại rất có hiệu quả trong nghiên cứu tiến hóa và phát sinh loài của chi *Bursera* trong họ

Burseraceae (Becerra, Venable, 1999; Becerra, 2003; Becerra *et al.*, 2012). Gen *matK* hiện được đề xuất như mã vạch DNA ở thực vật (Peter *et al.*, 2009) nhưng lại không thành công trong nghiên cứu này (vì đã không phân tách được loài KR530512 *C. acutifolium*, KR530509 *C.*

subulatum và VNMN000863 *C. nigrum*). Như vậy mỗi vùng gen đều có những đặc trưng riêng và có hiệu quả phân biệt đối với các loài khác nhau. Qua kết quả trong nghiên cứu chúng tôi nhận thấy có thể sử dụng vùng gen *rpoC1* cho việc nhận dạng loài Trám đen ở một số tỉnh phía Bắc Việt Nam.



Hình 5. Cây phát sinh chủng loại của loài Trám đen (VNMN000863 *C. nigrum*) với các loài trong chi *Canarium* trên GenBank bằng phần mềm MEGA 4.0 trên cơ sở phân tích vùng gen *rpoC1*.

KẾT LUẬN

Không có sự sai khác nucleotide ở cả 9 cá thể đại diện loài Trám đen khi phân tích 4 vùng gen *ITS*, *matK*, *rbcL* và *rpoC1*. Mức độ đa dạng nucleotide trung bình (π) giữa các loài Trám cao nhất (0,02) đối với vùng gen *ITS*, thứ hai là vùng gen *matK* (0,007), thứ ba là vùng gen *rbcL* (0,003). Cây phát sinh chủng loại của loài Trám đen nghiên cứu với các loài trong chi *Canarium* đã chỉ ra khả năng phân nhánh giữa các loài trong chi ở cả 4 vùng gen với giá trị bootstrap tại các điểm đầu nút chính dao động từ 98 đến 99% đối với vùng gen *ITS*, từ 65 đến 98% đối với vùng gen *matK*, từ 35 đến 67% đối với vùng gen *rbcL* và từ 98 đến 99% đối với vùng gen *rpoC1*. Loài Trám đen có mức độ tương đồng nucleotide gần gũi nhất với loài *C. tramdenum* (KP093200) và *C. album* (KP093198) (đều có nguồn gốc ở Trung Quốc) khi phân tích vùng gen *ITS* và loài *C. subulatum* (KR530509), *C. acutifolium* (KR530512) (đều có nguồn gốc ở Trung Quốc) khi phân tích vùng gen *matK*. Vùng gen *rpoC1* (độ dài 522 bp) có thể sử dụng cho việc nhận dạng các loài thuộc chi *Canarium*.

Lời cảm ơn: Công trình được hoàn thành nhờ kinh phí của Nhiệm vụ Quỹ gen cấp quốc gia năm 2016, mã số NVQG-2016-14. Chủ nhiệm đề tài xin chân thành cảm ơn các thành viên và các cơ quan địa phương tham gia thực hiện đề tài.

TÀI LIỆU THAM KHẢO

Becerra JX (2003) Evolution of Mexican *Bursera*

(Burseraceae) inferred from *ITS*, *ETS*, and 5S nuclear ribosomal DNA sequences. *Mol Phylogenet Evol* 26: 300–309.

Clarkson JJ, Chase MW, Harley MM (2002) Phylogenetic relationships in Burseraceae based on plastid *rps16* intron sequences. *Kew Bull* 57: 183–193. doi:10.2307/4110826.

Doyle JJ., Doyle JJ (1990) Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.

Fay MF, Cameron KM, Prance GT, Lledo MD, Chase MW (1997) Familial relationships of *Rhabdodendron* (Rhabdodendraceae): plastid *rbcL* sequences indicate a caryophyllid placement. *Kew Bull* 52: 923–932.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–98.

http://thongtinkhcn.com.vn/vn/tin-tuc/detail.php?ELEMENT_ID=2883

<http://www.kew.org/barcoding/protocols.html>

<http://www.ncbi.nlm.nih.gov/nucleotide>

Huang XC, Ci XQ, Conran JG, Li J (2015) Application of DNA barcodes in Asian tropical trees – a case study from Xishuangbanna Nature Reserve, southwest China. *PLoS ONE* 10(6): e0129295. doi:10.1371/journal.pone.0129295.

Liu J, Yan HF, Newmaster SG, Pei N, Ragupathy S, Ge XJ (2014) The use of DNA barcoding as a tool for the conservation biogeography of subtropical forests in China. *Divers Distrib* 1–12.

Liu Y, Yan HF, Ge XJ (2010) Evaluation of 10 plant barcodes in Bryophyta (Mosses). *J Syst Evol* 48: 36–46

- Newmaster SG, Ragupathy S (2009) Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Mol Ecol Res* 9: 172–180.
- Peter MH, Laura LF, John LS (2009) A DNA barcode for land plants. *Proc Natl Acad Sci USA* 106: 12794–12797.
- Phong DT, Tang DV, Hien VTT, Ton ND, Hai NV (2014) Nucleotide diversity of a nuclear and four chloroplast DNA regions in rare tropical wood species of *Dalbergia* in Vietnam: A DNA barcode identifying utility. *Asian J Appl Sci* 2 (2): 116–125.
- Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am J Bot* 84(8): 1120–1136.
- Schoch CL, Spouge JL, Seifert KA, Huhndorf S, Robert V, et al., (2012) Nuclear ribosomal internal transcribed spacer (*ITS*) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci USA* 109 (16): 6241–6246.
- Sun Y, Xia N, Lin R (2005) Phylogenetic analysis of *Bambusa* (Poaceae: Bambusoideae) based on internal transcribed spacer sequences of nuclear ribosomal DNA. *Biochem Genet* 43(11-12): 603–612.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
- Becerra JX, Noge K, Olivier S, Venable DL (2012) The monophyly of *Bursera* and its impact for divergence times of Burseraceae. *Taxon* 61(2): 333–343.
- Vũ Thị Thu Hiền, Đinh Thị Phòng, Nguyễn Tường Vân, Nguyễn Khắc Khôi (2014) Đa dạng nucleotide bốn vùng gen của tám loài tre thuộc chi *Bambusa* Schreb. chưa xác định tên khoa học. *Tạp chí Khoa học và Công nghệ* 52 (2D): 285–291.
- Becerra JX, Venable DL (1999) Nuclear ribosomal DNA phylogeny and its implications for evolutionary trends in Mexican *Bursera* (Burseraceae). *Am J Bot* 86(7): 1047–1057.
- Vũ Thị Thu Hiền, Trần Thị Việt Thanh, Nguyễn Khắc Khôi, Đinh Thị Phòng (2012) Lâm sàng tổ tên khoa học cho một số loài thuộc chi Tre (*Bambusa* Schreb.) ở Việt Nam do biến đổi hình thái trên cơ sở giải mã trình tự gen *trnL-trnF*, *psbA-trnH* và *matK*. *Tạp chí Khoa học và Công nghệ* 50(4): 463–473
- Weeks A, Simpson BB (2004) Molecular genetic evidence for interspecific hybridization among Hispaniolan *Bursera* (Burseraceae). *Am J Bot* 91: 975-983. doi:10.3732/ajb.91.6.976
- Weeks A, Simpson BB (2007) Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution and historical biogeography of an “impossible” genus. *Mol Phylogenet Evol* 42: 62-79. doi:10.1016/j.ympev.2006.06.015.
- Weeks A (2009) Evolution of the pili nut genus (*Canarium* L., Burseraceae) and its cultivated species. *Genet Resour Crop Evol* 56: 765–781. DOI 10.1007/s10722-008-9400-9404.
- Weeks A, Daly DC, Simpson BB (2005). The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. *Mol Phylogenet Evol* 35: 85–101.
- White TJ, Bruns T, Lee S, Taylor DJ (1990) *Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics*. In Gelfand D, Sminsky J, White T, eds. *PCR protocols: a guide to methods and applications*. Academic Press, San Diego, California, USA: 315–322.
- Yang HQ, Sheng P, Zhu LD (2007) Generic delimitations of *Schizostachyum* and its allies (Gramineae: Bambusoideae) inferred from GBSSI and *trnL-F* sequence phylogenies. *Taxon* 56(1): 45–54.

NUCLEOTIDE DIVERSITY OF A NUCLEAR GENE *ITS* REGION AND CHLOROPLAST GENES (*matK*, *rbcl*, *rpoC1*) OF *CANARIUM NIGRUM* IN SOME PROVINCES IN NORTHERN VIETNAM

Đinh Thị Phong^{1,3}, Trần Thị Liễu¹, Vũ Thị Thu Hiền¹, Hoàng Thanh Lộc²

¹Vietnam National Museum of Nature, Vietnam Academy of Science and Technology

²Institute for Improvement of Forest Genetic Resources and Products Development

³Graduate University of Science and Technology, Vietnam Academy of Science and Technology

SUMMARY

In this study, three chloroplast genes (*matK*, *rbcl* and *rpoC1*) and one nuclear gene (*ITS*) were used to assess the nucleotide diversity for nine individuals of *Canarium nigrum* species in Bac Giang, Hoa Binh and Phu Tho provinces (three individuals from each province). The nucleotide sequence of the four regions (*ITS*, *matK*, *rbcl* and *rpoC1*) of *C. nigrum* were determined with the results showing their size to be 696 bp, 798 bp,

702 bp and 522 bp, respectively. Results of nucleotide sequence comparison for the studied samples exhibited 100% similarity for all the four gene regions. Sequence comparison with other species in the *Canarium* genus available in the GenBank revealed that the nucleotide diversity level (π) was the highest for the *ITS* gene (0.02), followed by *matK* (0.007), and the lowest for *rbcL* (0.003). The phylogenetic tree of *C. nigrum* with the species in *Canarium* genus indicated that the separation of species was the clearest for the *rpoC1* gene, followed by *rbcL*, *matK* and *ITS* gene, with the bootstrap values obtained from the branching nodes of each species ranging from 98 to 99%, 35 to 67%, 65 to 98% and 98 to 99%, respectively. The species *C. nigrum* had the closest nucleotide similarity to *C. trandenum* (KP093200) and *C. album* (KP093198) for the *ITS* gene and to species *C. subulatum* (KR530509), *C. acutifolium* (KR530512) for *matK* gene region. These results suggests the *rpoC1* gene region could be used as barcode for the species in genus *Canarium*.

Keywords: *Canarium nigrum*, nucleotide diversity, *ITS*, *matK*, *rbcL*, *rpoC1*