

# Evaluating the systematic position of *Ehretia asperula* Zoll. & Moritzi based on ITS1, *matK* and *trnL-trnF* DNA sequences

Thuy Linh Nguyen<sup>1</sup>, Thi Hang Pham<sup>1</sup>, Van Truong Do<sup>2</sup>, Thi Thu Hue Huynh<sup>1\*</sup>

<sup>1</sup>Institute of Genome Research, Vietnam Academy of Science and Technology

<sup>2</sup>Vietnam National Museum of Nature, Vietnam Academy of Science and Technology

Received 31 July 2017; accepted 7 December 2017

## **Abstract:**

*Ehretia asperula*, popular as a medicinal herb that has potential on cancer treatment, has a limited research about its phylogeny relationships. With many innovative advancements in molecular biology, it is more easy and reliable to identify the taxonomic position of a species by molecular markers- DNA barcode. In this study, we used three different markers, ITS1, *trnL-trnF*, and *matK*, to evaluate *E. asperula*'s systematic position. Based on ITS1 sequence, *E. asperula* belongs to clade *Ehretia* I and is a close relative of *E. resinosa*. Moreover, ITS1 was suggested to be use as a suitable DNA barcode in order to identify *E. asperula*.

**Keywords:** *Ehretia asperula*, ITS1, *matK*, phylogenetics, *trnL-trnF*.

**Classification number:** 3.5

## **Introduction**

*Ehretia asperula* Zoll. & Moritzi is an accepted name of a species from *Ehretia* genus (Ehretiaceae family). It was first described by Zollinger and Moritzi in mid 1840s [1]. In Vietnam, it is present mainly on the mountainous area of the North [2]. Historically, *E. asperula* was used in ethnopharmacology and folklore. The ethnic minorities have been using *E. asperula* for treatments of various ailments, especially for liver diseases, such as hepatitis, liver cirrhosis and even liver cancer [2]. Besides, *E. asperula* appears to be effective in prevention of acne, jaundice, hypertension, and diabetes [2]. A research recently revealed the pharmacological potential of *E. asperula* in cancer treatment [3].

Although many patients who applied folklore procedure using *E. asperula* in combination with other medicinal herbs or with modern treatments recovered from cancer, and prolonged their life [3], no clinical trial has been published using *E. asperula* in cancer therapy. Likewise, the underlying mechanism responsible for its inhibitory effect on cancer is still unclear. Yet there is no study of *E. asperula* molecular and its phylogenetic relationship has not been reported.

The *Ehretia* genus has nearly 50 species distributed mainly in high altitude areas of Asia, Africa, and Australia [1]. In Vietnam, they represent seven species [3], in which, *E. asperula* shares many characteristics in common with other *Ehretia* species. *E. asperula* is

a climbing shrub, having bristle-covered reddish to greyish brown branches. Leaves are blade lanceolate in which the base is narrowly rounded and 5-7 mm in length, and have 7 nerves on each side. Inflorescences are terminal and lateral on short branches, and 5-10 cm in length, where individual flower's base is 2-4 mm long. Flowers have separated five calyx-lobes and white petals. Fruits are 1 cm-long globoses with 3 separate parts. Seeds have pink skin. *E. asperula* produces flowers from March to May and bears fruits from August to December [2].

The taxonomic relationship of *Ehretia* species was clarified by studies based on molecular data. On the basis of the ITS1 information, *Ehretia* was a sister group of *Bourreria* (Ehretiaceae family) and is composed of three major clades. Among them, *Ehretia* III has a closer relationship with *Ehretia* II compared to *Ehretia* I [4]. Additionally, the ITS1 secondary structure was applied to build phylogenetic trees at higher taxonomic levels, and seems to be good at giving a well-resolved tree. Recently, Gottschling, et al. [5] elucidated the phylogeny relationship of Boraginales by using concatenated ITS nuclear and plastid rps16 *trnL-trnF*, *trnS-trnG* sequences. An additional clade

\*Corresponding author: Email: hthue@igr.ac.vn

*Ehretia* IV, in which *E. microphylla* is a presentative, was inferred from extent analysis.

In this study, ITS1, *trnL-trnF*, and *matK* regions were used to investigate the systematic position of *E. asperula*. As inferred from the data, *E. asperula* belongs to *Ehretia* I and is a close relative of *E. resinosa*. Moreover, we contributed three sequences to DNA barcode database of *Ehretia* genus, and suggested ITS1 as an appropriate DNA barcode.

### Materials and methods

Three leaf specimens of *E. asperula* were collected at three different locations in the Hoa Binh province by a colleague at Vietnam national museum of nature, and were preserved in silica gel. Genomic DNA was isolated using CTAB (Cetyltrimethylammonium bromide) extraction protocol [6] from approximately 100 g of leaf tissue. Three DNA fragments were amplified using Thermo scientific phusion high-fidelity DNA polymerase with universal primer pairs (Table 1). PCR products were purified with Thermofisher scientific PCR clean-up purification kit. These DNA were sequenced by Applied biosystems 3500 genetic analyzer system using BigDye™ terminator v3.1 cycle sequencing kit. Additionally, GenBank accessions were downloaded to complete dataset for molecular investigation comprising of 37 sequences of *Ehretia* species, and three sequences of outgroup representatives (Table 2). Sequences were aligned automatically by BioEdit v7.1.9 [7]. Phylogenetic analysis was performed by PAUP\*4.0a152 [8]. Likelihood trees were built using heuristic search. Bootstrap analyses (criterion=parsimony, with full heuristic search: PBS; criterion=distance, with neighbor-joining search and maximum

**Table 1. List of primer pairs used in the study.**

| Primer names | DNA regions      | Primer sequences (5' → 3')            |
|--------------|------------------|---------------------------------------|
| ITS-AB-101   | ITS              | ACG AAT TCA TGG TCC GGT GAA GTG TTC G |
| ITS-AB-102   |                  | TAG AAT TCC CCG GTT CGC TCG CCG TTA C |
| TrnL-F       | <i>trnL-trnF</i> | ATT TGA ACT GGT GAC ACG AG            |
| TrnL-C       |                  | CGA AAT CGG TAG ACG CTA CG            |
| MatK-F1A     | <i>matK</i>      | ACY GTA TTT TAT GTT TAC GAC G         |
| MatK-R1A     |                  | TCC ATH TDG AAA TCT TGG TTC A         |

**Table 2. Species lists with ID number of DNA sequences on GenBank.**

| Name                 | ITS1       | <i>trnL-trnF</i> | <i>matK</i> | Name                   | ITS1       | <i>trnL-trnF</i> | <i>matK</i> |
|----------------------|------------|------------------|-------------|------------------------|------------|------------------|-------------|
| <i>E. asperula</i>   | KY320205   | KY320206         | KY320207    | <i>E. obtusifolia</i>  | AY331401.1 |                  |             |
| <i>E. acuminata</i>  | AF385799.2 | AY376167.1       | HQ427413.1  | <i>E. macrophylla</i>  | AF385802.2 | KF673271.1       |             |
| <i>E. amoena</i>     |            |                  | JX518091.1  | <i>E. microphylla</i>  | AY463160.1 | KF158204.1       |             |
| <i>E. anacua</i>     | AF385796.2 |                  | DQ197228.1  | <i>E. monopyrena</i>   | AF385792.2 |                  |             |
| <i>E. aquatica</i>   | AF385791.2 |                  | EU599659.1  | <i>E. resinosa</i>     | AY463161.1 |                  |             |
| <i>E. cortesia</i>   | AY463159.1 | KF673292.1       |             | <i>E. rigida</i>       | AF385789.2 |                  | JX518014.1  |
| <i>E. coerulea</i>   | KF673249.1 |                  |             | <i>E. saligna</i>      | AF385786.2 | KF673272.1       | KM894705.1  |
| <i>E. cysmosa</i>    | AF385790.2 |                  |             | <i>E. tinifolia</i>    | AF385793.2 | HQ286270.1       |             |
| <i>E. grahamii</i>   |            |                  | KU564569.1  | <i>E. thyrsoiflora</i> |            | EU600007.1       | EU599655.1  |
| <i>E. laevis</i>     | AF385787.2 | KF673273.1       |             | <i>E. wallichiana</i>  | AY331402.1 |                  |             |
| <i>E. latifolia</i>  | AF385797.2 | KF673282.1       |             | <i>B. succulenta</i>   | AF385776.2 |                  | DQ197229.1  |
| <i>E. longiflora</i> | AY331400.1 | EU600010.1       | KJ687555.1  | <i>B. petiolaris</i>   |            | KF673275.1       |             |

likelihood setting: DBS) were estimated based on 1,000 replicates, in which each was performed with 100 random-addition-sequence replicates, and the starting tree obtained by neighbor-joining.

### Results

All three DNA regions (ITS, *trnL-trnF*, and *matK*) were amplified by using the universal primer pairs (Table 1), and

the products obtained were 848 bp, 950 bp, and 811 bp in length, respectively (Fig. 1). ITS is a nuclear sequence from 3' end of 18S to 5' end of 26S. However, we used a portion of ITS-ITS1 for phylogenetic calculation in this study. While *trnL-trnF* is the intergenic region between two coding regions, *matK* spans from codon 171 to codon 440 of the open reading frame. The nucleotide sequences of each DNA region of the

three samples are identical. Therefore, sequences from only one sample were used for the phylogenetic calculation. DNA sequences of all three DNA regions (ITS, *trnL-trnF*, and *matK*) were deposited on GenBank with ID number KY320205, KY320206, and KY320207, respectively.

The ITS1, *trnL-trnF*, and *matK* datasets were sorted by BioEdit software and then, re-aligned manually to be more precise. The length of these alignment datasets and the number of

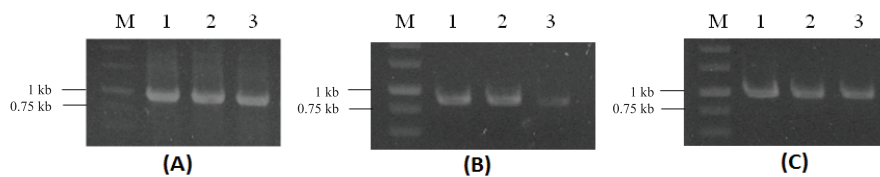
DNA sequences in each dataset vary depending on the DNA regions. Despite being the shortest alignment (285 bp), ITS1 database is the largest collection (20 sequences). In contrast, alignments of *trnL-trnF* region (463 bp in length) and *matK* region (678 bp in length) have 12 and 11 sequences, respectively. Indeed, many *Ehretia* species do not comprise all three DNA sequences. Additionally, it was observed from these alignments that nuclear marker ITS1 contains more variable sites (132) compared to the two

plastid markers (*trnL-trnF* of 33 and *matK* of 34) (Fig. 2).

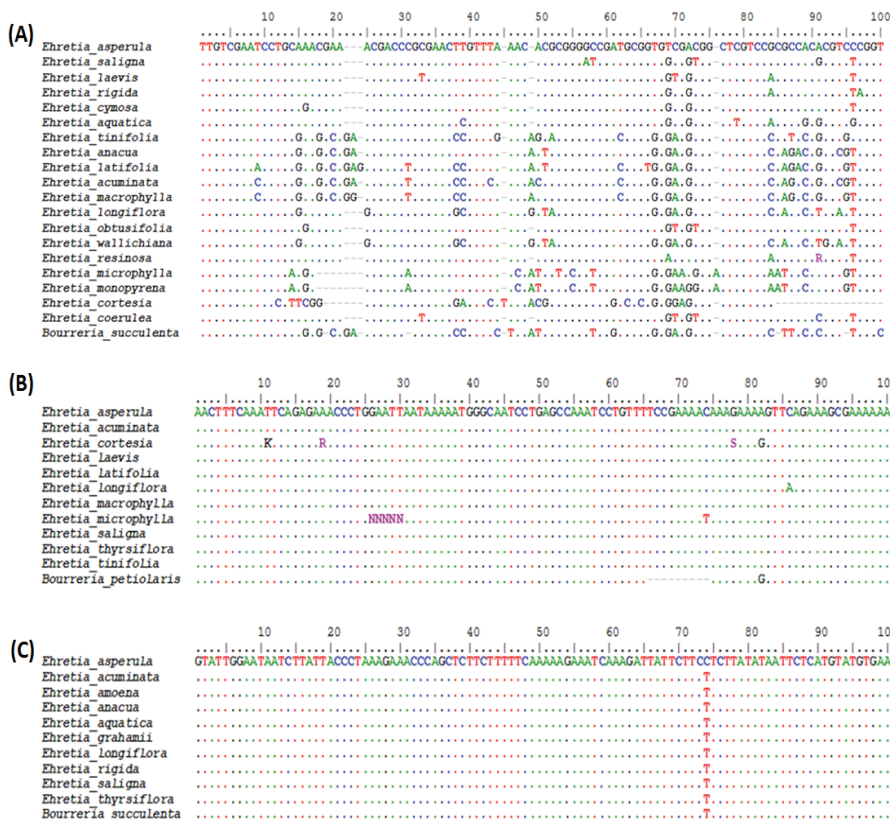
The best likelihood trees with bootstrap support values of all three DNA markers were indicated in Fig. 3. Based on the ITS1 region (Fig. 3A), the phylogenetic relationship of *Ehretia* genus is in agreement with the previous study [5], in which this taxon was divided into 4 main clades: *Ehretia* I (100 PBS, 100DPS), *Ehretia* II (even PBS was low 54, DBS was high 93), *Ehretia* III (96 PBS, 100 DBS), and *Ehretia* IV (BS under 50, but 100 PBS, 100 DPS). The phylogenetic trees derived from the two plastid markers differ to some degree. The species belonging to the same clade at ITS1 tree were still grouped, though with relatively low branch support. Similarly, some species were changed their phylogeny positions. For instance, *E. anacua*, which was initially subjected to *Ehretia* II clade at ITS1 tree, was separated away from other *Ehretia* II species at *matK* tree (Fig. 3C). Analysis of *matK* data also suggested the systematic relationship of three species in which ITS1 sequences were not available. *E. amoena* and *E. grahamii* were aligned with *Ehretia* I species with high bootstrap value of 92 PBS and 94 DPS. Meanwhile, *E. thyrsoiflora* relates to *E. acuminata* with low supported bootstrap in *matK* tree, but with high supported bootstrap in *trnL-trnF* tree (Fig. 3B). In the phylogenetic tree constructed from ITS1 data, which have a significant amount of data so far, *E. resiona* is a close relative of *E. asperula* even with low bootstrap support (54PBS, 69DBS).

## Discussions

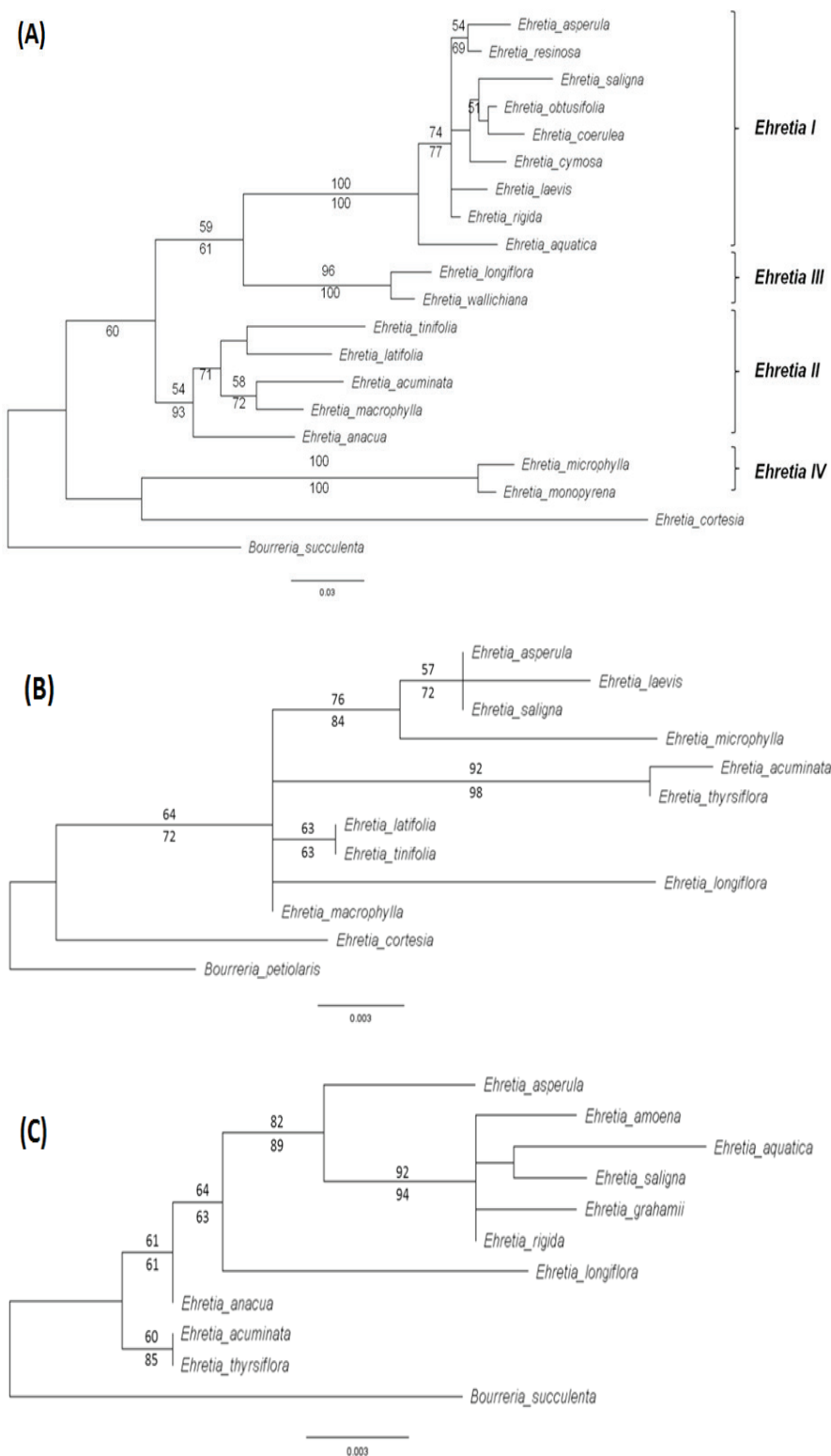
The molecular phylogenetics has been resolving the evolutionary relationship between related species over the few past decades [9]. Many regions in plant nuclear and plastid genome have been assessed and evaluated, such as ITS, *trnL-trnF*, *matK*, and *rbcL* [10]. ITS from the nuclear ribosomal DNA is obviously the only nuclear region and a proper choice for phylogeny analysis. It



**Fig. 1.** PCR results of three DNA regions: ITS (A), *trnL-trnF* (B), and *matK* (C). M: 1 kb marker; 1-3: PCR products of three different *E. asperula* DNA samples.



**Fig. 2.** Alignments (partial) of *E. asperula* with other *Ehretia* species and an outgroup species of three DNA regions: ITS (A), *trnL-trnF* (B), and *matK* (C).



**Fig. 3. Maximum likelihood trees of *Ehretia* based on ITS1 (A), *trnL-trnF* (B), and *matK* (C) sequences (-ln = 1733.905, 819.1461 and 1163.316 respectively).** Branch lengths are to scale. Major clades at ITS1 tree are indicated. Number on the branches are bootstrap support values, where values under 50% are not shown (above: parsimony, below: distance).

has greater species discrimination than other plastid regions at lower taxonomic levels and has efficiency even at intraspecies levels [10]. In fact, scientists sometimes use a portion of ITS, ITS1, or ITS2 to calculate the distance between species [11]. On the other hand, plastid genome contributes many candidates for phylogeny calculation. They have been used by means of separated sequence or in combination with others [10]. However, it is difficult to pick a combination that meets all requirements: good discriminatory power, good sequence quality, and universality. With regard to the Ehretiaceae family, both primary and secondary structures of ITS1 region were adopted successfully [4, 12, 13]. Similarly, *trnL-trnF* intergenic sequence can discriminate species when being combined with other plastid or nuclear DNA markers [14, 15]. On the other hand, *matK* has never been seen in any phylogeny study of the Boraginales order.

In our study, we used ITS1 to figure out the systematic position of *E. asperula* since ITS1 was analyzed in most researches about Ehretiaceae in general and *Ehretia* species in specific. Our analysis of ITS1 divided the *Ehretia* genus into 4 clades, which were numbered according to a previous report [4]. However, the position of a clade in relation to others differs from the previous studies due to the difference in the kind of data input. Yet, this result was acceptable for our purpose and discussion. Besides ITS1, two other plastid DNA markers were applied to investigate their efficiency on species discrimination. As expected, ITS1 could distinguish efficiently all the 20 species (Fig. 3A). In contrast, as it can be seen in Fig. 1C, *matK* could not differentiate between *E. acuminata* and *E. thyrsoiflora*. A similar situation was observed at *trnL-trnF* maximum likelihood tree (Fig. 3B). Perhaps, *trnL-trnF* and *matK* are effective when dealing with higher taxon level or in combination with other plastid and/or nuclear markers. Inferred from the ITS1 maximum likelihood tree (Fig. 3A), *E. asperula* and *E. resinosa*



are sister species even though the molecular data provides low bootstrap support (54PBS, 69DBS).

Based on three phylogenetic trees, *E. asperula* was grouped with *Ehretia* I species, which also have been used for medical purposes for a long time. For instances, *E. saligna*'s decoction of wood is drunk for aches and pain [16]. Moreover, the paste of the *E. laevis* leaves has remarkably wound healing properties [17]. Furthermore, *E. cymosa* leaves were used to treat toothaches [18] and stomach ulcers [19]. In order to treat many ailments, dried leaves and wood of *E. asperula* were normally added in hot water, and sometimes with other Vietnamese herbs [2].

Many cancer patients in Vietnam cannot afford to get some expensive treatments, such as surgery, radiation therapy, and chemotherapy. Furthermore, these treatments are usually coupled with unwanted side effects [20]. Traditional medicine, which uses many kinds of Vietnamese herbs, offers an alternative potential cost-effective and harmless treatment. It is likely to lead to a high demand for traditional medical herb, in general, and *E. asperula*, in specific, than ever before. Even though scientists utilized tissue culture for the massive production of *E. asperula* [21], most of this herb in medical plant market was exploited from its limited natural habitat. This paradoxical situation brings many concerns as *E. asperula* market is expanding. The concerns are outright substitution, contamination, and adulteration with some non-effective, less effective, or even some allergic, lethal herbs, and mislabeling fillers [22]. Initially, the conventional procedures to identify a plant are morphological and anatomical methods, which are not always successful for some reasons. Fortunately, advance in molecular technology has offered researchers a simple, cost-effective, and rapid approach to species identification based on DNA sequences, and DNA barcoding [22]. According to this study, we suggest that the ITS1 region is a suitable DNA

barcode for the classification of *Ehretia* genus and the identification of *E. asperula*.

## Conclusions

In short, we sequenced three DNA markers (ITS, *trnL-trnF*, and *matK*) of *E. asperula* and deposited these sequences on the Genbank with ID number KY320205, KY320206, and KY320207, respectively. *E. asperula*'s systematic position was evaluated by three phylogenetic trees. It is concluded that *E. asperula* belongs to *Ehretia* I clade and has a close relation with *E. resinosa*. Additionally, we presented ITS1 as a potential DNA barcode for identification, which could be further assessed to monitor the non-authentic medical herb market.

## REFERENCES

- [1] H. Riedl (1997), "Boraginaceae", *Flora Malesiana, Series I*, **13**, pp.91-99.
- [2] The Asia Foundation (2012), *Medicinal Plant Index of the Daos in Ba Vi*, <https://asiafoundation.org/resources/pdfs/MedicinalPlantIndexoftheDaosinBaVi>.
- [3] Q.H. Hoang (2010), *Study some medicinal plants of Ehretia PBR., be long to family Boraginaceae in the Northern provinces of Vietnam*, Dissertation, Hanoi University of Pharmacy.
- [4] M. Gottschling, H.H. Hilger (2004), "Characterisation of a novel fruit type found in *Ehretia* (Ehretiaceae, Boraginales)", *Blumea-Biodiversity, Evolution and Biogeography of Plants*, **49(1)**, pp.145-153.
- [5] M. Gottschling, F. Luebert, H.H. Hilger, J.S. Miller (2014), "Molecular delimitations in Ehretiaceae (Boraginales)", *Mol. Phylogenet. Evol.*, **72**, pp.1-6.
- [6] J.J. Doyle, J.L. Doyle (1990), "Isolation of-plant DNA from fresh tissue", *Focus*, **12**, pp.13-15.
- [7] T.A. Hall (1999), "BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT", *Nucl. Acids Symp. Ser.*, **41**, pp.95-98.
- [8] D.L. Swofford (2002), *PAUP\*, Phylogenetic analysis using parsimony\* and other methods, version 4.0 beta version*.
- [9] Z. Yang, B. Rannala (2012), "Molecular phylogenetics: principles and practice", *Nat. Rev. Genet.*, **13**, pp.303-314.
- [10] K. Vijayan, C.H. Tsou (2010), "DNA barcoding in plants: taxonomy in a new perspective", *Current science (Bangalore)*, **99(11)**, pp.1530-1541.
- [11] G.N. Feliner, J.A. Rossello (2007), "Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants", *Mol. Phylogenetic Evol.*, **44(2)**, pp.911-919.
- [12] M. Gottschling, H.H. Hilger (2001), "Phylogenetic analysis and charater evolution of *Ehretia* and *Bourreria* (Ehretiaceae, Boraginales) based on ITS1 sequences", *Bot. Jahrb. Syst.*, **123**, pp.249-268, doi: 10.1127/0006-8152/2002/0124-0149.
- [13] M. Gottschling, H.H. Hilger, M. Wolf, N. Diane (2001), "Secondary Structure of the ITS1 transcript and its application in a reconstruction of phylogeny of Boraginales", *Plant Biol.*, **3(6)**, pp.629-636.
- [14] F. Luebert, C. Brokamp, J. Wen, M. Weigend, H.H. Hilger (2011), "Phylogenetic relationships and morphological diversity in Neotropical Heliotropium (Heliotropiaceae)", *Taxon*, **60(3)**, pp.663-680.
- [15] F. Luebert, J. Wen (2008), "Phylogenetic analysis and evolutionary diversification of Heliotropium sect. *Cochranea* (Heliotropiaceae) in the Atacama Desert", *Syst. Bot.*, **33(2)**, pp.390-402.
- [16] L.J. Webb (1959), "The use of plant medicines and poisons by Australian Aborigines", *Mankind*, **7(2)**, pp.137-146.
- [17] T. Rushikesh, S. Bhutada, B. Chouragade, P. Khobragade, H. Ketaki (2016), "Unexplored wound healing property of *Ehretia laevis* Roxb. (Khandu Chakka) Plant", *Int. J. Res. Ayurveda Pharm.*, **7(4)**, pp.54-57.
- [18] T. Bahru, Z. Asfaw, S. Demissew (2013), "Wild edible plants: sustainable use and management by indigenous communities in and the buffer area of Awash National Park, Ethiopia", *SINET: Ethiopian Journal of Science*, **36(2)**, pp.93-108.
- [19] A. Dansi, A. Adjatin, H. Adoukonou-Sagbadja, V. Faladé, H. Yedomonhan, D. Odou, B. Dossou (2008), "Traditional leafy vegetables and their use in the Benin Republic", *Genet. Resour. and Crop Evol.*, **55(8)**, pp.1239-1256.
- [20] K.D. Miller, R.L. Siegel, C.C. Lin, A.B. Mariotto, J.L. Kramer, J.H. Rowland, K.D. Stein, R. Alteri, A. Jemal (2016), "Cancer treatment and survivorship statistics, 2016", *CA Cancer Journal Clinicians*, **66(4)**, pp.271-289.
- [21] T.T.T. Le, V.M. Tran (2015), "Tissue cultures of *Xa den* (*Ehretia asperula* Zollinger et Moritzi)", *Journal of Science*, An Giang University, **3(3)**, pp.113-123.
- [22] P. Mishra, A. Kumar, A. Nagireddy, D.N. Mani, A.K. Shukla, R. Tiwari, V. Sundaresan (2016), "DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market", *Plant Biotechnol. J.*, **14(1)**, pp.8-21.