https://doi.org/ 10.33472/AFJBS.6.Si2.2024.1500-1511

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Assessment of Genetic Diversity and Genetic Relationship of Vietnamese Pepper Samples (*Capsicum annuum* L.) Collected in Central regions of Vietnam using Molecular Markers

Tran Dang Khanh[1,2], Nguyen Duy Anh[3], Pham Van Dan[4], Nguyen Thi Ngoc Minh[5], Nguyen Thanh Nhung[5], Nguyen Ngoc Truyen[6] Duong Van Thao[7*]

[1] Vietnam National University of Agriculture, Hanoi, Vietnam
[2] Edulight Vietnam Ltd., South West 9th Pl, Yukon, Oklahoma, USA
[3]Vietnam Australia International School
[4]Center for Technology Development and Agricultural Extension (CETDAE), Vinh Quynh, Thanh Tri, Hanoi, Vietnam
[5]Agricultural Genetics Insitute, Hanoi, Vietnam
[6] Hue University of Agriculture and Forestry
[*7]Thai Nguyen University of Agriculture and Forestry, Thai Nguyen City, Vietnam
Corresponding authors: duongvanthao@tuaf.edu.vn

Article History

Volume 6,Issue Si2, 2024

Received:27 Mar 2024

Accepted : 28 Apr 2024

doi: 10.33472/AFJBS.6.Si2.2024.1500-1511

Abstract

Pepper (*Capsicum annuum* L.) has been widely used for multiple purposes, including medicine, culinary, and flavoring ingredients. This plant is extensively grown in Vietnam, encompassing diverse climatic conditions and ecosystems. However, there is a lack of understanding regarding the distinctions between most of the local species. Therefore, this study aimed to assess the genetic diversity and relationships of 25 Vietnamese native pepper species collected from two provinces, Quang Nam and Gia Lai, using molecular markers. Specifically, genome sequencing was conducted on the sequences of the internal transcribed spacer (ITS) regions ITS1-5.8S-ITS2 and *rbc*L regions using two pairs of specific primers: ITS1/ITS8 and *rbc*L-F/*rbc*L-R. The results revealed that all the native pepper species are diversified within the genus Capsicum, exhibiting species adaptation and locality-dependent characteristics. The similarity coefficient ranged from 81.27% to 98.99% for the ITS regions and from 97.91% to 100% for the rbcL region. These findings provide a foundation for identifying valuable genetic resources that can contribute to the development, breeding, and conservation of different native types of pepper in this country.

Keywords: *Capsicum annuum*, pepper, ITS region, *rbc*L, genetic diversity

Introduction

Pepper (*Capsicum annuum* L.) belongs to the family of Solanaceae and is an increasingly important vegetable crop worldwide with high economic value. This plant is used as an important spice with high economic value in many countries. Vietnam is currently ranked fifth in the world for its primary export goods, dried pepper and pepper powder, and seventh for the pepper growing areas. Most dried peppers are mainly exported to Japan, Hong Kong, and Singapore markets. Meanwhile, pepper powders are exported to Russia, Czechoslovakia, Hungary, and Bulgaria, and bringing a large source of revenue. According to the latest data from the Vietnam Pepper Association (VPA) reported that the chili pepper export turnover in 2023 reached over 20 million USD and, an increase of 107.4% compared to 2020 and growing areas are significantly expanded (VPA, 2023).

The genus *Capsicum* consists of about 22 wild species, of which only five are domesticated *C. annuum, C. baccatum, C. chinense, C. frutescens,* and *C. pubescens.* They belong to three distinct clades: Annuum, Baccatum, and Pubescens (Pandey et al., 2012; Liu et al., 2023). All hot peppers belong to the species *C. frutescens*, which is widespread in both tropical and subtropical areas and includes both wild-type and cultivated varieties. Two genera of *C. annuum* and *C. frutescense* are cultivated worldwide, with *C. annuum* being the most common. According to a description by Greenleaf, (1986), the plants of *C. annuum* are one and up to five flowers per node, with green or greenish-white corollas. They have oval anthers and brownish-yellow seeds. Calyx teeth and a calyx constriction appear in *C. Annuum,* and its fruits have a pungent taste.

The chloroplast genomes of most common land plants are characterized by the plasmids. The composition, size, structure, and sequence of the chloroplast genome are highly conserved in numerous evolutionary studies. This high conservation indicates that any changes in the structure, arrangement, or composition are strongly related to phylogenetic relationships. Different parts of the genome evolve at different rates and relatively slowly at the nucleotide sequence levels. Non-coding regions of the chloroplast genome evolve faster than coding regions. Mutations in chloroplast DNA are often single nucleotide polymorphisms (SNP). Insertions and deletions in non-coding regions accumulate at the same rate as nucleotide substitutions, which is the type of mutation that enhances the diversity of non-coding regions. Most chloroplast genes are single-copy genes, while nuclear genes are of multigene families. The high conservation of chloroplast DNA is an advantage when using it to reconstruct phylogenetic relationships and evolutionary processes at the levels from species to genus to family of plants. Among plastid genes, *rbc*L is the most characteristic gene sequence, encoding the large subunit of ribulose - 1,5 - bisphosphate carboxylase/ oxygenase (RUBISCO). The *rbc*L gene has been widely used in numerous phylogenetic studies and plant taxonomy with more than 10000 *rbc*L sequences available in GenBank. Due to the ease of PCR amplification in several plant groups, the Consortium for the Barcode of Life (CBOL) recently recognized *rbc*L as one of the most promising gene sequences for DNA barcoding studies in plants. However, due to the low discrimination ability between species, most people believe that *rbc*L should be used in combination with other barcode markers, such as *matK*, which are the two standard barcode loci for plants (Cuenoud, 2002; Huong et al., 2022). *Mat*K gene region, coding for maturase K, is one of the fastest evolving genes among the chloroplast genes. The *mat*K is about 1550 bp in size and codes for the maturase enzyme involved in the process of removing group II intron in RNA transcription. Since *mat*K evolves rapidly and presents in almost all plants, it has been used as an indicator in studying relationships between species and phylogeny in plants. CBOL tested *mat*K on nearly 550 plant species and found that 90% of angiosperm samples easily amplified the sequence using a single primer pair and recommended the use of *mat*K as one of the standard barcode loci for plants (Gao, 201; Huong et al., 2021).

On the other hand, a recent study showed that the *rbc*L gene alone has higher distinguishing effects than *mat*K or the combination of both genes (Bafeel et al., 2012).
The Internal Transcribe Spacer (ITS) region includes ITS1, separating the 18S rDNA gene from the 5.8S rDNA gene, and ITS2, separating the 5.8S rDNA gene from the 26S rDNA gene. ITS region widely exists below 700bp in angiosperms. The ITS region contains highly conserved sequences, so a variety of primers can be designed for amplification and sequencing. ITS1 and ITS2 sequences are inherently GC-rich, making sequencing difficult. ITS sequences from multiple angiosperm families indicate that the ITS1 and ITS2 sequences are more diverse than the sequences of the rDNA genes. The variance of ITS sequences is an effective way to resolve questions of genetic relationships that arise in recently related taxa. Accordingly, many reports have been able to successfully reconstruct evolutionary history using ITS sequences (Trang et al., 2021; Ha et al., 2020a; Bellarosa et al., 2005). Furthermore, the ITS region is inherited by both parents, which can also be used to probe for cross-hybridization (Whittall et al., 2000).
The objective of this study was to investigate genetic variability and relationships among several Vietnamese native pepper species by sequencing the internal transcribed spacer (ITS) region and the *rbc*L region. We obtained the sequences of ITS and *rbc*L regions of *Capsicum* in Vietnam as initial steps in building a molecular database of Vietnamese pepper. We have observed in aligned sequences certain highly variable areas with many insertion-deletions (indels). Hence, we also assessed the effects of these highly variable areas in the ITS region in analyzing the relationships among the studied species.

Materials & Methods

Plant materials

A total of 25 pepper samples of *C. annuum* were collected from 2 provinces (Quang Nam and Gia Lai) in Vietnam. The detail information is shown in Table 1 and Figure 1.

Table 1: *List of the collected C. annuum samples used in this study*

No.	Origin	Trait and local name	Code	No.	Origin	Trait and local name	Code
1	QN	Small Siem Pepper	XN1	14	QN	Large Siem Pepper	XL4
2	QN	Small Siem Pepper	XN2	15	QN	Large Siem Pepper	XL5
3	QN	Small Siem Pepper	XN3	16	GL	“Bay” Pepper(GL)	B1
4	QN	Small Siem Pepper	XN4	17	GL	“Bay” Pepper(GL)	B2
5	QN	Small Siem Pepper	XN5	18	GL	“Bay” Pepper(GL)	B3
6	QN	Medium Siem Pepper	XTB1	19	GL	“Bay” Pepper(GL)	B4
7	QN	Medium Siem Pepper	XTB2	20	GL	“Bay” Pepper(GL)	B5
8	QN	Medium Siem Pepper	XTB3	21	QN	“A.Rieu”Pepper(QN)	AR1
9	QN	Medium Siem Pepper	XTB4	22	QN	“A.Rieu”Pepper(QN)	AR2
10	QN	Medium Siem Pepper	XTB5	23	QN	“A.Rieu”Pepper(QN)	AR3
11	QN	Large Siem Pepper	XL1	24	QN	“A.Rieu”Pepper(QN)	AR4
12	QN	Large Siem Pepper	XL2	25	QN	“A.Rieu”Pepper(QN)	AR5
13	QN	Large Siem Pepper	XL3				

Chemicals used

Some common molecular biology chemicals from brands Sigma, and Merck. CTAB, Tris base, Boric acid, NaCl, dNTPs, EDTA, 6X Orange Loading Dye solution, Taq Polymerase, Ethanol, 2-propanol, Acetic acid glacial, Phenol, Chloroform, isoamyl alcohol, Agarose, ITS and *rbc*L primers and chloroplasts. Chemicals for PCR reaction: Four types of deoxynucleotide

triphosphates (dNTPs) from Sigma, and Taq-DNA polymerase from Fermentas. Agarose, ethidium bromide, TAE 1X, TBE 1X, TBE 10X, and TEMED.

Total DNA extraction

Total genomic DNA was extracted using the method of Doyle and Doyle (1987) method with some slight improvement. Genomic DNA was isolated from 0.3g fresh leaf samples and was further used as a template for polymerase chain reaction (PCR) amplifications following the previously described by Hoang et al. (2023) with minor modifications. In brief, a 15µl PCR reaction mixture contained 1.5 µl 25Mn PCR reaction buffer (including MgCl2), 0.3 µl dNTPs mix (10 mM/µl), 0.2 µl Taq ADN polymerase (5 U/µl), and 1.5 µl of each forward and reverse primers. The primers used for the amplification of the ITS region were two distinctive pairs of primers, ITS1/ITS8 and *rbc*L-F/ *rbc*L-R amplified ITS and *rbc*L regions. The mixture was denatured at 94°C for 5 min and subjected to 35 cycles at 94°C for 1 min, 57°C for 45s, and 72°C for 50s, and the final extension step of 72°C for 7 min. The PCR products were added 4 µl loading dye and then subjected to 1,5% agarose gel electrophoresis and detected by Ethidium bromide staining under UV.

Figure 1. *Pepper samples were collected from two provinces of Vietnam. The red stars from upper North to South are Quang Nam and Gia Lai provinces.*

Table 2. *List of the primers used in this study*

Primer	Nucleotide sequence (5'>3')
ITS1	GGTTCAAGTCCCTCTATCCC
ITS8	ATTTGAACTGGTGACACGAG
*rbc*L-F	ATGTCACCACAAACAGAGACTAAAGC
*rbc*L-R	CTTCTGCTACAAATAAGAATCGATCTC

Gel elution and Sequencing and data analysis

The specific DNA products were then purified using the DNA Purification System (Qiagen Kit) according to the perspective manufacturer's instructions. The purified PCR of ITS/*rbc*L products was directly sequenced in Macrogen company (Korean). The sequences were aligned and compared to similar sequences in the NCBI database. Then, these sequences were gathered and analyzed using the MEGA v6 programs to create a phylogenetic tree.

Results and Discussion

Sequencing the ITS-rDNA region of 25 *C. annuum* samples

The results of total genomic DNA extraction demonstrated that all DNA bands were sufficiently clear, providing sufficient DNA for further experiments. The purified PCR samples were sequenced in Macrogen company (Korea). The analyzed length of the sequenced ITS1-5,8S-ITS2 region varies briefly from 585 to 597 nucleotides. The nucleotide composition is indicated in Table 3. The results of these ITS1-5,8S-ITS2 sequences were parallelly compared using ClustalW of Mega 6.0 programs.

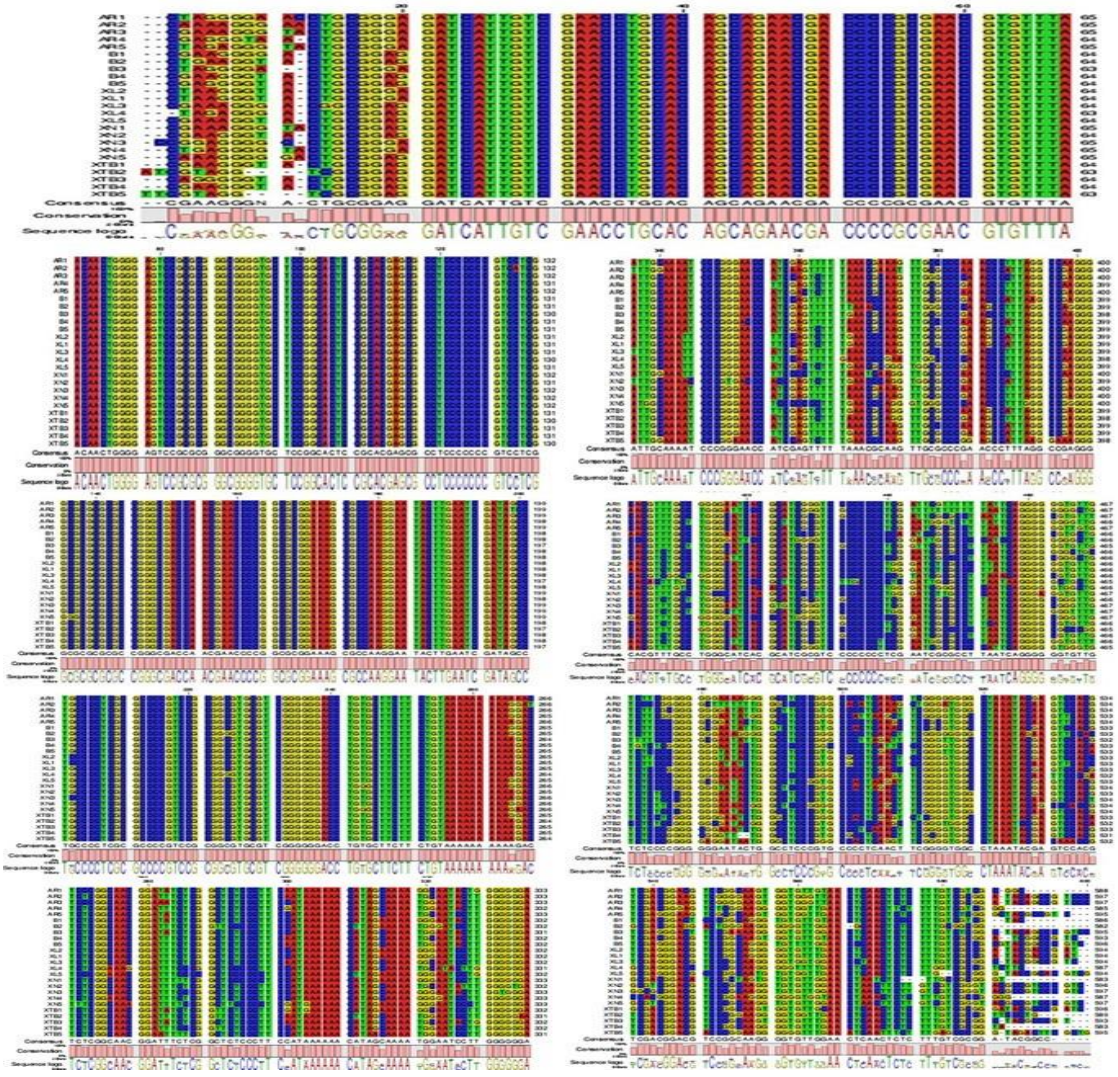

Figure 2. *Alignment on 25 ITS1-5,8S-ITS2 sequences of 25 Vietnamese native pepper species*

As a result of Figure 2, the 25 pepper samples have many differences in the sequence of the ITS1-5,8S-ITS2 region and sequences between samples in terms of nucleotide errors. The ITS1-5,8S-ITS2 sequence region has the most significant variants in the first 20 nucleotides and the last 300 nucleotides. These empty regions do not carry coding sequences on both sides of the 5.8S gene, which are the non-coding regions ITS1 and ITS2 region. Besides, there are some InDel mutations between sequences. This may be due to fluctuations in the direction of deletion and insertion of one or several nucleotides in the ITS1-5,8S-ITS2 sequence region of the 25 pepper samples surveyed.

In other words, in the pepper samples in this study, sequence variation occurred strongly in the ITS1 and ITS2 regions but less in the 5.8S gene region. Variation in these sequences of the 25 researched pepper samples clearly shows the genetic relationship. The less variation in the coding region of 5.8S indicates that the gene is conserved among the samples. And at the same time, the differences in the sequences of the samples also show diversity, for classifying, selecting, and preserving pepper genetic resources.

Table 3. *Composition of four types of nucleotides of the samples*

No	Code	T(U)	C	A	G	%GC	%AT	LS (bp)
1	AR1	18.2	30.8	21.3	29.8	60.5	39.5	596
2	AR2	20.1	29.1	21.8	29.0	58.1	41.9	589
3	AR3	18.4	31.3	20.9	29.3	60.6	39.4	593
4	AR4	19.8	29.2	20.2	30.8	60.0	40.0	583
5	AR5	18.7	32.6	20.5	28.2	60.8	39.2	595
6	B1	18.1	32.8	21.5	27.6	60.4	39.6	583
7	B2	17.5	30.6	19.6	32.3	62.9	37.1	596
8	B3	19.7	30.9	20.3	29.1	60.0	40.0	597
9	B4	17.7	32.9	20.7	28.7	61.6	38.4	587
10	B5	18.0	33.1	19.8	29.2	62.2	37.8	597
11	XL1	18.9	32.2	20.0	29.0	61.1	38.9	594
12	XL2	18.7	32.3	19.9	29.1	61.4	38.6	594
13	XL3	17.0	33.0	20.2	29.8	62.8	37.2	595
14	XL4	18.6	32.5	20.3	28.6	61.2	38.8	587
15	XL5	16.8	32.5	20.0	30.6	63.1	36.9	594
16	XN1	18.7	31.9	20.6	28.8	60.7	39.3	588
17	XN2	18.5	32.9	19.5	29.2	62.1	37.9	597
18	XN3	18.4	32.2	20.1	29.3	61.5	38.5	597
19	XN4	17.9	31.0	19.8	31.3	62.4	37.6	585
20	XN5	16.9	32.7	18.9	31.5	64.2	35.8	595
21	XTB1	17.8	31.9	20.0	30.4	62.2	37.8	586
22	XTB2	17.7	33.6	19.5	29.2	62.8	37.2	582
23	XTB3	17.9	32.5	20.7	28.8	61.4	38.6	595
24	XTB4	18.9	31.7	19.7	29.7	61.4	38.6	593
25	XTB5	17.8	30.4	21.7	30.1	60.5	39.5	596
Avg		**18.3**	**31.9**	**20.3**	**29.6**	**61.4**	**38.6**	**591.76**

Avg: Average; LS: Length of Sequencing; A: Adenine; C: Cytosine; G: Guanin; T: Thymine

The results obtained from Table 3 show that the length of the ITS1-5,8S-ITS2 region varies between survey samples, ranging from 582 to 597 nucleotides. Generally, the samples have a distinctive composition of Guanine (G), Cytosine (C), Adenine (A) and Thymine (T), which

confirms the differentiative of 25 samples based on ITS1-5,8S-ITS2 region. A higher ratio of Cytosine (C) and Guanin (G) is manifested compared to the ratio of Adenine (A) and Thymine (T). In other words, they have % CG content greater than the % AT component. The XN5 sample had the highest (C+G) component (64.2%) and the lowest (A+T) component (35.8%). The average % (C+G) in all 25 samples was 61.4% and the percentage (A+T) averaged 38.6%. The difference in the sequence of the ITS1-5,8S-ITS2 region between survey samples is expressed through the similarity coefficient of each pair of samples, calculated using the genetic distance measurement tool of CLC v8 .02 software interpreted in Figure 3.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
AR1	1		0.12	0.10	0.09	0.10	0.10	0.14	0.08	0.09	0.10	0.11	0.12	0.09	0.16	0.15	0.15	0.13	0.11	0.12	0.17	0.11	0.12	0.08	0.11	0.17
AR2	2	87.77		0.10	0.13	0.13	0.14	0.18	0.13	0.12	0.14	0.14	0.14	0.14	0.19	0.18	0.17	0.17	0.16	0.15	0.21	0.14	0.16	0.13	0.13	0.20
AR3	3	89.45	90.79		0.09	0.07	0.09	0.13	0.09	0.10	0.11	0.10	0.11	0.11	0.16	0.13	0.14	0.12	0.12	0.11	0.16	0.10	0.11	0.10	0.09	0.15
AR4	4	90.83	85.95	89.13		0.09	0.11	0.14	0.09	0.11	0.10	0.11	0.11	0.11	0.16	0.12	0.15	0.12	0.12	0.12	0.17	0.12	0.11	0.10	0.10	0.13
AR5	5	89.75	87.63	92.31	89.75		0.09	0.14	0.08	0.10	0.09	0.09	0.09	0.09	0.13	0.11	0.14	0.12	0.11	0.11	0.16	0.10	0.10	0.08	0.09	0.14
B1	6	90.14	85.93	89.78	88.61	89.92		0.15	0.09	0.10	0.08	0.09	0.09	0.10	0.16	0.11	0.12	0.11	0.11	0.13	0.15	0.09	0.12	0.10	0.09	0.13
B2	7	86.56	82.24	85.76	86.67	85.38	86.01		0.14	0.14	0.13	0.13	0.12	0.13	0.15	0.14	0.16	0.16	0.15	0.14	0.16	0.14	0.14	0.13	0.12	0.15
B3	8	90.28	87.44	91.46	89.78	91.64	90.10	85.40		0.10	0.08	0.09	0.10	0.09	0.14	0.11	0.12	0.11	0.11	0.11	0.16	0.10	0.11	0.08	0.09	0.15
B4	9	90.42	87.94	89.78	88.22	90.10	89.06	86.00	90.44		0.08	0.09	0.10	0.08	0.14	0.13	0.14	0.10	0.11	0.11	0.17	0.08	0.09	0.07	0.07	0.15
B5	10	89.45	86.79	89.63	88.76	90.95	91.28	85.91	92.29	92.28		0.08	0.09	0.07	0.14	0.11	0.13	0.09	0.10	0.11	0.15	0.09	0.08	0.07	0.07	0.15
XL2	11	88.40	86.43	89.95	88.07	90.94	90.57	86.20	91.28	91.25	92.45		0.01	0.09	0.13	0.11	0.11	0.11	0.11	0.11	0.15	0.08	0.10	0.09	0.06	0.14
XL1	12	88.07	86.43	89.45	88.07	90.77	89.90	86.70	90.60	90.57	91.44	98.99		0.10	0.13	0.12	0.12	0.12	0.12	0.11	0.15	0.08	0.11	0.09	0.06	0.15
XL3	13	90.25	86.77	89.45	88.24	90.94	89.73	86.03	91.28	92.59	92.95	91.25	90.91		0.12	0.12	0.12	0.10	0.12	0.11	0.15	0.09	0.10	0.07	0.08	0.15
XL4	14	85.40	81.44	84.11	84.86	86.72	85.54	85.71	85.59	85.71	86.24	86.39	86.39	87.39		0.14	0.17	0.16	0.16	0.16	0.18	0.15	0.16	0.13	0.14	0.18
XL5	15	84.92	83.58	87.60	86.26	88.46	87.92	84.56	88.93	86.91	89.45	89.26	88.59	88.26	85.43		0.11	0.12	0.14	0.13	0.15	0.12	0.12	0.12	0.11	0.15
XN1	16	85.06	82.44	85.28	85.84	85.71	87.78	84.47	85.95	85.55	85.76	87.75	86.74	86.74	83.87	87.12		0.14	0.14	0.14	0.14	0.12	0.15	0.14	0.13	0.16
XN2	17	86.60	84.11	88.13	86.91	88.27	87.92	83.72	89.45	90.44	91.28	89.60	88.93	90.27	84.06	87.94	84.92		0.06	0.13	0.09	0.11	0.12	0.10	0.09	0.16
XN3	18	88.29	84.97	88.48	87.10	88.96	87.77	84.25	88.96	89.11	90.62	88.94	88.61	88.78	83.92	86.79	85.12	93.63		0.11	0.10	0.13	0.13	0.11	0.11	0.17
XN4	19	88.79	84.28	87.96	88.59	88.24	87.27	86.88	87.79	88.09	88.27	88.26	88.26	87.92	85.23	85.62	86.37	86.77	88.29		0.17	0.13	0.13	0.12	0.11	0.16
XN5	20	83.42	81.27	85.28	82.91	84.92	84.76	83.42	84.78	84.25	86.43	86.10	85.76	85.76	82.58	85.62	85.09	91.12	90.64	83.58		0.15	0.16	0.15	0.15	0.20
XTB1	21	88.11	86.62	89.80	87.08	90.45	89.60	85.23	89.95	91.61	91.61	91.95	91.78	90.77	84.73	88.44	86.43	90.10	88.11	86.77	86.43		0.11	0.07	0.08	0.15
XTB2	22	87.52	83.50	87.33	87.84	89.11	87.67	85.64	87.79	89.78	90.97	88.61	87.77	89.11	84.46	86.48	83.98	87.29	86.96	86.17	83.64	88.13		0.10	0.09	0.14
XTB3	23	91.08	87.60	90.45	88.89	91.76	89.88	86.17	91.95	93.27	92.79	91.58	91.58	93.43	87.21	87.92	85.38	90.44	89.28	87.06	85.59	92.62	89.43		0.06	0.15
XTB4	24	88.79	85.79	88.80	89.91	89.58	90.31	88.21	89.45	91.41	91.44	92.27	91.93	90.25	86.56	87.44	87.20	89.09	87.94	88.76	84.76	90.77	89.86	92.26		0.11
XTB5	25	82.64	81.80	85.14	85.64	85.50	85.62	83.61	85.76	85.62	85.64	85.95	85.28	85.45	81.64	85.12	82.50	84.64	83.81	83.33	81.20	85.14	85.91	85.62	86.98	

***Figure 3**. Genetic distance between 25 samples C. annuum based on the sequence of ITS1-5,8S-ITS2 region. The order of samples from 1 to 25 in horizontal rows is similar to the order of samples in vertical rows*.

Sequencing analysis based on ITS1-5,8S-ITS2 region shows that the populations of Siem Pepper, "A.Rieu" Pepper (Quang Nam) and "Bay" Pepper (Gia Lai) are genetically diverse, and there was a high homologous coefficient of some samples in the 25 sequences. The highest homologous coefficient was 98.99% and the lowest was 81.27%. The closest genetic distance was 0.01 and the furthest was 0.21 (Figure 3). The closeness of some pepper samples represents the same origin, evolution, and arising relationships. Thereby, we found similarities in terms of communication and the original discovery source of the researched pepper samples. Transmission fluctuations in the ITS1-5,8S-ITS2 region can generate a diversity of sample studies. From these genetic resources, high-yield and high-quality genes could be able to be identified to serve breeding, developing, and preserving Vietnam's precious pepper genetic resources.

The Small Siem pepper in Quang Nam has the highest level of diversity transmission, with the homologous coefficient ranging from 83.58 to 93.63%. The Large Siem pepper has a homologous coefficient ranging from 90.91 to 98.99%. The genetic homologous coefficient of "Arieu" pepper (Quang Nam) ranges from 85.95 to 92.31%. "Bay" pepper (Gia Lai) has a homologous coefficient ranging from 86.01 to 92.29%, and similar to Medium Siem pepper (Quang Nam), the transmission diversity level is the lowest, ranging from 88.13 to 92.62%. Our recent study showed that the genetic dissimilarity of some Xiem and A Rieu peppers

collected in Quang Nam province ranged from 0.56 to 1.0 by RAPD markers (Truyen et al., 2020)

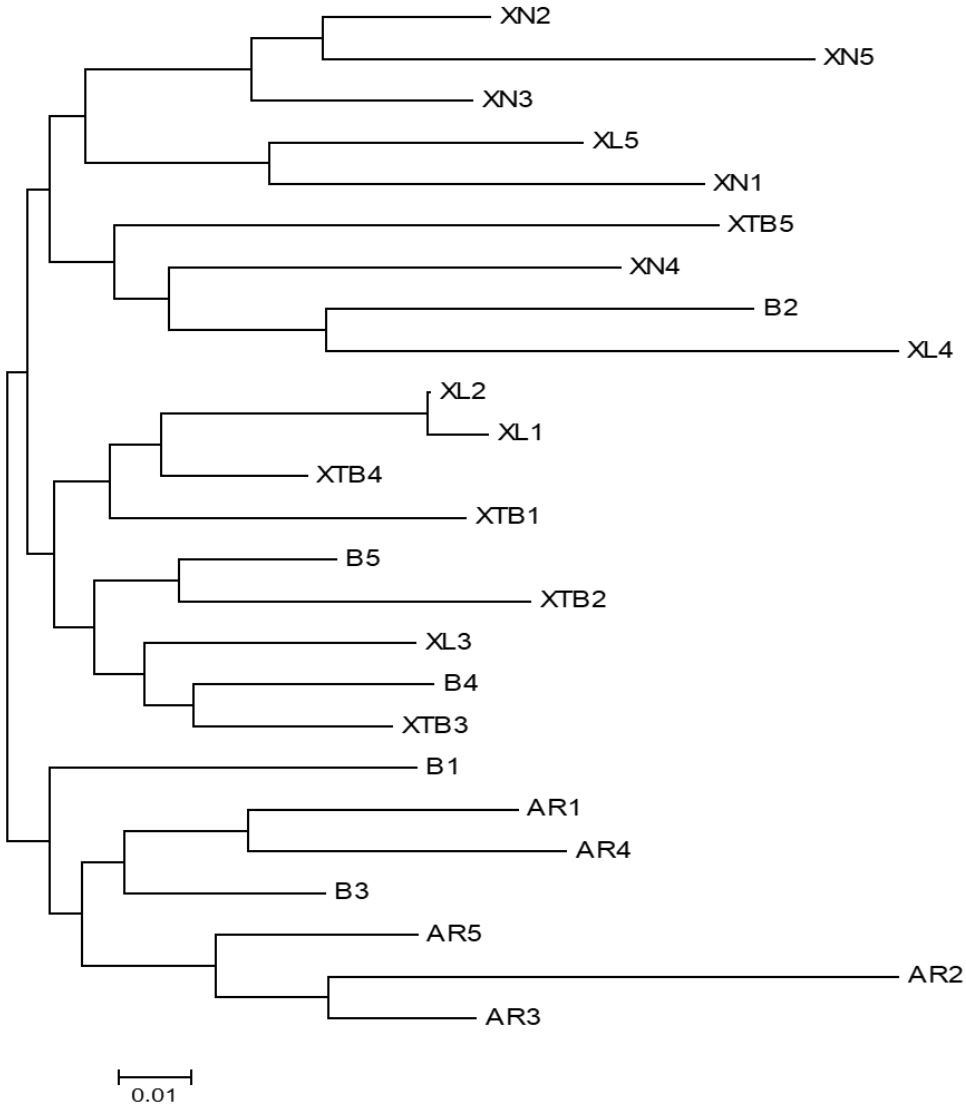

***Figure 4**. The phylogenetic tree generated among the 25 samples*

After determining the nucleotide sequence of the ITS1-5,8S-ITS2 region, the construction of the relation tree was generated by Mega 6.0 software under the Maximum likelihood method (Figure 4). Based on the classification tree which was expressed by the sequence of ITS1-5,8S-ITS2 region, 25 *C. annuum L.* samples were divided into 2 main groups. Group 1 consisted of 7 taxa: AR1, AR2, AR3, AR4, AR5, B1 and B3. Meanwhile, group 2 included 18 taxa and was clustered into 2 subgroups as follows: Subgroup 2.1 was 9 taxa (XN2, XN5, XN3, XL5, XN1, XN4, XTB5, B2 and XL4). Subgroup 2.2 comprised of 9 taxa (XL2, XL1, XTB4, XTB1, B5, XTB2, XL3, XTB3 and B4.

The results of the phylogenetic tree present a diversity of studied samples in terms of molecular, and geographical distribution and among the same variety samples. The distance value is not the same between the taxa of a group, which indicates the different times of genetic divergence between different taxa. Quang Nam Siem peppers are diverse in fruit shapes and the phylogenetic tree also shows the relationship between different Siem pepper varieties and the locality of collection. Pepper samples from B1 to B5, which are collected from the same area (Gia Lai), fall into different main groups, and this occurs as well with the samples collected

from Quang Nam. Moreover, pepper samples of the same local variety are not distributed in the same, except for XN and AR (Quang Nam). The level of genetic diversity shows differences between varieties that can be used for breeding new varieties.

Constructing the relationship on five representative sequence samples based on the *rbc*L region

Representative samples are important in phylogenetic studies because they help to make inferences about the larger population based on a small representative group. The 5 representative samples are chosen in a way that represents the diversity of the 5 study groups (XN2, XTB2, XL2, B2, AR2), in which one sample is selected from each major clade within the group. Overall, the 5 representative pepper samples studied had very little difference in the sequence of the *rbc*L region, since the *rbc*L region is highly conserved and relatively stable over time in most plant species (Ha et al., 2020b). The sequence variation between samples was often due to nucleotide errors.

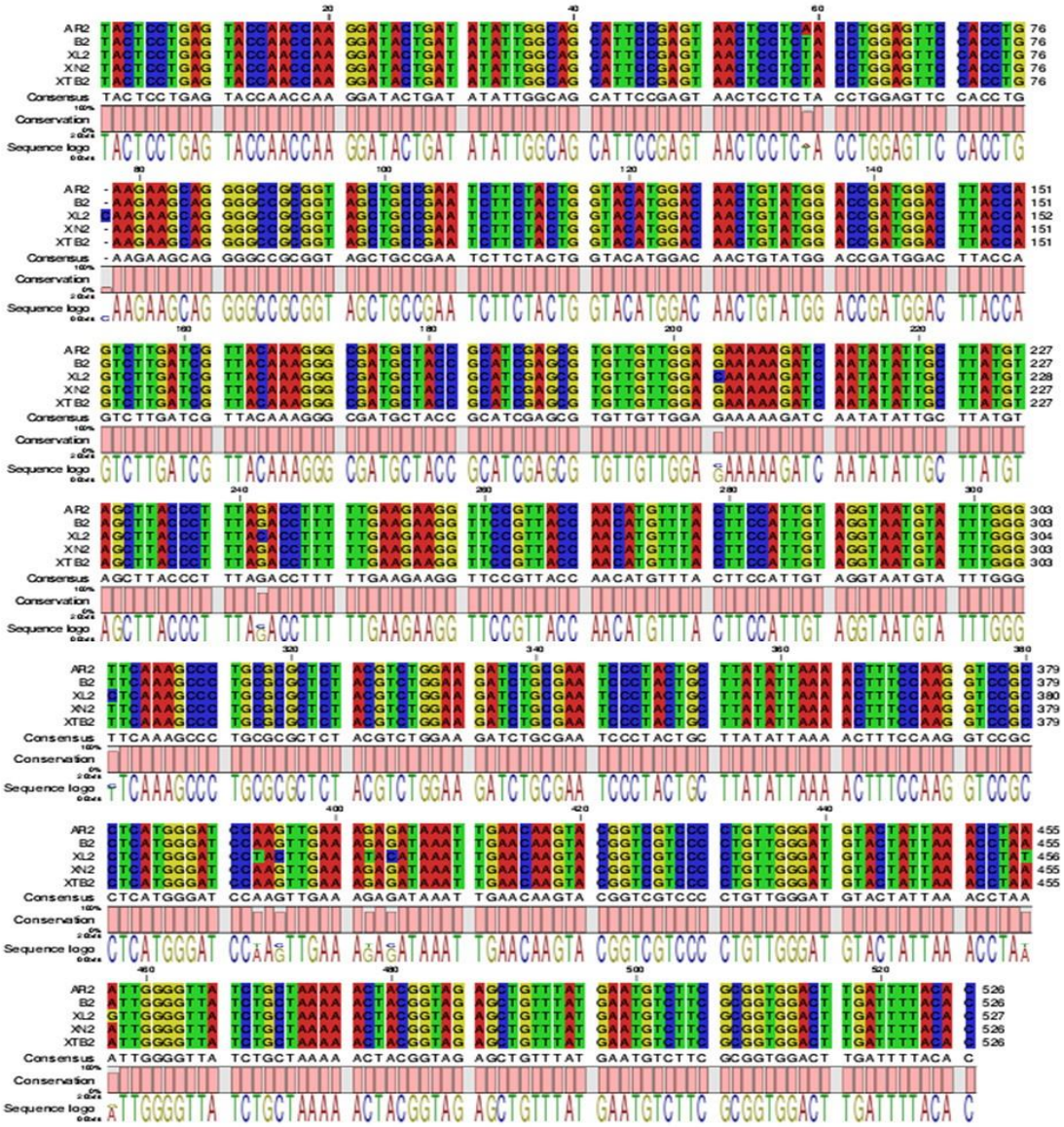

Figure 5. *Alignment on rbcL sequences of 5 representative pepper species*

Based on the results of comparing the nucleotide sequences in Figure 5, it shows that sample XL2 has the most differences between nucleotides, whereas sample AR2 only has one SNP. In particular, sample XL2 has an insertion of Cytosine at position 77 and three substitutions at positions 203, 242 and 304. Similarly, in other positions of sample XL2, there are also changes, 393 T/G, 395 C/G, 402 T/C, 404 C/ G, 455 T/A and 456 G/A.

***Table 4**. Composition of four types of nucleotides of five representative samples*

No.	Code	T(U)	C	A	G	%GC	%AT	LS (bp)
1	AR2	28.9	21.7	26.6	22.8	48.3	55.5	526
2	B2	29.1	21.7	26.4	22.8	48.1	55.5	526
3	XL2	29.4	22.8	25.8	22.0	48.6	55.2	527
4	XN2	29.1	21.7	26.4	22.8	48.1	55.5	526
5	XTB2	29.1	21.7	26.4	22.8	48.1	55.5	526
Avg		**29.1**	**21.9**	**26.3**	**22.7**	**48.2**	**55.5**	**526.2**

Avg: Average; LS: Length of sequencing

The results obtained from Table 4 show that the length of the *rbc*L region varies between survey samples, ranging from 526 to 527 nucleotides. The data show the composition differences of Guanin, Cytosine, Adenine, and Thymine, which are also characteristics that show differences between the survey samples based on the *rbc*L region. In general, the research samples have a lower ratio of %GC composition than %AT composition. Sample XL2 and AR2 have different remaining AT and GC compositions compared to the remaining 3 survey samples.
The difference in the sequence of the *rbc*L region between the five survey samples is statistically shown through the similarity coefficient of each pair of samples, calculated using the genetic distance measurement tool of CLC v8.02 software (Figure 6).

		1	2	3	4	5
AR2	1		0.00	0.02	0.00	0.00
B2	2	99.81		0.02	0.00	0.00
XL2	3	97.91	98.10		0.02	0.02
XN2	4	99.81	100.00	98.10		0.00
XTB2	5	99.81	100.00	98.10	100.00	

***Figure 6.** Genetic similarity coefficient between 5 representative samples of rbcL sequence (Notes: Sample order from 01 to 05 in the horizontal row is similar to the vertical row)*

Analysis results based on the *rbc*L sequence (Figure 6) show that the samples of Siem peppers, "Arieu" peppers (Quang Nam) and 'Bay" peppers (Gia Lai) have little genetic diversity, high genetic similarity, and high similarity coefficients. The highest coefficient is 100.00%, while the lowest coefficient is 97.91%; The closest genetic distance is 0.00 and the furthest is 0.02.
After determining the nucleotide sequence of the *rbc*L region, a phylogenetic relationship tree is built by Mega 6.0 software using the Maximum likelihood method. The results are shown in Figure 7. Based on the results of analyzing the generative relationship tree in the figure, the 05 samples were divided into 2 main branches:

* Group I: included 04 research taxa: AR2, XN2 (Quang Nam), B2 (Gia Lai) and XTB2 (Quang Nam). In which AR2, XN2 and B2 are 100% similar (possibly the same species), the genetic distance is 0.00. This proves that these samples are very closely related to each other and may have the same origin.
* Group II consisted of one remaining research taxon, XL2, collected in Quang Nam. Its similarity coefficient was 97.91% compared to sample AR2 and 98.10% compared to sample B2. The furthest genetic distance is 0.02.

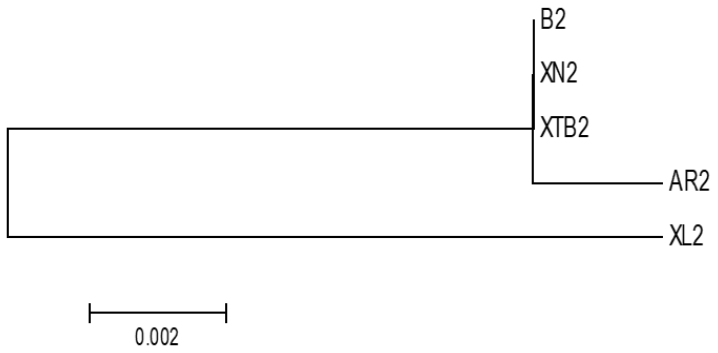

Figure 7. *Phylogenetic tree generated among the 5 representative C. annuum samples*

In general, the phylogenic tree points out that two types of Siem pepper varieties (XN and XTB) are closely related to each other and the Bay pepper (Gia Lai), as they share a common ancestor. ARieu pepper belongs to the same clade as XN, XTB and B varieties, as the result aligns with pepper varieties from the same local are not distributed similarly.

Conclusions

In summary, a total of 25 pepper samples can be identified using the ITS1-5,8S-ITS2 and rbcL sequencing regions. Among the 5 samples that used the rbcL sequence, there were fairly high homologous coefficients ranging from 97.91% to 100%. This provides a foundation for the discovery of valuable genetic resources that can aid in the development, breeding, and conservation of different types of pepper. To ensure a more objective evaluation, it is recommended to conduct further research and collect pepper varieties from various locations. Integrative analysis with gene bank data from NCBI can also be carried out in future studies. Moreover, it is necessary to select other markers, such as genes in the *mat*K region and *trn*H-*psb*A regions to identify closely related species in local pepper varieties.

References

Bafeel, S. O., Arif, I. A., Bakir, M. A., Al Homaidan, A. A., Al Farhan, A. H., & Khan, H. A. (2012). DNA barcoding of arid wild plants using *rbc*L gene sequences. *Genetic and Molecular Research*, 11(3):1934-41.
Bellarosa R., Simeone M. C., Papini A. and Schirone., (2005). Utility of ITS sequence data for phylogenetic reconstruction of Italian Quescus spp. *Plant Physiology and Evolution*. 34: 355-370.
Cuenoud P., S.V., Chatrou L. W., et al. (2002). "Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbc*L, *atp*B, and *matK* ADN sequences, *American Journal of Botany*, 89:132-144.

Doyle J.J., Doyle J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19:11-15.
Gao T., S.Z., Yao H., Song J., Zhu Y., *et al*. (2011). Identification of Fabaceae plants using the DNA barcode *mat*K, *Planta Medica*, 77: 92-94.
Greenleaf W.H., (1986). Pepper breeding. In Bassett M.J. (ed). Breeding Vegetable Crops, AVI Publ. Co., p. 67-134.
Ha, T. T. T., Hai, N. M., Hien, P. H., Khanh, T. D., & Trung, K. H. (2020a). Genetic Diversity and Nutritional Values of *Dendrocalamus yunnanicus* Species in the Northern Mountainous Regions of Vietnam. *Advanced Studies in Biology*, 12(1): 37-46.
Ha, T.T.T., Khanh, T.D., Trung, K.H. (2020b). Evaluation of Genetic Diversity and Identification of Huperzia Species Collected in Some Different Areas in Vietnam by Molecular Markers. *International Letters of Natural Sciences*, 80: 13-23.
Hoang, N., Nhung, T.N., Trang, V.T.H., Khanh, T.D., Hung, H.V. (2023). Molecular Analysis of Genetic Diversity and Genetic Relationship of *Polygonatum kingianum* Samples Collected in Northern Mountainous Regions in Vietnam. *European Chemical Bulletin*, 12(8):9296-9304.
Huong, N.T.T., Hung, N.Q., Lu, D.V., Khanh, T.D. (2021). Morphological characteristics and genetic diversity of some red dragon fruit samples in field trials and *mat*K marker sequencing. *Journal of Scientific and Engineering Research*, 8(1): 189-195.
Huong, B.T.T., Anh, D.X., Cuong, N.H., An, N.T., Gioi, D.H., Tuong, H.M., Ha, C.H., Ha, T.T.T., Trung, K.H., Khanh, T.D. (2022). Morphological characteristics and DNA barcoding in bach hop (*'Lilium poilanei'* Gagnep) in Vietnam. *Australian Journal of Crop Science*, 16(4): 471-478.
Liu, F., Zhao, J., Sun, H., Xiong, C., Sun, X., Wang, X., *et al.* (2023). Genomes of cultivated and wild *Capsicum* species provide insights into pepper domestication and population differentiation. *Nature Communications*, 14(1): 5487.
Pandey, S. K., Yadav, S. K., & Singh, V. K. (2012). An overview on *Capsicum annuum* L. *Journal of Pharmaceutical Science and Technology*, 4 (2): 821 - 828.
Trang, L. T. T., Ha, D. T. T., Diep, N. T., Nghia, L. T., Duc, D. X., Khoa, N. T., & Hue, H. T. (2021). Genetic Diversity and Genetic Relationship of Vietnamese Citrus Varieties Using Internal Transcribed Spacer Region (ITS). *Advanced Studies in Biology*, 13(1), 1-9.
Truyen, N.N., Co, N.Q., Tam, N.T., Uy, T.C., Phuoc, D.N., Duc, N.V., Thuy, D.T. (2020). Analyzing genetic diversity of local Xiem chili of Quang Ngai province by RAPD markers. *Journal of Agricultural Science & Technology*, 5(1): 2237-2242.
VPA- Vietnam Pepper Association (2023). Annual report in 2023.
Whittal, J., Liston, A., Gisler, S., Meinke, R.J. (2000). Detecting Nucleotide Additivity from Direct Sequences is a SNAP: An Example from Sidalcea (Malvaceae). *Plant Biology*, 2(2): 211-217