Phylogenetic Relationships Among Some Species of The Genus *Hypericum* (Hypericaceae) in Kurdistan-Region /Iraq

1Hasan H. H.
2Husain S. A.
2Sardar A. S.
2Hamasalih R. M.

1Department of Biology, College of Education-Shaqlawa, Salahaddin University-Erbil, Iraq
2Department of Biology, College of Education, Salahaddin University-Erbil, Iraq

**Abstract:** Phylogenetic relationships for the widely distributed genus *Hypericum* (St. John's wort) had already been focused on morphology, with just a few species having been studied molecularly. We're currently in the midst of a long-term study of the genus *Hypericum*, and we've been analyzing some pretty interesting data. Specifically, we've been looking at the nuclear ribosomal DNA internal transcribed spacer sequences and 28SrRNA of eight different species of *Hypericum*, as well as one species of *Thornea calcicola* (which serves as our outgroup). It's all part of our larger effort to better understand the evolutionary relationships between these species. The subsequent parsimony analysis and Bayesian inference support the actual categorization and interspecific affinities. The molecular evidence supports the monophyletic assemblage of the sections in the ITS and 28SrRNA region trees, which indicate four distinct clades with little variation in the positions of the individual species in both trees. We identified traits that support significant clades within the genus *Hypericum* using Maximum Parsimony and Bayesian approaches to reconstruct ancestral states of the chosen features. The findings support using ITS and 28SrRNA sequencing analysis to clarify the evolutionary connections within the numerous members of the genus *Hypericum*. It's great to know that the phylogenetic relationships among *Hypericum* species have been mostly resolved via ITS phylogeny. This means that we have a better understanding of the evolutionary relationships between these species. With this information, we can make more informed decisions about how to classify and study these plants.

**Keywords:** nrDNA ITS gene, 28SrRNA gene, Phylogenetic relationship, *Hypericum*
1. Introduction

*Hypericum* L., the prevalent genus in Hypericaceae family. The appellation of a particular botanical genus, *Hypericum prepartum* was originally assigned by the Greeks to plants that were suspended above their sacred figures as a measure to protect against malevolent entities. Currently, this genus includes a plant species commonly referred to as Saint John's Wort, which is putted in the clusioid clade of Malpighiales [1].

The genus *Hypericum* encompasses an extensive variety of over 500 plants, including herbs, shrubs, and small trees. This genus constitutes more than 80% of the identified species within the plant family Hypericaceae. Additionally, the Hypericaceae family comprises three tribes that are recognized based on molecular findings. These are Cratoxyleae, Hypericeae, and Vismieae. The Cratoxyleae tribe, which is classified into *Cratoxylum* and *Eliea*, comprises seven species. The Hypericaceae tribe consists of about 494 species, which are classified into *Hypericum, Lianthus, Santomasia, Thornea*, and *Triadenum*. The Vismieae tribe, on the other hand, comprises about 102 species, which are classified into *Harungana* and *Vismia*. It is noteworthy that *Hypericum*, which is part of the clusioid clade, is the only group of this clade that occurs in temperate regions. This genus occurs alongside the monotypic genus *Lianthus* and the genus *Triadenum*. All other members of the clusioid clade are native to pantropical or subtropical lowland regions of the world. Overall, the molecular findings support the recognition of the three tribes within the Hypericaceae family [2]. In a recent publication, the authors [3, 4] undertook a review and revision of the previously published classifications for *Hypericum* at the intrageneric level by [5-10]. Through a thorough investigation of various factors, including anatomical and morphological characteristics, ecological distribution, cytological data, and phytochemical analyses, the authors suggested a subdivision of the genus into 30 primary sections and 6 additional subsections. (summarized in [7]). *Hypericum* is one of the most prevalent toxic plants in Iraq, with just sixteen species identified, and the most abundant species are *H. triquetrifolium* and *Hypericum perforatum* [8]. Despite the fact that numerous molecular analyses have been conducted using various techniques for various plants, there is a dearth of literature on molecular phylogenetic studies in *Hypericum*. The rationale behind this is that phylogenetic investigations in a vast genus such as *Hypericum* frequently encounter the issue of inadequate sampling. Numerous molecular phylogenetic studies have involved *Hypericum*, a genus of flowering plants. However, these studies were either geared towards family or genus-level connections [1,9,10], or relied on an insufficient sample size that did not encompass the entire genus. The latter studies were also characterized by a lack of representative outgroups, which were either few in number or distantly related to the genus in question. These limitations suggest that a comprehensive analysis of *Hypericum* is necessary to fully understand its phylogenetic relationships [11,15]. Just one of these research has examined the connections between the main lineages of *Hypericum*. [10], evaluating one mitochondrial and three chloroplast gene, showed support for two broad clades within *Hypericum*. As a result, several issues remain unclear, including the connections between key lineages and the monophyly of sections within the *Hypericum* genus. On the other hand, to delve into the causes of the significant diversity observed within the *Hypericum* family, which constitutes 80% of its members, a comprehensive phylogenetic hypothesis is deemed necessary. Such an approach will enable the identification of biogeographic patterns and the evolution of traits exhibited within the genus. Phylogenetic, which involves the study of evolutionary relationships among organisms, has been widely utilized to establish the evolutionary history of various taxa. In this
In this study, the noncoding regions of nuclear ribosomal DNA internal transcribed spacer and 28SrRNA were amplified by utilizing specific primers that were obtained from Macrogen Company located in Seoul, Korea represented in table (2). The amplification reactions were conducted in a total volume of 25 μL, where the Master Mix consisted of 12. μL 3 μL genomic DNA extract, 2 μL of each primer, and 5. μL free nuclease water. The process of Polymerase Chain Reaction (PCR) involved a series of steps performed at specific temperatures and time intervals for ITS as follow, the initial step, referred to as denaturation, involved heating the sample to 95 degrees Celsius for three minutes. This was followed by 38 cycles of heating the sample to 95 degrees Celsius for 30 seconds, cooling it to 53 degrees Celsius for 45 seconds, and then heating it to 68 degrees Celsius for one minute. Finally, the sample was heated to 70 degrees Celsius for eight minutes in a process referred to as the final step. On the other hand, the PCR program for 28SrRNA gene started with the initial step, referred to as denaturation, involved heating the sample to 94 degrees Celsius for 20 seconds. This was followed by 35 cycles of heating the sample to 94 degrees Celsius for 30 seconds, cooling it to 56 degrees Celsius for 20 seconds, and then heating it to 68 degrees Celsius for one minute. Finally, the sample was heated to

2. Methodology

Taxon Sampling

The current study utilized various plant taxa that were gathered from different physiogeographical districts of Kurdistan region-Iraq, in addition to preserved herbarium specimens found in the Herbaria of our College. The analysis involved 9 separate taxa, comprising of eight ingroup taxa and one outgroup Thornea calcicola. The outgroup sequence was acquired from gene bank, with Accession numbers LT904679 and AY573028 for ITS region and 28SrRNA region, respectively. Table (1) provides further details on these taxa.

DNA Extraction

During the DNA extraction process, we experimented with various amounts of plant tissue and different extraction methods. After testing, we found that two methods worked best when extracting DNA from Hypericum plants. These methods included a CTAB approach (modified from Doyle and Doyle) [16] and the Invisorb® Spin Plant Mini Kit (Invitek, Berlin, Germany). For fresh plant tissue, we followed the manufacturer's protocol for the Invisorb® Spin Plant Mini Kit. However, for old and poorly preserved tissues from herbarium sheets the CTAB method is suitable method, we added 2% PVP40 (polyvinylpyrrolidone). We made sure to include no more than 10 mg of plant tissue per extraction to maintain DNA quality and yield. The resulting DNA pellet was thoroughly washed twice using 0.5 mL of 80% ethanol. Subsequently, the DNA was dissolved in 25 μl of TE-buffer.

PCR and DNA Sequencing

In this study, the noncoding regions of nuclear ribosomal DNA internal transcribed spacer and 28SrRNA were amplified by utilizing specific primers that were obtained from Macrogen Company located in Seoul, Korea represented in table (2). The amplification reactions were conducted in a total volume of 25 μL, where the Master Mix consisted of 12. μL 3 μL genomic DNA extract, 2 μL of each primer, and 5. μL free nuclease water. The process of Polymerase Chain Reaction (PCR) involved a series of steps performed at specific temperatures and time intervals for ITS as follow, the initial step, referred to as denaturation, involved heating the sample to 95 degrees Celsius for three minutes. This was followed by 38 cycles of heating the sample to 95 degrees Celsius for 30 seconds, cooling it to 53 degrees Celsius for 45 seconds, and then heating it to 68 degrees Celsius for one minute. Finally, the sample was heated to 70 degrees Celsius for eight minutes in a process referred to as the final step. On the other hand, the PCR program for 28SrRNA gene started with the initial step, referred to as denaturation, involved heating the sample to 94 degrees Celsius for 20 seconds. This was followed by 35 cycles of heating the sample to 94 degrees Celsius for 30 seconds, cooling it to 56 degrees Celsius for 20 seconds, and then heating it to 68 degrees Celsius for one minute. Finally, the sample was heated to
70 degrees Celsius for eight minutes in a process referred to as the final. The PCR products were then analyzed on 1.5 % agarose gels to confirm the presence of the desired DNA sequences and visualized by staining with Safe red dye and photographed under UV transilluminator. Following this, the purified PCR products were subcontracted to the National Science and Technology Development Agency (NSTDA) in Thailand for sequencing. The PCR products were purified using kits obtained from Promega Company located in Madison, USA.

Table 1: Specimen ID of Hypericum species in the herbarium of College of Education-Salahaddin University-Erbil with their date of collection and locality in Kurdistan region.

<table>
<thead>
<tr>
<th>Taxa</th>
<th>Specimen ID and Herbarium symbol</th>
<th>Locality in Kurdistan</th>
<th>Date of collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. amblysepalum</td>
<td>8042 ESUH</td>
<td>Kory valley</td>
<td>7.11.202</td>
</tr>
<tr>
<td>H. asperulum</td>
<td>8020 ESUH</td>
<td>Halgard M.</td>
<td>7.7.2011</td>
</tr>
<tr>
<td>H. hirtellium</td>
<td>8057 ESUH</td>
<td>Sakran M.</td>
<td>13.7.201</td>
</tr>
<tr>
<td>H. tetrapterum</td>
<td>8038 ESUH</td>
<td>Rowanduz</td>
<td>1.5.2005</td>
</tr>
<tr>
<td>H. perforatum</td>
<td>8049 ESUH</td>
<td>Hasarost M.</td>
<td>14.9.202</td>
</tr>
<tr>
<td>H. scabrum</td>
<td>8053 ESUH</td>
<td>Halgard M.</td>
<td>26.7.201</td>
</tr>
<tr>
<td>H. triquetrifolium</td>
<td>8072 ESUH</td>
<td>Gara M.</td>
<td>7.11.202</td>
</tr>
<tr>
<td>H. amblysepalum</td>
<td>8065 ESUH</td>
<td>Hasan Bag M.</td>
<td>7.7.2011</td>
</tr>
</tbody>
</table>

Table 2: The primer list with its sequences that are utilized in the current study.

<table>
<thead>
<tr>
<th>Primer regions</th>
<th>Product size</th>
<th>Sequence 5'---3'</th>
<th>Foreword</th>
<th>Reverse</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>28SrNA</td>
<td>700 bp</td>
<td>TCTGACATGTG GCGAGTCA</td>
<td>GATTCGGCA GGTGAGTGT TT</td>
<td>[17]</td>
<td></td>
</tr>
<tr>
<td>ITS</td>
<td>400 bp</td>
<td>ATCGGATACTT GGTGGAAT</td>
<td>TCTCCGCCT ATTGATATGC</td>
<td>[18]</td>
<td></td>
</tr>
</tbody>
</table>

Phylogenetic Study and Alignment of Sequences

The purpose of the phylogenetic analyzes was to investigate the monophyly of Hypericum, examine the likelihood of hybridization events among major groups, and create a broad hypothesis of lineage relationships within the genus. To accomplish this, a dataset consisting of nine sequences was utilized, which were directly sequenced from eight different Hypericum species and one outgroup. Both Bayesian and parsimony analyzes were conducted on the dataset to further explore the genetic relationships within the genus. The DNA sequences underwent editing and alignment using the ClustalW feature in BioEdit, Version 7.0.4.1 [19] with manual adjustment. The ITS and 28SrRNA datasets, including the out-group species, consisted of 9 accessions each. Bayesian inference (BI) and maximum parsimony (MP) studies were distinctly assessed for each sequences made from the two indicators, using 9 limittaxa with all available sequences. The software program PAUP_ 4. a164 [20] was utilized in the MP analysis. This method sophisticated a heuristic search approach with 100 repeats of adding random taxa, Tree-Bisection-Reconnection (TBR) branch swapping, and MulTrees turned on while steepest descent was turned off. The number of trees stored was 100 per replica, and the bootstrap value was calculated from 100 replicas. Additionally, the subsequent factors were considered: (CI), (RI), (RC), and (HI). These calculations were achieved as outlined in reference [21].

Prior to performing the BI (Bayesian Inference) study, it is compulsory to measure the ideal substitution model. This is accomplished by utilizing the Akaike Information Criterion (AIC) in MrModeltest2 version 2. [22]. The results of this evaluation indicated that the (GTR+I+G) model was the most appropriate model for the ITS region. Similarly, the (GTR+G) model was selected as the perfect substitution model for 28SrRNA. To conduct the Bayesian Inference examination, MrBayes v.3.2 [23] was utilized. The program automatically estimated priors on state frequencies and rates and variation across sites. We run four Markov chains beginning with random
trees concurrently, using 500,000 generations for the internal transcribed spacer and one million generations for the 28S rRNA dataset, with four chains (one cold and three heated) per generation. was used and two independent analyses were performed with the temperature parameter set to 0.1. Trees were selected every 100th generation. In the course of the experiment, a certain number of initial tree samples were obtained. To ensure the accuracy and reliability of the results, a burn-in period was implemented, during which 25% of the initial tree samples were removed. Subsequently, a tree with a maximum of 50% was plotted using a majority rule consensus tree. The value of posterior probability (PP) was calculated to further validate the results. Finally, the resulting tree was plotted using the FigTree software, version 1.4.3. This software is a commonly used tool in the field of molecular evolution for the visualization and manipulation of phylogenetic trees. The aforementioned methodology was employed to obtain a comprehensive understanding of the evolutionary relationships between the organisms under investigation [24].

3. Results and discussion:

The PCR for ITS and 28SrRNA gene in all investigated species produced ~300-350 and ~600-890 bp monomorphic fragment respectively. The analogue of the sequenced products was recognized using the BLAST on NCBI server (http:// blast.ncbi.nlm.nih.gov/Blast.cgi). All the sequences showed 95 to 99% identity to available Hypericum species.

Sequence alignment and phylogenetic analyses

The ITS and 28SrRNA length of all sequences were 370 and 810 bp respectively. The sequence lengths were different in the numbers of characters included because of ambiguity at the beginning and end of sequences.

Nuclear ribosomal DNA ITS gene

A data set consisted of 8 ingroups and one outgroup taxa with 324 aligned DNA characters (including gaps) were used, from them only 42 character was parsimony informative. The maximum parsimony analysis showed 100 trees from which a solitary most parsimonious tree was retained with a tree length of 279 steps. The CI, RI, RC and HI were 0.935, 0.710, 0.664 and 0.065, respectively, with topology identical between MP and BI analyses. The summary of the analysis showed in (Table 4). Maximum parsimony and Bayesian Inference based phylogenetic tree shows the same tree topology, the strict consensus tree generated by summarizing the entire most parsimonious tree displays four clades described by color Figure (1): Clade A consists of four species, two sister clade H. asperulum and H. seabrnum with very strongly bootstrap support and posterior probability (bs=100 and pp=0.1) and H. lysisimachaides and H. triquetrifolium with moderately bootstrap support and posterior probability (bs=75 and pp=0.89); Clade B Contains only one species which is H. tetrapterum with strongly support (bs=88 and pp=0.93); Clade C involves two sister species H. amblysepalum and H. perforatum with very strongly support (bs=100 and pp=0.1) and finally clade D comprise the H. hirtellum with very strongly support (bs=100 and pp=0.1).

28SrRNA gene

A data set consisted of 8 ingroups and one outgroup taxa with 635 aligned DNA characters (including gaps) were used, from them only 39 character was parsimony informative. The maximum parsimony analysis showed 100 trees from which a solitary most parsimonious tree was retained with a tree length of 438 steps. The CI, RI, RC and HI were 0.966, 0.727, 0.702 and 0.034, respectively, with topology identical between MP and BI analyses. The summary of the analysis showed in (Table 4).

Maximum parsimony and Bayesian Inference based phylogenetic tree shows the same tree topology, the strict consensus tree generated by summarizing the entire most parsimonious tree displays four clades described by color tree topology was more
clear compared to ITS gene. A monophyletic clade of genus Hypericum was observed against outgroup from the genus Thornea calcicola Figure (2), the bootstrap support and posterior probability (pp) value ranges were between (bs=75-100) and (pp=0.89-0.100), with moderately and strongly support respectively.

Table 4: A list of statistics of phylogenetic tree and alignment of nrDNA ITS and 28SrRNA gene studies.

<table>
<thead>
<tr>
<th>Parameters / Regions</th>
<th>ITS</th>
<th>28SrRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligned length</td>
<td>324</td>
<td>835</td>
</tr>
<tr>
<td>Number of parsimony informative characters</td>
<td>42</td>
<td>39</td>
</tr>
<tr>
<td>Number of variable parsimony uninformative characters</td>
<td>159</td>
<td>341</td>
</tr>
<tr>
<td>Number of constant characters</td>
<td>123 (0.37)</td>
<td>255 (0.40)</td>
</tr>
<tr>
<td>Tree length (steps)</td>
<td>279</td>
<td>438</td>
</tr>
<tr>
<td>Consistency Index (CI)</td>
<td>0.935</td>
<td>0.966</td>
</tr>
<tr>
<td>Retention Index (RI)</td>
<td>0.710</td>
<td>0.727</td>
</tr>
<tr>
<td>Rescaled Index (RC)</td>
<td>0.664</td>
<td>0.702</td>
</tr>
<tr>
<td>Homoplasy index (HI)</td>
<td>0.65</td>
<td>0.034</td>
</tr>
<tr>
<td>Substitution Model</td>
<td>GTR+I+G</td>
<td>GTR+G</td>
</tr>
</tbody>
</table>

It's really important for this study to consider the possibility of paralogous ITS sequences or pseudogenes that could mess up the phylogenetic reconstruction. But, we didn't find any evidence in our dataset that suggests the existence of paralogous loci. This could be due to technical reasons like PCR drift or not considering enough clones, or it could be because there's almost complete concerted evolution of rDNA in *Hypericum*, meaning there are no paralogous sequences at all.

Robson's evolutionary tree is the only currently available study on infrageneric hereditary affairs of *Hypericum*. This tree was built on morphological and geographical associations among taxa, and it shows the affinities among sections. However, in this study, only one species, *H. perforatum*, was mentioned in the results. Another study [25] identified the presence of our species in two different sections. The first section is Section. 9. *Hypericum* subsection. 1. Hypericum series 2. Senanensia, which includes *H. perforatum*, *H. triquetrifolium*, and *H. tetrapterum*. The second section is Sect. 17. Hirtella group, which contains the remaining species. The tree gathered from 28S rRNA sequences shows a excellent resolution of most taxonomic interconnections. The agreement with the prevailing general taxonomy of *Hypericum* is very remarkable. Our ITS and 28SrRNA phylogeny suggests that *Hypericum* is a monophyletic group [26]. Although there is definite disagreement between Robson's description of the sections and the clades distinguished in this analysis, the association between the sections differ considerably from those advances by Robson.

This amended theory about phylogenetic communications between major parts within a genus (that is, the clade/grade above) is relevant to the clarification of biogeographic patterns and the reconstruction of historic trait states [27].

Figure 1: Strict consensus tree of most parsimonious tree represented in a maximum parsimony study of the nuclear ribosomal DNA internal transcribed spacer (ITS) region with heuristic search. (Tree length of 279 steps, CI = 0.935, RI = 0.719, RC = 0.664 and HI =0.065). Numbers above the branches which in green colour provide bootstrap values for the nodes and numbers in blue colour give Bayesian posterior probability values and clades are identified by colours.
Figure 2: Strict consensus tree of most parsimonious tree represented in a maximum parsimony study of the nuclear ribosomal DNA 28SrRNA region with heuristic search. (Tree length of 438 steps, CI = 0.966, RI = 0.727, RC = 0.702 and HI =0.034). Numbers above the branches which in green colour provide bootstrap values for the nodes and numbers in blue colour give Bayesian posterior probability values and clades are identified by colours.

Conclusions

The findings presented in this study support previous classifications of the Hypericum species while also revealing new relationships that warrant further investigation within this genus. Both the ITS and 28SrRNA trees illustrate four distinct clades, albeit with the 28SrRNA tree demonstrating better resolution of intraspecific separation. Future investigations in this area should focus on expanding the taxonomic specimen inside the cosmopolitan Hypericum genus. To fully assess the phylogeny of the whole genus, it is critical to increase the sample size to include all taxa. Additionally, revisiting the general range and interspecies relationships of Hypericum through additional morphological revisions to assess the taxonomic importance of traits conventionally used to delineate subdivisions and species may be useful. It would be worth it. To precisely place section relationships within Hypericum, additional molecular and morphological comparisons and an expanded sample size are necessary.

References


