RESEARCH ARTICLE

The Journal of Agriculture and Environment, Vol: 24, June, 2023 | (167-178)

MOLECULAR IDENTIFICATION OF *Coffea arabica* VARIETIES IN NEPAL: INSIGHTS FROM PHYLOGENETIC ANALYSIS

Shreejan Pokharel[1], Samsher Basnet[2], Anish Basnet[3], Ashmita Mainali[3], Sadikshya Rijal[3], Asmita Shrestha[3], Bignya Chandra Khanal[1] and Gyanu Raj Pandey[3,*]

[1] National Biotechnology Research Center, Nepal Agricultural Research Council, Lalitpur, 44700, Nepal
[2] National Entomology Research Center, Nepal Agricultural Research Council, Lalitpur, 44700, Nepal
[3] Shubham Biotech Nepal Pvt. Ltd., Bharatpur-29, Chitwan, 44200, Nepal

ARTICLE INFO

Keywords:
DNA Isolation,
gene sequencing,
internal transcribed
spacers (ITS),
polymerase chain reaction (PCR),
phylogenetic tree

*Correspondence:
gyanupandey9@gmail.com
Tel: +977-9860470938

ABSTRACT

This study aimed to identify and characterize the coffee varieties cultivated in Nepal using molecular phylogenetic analysis. The molecular identification and genetic relationship of twenty five coffee varieties were collected from the Nepal Coffee Development Center, Gulmi, Nepal. DNA was isolated from leaf tissue, and Internal Transcribed Spacers Region (ITS)-specific PCR was performed, followed by sequencing and phylogenetic tree construction. BLASTN was performed to identify the similarities with the sequences of the National Center for Biotechnology Information (NCBI) Database. Evolutionary divergence between the sequences was computed using Maximum Composite Likelihood Model. Sequences were analyzed using Maximum Likelihood Model and Tamura-Nei model to construct molecular phylogeny. BLASTN and molecular phylogeny confirm all the samples to be Coffea arabica. Evolutionary divergence in pairwise comparison was found to be 0% to 4.3%. Divergence of 4.3% was detected between CDC-S21 and CDC-S73. With this, we identified the coffee samples to be C. arabica and we also computed relatedness among our varieties.

1. INTRODUCTION

Coffea arabica L. and *Coffea caneophora*, commonly known as Arabica and Robusta, are the most popular genus, famous for their consumption as popular drinks globally. Coffee belongs to the genus Coffea and the family Rubiaceae comprising almost 124 species (Razafinarivo *et al*., 2013). Coffee is mainly grown in tropical and subtropical regions (Berthaud & Charrier, 1988). The three species of coffee with primary importance for worldwide commercial production are *Coffea arabica, C. canephora*, and *C. liberica* (Farah & Dos Santos, 2015), (Ferreira *et al*., 2019). *C. arabica* is the only tetraploid (2 n = 4x = 44) and self-fertile species in the genus, whilst the others are diploid (2 n = 2x = 22) and genetically self-incompatible (Clarindo & Carvalho, 2008). Among the popular ones, *C. arabica* is highly cultivated, as it has superior taste, rich aroma, and low caffeine content is responsible for 70% of the world's coffee production. On the other hand, C. canephora accounts for 24% and *C. liberica* for 1% of the world's coffee production (Rohwer, 2002), (Kiran *et al*., 2019).

In Nepal, coffee is commercially grown in around 44 districts, while Illam, Kavre, Nuwakot, Gulmi, Arghakhanchi, Palpa, Syangja, Kaski , Sindhupalchok, Sankhuwasabha, Parbat are the largest producers (MoALD, 2022). The Prime Minister Agriculture Modernization Project (PMAMP) has begun laying the groundwork for coffee plants in five districts designated as coffee-super zones, with the goal of increasing domestic coffee output at a time when domestic production is insignificant compared to surging demand (Prasain K, 2019). Nepali coffee is recognized for its distinctive aroma and flavor, being cultivated at higher altitudes (800-1600m), away from the usual coffee-growing region in other parts of the world. As a result, Coffee produced in Nepal is labeled as "Specialty Coffee" in several foreign markets and also noted for its organic certification and fair trade practices (NTCDB, 2020/21).

Despite the long history of Coffee plantations in Nepal, the taxonomical study has not been conducted to date, and genetic diversity is unknown. This study aims at

molecular identification of the coffee species found in Nepal and extrapolate genetic relationship among them. The implication of this study will be useful in crop improvement, along with unraveling taxonomical ambiguity. The conventional method of identification based on morphological characteristics like growth habitat, leaf type, floral characteristics, and fruit morphology is insufficient to explain the enormous variety found in allogamous wild coffee communities (Charrier & Berthaud, 1985).

On the other hand, a molecular method involving short sequences known as barcodes that are independent of the environment and life stage of the coffee plant is more efficient in coffee species identification (MoALD. 2022) ITS (Internal Transcribed Spacer) region of nuclear ribosomal cistron shows a high level of intra-specific variability and, thus, is often used for intra-specific discrimination in angiosperms (Kress *et al*., 2005). Two spacers, ITS1 and ITS2, are located in between the small subunit 18S ribosomal gene and the large subunit 28S ribosomal gene and are separated by the 5.8S gene (Álvarez & Wendel, 2003). As ITS rDNA is often a target in metagenomic studies, an array of resources is available to facilitate sequence-based identifications. Nuclear ribosomal DNA (rDNA) is composed of the 18s, 5.8s, and 36s regions separated by intergenic spacers. As, the ribosomal DNA is ubiquitous throughout plants, nuclear rDNA has proven to be a powerful tool for phylogenetic analysis, techniques for determining the primary nucleotide sequence and the diversity of rates of evolution of the subunits and spacers within the molecule have been developed (Jansen, 1994). Whereas the 18s and 26s coding regions have been used to address phylogenetic questions at the family level or higher taxonomic levels, the internal transcribed spacers (ITS 1 and ITS 2) appear to be useful for assessing relationships at lower taxonomic levels because the sequences of spacer regions generally evolve more rapidly than the coding regions. Recently, the ITS region has been shown to be useful for resolving phylogenetic relationships among several plant genera, including Antennaria, calycodenia (Hsaio *et al*., 1994) and Krigio (Tamura & Nei, 1993). In the present study, we have sequenced the internal transcribed spacer (ITS 2) region of 25 Coffea taxa. The main purpose was to use the ITS sequences to attempt to resolve phylogenetic relationships among the closely related but highly diversified taxa associated within the genus Coffea found in Nepal.

2. MATERIALS AND METHODS

2.1 Sample collection

Twenty five young coffee leaves samples (Table 1) were harvested from Coffee Nepal Coffee Development Center, Aanpchaur, Gulmi, Nepal (Latitude: 27 56′17.85″ N to 27 5′44.87″ N; Longitude: 8 25′29.2′′1 E to 83 25′30.20″ E) (Figure 1) stored in poly bags with zipper. The leaves surfaces were cleaned with running water and stored at -80 ^{0}C for DNA extraction.

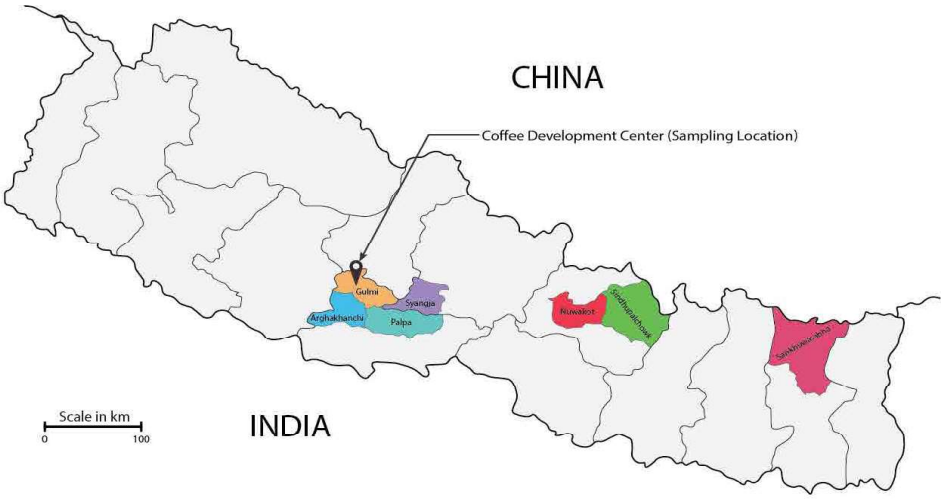

Figure 1: Map of Nepal showing districts with commercial coffee productions and sampling site.

(G = Green, Y = Yellow and R = Red; NA = Not Available)

Table 1. List of Coffee accessions with its features used in this study.

Code	Sample Type	Year of Introduction (A.D.)	Young Leaf Color	Fruit Color
CDC-S1	Brazillian	2002	Light green	Maroon red (G to R)
CDC-S2	Caturra Amarillo	2002	Light green	Yellow (G to Y)
CDC-S3	Brazillian	2002	Light green	NA
CDC-S4	Bourbon Amarelo	2002	Light green	Red (G to Y to R)
CDC-S5	Brazillan	2002	Light bronze	Red (G to Y to R)
CDC-S6	Bourbon Vermelo	2022	Light bronze	Red (G to R)
CDC-S21	Local	1998	Light green	Red (G to R)
CDC-S23	Tekisic/Catisic	2022	Green	Red (G to R)
CDC-S36	Local	1998	Green	Red (G to R)
CDC-S48	Tekisic/Catisic	2022	Light bronze	Maroon red (G to R)
CDC-S51	Local	1989	Green	Maroon red (G to R)
CDC-S54	Local	1989	Green	NA
CDC-S58	Local	1989	Light bronze	Maroon red (G to R)
CDC-S65	Local	1989	Green	Maroon red (G to R)
CDC-S66	Local	1989	Green	Red (G to R)
CDC-S70	Local	1989	Green	NA
CDC-S73	Local	1989	Green	Red (G to R)
CDC-S75	Local	1989	Light green	Red (G to R)
CDC-S78	Local	1989	Bronze	Maroon red (G to R)
CDC-S79	Local	1989	Green	Red (G to R)
CDC-S82	Local	1989	Green	Red (G to R)
CDC-S85	Local	1989	Green	Red (G to R)
CDC-S88	Local	1989	Green	Red (G to R)
CDC-S91	Local	1993	Light green	Red (G to R)
CDC-S92	Phirphire	1995	Light bronze	Red (G to R)

2.2 DNA extraction

DNA samples were extracted using Doyle and Doyle method (Doyle & Doyle, 1987) with some optimizations. Leaf samples (1g) were taken and grounded in liquid nitrogen using a mortar and pestle. The pulverized paste was immediately transferred to micro-centrifuge tubes and 700µl extraction buffer (2% CTAB (w/v), Tris HCL pH 8.0 (0.1M), EDTA pH 8.0 (20mM), NaCl (1.4M), 2% PVP (w/v), 1% -mercaptoethanol) was added. The tubes were incubated at 65 0C in a water bath for 1 hour, and then centrifuged at 15,000 rpm for 15 minutes. The aqueous phase was carefully transferred to fresh tubes, and an equal volume of Chloroform: Isoamyl alcohol (24:1) was added and mixed properly by inversion for a couple of minutes. The mixture was centrifuged at 15,000 rpm for 15 min at 25 0C. After the phase separation, the supernatant was transferred into a new tube. To the supernatant, 0.2ml sodium acetate was added in order to enhance the quality of DNA. An equal volume of isopropanol was added to each tube. The tubes were kept at -20 0C for 30 minutes and centrifuged at 15000 rpm for 7 minutes for precipitation. Then, the supernatant was removed, and the pellets were washed with 96% and 70% ethanol twice, respectively. The pellets were air-dried and resuspended in 1X TE buffer (Tris-HCl 10 mM, EDTA 1 mM, pH 8.0).

2.3 Gel electrophoresis

The quality of extracted DNA was also assessed by using 0.8% agarose gel electrophoresis (Cleaver Scientific, UK) in 1X TAE (50X TAE; 242gm Tris-base, 57.1 ml acetic acid (or 100% glacial acid) and 100ml of 0.5 M EDTA (pH-8.0) at 70V for 1 hr. The PCR amplification products were analyzed using 1.5% agarose gel at 70V for 1 hr. using the same buffer system. The gel was stained with ethidium bromide and photographed using a Gel Documentation system (VWR®Genosmart 2, UK). 100 bp ladder (Thermo scientific), was used as a molecular marker for the size comparison of the visible fragments.

2.4 PCR amplification and sequencing

Primer Pairs ITS (ITS1-5.8s-ITS2); ITSL TCGTAACAAGGTTTCCGTAGGTG, ITSR TATGCTTAAAYTCAGCGGG; designed by (Hsaio *et al*., 1994) were used to amplify the ITS region of the extracted DNAs. The optimized PCR mixture contained 25 of the following; 2.5 10x Taq Buffer, 1.5 25 mM Mgcl2, 0.5 0.2mM dNTPs, 0.5 0.2U of Taq Polymerase, 2 DNA, 1mg/ml BSA and 7% DMSO. The PCR amplification was performed using MiniAmp Thermal cycler (AppliedBiosystems by Thermo Fisher Scientific) as follows, initial denaturation at 94 0C for 3 minutes, followed by 35 cycles of annealing at 53.8 0C for 1.15 minutes, followed by a final extension at 72 0C for 15 minutes.

The PCR products were sent to Macrogen Inc., Seoul, Korea, and sequencing was performed using sequencing primer ITSL (5'-TCGTAACAAGGTTTCCGTAGGTG – 3') and ITSR (5'- TATGCTTAAAYTCAGCGGG-3').

2.5 Sequence analysis

Raw sequences were assembled and trimmed using Codon Code Aligner. Contig sequences generated were subjected to BLASTN, and the database "Standard databases (nr etc.)" was selected. Highly similar sequences were taken in FASTA format for phylogenetic analysis. All sequences were deposited to the NCBI genebank, and accession numbers were received as listed in Table 3. Pairwise comparison of these ITS-2 sequence was performed using the Maximum Composite Likelihood model (Tamura *et al*., 2004) to estimate the evolutionary divergence between the sequence.

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model (Tamura & Nei, 1993). The bootstrap consensus tree inferred from 2000 replicates (Felsenstein, 1985) is taken to represent the evolutionary history of the taxa analyzed. Initial tree(s) for the heuristic search were obtained automatically by applying the 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. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.2731)). This analysis involved 75 nucleotide sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). Evolutionary analyses were conducted in MEGA X (Kumar *et al*., 2018).

2.6 Accession numbers of the nucleotide sequence used:

2.6.1 Accession numbers of the samples of our study:

MZ678246, MZ734310, MZ734319, MZ734320, MZ734321, MZ734322, MZ734325, MZ734324, MZ734326, MZ734331, MZ734332, MZ734333, MZ734334, MZ734336, MZ734335, MZ734337, MZ734338, MZ734339, MZ734344, MZ734345, MZ734346, MZ734347, MZ734348, MZ734323 and MZ734349.

2.6.2 Accession numbers used from the NCBI database:

AY853406.1, MF417757.1, MF417758.1, MK615726.1, MK611791.1, MK615727.1, MK615729.1, MN719947.1, MK615731.1, MK615728.1, DQ153609.1, MK611792.1, MN719952.1, MK615737.1, MK615732.1, AY780425.1, MK615738.1, MK615730.1, MK615733.1, DQ153593.1, MF417756.1, MF417755.1, DQ153591.1, MK615734.1, DQ153594.1, DQ153638.1, MN719950.1, DQ153632.1, DQ153592.1, DQ153616.1, DQ153614.1, DQ153596.1, DQ153615.1, DQ153626.1, DQ153563.1, DQ153617.1, MN719972.1, MT250045.1, MN719966.1, AF542982.1, AF543006.1, AF543002.1, AF542985.1, AF542983.1, AF542981.1, AF543004.1, AF542988.1, AF542992.1, AF543007.1 and AF543003.1.

3. RESULTS AND DISCUSSION

The DNAs of all the twenty five coffee samples from Coffee Development Center from Gulmi, Nepal, taken in this study were successfully extracted. The presence of DNA for 25 samples from extraction was verified by gel electrophoresis at 0.8% agarose (Figure 2). Subsequently, the ITS-based PCR mixture was optimized, modifying suitably the concentration of the sample DNAs, dNTPs, MgCl2, and ITS. Following that, ITS-based PCR was performed for each coffee sample, and the completion of PCR was verified by conducting 1.5 % agarose gel electrophoresis against 100bp ladder, as shown in Figure 3 with a size of range of 600-700 bp.

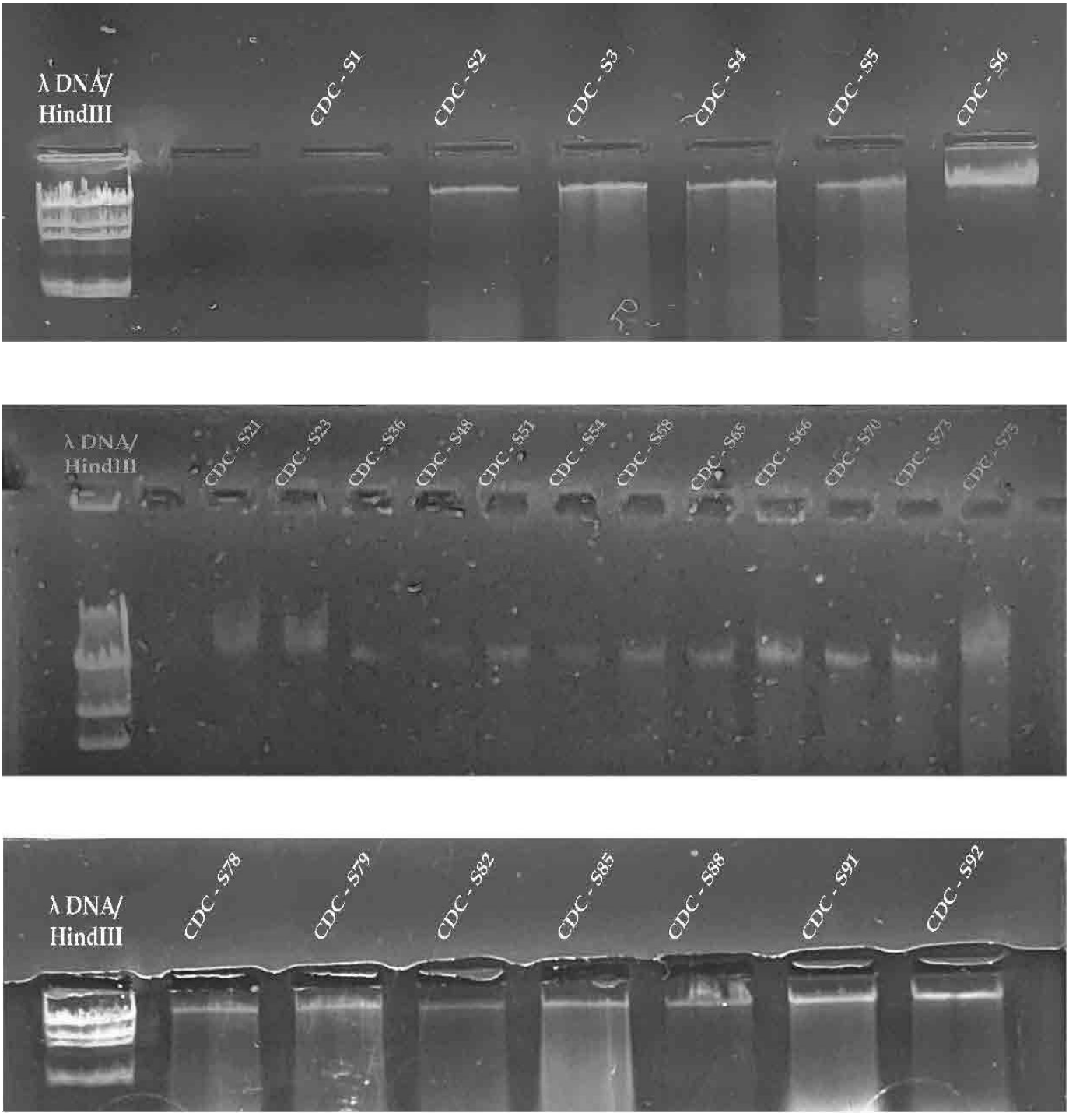

Figure 2. Gel electrophoresis of extracted 25 DNA samples with 0.8% Agarose ran against HindIII digested λDNA.

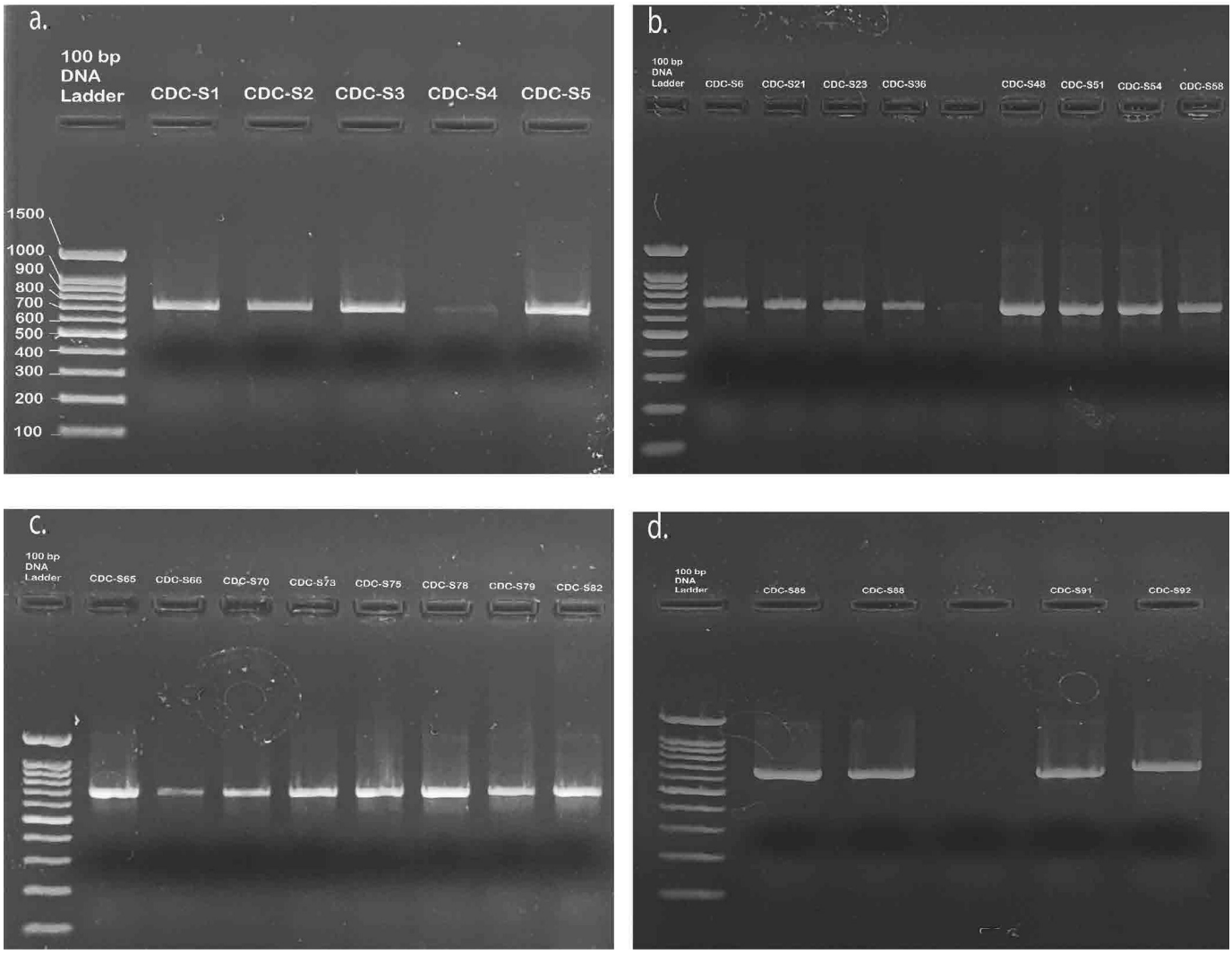

Figure 3. Get electrophoresis of PCR with amplification of ITS region of the 25 coffee samples used in the study.The rightmost lane in each figure a, b, c, and d represents the reference of 100bp ladder. Where the subsequent samples read from right to left. PCR Amplification products were fractioned in 1.5% agarose gel. Amplified from genomic DNA using ITS region (ITS1-5.8s-ITS7) was amplified with primers ITSL (5'- TCGTAACAAGGTTTCCGTAGGTG-3'and ITSR (5'- TATGCTTAAAYTCAGCGGG-3') (Hsaio *et al*., 1994). From right to left Figure a represents the samples serially as CDC-S1, CDC-S2, CDC-S3, CDC-S4, and CDC-S5. Similarly, Figure b represents the samples as CDC-S6, CDC-S21, CDC-S23, CDC-S36, CDC-S48, CDC-S51, CDC-S54, and CDC-S58, while the Figure c reads the sample as CDC-S66, CDC-S70, CDC-S73, CDC-S75, CDC-S78, CDC-S79, and CDC-S82. Lastly, the samples of Figure d precedes as CDC-S85, CDC-S88, CDC-91, and CDC-S92 from right to left. The length of ITS regions of our 25 samples ranges between 600bps to 700bps.

3.1 ITS-based sequencing and BLASTN analysis of 25 coffee samples

The sequencing results of the ITS-PCR amplifications product of 25 coffee samples are presented in Table 2. The ITS regions of the coffee samples were approximately 100% amplified and sequenced. The nucleic acid sequence of samples was analyzed using the BLASTN program to determine the identity of the coffee samples. The BLASTN analysis showed that the coffee samples had 100% similarity to *C. arabica*. The cultivar of *C. arabica* mainly comprised of Typica, Bourbon, and near to Java and Typica. The sequences are deposited in the GenBank with the accession numbers, as shown in Table 2. The length of ITS regions of our 25 samples ranged from 650 bp to 704 bp.

Table 2. Sequencing data after BLAST in NCBI showing Organism, Similarity percentage, Cultivar.

Code	Organism	Query Coverage (%)	Similarity Percentage	Genbank Accession Number	Sequence Lengths (bp)	Cultivar
CDC-S1	*Coffea arabica*	99	99.08	MZ678246	650	Typica
CDC-S2	*Coffea arabica*	97	99.41	MZ734310	699	Typica
CDC-S3	*Coffea arabica*	99	99.42	MZ734319	694	NA
CDC-S4	*Coffea arabica*	100	99.26	MZ734320	678	Bourbon
CDC-S5	*Coffea arabica*	98	99.13	MZ734321	703	Bourbon
CDC-S6	*Coffea arabica*	96	100	MZ734322	690	Bourbon
CDC-S21	*Coffea arabica*	96	99.41	MZ734325	701	Bourbon
CDC-S23	*Coffea arabica*	97	99.58	MZ734324	695	Bourbon
CDC-S36	*Coffea arabica*	96	100	MZ734326	687	Bourbon
CDC-S48	*Coffea arabica*	97	100	MZ734331	669	Bourbon
CDC-S51	*Coffea arabica*	97	99.88	MZ734332	686	Typica
CDC-S54	*Coffea arabica*	96	100	MZ734333	690	Bourbon
CDC-S58	*Coffea arabica*	96	99.88	MZ734334	682	Bourbon
CDC-S65	*Coffea arabica*	96	100	MZ734336	680	Bourbon
CDC-S66	*Coffea arabica*	96	99.85	MZ734335	702	Bourbon
CDC-S70	*Coffea arabica*	99	99.81	MZ734337	686	Near to Java and Typica
CDC-S73	*Coffea arabica*	96	99.7	MZ734338	698	Bourbon
CDC-S75	*Coffea arabica*	94	99.71	MZ734339	691	Bourbon
CDC-S78	*Coffea arabica*	97	100	MZ734344	681	Bourbon
CDC-S79	*Coffea arabica*	98	98.49	MZ734345	704	Bourbon
CDC-S82	*Coffea arabica*	98	99.7	MZ734346	687	Bourbon
CDC-S85	*Coffea arabica*	97	99.85	MZ734347	692	Bourbon
CDC-S88	*Coffea arabica*	96	99.85	MZ734348	701	Bourbon
CDC-S91	*Coffea arabica*	96	99.56	MZ734323	702	Bourbon
CDC-S92	*Coffea arabica*	97	99.41	MZ734349	691	Bourbon

Table 3. Pairwise comparisons of ITS-2 sequences obtained from 25 coffee-tree accession listed in Table 2. The numbers of base substitutions per site from between sequences are shown. Analyses were conducted using the Maximum Composite Likelihood model (Tamura *et al*., 2004)

		1	2	3	4	5	6	7	8	9	10	11	12
1	**CDC-S1**												
2	CDC-S2	0.011											
3	CDC-S3	0.008	0.010										
4	CDC-S4	0.011	0.007	0.007									
5	CDC-S5	0.014	0.013	0.010	0.006								
6	CDC-S6	0.006	0.023	0.016	0.003	0.016							
7	CDC-S91	0.014	0.018	0.015	0.007	0.012	0.019						
8	CDC-S23	0.011	0.015	0.015	0.004	0.010	0.009	0.012					
9	CDC-S21	0.013	0.009	0.012	0.006	0.012	0.021	0.015	0.012				
10	CDC-S36	0.017	0.016	0.012	0.003	0.012	0.011	0.013	0.012	0.012			
11	CDC-S48	0.000	0.008	0.006	0.000	0.006	0.003	0.011	0.005	0.006	0.005		
12	CDC-S51	0.013	0.013	0.009	0.004	0.009	0.001	0.010	0.006	0.012	0.010	0.000	
13	CDC-S54	0.013	0.018	0.012	0.004	0.015	0.004	0.015	0.012	0.013	0.013	0.002	0.006
14	CDC-S58	0.013	0.012	0.009	0.004	0.006	0.008	0.007	0.007	0.010	0.010	0.003	0.009
15	CDC-S66	0.009	0.015	0.016	0.004	0.015	0.018	0.018	0.013	0.009	0.013	0.005	0.009
16	CDC-S65	0.000	0.007	0.009	0.000	0.006	0.013	0.009	0.004	0.000	0.005	0.006	0.002
17	CDC-S70	0.005	0.013	0.012	0.000	0.012	0.009	0.013	0.007	0.013	0.005	0.003	0.002
18	CDC-S73	0.013	0.041	0.038	0.023	0.039	0.038	0.036	0.032	0.043	0.035	0.033	0.029
19	CDC-S75	0.006	0.013	0.012	0.000	0.013	0.023	0.007	0.010	0.010	0.007	0.011	0.006
20	CDC-S78	0.011	0.013	0.010	0.006	0.009	0.003	0.012	0.009	0.010	0.010	0.002	0.004
21	CDC-S79	0.011	0.018	0.016	0.004	0.017	0.021	0.012	0.012	0.016	0.013	0.008	0.006
22	CDC-S82	0.008	0.016	0.010	0.006	0.015	0.004	0.015	0.009	0.016	0.015	0.002	0.004
23	CDC-S85	0.013	0.013	0.009	0.006	0.010	0.006	0.010	0.007	0.012	0.012	0.002	0.003
24	CDC-S88	0.009	0.019	0.016	0.004	0.016	0.021	0.007	0.012	0.015	0.013	0.011	0.007
25	CDC-S92	0.014	0.018	0.013	0.003	0.010	0.012	0.015	0.010	0.016	0.012	0.005	0.010
		13	14	15	16	17	18	19	20	21	22	23	24
14	CDC-S58	0.012											
15	CDC-S66	0.012	0.007										
16	CDC-S65	0.003	0.003	0.001									
17	CDC-S70	0.004	0.002	0.010	0.010								
18	CDC-S73	0.037	0.027	0.041	0.032	0.031							
19	CDC-S75	0.007	0.003	0.015	0.009	0.015	0.034						
20	CDC-S78	0.006	0.009	0.009	0.003	0.001	0.035	0.006					
21	CDC-S79	0.010	0.009	0.022	0.010	0.015	0.035	0.012	0.010				
22	CDC-S82	0.006	0.010	0.012	0.005	0.003	0.030	0.011	0.007	0.010			
23	CDC-S85	0.006	0.009	0.009	0.003	0.004	0.037	0.006	0.006	0.009	0.006		
24	CDC-S88	0.012	0.009	0.016	0.007	0.013	0.035	0.003	0.010	0.015	0.012	0.010	
25	CDC-S92	0.013	0.004	0.010	0.006	0.004	0.035	0.007	0.010	0.013	0.013	0.010	0.013

3.2 ITS-2 sequence divergence

The pairwise nucleotide-sequence divergence (Maximum Composite Likelihood Model) among the coffee taxa (Table 3) ranged from 0% to 4.3% within different taxa of *C. arabica*. Divergence of 4.3% was detected between CDC-S21 and CDC-S73. CDC-S48 and CDC-S65 were detected with no divergence with CDC-S1; CDC-S48, CDC-S65 and CDC-S70 were also detected with no divergence with CDC-S4. No divergence was observed between CDC-S21 and CDC-S65; and CDC-S48 and CDC-S51. These divergence were attributable to deletion and insertion events, and gaps were introduced to align the sequence (Charrier, 1997). Previously, intraspecific divergence of 4.3% was observed in *Calycadenia truncate* using ITS 2 sequences (Balwdin, 1993). However, this phenomenon appeared particularly important within coffee-tree species (T. A. Charrier, 1997). Such results are expected when the rate of nucleotide divergence exceeds the homogenization rate of the gene copies within a multigene family, a situation that could arise in cases of explosive radiation or interspecific hybridization (Hillis & Davis, 1988).

3.3 Validation of molecular identification of 25 coffee samples by ITS phylogeny

The entire ITS sequence of individual sequenced results was analyzed and a phylogenetic tree (Figure 4) was constructed by the Maximum Likelihood method to validate the molecular identity of the 25 coffee samples as implied by the BLASTN. The similarities were assessed based on the genetic distances among the sample sequences. The molecular identity was corroborated through phylogenetic analysis by comparing the results of alignments of nucleotide sequences of coffee samples with some standard sequences selected such as *C. arabica, C. canephora, C. congensis, C. dewevrei, C. liberica, C. stenophylla, C. dubardii, C. affinis, C. kapakata, C. brevipes, C. arabica* C. canephora, C. mayombensis, C. heterocalyx, Coffea sp.,* and *C. stenophylla* C. liberica*. Moreover, the coffee species of Gulmi, Nepal, were assumed to be *C. arabica* until now. This study helped prove and document this speculation.

The distinct clad separation of *C. arabica*, along with our sample species, and the remaining species in the Neighbor-Joining phylogenetic tree, (Figure 4) , validates that all of the samples are *C. arabica.* This paper documents the *Arabica* coffee for the first time in Nepal with validation. Phylogenetic analysis also demonstrated the relatedness among the coffee varieties, indicating origin of varieties.

The overall sequence homogeneity among members of a gene family such as the nuclear rDNA is assumed to be maintained by homogenization mechanisms associated with concerted evolution (Arnheim, 1983). As a result, rDNA repeats are usually very similar within individuals and species, although differences may accumulate between species (Hilis & Dixon, 1991) as we observed in pairwise comparison. An unusual feature of the ITS 2 sequence in *Coffea* was the importance of intraspecific variation. Even the presence of ITS variants within individuals was observed. Similar intraspecific variation has already been reported. The generation times of coffee-trees have been estimated as between 20 and 30 years. The observed deficiency in the homogenization mechanisms *may* be related to the long life cycles of coffee-trees, in the same manner that nucleotide-substitution rates have been reported to be related to the length of the reproductive cycle (Li & Graur, 1991). Moreover, it is most likely that spontaneous interspecific hybridization occurs between taxa and has been involved in speciation. Given their important role in post-translational processing, ITS regions are considered to be quite conserved (Hamby & Zimmer, 1992). However, in this study, ITS phylogeny has been successful in discriminating the coffee species. This finding adds more credibility to the use of ITS as a phylogeny reconstruction tool for intra-species discrimination in coffee species.

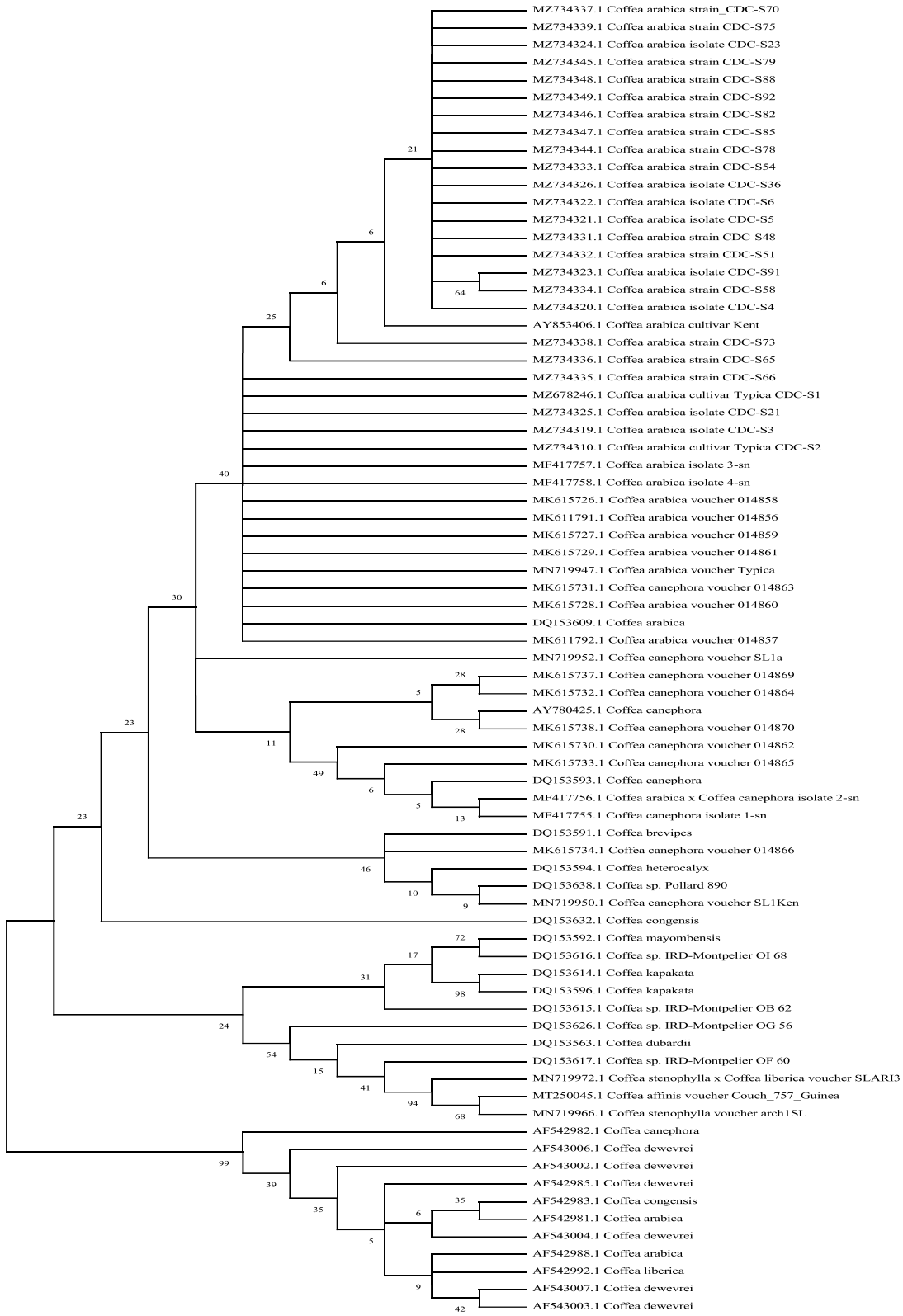

Figure 4. Maximum Likelihood tree generated from ITS-2 sequence data. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to the branches (Felsenstein, 1985). There were a total of 557 positions in the final dataset. The coffee samples of our study were distributed mainly into three clads, while the distinct clad separation of *C. arabica*, along with our sample species, and the remaining species, elucidates that all of the samples are *C. arabica*.

4. CONCLUSION

This study used molecular phylogenetic analysis to identify and describe the different coffee cultivars grown in Nepal. The analysis also showed that there was only a little amount of evolutionary divergence across the samples, ranging from 0% to 4.3%, with the biggest difference found between CDC-S21 and CDC-S73. These findings contribute to our understanding of the genetic links among the many coffee varietals found in Nepal. The use of a gene-based identification strategy in this work provides a better and more efficient substitute for identifying coffee species using traditional morphological traits. The findings of this study have significance for improving crops and clearing up taxonomical uncertainty. On the basis of these results, future research might investigate the genetic variety of coffee kinds in Nepal and possibly improve the region's approaches to producing coffee.

ACKNOWLEDGEMENTS:

The authors thankfully acknowledge "Coffee Development Center" for research funding with management support at field and "Shubham Biotech Nepal Pvt. Ltd." for research facilities and space. Authors are thankful to "National Biotechnology Research Center: Nepal Agricultural Research Council" for their support during the research period. Authors are grateful to Mr. Raj Kumar Shrestha, Mr. Dinesh Tiwari, Mr. Sudip Silwal, Ms. Shristi Khanal and Mr. Dipak Raj Pandey.

REFERENCES

Álvarez, I., & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*. https://doi.org/10.1016/S1055-7903(03)00208-2

Arnheim, N. (1983). Concerted evolution of multigene families. In *Evolution of genes and proteins* (pp. 38–61).

Balwdin, B. G. (1993). Molecular phylogenetics of Calycadenia based on ITS sequences of nuclear ribosomal DNA: chromosomal and morphological evolution re-examined. *American Journal of Botany*, *80*(2), 222–238.

Berthaud, J., & Charrier, A. (1988). Genetic resources of Coffea. *Coffee, Vol. 4. Agronomy*, *November*, 1–42.

Charrier, A., & Berthaud, J. (1985). Botanical Classification of Coffee. In *Coffee*. https://doi.org/10.1007/978-1-4615-6657-1_2

Charrier, T. A. (1997). Phylogenetic relationships of Coffee-tree species (Coffee L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theor Applied Genetics*, *94*, 947–955.

Clarindo, W. R., & Carvalho, C. R. (2008). First Coffea arabica karyogram showing that this species is a true allotetraploid. *Plant Systematics and Evolution*. https://doi.org/10.1007/s00606-008-0050-y

Doyle, J. J., & Doyle, J. L. (1987). Doyle_plantDNAextractCTAB_1987.pdf. In *Phytochemical Bulletin* (Vol. 19, Issue 1, pp. 11–15).

Farah, A., & Dos Santos, T. F. (2015). The Coffee Plant and Beans: An Introduction. *Coffee in Health and Disease Prevention*, 5–10. https://doi.org/10.1016/B978-0-12-409517-5.00001-2

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*. https://doi.org/10.2307/2408678

Ferreira, T., Shuler, J., Guimarães, R., & Farah, A. (2019). Chapter 1. Introduction to Coffee Plant and Genetics. In *Coffee*. https://doi.org/10.1039/9781782622437-00001

Hamby, R. K., & Zimmer, E. A. (1992). Ribosomal RNA as a Phylogenetic Tool in Plant Systematics. In *Molecular Systematics of Plants*. https://doi.org/10.1007/978-1-4615-3276-7_4

Hilis, D. M., & Dixon, M. T. (1991). Ribosomal DNA: Molecular Revolution and Phylogenetic Inference. *The Quarterly Review of Biology*, *66*. https://doi.org/https://doi.org/10.1086/417338

Hillis, D. M., & Davis, S. K. (1988). Ribosomal DNA: intraspecific polymorphism, concerted evolution and phylogeny reconstruction. *Systematic Zool*, *37*, 63–66. https://doi.org/https://doi.org/10.2307/2413191

Hsaio, C., Chatterton, N. J., Asay, K. H., & Jensen, K. B. (1994). Phylogenetic relationships of 10 grass species: An assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. *Genome*, *37*(1), 112–120. https://doi.org/10.1139/g94-014

Kiran, V. S., Asokan, R., Revannavar, R., Hanchipura Mallesh, M. S., & Ramasamy, E. (2019). Genetic characterization and DNA barcoding of coffee shot-hole borer, Xylosandrus compactus (Eichhoff) (Coleoptera: Curculionidae: Scolytinae). *Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis*. https://doi.org/10.1080/24701394.2019.1659249

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0503123102

Krishana Prasain. (2019). Ground work begins for coffee super zone in five districts. *The Kathmandu Post*.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Li, W.-H., & Graur, D. (1991). *Fundamentals of Molecular Evolution*.

Razafinarivo, N. J., Guyot, R., Davis, A. P., Couturon, E., Hamon, S., Crouzillat, D., Rigoreau, M., Dubreuil-Tranchant, C., Poncet, V., De Kochko, A., Rakotomalala, J. J., & Hamon, P. (2013). Genetic structure and diversity of coffee (Coffea) across Africa and the Indian Ocean islands revealed using microsatellites. *Annals of Botany*, *111*(2), 229–248. https://doi.org/10.1093/aob/mcs283

Rohwer, J. G. (2002). A trópusok növényei. *Magyar Könyvklub, Budapest*, 148.

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tamura, K., Nei, M., & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *PNAS*, *101*(30), 11030–11035.