Anaphalis longifolia is a member of the Asteraceae family found throughout the highlands of Europe, America, and Asia. Studies on this plant are still limited to their habitat, whereas its molecular identification is largely unknown. This study aims to analyze the DNA barcode of A. longifolia using the matK gene sequence. DNA of the samples obtained from North Sumatra were isolated and then amplified using specific primers, and eventually sequenced. The results of sequencing were analyzed using the Molecular Evolution Genetics Analysis (MEGA) program Version X. The results showed that the matK gene sequence was successfully amplified at a length of 800-850 kb. The results of the phylogenetic tree analysis show that the matK gene sequence is capable of classifying A. longifolia. In the A. longifolia matK gene sequence, the AT content was
higher than GC. The genetic distance obtained from the sequencing ranges from 0-0.0014. The results of matK gene sequence alignment analysis show that there were 1521 observable characteristics, 1403 conserved site characteristics, 118 site variable characteristics, 9 parsimony informative site characteristics, and 7 single nucleotide polymorphism (SNP) site characteristics. The matK gene sequence can be used in DNA barcoding to identify *A. longifolia*. The results of this study are expected to provide important information in *A. longifolia* conservation attempts.

**Keywords:** *Anaphalis longifolia*, matK, DNA barcoding, North Sumatra

**Introduction**

*Anaphalis* is a member of the Asteraceae family (Tjitosoedirdjo, 2002) widespread across mountainous areas in the continents of Europe, America, and even Asia (Chanchani et al., 2011). *Anaphalis* thrives at an altitude between 800 to 3400 asl (Backer & van den Brink, 1965). Most of this genus can be found in the highlands and mountains (Prakasa et al., 2018). Due to its ability to thrive in a nutrient-poor environment, *Anaphalis* is considered to have high ecological value (Aliadi et al., 1990).

From 2001 to 2019, North Sumatra has lost 23% (1.33 Mha) of its tree coverage, which is equivalent to the ability to absorb 549 Mt of CO₂ emissions. Mandailing Natal Regency is the region with the most reduction in tree coverage (147 Kha) (Global Forest Watch, 2020). Forest degradation and climate change are the main causes of the increasing difficulty of finding *Anaphalis*. In addition to its habitat in critical environments, *A. longifolia* is also a plant with low seed viability, making it difficult to conserve. Besides that, *A. longifolia* also has very slow growth.

Plant identification using barcoding DNA is one tool that can be used in conservation efforts. Barcoding DNA is used to identify, inventory and study specimens to understand species diversity and evaluate the genetic variability of species (Krishna Krishnamurthy & Francis, 2012). With barcoding DNA, researchers can identify species more quickly and thoroughly in order to take the appropriate action for establishing the right scale for conservation (Francis et al., 2010).

The matK gene is a plant DNA barcode recommended by Barcode of Life (CboL). It is one of the fastest growing plastid encoders and consistently shows high levels of discrimination towards angiosperm species. (Fazekas et al., 2008; Lahaye et al., 2008a). (Li et al., 2011) stated that a combination of plastid markers including rbcl, matK and trnH-psbA could be used as DNA barcoding with a success rate of 87.1-92.7%. According to (Hollingsworth et al., 2009), matK markers can be used to identify angiosperms (90%), gymnosperms (83%), and cryptogams (10%). In Indonesia, the matK marker has been used for identifying Aeridinae (Orchidaceae) subtribe. (Topik et al., 2005), *Myristica fragrans* (Tallei & Kolondam, 2015), Dipterocarpaceae (Harnely et al., 2018), palm (Abbas et al., 2020), and Andrographis paniculata (Arif et al., 2019).

*Anaphalis longifolia* is a member of the genus *Anaphalis* (Koster, 1941). IUCN Redlist (2008) classified *Anaphalis spp* as threatened or endangered plants. Research related to *A. longifolia* is still limited to its ecological status and distribution (Taufiq et al., 2013). The use of matK gene DNA barcoding to identify this species has not yet been done before. This study aims to analyze the potential of the matK gene as barcoding DNA of *A. longifolia* plants from North Sumatra. This research is expected to provide important information on how to identify *A. longifolia* in its conservation effort in North Sumatra.

**Materials and Methods**

**A. longifolia Samples**

2 samples of *A. longifolia* were obtained from Taman Eden in Toba Samosir regency and Bukit Sipiso Piso in Simalungun regency, North Sumatra. A total of 8 data from NCBI were used ingroup, namely that of *Anaphalis marginata* (HQQ94564), *Anaphalis aureopunctata* (MH659925.1), *Anaphalis aureopunctata* (MH714290.1), *Anaphalis sinica* (MH659676.1), *Anaphalis sinica* (KX148081.1), *Anaphalis marginata* (MG224815), *Anaphalis marginata* (HM445632), and *Helianthus annuus* (AY215805.1) outgroup.

**DNA Extraction**

Fresh leaves from *A. longifolia* were extracted using the Geneaid Plant DNA Isolation Kit following the kit protocol. A total of 100 mg of plant leaf tissue was crushed and then put into a 1.5 ml microcentrifuge tube and added 400 μl of lysis buffer GP1 and 5 μl of RNase A. After homogenizing and incubating at 60°C for 10 minutes, 200 l of elution buffer and 100 μl of GP2 buffer were added. The mixture was then transferred to a column filter in a 2 ml collection tube and then centrifuged at a rate of 1,000xg for 1 minute. The column filter containing
the supernatant was removed and replaced with a new column filter. The solution in the collection tube was moved to a new filter column and 150% buffer GP3 was added to the solution volume. GD Column and the collection tube were then centrifuged at a speed of 16,000 x g for 2 minutes. DNA was washed 2 times using W1 buffer and wash buffer. The DNA in the GD column was then eluted using a 100 µl elution buffer which had been heated at 60°C. The collection tube was then replaced with a microcentrifuge tube. After being centrifuged again at a speed of 16,000 x g for 30 seconds, the DNA that had entered the 1.5 ml microcentrifuge tube was then stored at -20°C.

**DNA Amplification**

The matK sequence was amplified using matK-F 5'-ACC CAG TCC ATC TGG AAA TCT TGG TTC-3’ and matK-R 5'-CGT ACA GTA CTT TTG TGT TTA CGA G-3’ primers (Ki-Joong Kim, School of Life Sciences and Biotechnology, Korea University, Korea, unpublished). Amplification was carried out using the MyTaq HS Red Mix (Bioline) kit with a total reaction of 25 µl (2.5 µl of DNA template; 2.5 l matK-F primer; 2.5 µl of matK-R primer; 5 µl of distilled water; 12.5 µl of PCR Mix). Amplification of the matK sequence was carried out at the predenaturation stage at 97°C for 5 minutes, denaturation at 94°C for 1.5 minutes, annealing at 52°C for 1 minute and extension at 72°C for 1 minute. PCR results were visualized using agarose gel. PCR products which showed clear DNA bands were to be sent to FirstBase DNA Sequencing Service in Singapore for sequencing.

**Data Analysis**

The sequencing results in the form of a chromatogram were edited using Bioedit 7.0.1 to obtain a consensus sequence based on the conservative sequences generated from the primary sequencing results of matK-F and matK-R. The consensus sequence that has been obtained was then aligned using Basic Local Alignment Search Tool (BAST) from the National Center for Biotechnology Information. Data with high similarity to the sample were included in the phylogenetic tree analysis. Phylogenetic trees were constructed using the Molecular Evolutionary Genetics Analysis (MEGA) X program (Kumar et al., 2018). Analyzes were performed to calculate the percentage of similarity, GC content, and genetic distance.

**Result and Discussion**

We succeeded in amplifying the matK gene sequence from the total genome of *A. Longifolia*, which would then be analyzed as DNA barcode (Figure 1). The amplification results ranged from 800-850 kb.

**Figure 1.** The results of visualization of *A. longifolia* matK gene PCR product using 1% agarose gel with 1 kb marker.

PCR products that have shown positive visualization results using agarose gel were then sequenced. After the sequencing results were analyzed using BLAST on NCBI, it was found that data with high similarity to the sample were primarily from the Asteraceae family. Species with a high level of similarity according to BLAST analysis were *Anaphalis margitacea* (99.87%), *Anaphalioidees mariae* (99.87%), *Anaphalis hancockii* (99.75%), *Helichrysum felinum* (99.62%), *Anahalis aureopunctata* (99.40%), and *Anaphalis sinica* (99.15%). Phylogenetic studies show that Anaphalis is very close to Helichrysum and Pseudognaphalium (Smissen et al., 2011; Ward et al., 2009). The results of the BLAST analysis, which showed species variation, indicated that the matK marker is not every effective when used as DNA barcoding on *A. longifolia*. Some researchers suggested a combination of the matK and rbc markers in determining DNA barcode (Saarela et al., 2013; Wattoo et al., 2016; Hollingsworth et al., 2009; Techen et al., 2014).

In a previous study using the same primer as this one, the said primer has succeeded in identifying 35% herbarium specimens and 45% fresh specimens consisting of 900 vascular plants from 51 families, 24 orders, 147 genera, and 312 species. (Kuzmina et al., 2012). The use of this primer on Indonesian plants has been performed on Annonaceae (18 samples), Apocynaceae (7 samples), Dipterocarpaceae (12 samples), Lauraceae (6 samples), Meliaceae (7 samples), Moraceae (7 samples), Myristicaceae (12 samples), Phyllanthaceae (1 sample), Primulaceae (4 samples), and Rubiaceae (7 samples) plants (Dean et al., 2018). Especially in Sumatra, this primer has
successfully amplified 441 species, 97 families and 40 orders of flowering plants (Amandita et al., 2019).

Figure 2. Phylogenetic tree based on the matK gene sequence of Anaphalis longifolia with the Helianthus annuus outgroup reconstructed using the Neighbor Joining method, with the evolutionary distance calculated using the Kimura 2-parameters method (Kimura, 1980). Percentage of species-replicating trees under bootstrap test (1000 replicates) (Felsenstein, 1985).

The phylogenetic tree in Figure 2 shows that the matK gene sequence in A. longifolia can distinguish this species from other Anaphalis genus and from Helianthus annuus as an outgroup of the Asteraceae family. This shows that the matK gene has the potential to be used as DNA barcoding for A. longifolia, but it is less effective when used to distinguish between species in the anaphalis genus. Analysis using clustal W in the MEGA X application shows that there are 1521 characteristics that can be observed. From this data, there are 1403 conserved sites characteristics, 118 site variables characteristics, and 9 parsimony informative sites characteristics. Phylogenetic analysis using the matK gene sequence has more parsimony informative sites than genes in other chloroplasts (Müller et al., 2006; Barthet & Hilu, 2007).

Table 1. Composition of nucleotides, AT and GC content of the matK gene sequence in Anaphalis longifolia

<table>
<thead>
<tr>
<th>Species</th>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
<th>Total</th>
<th>GC (%)</th>
<th>AT (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaphalis margaritacea</td>
<td>36.34</td>
<td>17.41</td>
<td>29.48</td>
<td>16.77</td>
<td>787</td>
<td>34.18</td>
<td>65.82</td>
</tr>
<tr>
<td>Helianthus annuus</td>
<td>36.92</td>
<td>16.60</td>
<td>29.48</td>
<td>17.00</td>
<td>1506</td>
<td>33.60</td>
<td>66.40</td>
</tr>
<tr>
<td>Anaphalis acaeropunctata</td>
<td>36.88</td>
<td>17.36</td>
<td>29.28</td>
<td>16.48</td>
<td>789</td>
<td>33.84</td>
<td>66.16</td>
</tr>
<tr>
<td>Anaphalis sinica</td>
<td>36.38</td>
<td>17.49</td>
<td>29.58</td>
<td>16.55</td>
<td>852</td>
<td>34.04</td>
<td>65.96</td>
</tr>
<tr>
<td>Anaphalis sinica</td>
<td>37.01</td>
<td>17.52</td>
<td>28.65</td>
<td>16.82</td>
<td>862</td>
<td>34.34</td>
<td>65.66</td>
</tr>
<tr>
<td>Anaphalis acaeropunctata</td>
<td>36.28</td>
<td>16.24</td>
<td>30.05</td>
<td>16.44</td>
<td>1521</td>
<td>32.68</td>
<td>67.32</td>
</tr>
<tr>
<td>Anaphalis margaritacea</td>
<td>36.53</td>
<td>17.24</td>
<td>29.63</td>
<td>16.60</td>
<td>783</td>
<td>33.84</td>
<td>66.16</td>
</tr>
<tr>
<td>Anaphalis margaritacea</td>
<td>36.89</td>
<td>16.34</td>
<td>30.04</td>
<td>16.73</td>
<td>1518</td>
<td>33.07</td>
<td>66.93</td>
</tr>
<tr>
<td>Anaphalis longifolia</td>
<td>36.32</td>
<td>17.68</td>
<td>29.51</td>
<td>16.49</td>
<td>837</td>
<td>34.17</td>
<td>65.83</td>
</tr>
<tr>
<td>Anaphalis longifolia</td>
<td>36.24</td>
<td>17.53</td>
<td>29.65</td>
<td>16.59</td>
<td>850</td>
<td>34.12</td>
<td>65.88</td>
</tr>
<tr>
<td>Average</td>
<td>36.75</td>
<td>16.99</td>
<td>29.60</td>
<td>16.66</td>
<td>1030.5</td>
<td>33.65</td>
<td>66.35</td>
</tr>
</tbody>
</table>

The results of the matK sequence analysis on A. longifolia show that the AT content was higher than the GC content in the Asteraceae family (Table 1). Variation in GC content is a key genome feature due to being closely related to the fundamental elements of genome organization in an organism (Eyre-Walker & Hurst, 2001; Mukhopadhyay et al., 2007). Genomes rich in GC show higher gene density, higher mutation rates conservation level, and higher rates of recombination level compared to regions lacking in GC (Niu et al., 2017). GC content from 65 accessions of Edelweiss (Leontopodium) from the Himalayan/Tibet centre using nuclear ribosomal (ITS and ETS) and plastid (matK and trnL, F) sequences ranged from 43-52% (Blöch et al., 2010). Anapalis selengensis genome has 37.46% GC content and 62.54% AT content (Meng et al., 2019).

Table 2. Genetic distance between Anaphalis longifolia and species from the Anaphalis genus and the Asteraceae family
The genetic distance analysis of Anaphalis longifolia with species from the Anaphalis genus and the Asteraceae family showed that the distances between species in the genus Anaphalis ranged between 0-0.014 (Table 2). The highest variation was found in Anaphalis sinica and the lowest was in Anaphalis longifolia. The genetic distance between the Anaphalis genus and Helianthus annuus as an outgroup ranged from 0.067 to 0.074. The genetic distance in the Leontopodium ITS region ranged from 0.2% to 6.8% (Blöch et al., 2010). A study by Ade et al. (2019) who analyzed the genetic distance of Anaphalis spp (A. javanica, A. longifolia, and A. viscida) based on molecular characteristics (ITS, ETS, and EST-SSR markers) showed that the genetic distance was between 0.004 to 0.040, indicating small genetic distance between species in the Anaphalis genus.

Molecular-based research on Anaphalis has been successfully carried out using the internal transcribed spacer (ITS) marker (Glenny & Wagstaff, 1997; Blöch et al., 2010) and external transcribed spacer (ETS) (Blöch et al., 2010; Nie et al., 2013; Ade et al., 2019), EST-SSR markers (Ade et al., 2019), trnL (Blöch et al., 2010; Kurniawan et al., 2014), and matK (Blöch et al., 2010). Molecular-based research on Anaphalis using the matK marker is still very rare. The matK gene is a marker that can be used as a DNA barcode and is recommended by Barcode of Life (BOL) for its consistent ability to show high discrimination level between angiosperm species) (Lahaye et al., 2008a).

In this study, the results of alignment of the matK gene sequence from 9 data on Anaphalis genus indicated that there were 7 single nucleotide polymorphism (SNP) sites (Table 3), namely sites 492, 504, 505, 506, 1061, 1068, and 1176. In the A. longifolia species, no SNP site was detected. SNP for defined genetic location were determined in at least 1% of the population (Kim & Misra, 2007). SNP is one of the stable genetic polymorphisms in a genome and can be used to analyze differences between closely related species (Germarino & Klein, 1999; Yamamoto et al., 2010). matK and rbcL showed high sequence quality, but only provided a few SNP sites (Huang et al., 2014) and therefore highly suitable for use in identifying species (Hollingsworth et al., 2009).

Table 3. Single nucleotide polymorphism in the Anaphalis genus

<table>
<thead>
<tr>
<th>Sample</th>
<th>A. margaritacea (HQ594564)</th>
<th>A. aureopunctata (MH6599251)</th>
<th>A. aureopunctata (MH7142901)</th>
<th>A. sinica (MH6596761)</th>
<th>A. sinica (KX148081.1)</th>
<th>A. margaritacea (MG224815)</th>
<th>A. margaritacea (MH445632)</th>
<th>A. longifolia</th>
<th>A. longifolia (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>A</td>
<td>-</td>
<td>A</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

In this study, the matK gene sequence in A. longifolia has the potential to be used as DNA barcode. Lahaye et al. (2008) proposed that matK could potentially become DNA barcode in plants. The matK gene sequence is one of the fastest growing sequences of the plastid genome (Hilu & Liang, 1997) and possibly the closest plant analogue to the COI gene sequence in animal DNA barcodes (Hollingsworth et al., 2011). The matK gene sequence can be very difficult to amplify by PCR using existing primary sets, especially in non-angiosperm plants (Hollingsworth et al., 2011).

DNA barcoding, which is widely applied in taxonomic research today, is invaluable for understanding species boundaries, community ecology, evolution and biodiversity conservation (Kress et al., 2015). Currently, conservationists have adopted DNA barcodes as a tool in the field of conservation (Chakraborty et al., 2014; Joly et al., 2014). DNA Barcoding enables the identification species boundaries to be used as clues in determining target conservation habitats (Faith, 1992). When the DNA barcode for each species is complete, comparative measure of phylogenetic diversity will become the standard metric for assessment in determining conservation strategies (Kress et al., 2015).
In addition, the development of DNA barcode research is also used for the identification and detection of illegally-traded endangered species (Lahaye et al., 2008b). There will be increased use of DNA barcode in the future, mainly because the available technology will become simpler and cheaper (Kress et al., 2015).

**Conclusion**

The results showed that the matK gene sequence can be amplified at a length of 800-850 bp. In the *A. longifolia* matK gene sequence, the AT content was higher than the GC content. The resulting genetic distance ranged between 0.0014. The alignment results of the matK gene sequence showed that there are 1521 observable characteristics, including 1403 conserved site characteristics, 118 site variable characteristics, 9 parsimony informative site characteristics, and 7 single nucleotide polymorphism (SNP) site characteristics. This suggests that the matK gene sequence has the potential for development as DNA barcoding to identify *A. longifolia*.

**References**


