Phylogenetic Analysis of Black Peper (*Piper* Spp.) Population Collected in Different Locations of Vietnam Based on the $ITS_{U1\text{-}4}$ Gene Region

Sonexay Rasphone
Institute of Biotechnology, Hue University
Long Thanh Dang
Institute of Biotechnology, Hue University
Hoan Nguyen
Institute of Biotechnology, Hue University
Ngoc Quang Nguyen
Research and Developmnet Center, Pleiku, Gia Lai
Oanh Thi Duong
Research and Developmnet Center, Pleiku, Gia Lai
Hung Van Phi Nguyen
Institute of Biotechnology, Hue University
Kim-Cuc Thi Nguyen
Institute of Biotechnology, Hue University
Hai Thi Hong Truong (tthhai@hueuni.edu.vn)
Institute of Biotechnology, Hue University
Nhi Thi Hoang Ho
Institute of Biotechnology, Hue University

Research Article

Keywords: Phylogenetic, Piper spp., Black peper, ITS, Vietnam, Genetic analysis

Posted Date: September 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-844711/v1

Abstract

Background: The internal transcribed spacer (*ITS*) of nuclear ribosomal DNA is one of the most commonly used DNA markers in plant phylogenetic and DNA barcoding analyses, and it has been recommended as a core plant DNA barcode. To compare and find out the analysis genetic diversity difference some pepper individuals collected in different localities in Vietnam when using the *ITS* of nuclear ribosomal DNA. The *ITS* gene region from the nuclear genomes were tested for their suitability as DNA barcoding regions of thirty-nine pepper individuals. Universal primers were used, and sequenced products were analyzed using the Maximum Likelihood method and Tamura-Nei model in the MEGA X program.

Results: We did not observe high variability in intraspecific distance within the $ITS_{u1\text{-}4}$ gene region between individuals, ranged from 0.000 to 0.155 (mean = 0.033). The size of the gene region has fluctuated from 667 to 685 bp between different individuals with the percentage (G + C) contained in the $ITS_{u1\text{-}4}$ gene region was ranged from 54.776% to 60.805%, mean = 60.174%. The values of Fu's Fs, D, Fu and Li's D* and F* were negative as well (Fs = -0.209, D = -1.824; *P < 0.05*, D* = -1.205; *not significant, P > 0.10* and F* = -1.699; *not significant, 0.10 > P > 0.05*), indicating an excess of recently derived haplotypes and suggesting that either population expansion or background selection has occurred. The value Strobeck's S the obtained between individuals in a population is high (S = 0.684). The results of evolutionary relationships of taxa obtained 3 groups with the highest value of Fst is shown in the pairs of groups II and III (Fst = 0.151), and the lowest is in groups II and I (Fst = 0.015). All of the new sequences have been deposited in GeneBank under the following accession numbers MZ636718 to MZ636756.

Conclusions: This database is an important resource for researchers working on Species of pepper in Vietnam and also provides a tool to create $ITS_{u1\text{-}4}$ databases for any given taxonomy.

Background

The tropical plant family Piperaceae has provided many past and present civilizations with a source of diverse medicines and food-grade spice [1]. Piper, comprising more than 2000 species, is the largest genus in the family Piperaceae, with most species growing in the tropics, although some extend into the subtropical zone [2]. Most Piper species are famous due to their delicious taste and biological activities [3]. For example, the fruits of P. nigrum are one of the important flavorings in the world [4]. The largest number of Piper species are found in the Americas (about 700 species), with about 300 species from Southern Asia. There are smaller groups of species from the South Pacific (about 40 species) and Africa (about 15 species). The American, Asian, and South Pacific groups each appear to be monophyletic; the affinity of the African species is unclear [5].

For identification and classification of different taxa, rapid species identification techniques like DNA barcoding have been undertaken by different groups utilizing DNA regions from the mitochondrial, plastid, and nuclear genomes. Traditional morphophenology methods to identify Peper species are

mostly based on phenotypic characters, but morphological characteristics are subjected to be affected by developmental and environmental [1; 6; 7]. Therefore, DNA barcoding, a new method for the quick identification of any species based on extracting a DNA sequence from a tiny tissue sample of any organism, is now being applied to taxa across the tree of life. As a research tool for taxonomists, DNA barcoding assists in identification by expanding the ability to diagnose species by including all life history stages of an organism. As a biodiversity discovery tool, DNA barcoding helps to flag species that are potentially new to science. As a biological tool, DNA barcoding is being used to address fundamental ecological and evolutionary questions, such as how species in plant communities are assembled [8].

The internal transcribed spacer (*ITS*, or a part of it) of nuclear ribosomal DNA is one of the most commonly used DNA markers in plant phylogenetic and DNA barcoding analyses, and it has been recommended as a core plant DNA barcode [9]. *ITS* was first proposed as a barcode for flowering plants [10] but lost popularity for some time due to concerns about the incomplete concerted evolution of multiple copies, different alleles from paternal and maternal parents, DNA contamination of different species (e.g. through symbiosis) and some technical problems. It was demonstrated that these imperfections did not cause large problems, and it was reproposed as a core barcode for seed plants [11; 12; 13].

In recent classified botanical studies, the *ITS* genetic region is the most commonly decoded locus. The region is highly effective in the classification of a variety of plants and fungi (except ferns), and this is a locus used for short DNA sequencing [14]. At the species level, the *ITS* genetic region has a high diversity (about 13.6% between closely related species) and has been demonstrated in almost all studies. The *ITS* genetic region has also been shown to have low levels of variation within the species [15]. Today, in the presence of more than 100.000 *ITS* sequences (as of December 2016) published on Genbank, this is a valuable resource, opening great prospects for species identification researches. The amount of sequences continues to be added daily [16]. The purpose of this study was to test the utility of DNA barcoding for the identification of closely related Peper individuals in the population base on gene region *ITS*. In a conservation project, the individuals were collected from different locations in Vietnam. In this study, we used the internal transcribed spacer (*ITS*) of nuclear ribosomal DNA with the universal *ITS* primer pair referenced from Chen et al. (2016) [9] genetic diversity analysis of the Peper population collected in different locations, Vietnam.

Materials And Methods

Plant materials

In this study, thirty-nine black peper leaf samples were selected and collected in different localities of Vietnam (Table 6). They were washed with distilled water and then refrigerated in the dark for further experiments.

Table 6
The list of black peper samples were collected

Individuals	GenBank Sequence ID	Source of materials	Identifier
HUIB_PN27	MZ636718	Quang Tri, Vietnam	Truong,H.T., Rasphone,S., Dang,L.T., Nguyen,H.V., Nguyen,N.Q., Duong,O.T., Nguyen,C.T., Nguyen,H.V., Hoang,Q.T., Nhung,N.A., Triet,N.T. and Nguyen,L.T. Institute of Biotechnology, Hue University, Road 10, Hue, Thua Thien Hue 53000, Viet Nam
HUIB_PN29	MZ636719	Quang Tri, Vietnam	
HUIB_PN38	MZ636720	Quang Tri, Vietnam	
HUIB_PN42	MZ636721	Gia Lai, Vietnam	
HUIB_PN43	MZ636722	Gia Lai, Vietnam	
HUIB_PN45	MZ636723	Gia Lai, Vietnam	
HUIB_PN47	MZ636724	Quang Nam, Vietnam	
HUIB_PN35	MZ636725	Gia Lai, Vietnam	
HUIB_PN52	MZ636726	Binh Phuoc, Vietnam	
HUIB_PN54	MZ636727	Dong Nai, Vietnam	
HUIB_PR48	MZ636728	Quang Tri, Vietnam	
HUIB_PR41	MZ636729	Quang Tri, Vietnam	
HUIB_PH46	MZ636730	Quang Tri, Vietnam	
HUIB_PH30	MZ636731	Quang Tri, Vietnam	

Individuals	GenBank Sequence ID	Source of materials	Identifier
HUIB_PN21	MZ636732	Quang Nam, Vietnam	
HUIB_PN56	MZ636733	Quang Ngai, Vietnam	
HUIB_PN55	MZ636734	Phu Quoc, Vietnam	
HUIB_PN10	MZ636735	Gia Lai, Vietnam	
HUIB_PN20	MZ636736	Gia Lai, Vietnam	
HUIB_PN50	MZ636737	Gia Lai, Vietnam	
HUIB_PN69	MZ636738	Gia Lai, Vietnam	
HUIB_PN70	MZ636739	Gia Lai, Vietnam	
HUIB_PN84	MZ636740	Gia Lai, Vietnam	
HUIB_PN87	MZ636741	Gia Lai, Vietnam	
HUIB_PN89	MZ636742	Gia Lai, Vietnam	
HUIB_PN91	MZ636743	Gia Lai, Vietnam	
HUIB_PN93	MZ636744	Gia Lai, Vietnam	
HUIB_PN95	MZ636745	Gia Lai, Vietnam	
HUIB_PN96	MZ636746	Gia Lai, Vietnam	
HUIB_PN97	MZ636747	Gia Lai, Vietnam	
HUIB_PN101	MZ636748	Gia Lai, Vietnam	

Individuals	GenBank Sequence ID	Source of materials	Identifier
HUIB_PN102	MZ636749	Gia Lai, Vietnam	
HUIB_PN105	MZ636750	Gia Lai, Vietnam	
HUIB_PN113	MZ636751	Gia Lai, Vietnam	
HUIB_PN114	MZ636752	Gia Lai, Vietnam	
HUIB_PN115	MZ636753	Gia Lai, Vietnam	
HUIB_PN116	MZ636754	Gia Lai, Vietnam	
HUIB_PD36	MZ636755	Gia Lai, Vietnam	
HUIB_PN34	MZ636756	Dak Lak, Vietnam	

Methods

DNA extraction and purification

Genomic DNA of the 39 individuals black peper was extracted from leaves following the protocol described by Raz and Ecker (1997) [28]. Then, extracted DNA were incubated with a 1:10,000 dilution of the SYBR Green I nucleic acid gel stain (Invitrogen, USA) for 20 min and separated on a 1 % agarose gels using 0.5X TBE buffer for 30 minutes at 120 V and photographed under UV light (100 bp ladder was used as a molecular weight marker). DNA containing many impurities will be purified with a QIAquick gel extraction kit (Qiagen, Germany). Besides, DNA concentration was measured on a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, USA), the quality of the DNA was checked using agarose gel electrophoresis and absorption spectroscopy (ratio A260/A280).

PCR amplification and sequencing

The *ITS*$_{u1\text{-}4}$ gene region was amplified in a 25 µL reaction volume, using OneTaq® DNA Polymerase (Biolabs Inc., New England), 5 µL One Taq standard reaction buffer (5X), 5 mM dNTP, 5 µM of ITS$_{u1}$ primer, 5 µM of ITS$_{u4}$ primer and 100 ng DNA template (50 ng/µL), 0,125 µL (1.25Unit) OneTaq® DNA Polymerase and sterile distilled water to a final volume of 25 µl. PCR amplification was performed on Applied Biosystems-Life Technologies (Thermo Fisher Scientific Inc. United States). The *ITS*$_{u1\text{-}4}$ gene region is amplified with a pair of ITS$_{u1}$ primers: GGAAGKARAAGTCGTAACAAGG and ITS$_{u4}$: RGTTTCTTTTCCTCCGCTTA [9], and the following thermal cycle: 95°C/5 minutes; 30 cycles x (95°C/40

seconds; 56°C/1 minute; 72°C/ 1 minute); 72°C/10 minutes. PCR products are tested by electrophoresis on 1% agarose gel in TAE 1X buffer with Ethidium bromide dye and read electrophoresis images by direct UV reading system (UV-transilluminator, Model: DyNa Light). Samples showing a clear single band were sent to Maccrogen Company, Korea, and sequenced in both directions with the same primers used for PCR by the method the dideoxy terminator method on the ABI PRISM® 3100 Avant Genetic Analyzer (Applied Biosystems)

Data analysis

Raw sequences for the $ITS_{u1\text{-}4}$ gene region were assembled and edited using BioEdit v7.2.5. Edited sequences were then aligned by ClustalW in MEGA X and the non-overlapping sequence regions at the 5'- and 3'-ends were trimmed [22]. The seven parameters including the number of separate polymorphic sites (S), the total number of mutant sites (Eta), number of haplotypes (h), haplotype diversity (Hd), the average number of nucleotide differences (k), nucleotide diversity (Pi), Minimum number of recombination events (Rm) are considered as a polymorphic measurement in the population. Neutrality is tested based on five methods namely (Tajima's D test [19], Fs, Fu's statistic [21], D* and F*, Fu and Li's statistics [21]; S, Strobeck's statistic [29] were used to DNAsp 6.0 software [30].

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [31]. The tree with the highest log likelihood (-2082.26) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with a superior log-likelihood value. Evolutionary analyses were conducted in MEGA X [22]. The barcode sequences were queried against the GeneBank database (NCBI) using the Nucleotide BLAST algorithm. Haplotype network construction is a widely used approach for analyzing and visualizing the relationships among DNA sequences within individuals of the Peper population base on Network 10.2 software.

Results

Sequence characteristics and genetic diversity analysis of the ITS_{u1-4} gene region

The ITS_{u1-4} gene region of the Pepper population showed high success rates for PCR amplification and sequencing using a single primer pair specific (100%). The sequences characteristics of the gene region are presented in Table 1 shows that the genetic distances of the ITS_{u1-4} gene region of the total of 39 individuals ranged from 0.000 to 0.155 (mean = 0.033) (Table 1)

Table 1
The characteristics based on of ITS_{u1-4} gene region of Pepper populatation

Regions gene	PCR success (%)	Sequencing success (%)	Total aligned length (bp)	Number of monomorphic sites	Variable sites (%)	Intraspecific distance (mean)
ITS_{u1-4}	100	100	667–685	517–535	21.898	0.000-0.155 (0.033)

Table 2
Nucleotide components of ITS_{u1-4} gene region of Pepper population

Individuals	Compute nucleotide composition (%)					Total
	T(U)	C	A	G	G + C	(bp)
HUIB_PN27	19.970	28.912	19.672	31.446	60.358	671
HUIB_PN29	19.940	28.720	19.940	31.399	60.119	672
HUIB_PN38	19.940	28.720	19.940	31.399	60.119	672
HUIB_PN42	19.970	28.912	19.821	31.297	60.209	671
HUIB_PN43	19.970	28.912	19.672	31.446	60.358	671
HUIB_PN45	19.970	28.912	19.672	31.446	60.358	671
HUIB_PN47	19.821	29.061	19.672	31.446	60.507	671
HUIB_PN35	19.970	28.763	19.672	31.595	60.358	671
HUIB_PN52	19.970	28.763	19.821	31.446	60.209	671
HUIB_PN54	19.821	29.210	19.821	31.148	60.358	671
HUIB_PR48	19.640	29.085	21.739	29.535	58.621	667
HUIB_PR41	19.640	29.085	21.739	29.535	58.621	667
HUIB_PH46	18.394	31.825	18.832	30.949	62.774	685
HUIB_PH30	18.394	31.825	18.832	30.949	62.774	685
HUIB_PN21	19.970	28.763	19.672	31.595	60.358	671
HUIB_PN56	19.970	28.763	19.672	31.595	60.358	671
HUIB_PN55	19.970	28.763	19.821	31.446	60.209	671
HUIB_PN10	19.374	29.210	19.970	31.446	60.656	671
HUIB_PN20	20.119	28.614	19.821	31.446	60.060	671
HUIB_PN50	19.970	29.210	19.225	31.595	60.805	671
HUIB_PN69	19.970	28.763	19.821	31.446	60.209	671
HUIB_PN70	19.672	29.210	21.311	29.806	59.016	671
HUIB_PN84	19.672	29.508	20.268	30.551	60.060	671
HUIB_PN87	19.970	28.763	19.821	31.446	60.209	671
HUIB_PN89	19.523	28.912	19.970	31.595	60.507	671

Individuals	Compute nucleotide composition (%)					Total
	T(U)	C	A	G	G + C	(bp)
HUIB_PN91	19.851	28.955	19.851	31.343	60.299	670
HUIB_PN93	19.672	28.614	20.417	31.297	59.911	671
HUIB_PN95	19.523	29.061	20.119	31.297	60.358	671
HUIB_PN96	19.970	28.763	19.821	31.446	60.209	671
HUIB_PN97	19.970	28.763	19.821	31.446	60.209	671
HUIB_PN101	19.970	28.763	19.970	31.297	60.060	671
HUIB_PN102	20.119	28.614	19.821	31.446	60.060	671
HUIB_PN105	19.821	29.061	19.672	31.446	60.507	671
HUIB_PN113	19.970	28.763	19.821	31.446	60.209	671
HUIB_PN114	19.821	29.061	19.672	31.446	60.507	671
HUIB_PN115	19.523	29.061	19.821	31.595	60.656	671
HUIB_PN116	19.672	29.061	19.821	31.446	60.507	671
HUIB_PD36	20.597	28.209	24.627	26.567	54.776	670
HUIB_PN34	19.970	28.763	19.821	31.446	60.209	671
Avg.	**19.795**	**29.047**	**20.031**	**31.128**	**60.174**	**671.513**

The PCR products of the ITS_{u1-4} gene region were sequenced on ABI PRISM® 3100 Avant Genetic Analyzer (Applied Biosystems) by the dideoxy terminator method. The results of the ITS_{u1-4} gene region were 667 bp (HUIB_PR41 and HUIB_PR48), 670 bp (HUIB_PN36 and HUIB_PN91), 672 bp (HUIB_PN29 and HUIB_PN38), 685 bp (HUIB_PN46 and HUIB_PN30), and 671 bp for the remaining individuals Peper (Table 1). The BLAST result on NCBI was used to verify and compare with the sequences of the genus Peper showed that the nucleotide sequences obtained were highly similar to species of the *Piper nigrum* (accession number: MH493477-MH493487, KF924121, KF924111), *Piper retrofractum* (accession number: MH493562), *Piper hancei* (accession number: EF450274) and *Piper divaricatum* (accession number: DQ868714) ranging from 96 to 100%. The percentage of occurrence of each type of nucleotide in the ITS_{u1-4} gene region showed that Guanidin (G) accounts for the highest proportion, ranging from 26.567 to 31.595% (mean = 31.128%), followed by Cysteine (C) accounting for 28.209 to 28.955 % (mean = 29.047%) and the lowest was Timin (Uracin) accounting for 18.394 to 20.119% (mean = 19.795%).

Table 3
The results of DNA Polymorphism based on ITS_{u1-4} gene region of Pepper population

Regions gene	n	S	Eta	h	Hd	k	Pi ($x10^{-3}$)	Rm
ITS_{u1-4}	31	150	169	23	0.947	20.352	30.930	18

Note- *n: Number of samples; S: Number of variable sites; Eta: Total number of mutations; h: Number of Haplotypes; Hd: Haplotype (gene) diversity; Pi: Nucleotide diversity (per site); k: Average number of nucleotide differences; Rm: Minimum number of recombination events*

The percentage (G + C) contained in the ITS_{u1-4} gene region was the highest at 60.805% (HUIB_PN50) and there is a difference between different Pepper individuals, the differences ranged from 54.776–60.805% and reached an average of 60.174% (Table 2). All of the new sequences have been deposited in GeneBank under the following accession numbers: MZ636718 to MZ636756.

The results presented in Table 3 show that the nucleotide sequence of the ITS_{u1-4} gene region have contains 150 different nucleotide positions between 31 studied pepper individuals. Of these, there are 53 singleton variable sites (with 52 singleton variable sites containing two variants: 17, 21, 31, 36, 37, 41, 50, 63, 75, 130, 141, 146, 147, 148, 164, 165, 167, 173, 195, 224, 301, 318, 343, 367, 392, 395, 410, 420, 431, 432, 439, 451, 462, 463, 468, 469, 484, 487, 506, 507, 509, 538, 588, 597, 598, 601, 616, 622, 663, 670, 672, 677 and 1 singleton variable sites containing three variants: 262) and 97 Parsimony informative sites (with 79 Parsimony informative sites containing two variants: 10, 22, 47, 59, 64, 65, 74, 85, 86, 89, 91, 94, 97, 106, 111, 115, 121, 128, 129, 150, 168, 179, 180, 181, 186, 189, 192, 194, 201, 202, 211, 213, 229, 231, 241, 265, 351, 379, 386, 390, 402, 405, 411, 413, 424, 428, 436, 438, 440, 441, 444, 447, 448, 455, 461, 480, 497, 518, 523, 536, 539, 557, 569, 570, 574, 575, 587, 590, 593, 604, 606, 627, 630, 642, 650, 651, 655, 685, 690 và 18 Parsimony informative sites containing three variants: 108, 187, 188, 220, 352, 396, 414, 422, 430, 434, 452, 491, 510, 525, 558, 584, 632, 671), and ITS_{u1-4} gene region without the coding region assignation protein. The ITS_{u1-4} gene region contains 169 (S) total number of mutations, with the average number of nucleotide differences (k = 20.352), the nucleotide diversity coefficient accounts for 30.930 x 10^{-3} (Pi), created 23 types of haplotype (h) in 31 individuals with haplotype diversity coefficient accounting for 0.947 (Hd). The individuals in the pepper population with large differences based on the ITS_{u1-4} gene region showed up to 18 minimum recombination events predicted to occur in this population when analyzed on DNAsp 6.0 software. All indicators were processed with statistical significance $p < 0.05$ (Table 3).

Table 4
Neutrality tests results based on ITS_{u1-4} gene region of Pepper population

Regions gene	Fu's	Tajima's		Fu and Li's				S
	Fs	*D*	*P*	*D**	*P*	*F**	*P*	
ITS_{u1-4}	-0.209	-1.824	*Statistical significance: *, P < 0.05*	-1.205	*Not significant, P > 0.10*	-1.699	*Not significant, 0.10 > P > 0.05*	0.684

Note: *D, Tajima's statistic* [Error! Reference source not found.]; *Fs, Fu's statistic* [19]; *D* and F*, Fu and Li's statistics* [21], *S, Strobeck's S statistic* [29]

The mismatch distribution of pairwise nucleotide differences among ITS_{u1-4} sequences among all individuals in the population exhibited a smooth unimodal distribution characteristic of a large population expansion [17; 18]. Study-wide site frequency spectra reveal an excess of singleton mutations when compared with expected frequencies under neutrality and stable population size (Fig. 1).

Fu's Fs were negative (Fs = -0.209) which occurs when an excess of rare haplotypes is present and suggests that either population expansion or genetic hitchhiking has taken place [19]. The values of D, Fu and Li's D* and F* were negative as well (D = -1.824; *P < 0.05*, D* = -1.205; *not significant, P > 0.10* and F* = -1.699; *not significant, 0.10 > P > 0.05*), indicating an excess of recently derived haplotypes and suggesting that either population expansion or background selection has occurred [20; 21]. Strobeck's S, the probability of obtaining equal or fewer haplotypes based on gene frequency and mutation rate was high in between individuals in a population (S = 0.684). These results are consistent with deviation from neutrality due to either selection or population expansion (Table 4).

Phylogenetic analysis

Figure 2. Haplotype network of Pepper population based on ITS_{u1-4} gene region of the nucleus.

An ITS_{u1-4} haplotype network was constructed using statistical parsimony with a 95% connection limit (parsimony cutoff = 7 mutational steps) (Fig. 2). The evolutionary history based on ITS_{u1-4} gene region allele frequencies was inferred by using the Maximum Likelihood method and Tamura-Nei model [22], in which a constant rate of evolution across individuals in the population is assumed and genetic distances between individuals in the population are relative. The tree with the highest log likelihood (-2082.26) is shown. Evolutionary analyses were conducted in MEGA X [22]. The results presented in Figs. 2 and 3 show that have the 3 groups are present. Group I include: 1 individual of species *Piper divaricatum* (HUIB_PD36), group II include: 4 individuals of species *Piper retrofractum* (HUIB_PR41 and HUIB_PR48) and *Piper hancei* (HUIB_PH30 and HUIB_PH46), and group III include: 26 for the remaining individuals of species *Piper nigrum* (Figs. 2 and 3). The highest value of Fst is shown in the pairs of groups II and III (Fst = 0.151), and the lowest is in groups II and I (Fst = 0.015), between different groups, the Fst value based on ITS_{u1-4} haplotype is also different, frequencies ranged from 0.015 to 0.151 (Table 5).

Table 5
The F_{ST} values of the between group mean distance

Group	n	Group I	Group II	Group III
Group I	1	**	0.015	0.023
Group II	4	0.097	**	0.023
Group III	26	0.144	0.151	**

Note- *n: Number individual; Data allow the diagonal is Fst ; indicators were processed with statistical significance p < 0.05*

Discussion

The nuclear ribosomal internal transcribed spacer region or a part of it is one of the most frequently used nuclear markers for phylogenetic reconstructions at the species level or even below. Recently, the *ITS* region has been advocated as a barcode for plants, fungi and possibly protists and animals as well [12; 23; 24; 25]. The applicability of DNA barcoding to species identification relies heavily on the inclusiveness of reference sequence libraries. To generate as many sequences as possible, universal primers have to be used for taxa without reference sequences. Previous ITS primers for plants were neither specific enough to plants nor universal enough among plants, and the primer problem could be one of the most important factors that limits the extensive use of *ITS* as a barcode [10; 11]. For most nonmodel plants, there are hardly any nuclear markers available in many cases and *ITS* is the sole choice owing to the availability of universal primers [26]. In plant DNA barcoding, the assembly of reference libraries of barcode sequences of known species has been one of the most important goals [11; 27]. Furthermore, nuclear genes are indispensable to evolutionary inferences and DNA barcoding because uniparentally inherited chloroplast or mitochondrial markers can only reveal the evolution of one parent.

In this study, we used the *ITS* primer pair referenced from Chen et al. (2016) were experimentally tested for specificity and evaluate the genetic diversity of 39 pepper individuals collected from different localities in Vietnam [9]. We did not observe high variability in intraspecific distance within the ITS_{u1-4} gene region between individuals, ranged from 0.000 to 0.155 (mean = 0.033). The size of the gene region has fluctuated from 667 to 685 bp between different individuals with the percentage (G + C) contained in the ITS_{u1-4} gene region was ranged from 54.776–60.805%, mean = 60.174%. We determined nucleotide sequences of the nuclear rDNA internal transcribed spacer (*ITS*) region in 39 individuals of 4 Peper species (*Piper nigrum, Piper retrofractum*, *Piper hancei*, and *Piper divaricatum*) from different locations in Vietnam. The results of evolutionary relationships of taxa show that have the three groups are present. The highest value of Fst is shown in the pairs of groups II and III (Fst = 0.151), and the lowest is in groups II and I (Fst = 0.015), between different groups, the Fst value based on ITS_{u1-4} haplotype is also different, frequencies ranged from 0.015 to 0.151. Neutrality tests results show that the values of Fu's Fs, D, Fu, and Li's D* and F* were negative as well, with an excess of recently derived haplotypes and have suggested that either population expansion or background selection has occurred. According to research by Chen et

al. (2016) shows that a combination of ITS-u1 and ITS-u4 amplifies the entire *ITS* region, In accordance with the high coverage of these universal primers indicated by the in silico tests, these combinations all resulted in high PCR success rates of over 95%, improvements from 5–30% compared with common-used ones [9].

Conclusion

Pepper, is a valuable spice crop. Barcodes base on studies circumvents the traditional methods for identification of cultivars, which will take several years. Traditional morpho-phenology methods to identify Peper species are mostly based on phenotypic characters, but morphological characteristics are subjected to be affected by developmental and environmental, DNA Barcode has the potential to provide solutions to the presence of duplicates in the collection of Pepper varieties with economic potential. Thus the data obtained can also be utilized for further comparison and improvement of pepper cultivars, thereby ensuring a promising future by facilitating rational selection of parents from genetically divergent groups of cultivars. All the observations support the view that this database is an important resource for researchers working on Species of pepper in Vietnam and also provides a tool to create ITS_{u1-4} databases for any given taxonomy.

Abbreviations

ITS: Internal transcribed spacer, PCR: Polymerase Chain Reaction, TAE: Tris acetate EDTA, MCL: Maximum composite likelihood, NCBI: GeneBank database.

Declarations

Acknowledgements:

The study was supported by the Ministry of Science and Technology of Vietnam (Grant No. ĐTĐL.CN-08/20). The authors also acknowledge the partial support of Hue University under the Core Research Program (Grant No. NCM.DHH.2019.01)

Authors' contributions: Rasphone, S, Hoang, N.H.T and Long, D.T conducted the experiments and participated in interpretation of results and manuscript preparation. Truong, H.T.H participated in the design of studies and in the discussion for preparing the manuscript and did the final revision.; Nguyen, H.V; Nguyen, H.V.P; Nguyen, K-C.T; Nguyen, N.Q; Duong, O.T; collected and provided the plant materials used for the study. All authors read and approved the final manuscript.

Availability of data and materials

The sequence of plant materials and its annotations are deposited at GenBank under accession number MZ636718- MZ636756 (table 6).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Authors have declared that no competing interests exist

Author details

1. Institute of Biotechnology, Hue University, Hue city, Thua Thien Hue 49000, Vietnam
2. Black pepper Research and Developmnet Center, Pleiku, Gia Lai, Vietnam

* Corresponding author: Institute of Biotechnology, Hue University, Hue, 49000, Vietnam.

Email: Rasphone@hueuni.edu.vn (Rasphone, S.), dtlong@hueuni.edu.vn (Long, D.T.), nvhoan1974@gmail.com (Nguyen, H.V.), ngocvtn@yahoo.com.vn (Nguyen, N.Q.), oanhttgl@yahoo.com.vn (Duong, O.T.), nvphung@hueuni.edu.vn (Nguyen, H.V.P.), ntkcuc.huib@hueuni.edu.vn (Nguyen, K-C. T.), hohoangnhi0112@gmail.com (Ho, N.T.H.), tthhai@hueuni.edu.vn (Truong, H.T.H)

References

1. Scott IM, Jensen HR, Philogène BJR, & Arnason JT: **A review of** *Piper* **spp. (Piperaceae) phytochemistry, insecticidal activity and mode of action**. *Phytochem Rev.* 2008, 7:65.
2. Committee FoCE: **The Flora of China**. *Science Press* 1982, **21**:14–70.
3. Gutierrez RMP, Gonzalez AMN, Hoyo-Vadillo C: **Alkaloids from** *Piper*: **a review of its phytochemistry and pharmacology**. *Mini – Reviews in Medicinal Chemistry* 2013, **13**:163–193.
4. Lin RIS: **Pharmacological properties and medicinal use of peper (***Piper nigrum* **L.)**. *Developments in Food Science* 1994, **34**:469–481.
5. Jaramillo MA, Manos PS: **Phylogeny and patterns of floral diversity in the genus Piper (Piperaceae)**. *American Journal of Botany* 2001, **88**(**4**):706–716.
6. Schindel DE, Miller SE: **DNA barcoding a useful tool for taxonomists**. *Nature* 2005, **435**:17
7. Ao T, Deb CR, Rao SR: **Molecular strategies for identification and characterization of some wild edible mushrooms of Nagaland**. *India. Mol Biol Rep.* 2020, **47**(**1**):621–630.
8. Kress and Ericksion: **DNA barcodes: methods and protocols**. *Methods Mol Biol*. 2012, **858**:3–8.
9. Cheng T, Xu C, Lei l, Li CG, Zhang Y and Zhou S: **Barcoding the kingdom Plantae: new PCR primers for** *ITS* **regions of plants with improved universality and specificity**. *Molecular Ecology Resources*

2016, **16**:138–149.

10. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH: **Use of DNA barcodes to identify flowering plants**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:8369–8374.
11. Hollingsworth PM: **Refining the DNA barcode for land plants**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:19451–19452.
12. Li DZ, Gao LM, Li HT et al: **Comparative analysis of a large dataset indicates that internal transcribed spacer (*ITS*) should be incorporated into the core barcode for seed plants**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:19641–19646.
13. Song J, Shi L, Li D et al: **Extensive pyrosequencing reveals frequent intra-genomic variations of internal transcribed spacer regions of nuclear ribosomal DNA**. *PLoS ONE* 2012, **7**: e43971.
14. Stoeckle M: **Taxonomy, DNA and the bar code of life**. *BioScience* 2003, **53**:2–3.
15. Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ: **The *ITS* region of nuclear ribosomal DNA-A valuable source of evidence on Angiosperm phylogeny**. *Annals of the Missouri Botanical Garden* 1995, **82**:247–82
16. Long DT, Hong HTK, Tram LLT, Trang NTQ, Hiep NV and Tien NPT: **Genetic diversity analysis of lotus species (*Nelumbo nucifera* Gaertn.) in Thua Thien Hue based on *ITS4-5* genetic region**. *Plant Cell Biotechnology and Molecular Biology* 2019, **20**(**23**&**24**):1160–1171.
17. Slatkin M, Hudson RR: **Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations**. *Genetics* 1991, **129**:555–562. 25
18. Rogers AR, Harpending H: **Population growth makes waves in the distribution of pairwise genetic differences**. *Mol Biol Evol*. 1992, **9**:552–569. 26
19. Fu YX: **Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection**. *Genetics* 1997, **147**:915–925. 19
20. Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism**. *Genetics* 1989, **123**:585–595. 20
21. Fu YX, Li WH: Statistical tests of neutrality of mutations. *Genetics* 1993, **133**:693–709. 21
22. Kumar S, Stecher G, Li M, Knyaz C, and Tamura K: **MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms**. *Molecular Biology and Evolution* 2018, **35**:1547–1549. 22
23. Moniz MBJ, Kaczmarska I: **Barcoding of Diatoms: nuclear encoded *ITS* revisited**. *Protist* 2010, **161**:7–34. 27
24. Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Xu H, Zhu Y, Xiao P, Chen S: **Use of *ITS2* Region as the Universal DNA Barcode for Plants and Animals**. *PLoS ONE* 2010, **5**(**10**): e13102. 28
25. Schoch CL, Seifert KA, Huhndorf S et al: **Nuclear ribosomal internal transcribed spacer (*ITS*) region as a universal DNA barcode marker for Fungi**. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**:6241–6246. 29

26. Young I, Coleman AW: **The advantages of the *ITS2* region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a Drosophila example**. *Molecular Phylogenetics and Evolution* 2004, **30**:236–242. 30
27. Chase MW, Fay MF: **Barcoding of plants and fungi**. *Science* 2009, **325**:682–683. 31
28. Raz V, & Ecker J: **DNA isolation from *Arabidopsis thaliana***. In B. Birren, E. Green, S. Klapholz, R. Myers, & J. Roskams (Eds.), *A laboratory manual* 1997, **1**:24–25. New York: Cold Spring Harbor.
29. Strobeck C: **Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision**. *Genetics* 1987, **117**:149–153
30. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A, **DNAsp 6: DNA Sequence Polymorphism Analysis of Large Data Sets**. *Molecular Biology and Evolution* 2017, **34**(**12**):3299–3302.
31. Tamura K. and Nei M: **Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees**. *Molecular Biology and Evolution* 1993, **10**:512–526.

Figures

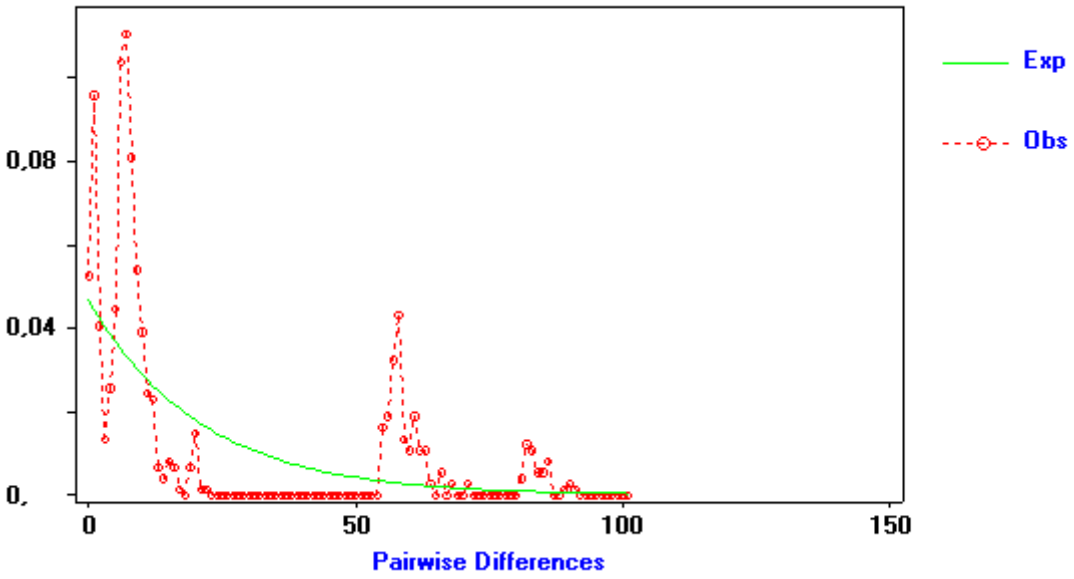

Figure 1

Population expansion signatures in nucleus ITSu1-4 sequence data. Site frequency spectrum indicating an excess of singleton mutations in the ITSu1-4 sequence. Spectrum compares observed frequencies of segregating sites to expected distribution under the null hypothesis of no population change.

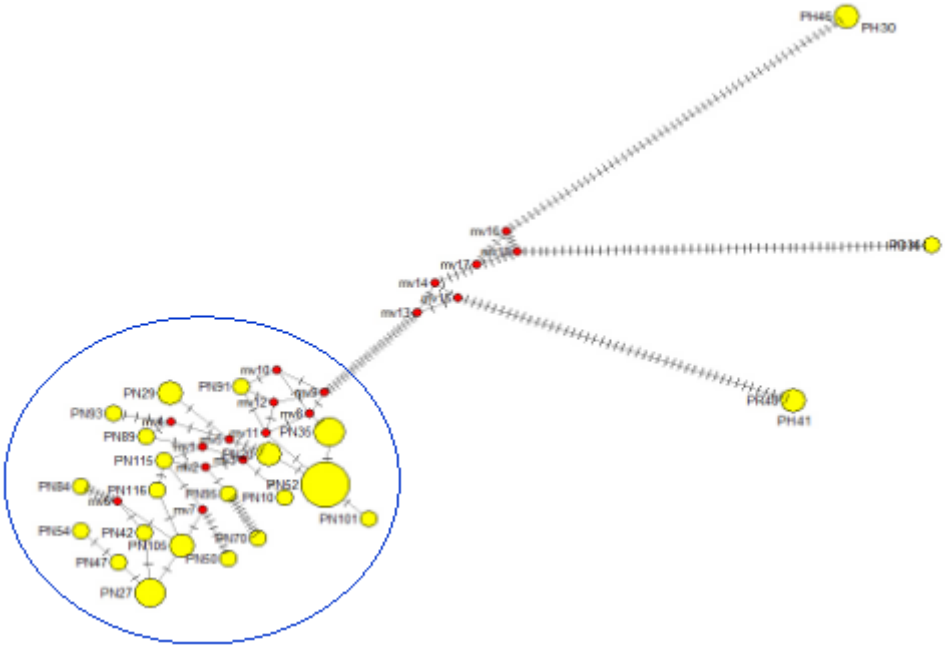

Figure 2

Haplotype network of Pepper population based on ITSu1-4 gene region of the nucleus.

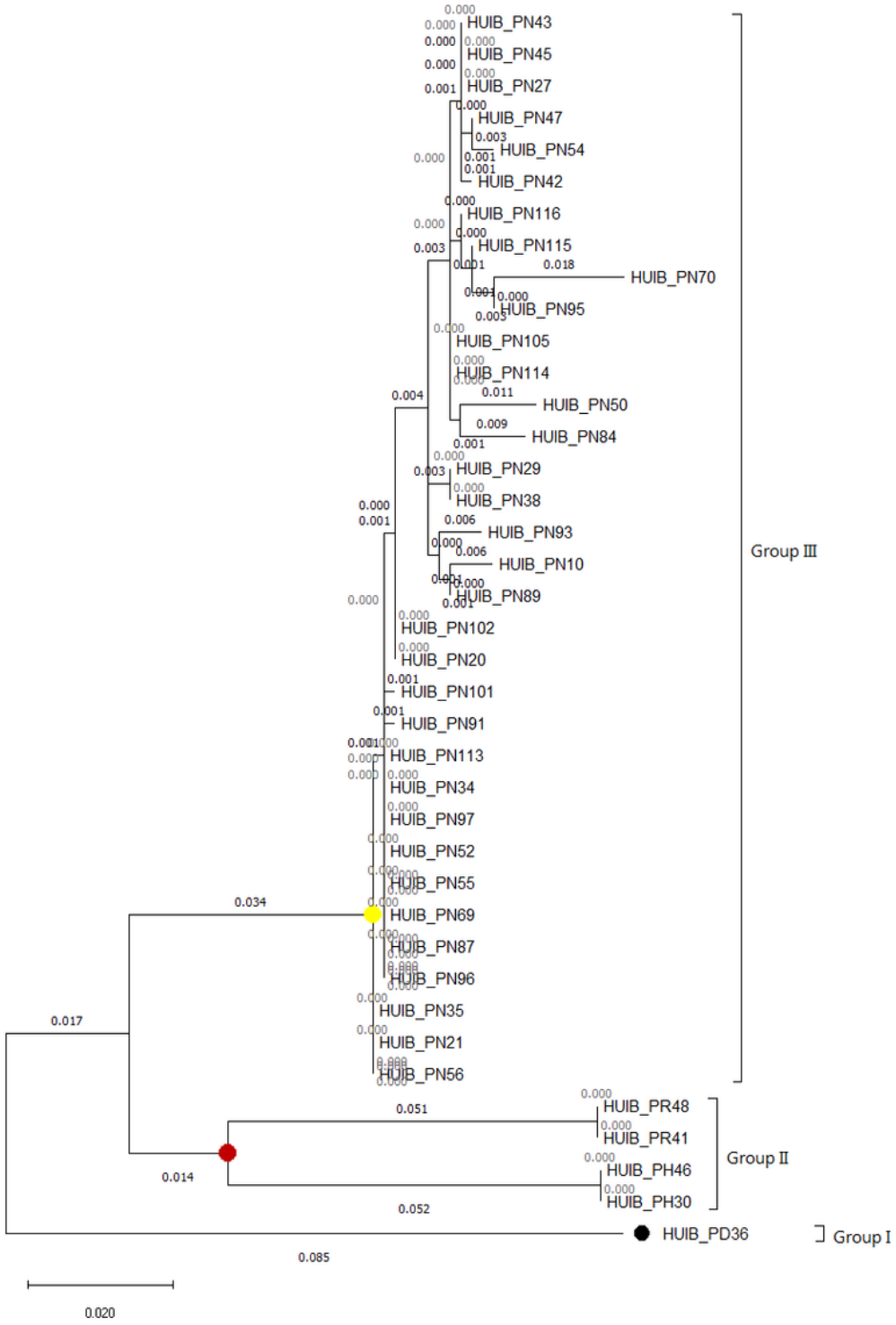

Figure 3

Evolutionary relationships of taxa based on ITSu1-4 gene region in the genetic nucleus of Pepper population