Review on molecular markers for identification of Orchids

Thi Huyen Trang Vu[1*], Thi Ly Le[2], Truong Khoa Nguyen[3], Duy Duong Tran[3], Hoang Dung Tran[4]

[1]*School of Agricultural Science and Biotechnology, Nguyen Tat Thanh University*
[2]*School of Biotechnology, Ho Chi Minh International University*
[3]*Genetic Engineering Division, Institute of Agricultural Genetics*
[4]*Nguyen Tat Thanh University*

Received 1 March 2017; accepted 2 June 2017

Abstract:

***Orchidaceae* is one of the most valuable plant groups all over the world, and is also an impressively large and complex family of flowering plants. Effective molecular tools used for the identification of orchid species should be developed to support traditional morphological approaches. This study reviews most of the DNA fragments that have been used as taxon identifiers in researches conducted on *Orchidaceae* in order to assess potential molecular markers and metric measurements for the identification of orchid species.**

***Keywords:* DNA barcode Orchids, DNA fingerprinting, molecular identification Orchidaceae, molecular markers, molecular phylogeny.**

***Classification number:* 3.5**

Introduction

Orchidaceae is one of the largest and most complex families of flowering plants, comprising of approximately 22,500 species belong to 736 currently recognised genera [1]. Orchids have many values ranging from the beauty of their flowers to therapeutic properties in some species. However, taxons of *Orchidaceae* are endangered, this is mainly because of over-collection and habitat destruction, and all species are included in Conventions on International Trade of Endangered Species of Fauna and Flora (CITES) I and II [2]. Illegal trade and imitations have also become increasing problems. Unfortunately, laws banning these issues and their enforcement have met obstacles mostly due to the imperceptible difference of species' morphology. So, it is very difficult to identify orchid species and their inter-species hybrids using traditional classification; even those with fertile parts. Besides this, the species can be transported in a vegetative state, as seeds or as fragments [3]. Accurate authentication of orchid species is critical for biodiversity conservation and effective utilisation of orchids as plant resources [4].

Many researchers have, therefore, tended to develop genetic tests that can cheaply and easily determine the present species. "DNA barcodes" tools are promising options in providing a practical, standardised, species-level identification approach that can be used for biodiversity assessment, life history, ecological studies, and forensic analysis [5]. DNA barcode refers to the use of a single segment of DNA to identify specific coding information that offers discriminating ability of the living taxa, even if only a small fragment of the organism at any stage of development is available [6]. The potential DNA regions used as barcodes should match some key criteria: i) The universality of amplification and sequencing; ii) The pattern of intraspecific vs. interspecific variation; and iii) The power to identify species [7].

The selection of a barcode locus is complicated due to the trade-off that arises between the need for universal application in a wide range of taxa and sequence substitution saturation [5]. The single region 5' end of cytochrome c-oxidase 1 (CO1) from the mitochondrial genome is quite successfully used for the identification of animals [8, 9]. However searching for DNA barcode in plants is far more challenging than in animals. Mitochondrial genes, including CO1, in plants have low rates of synonymous substitution [10, 11], a large structural rearrangement in the genome, and import of sequences from nucleus and chloroplast [12]. Because of these problems, they are not recommended to use for DNA barcodes for plants. So, the nuclear and chloroplast genomes are focused to look for identifying markers for plants. Until now, no single sequences can be sufficient to identify all plant species. Even the use of a combination of multi-locus barcodes also gives different levels of discrimination in different groups of plants (Table 1).

The aim of this paper is to assess potential molecular markers for the identification of orchids. We review most of the DNA fragments that have been used as taxon identifiers in researches on Orchidaceae and other land plants, including orchids. The capability of taxon discrimination is often evaluated along with the construction of a phylogenetic tree. So we also use phylogenetic articles as literature sources for seeking suitable sequences. Highly evaluated markers proposed by previous authors will be deeply discussed to summarise a database of molecular candidates for orchid authentication.

[*]*Corresponding author: Email: vuthihuyentrang.vu@gmail.com*

Table 1. Summary of advantages and disadvantages of molecular loci in plants.

Loci	Advantages	Disadvantages
Nuclear regions (such as ITS)	- High variability - High number copies in cell - Biparental inheritance → more information	- Universal across different groups of organisms → unexpected contamination - Intra-genomic variability, divergent paralogous copies (multiple functional copies), pseudogenes→ poor-quality sequences
Chloroplast gene	- High copy number - Maternal inheritance information → no reflection of complexity - Some genes are highly variable - The high variable regions → used for low-level identification (species, under-species) - The lower variable regions → used for high-level identification (genus, family, tribe…)	- Low evolutionary rate → few choices
Mitochondrial genes		- Low rate of sequence change - Genome structure of mitochondrial in plant rapidly change
Intergenic spacer and intron	- High variability (do not encode any products→ faster evolve than coding regions)	- Difficult to amplify, sequence and align - Too variable, even intra-species
Low-copy nuclear genes (such as *Xdh*)	- High variability	- Low copy number - The present and absent of introns, the size of introns, the substitution rate are greatly variable and poorly studied → no universal

Studied barcodes for Orchid taxa

Single locus barcodes

The ITS region: In plants, nuclear genes (particularly introns) and spacers exhibit the highest variability [13]. The internal transcribed spacers (ITS) of nuclear ribosomal regions were proposed as a variable molecular marker for detecting genetic variation among genera, species, and within species. The two internal transcribed spacers (ITS1 and ITS2) do not encode any product but permit it to evolve at a faster rate than the ribosomal coding regions. For example, the ITS length of the aligned sequence in *Holcoglossum* (*Orchidaceae*) was 567 bp and it contained 26 informative sites and 27 variable sites [14]. The ITS exhibits high resolution at the species level [5]. The ITS has been shown to have unparalleled species resolutions compared with candidate barcodes proposed thus far [15]. Besides that, they exist in cells with high numbers of copies [5, 9]. The flanked regions 5.8S at the middle and 18S, 26S at the two ends of the ITS fragment, are conserved sequences which are useful to develop primers [9] (Fig. 1). The high retrieval rate of amplicons of the ITS [5, 16] may be due to these characteristics. The ITS regions were evaluated and found to have high-quality bidirectional sequences [5]. In addition, the ITS with nuclear genes can provide more complex information which relates to biparental inheritance in comparison with plastid markers [17]. The length of the ITS is about 600 bp [18-20], which can satisfy the length requirements of barcoding.

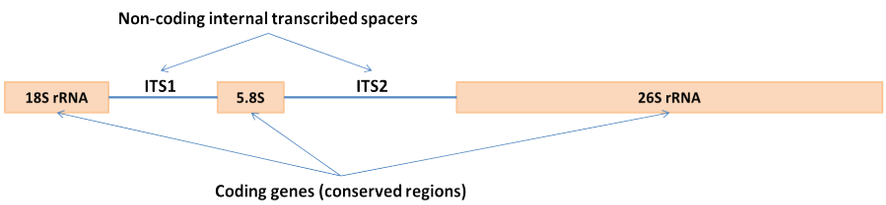

Fig. 1. Coding genes.

The ITS spacer, although often highly variable, also reached a number of limits for DNA barcoding [5]. Using the ITS as a barcode still has its challenges [21]. The high variability among intra-genomic systems, on the contrary, was a disadvantage of species discrimination [22]. The CBOL Plant Working Group did not recognize the ITS as a suitable locus for DNA barcoding because there were many factors that affected the quality of sequences from direct sequencing of PCR products, including such reasons as the presence of intra-genomic variability, the divergent paralogous copies (multiple functional copies) within individuals and pseudogenes [22]. The coexistence of variation orthologous copies of the ITS in the hybrid genomes of *Paphiopedilum* leads to poor-quality sequences which consisted of multiple peaks. Thus, the ITS was not found to be suitable neither for species resolution nor for getting an insight into the parentage of the hybrids (50% species resolution for eight natural species) [6]. Another disadvantage of the ITS was that it could not be amplified from some barium samples because the ITS is too variable to guarantee reliable alignments and contains variable indels (insertion/deletions) at the species level [23].

However, the internal transcribed spacer region (ITS) of ribosomal DNA has proved to be an effective marking barcoding progress. At the genus level, the ITS clearly distinguishes between the two genera *Paphiopedilum* and *Phragmipedium,* and also the *Cypripedioid* genera [24]. At the species level, Kress, et al. (2005) [5] has evaluated that the ITS has a much higher divergence value than any of the plastid regions studied and has an amplification success rate of 88%. Therefore this region was proposed as a potentially usable DNA region for the application of barcoding to flowering plants using the optional supplementary marker *trnH-psbA*. The ITS was also evaluated to be an effective candidate DNA barcode for *Orchidaceae* [16, 18, 25]. The PCR success rate of the ITS was high [16], 100% in 158 wild orchid samples [18], and in *Holcoglossum* [6, 14]. Although the combination of the ITS with another sequence showed a greater ability to identify species, the ITS sequence alone was still an effective barcode among proposed loci

[*rbcL*, *matK*, *atpF-atpH*, *rpoB*, *rpoC1*, *trnH-psbA*, *trnL-F,* and ITS] [25]. The ITS was employed successfully in *Dendrobium (Orchidaceae*). Specifically, the phylogenetic relationship and the differentiation of 11 medicinal *Dendrobium* spp. from one another and from two adulterant species *Pholidotaarticulata* and *Flickingeria comate* could be analysed using this locus [20]. The specific nucleotide sequences of the ITS is used for the identification and phylogeny of 20 *Dendrobium* species, in which ITS1 and ITS2 regions exhibit more variation than the 5.8S rDNA [19]. The single ITS barcode revealed to be the best DNA barcode affording 100% species resolution based on 129 congeneric species of *Dendrobium*, and 93% based on sets of sequences from both the experiment and the GenBank. This resolution value of the single ITS barcode was higher than other single barcodes *matK*, *rbcL*, *rpoB*, *rpoC1* and even combined barcodes [*matK+rbcL*] and [*matK+rpoB+rpoC1*] (33, 20, 18, 17, 80.77 and 92.31% respectively) in the study [16]. Using nuclear ribosomal ITS sequence data, genetic units in *Grammatophyllum speciosum* complex (*Orchidaceae*) were totally recognised at species level to be *G. speciosum* Blume*, G. wallisii* Rchb.f., *G. kinabaluense* Ames and C. Schweinf.*, G. pantherinum* Rchb.f., and *G. cominsii* Rolfe [26].

In orchid phylogenetic research, there have been many different markers used, among them is ITS, which has usually been a favourite choice of researchers [27-31]. The ITS tree of 16 *Paphiopedilum* species and two varieties found in Vietnam, received strong Jack-knife support for phylogenetics analysis [32]. However as discussed in the barcoding field, not any single or multiple-locus markers could totally identify the species, and so the combination of regions in phylogenetic studies could not get the strong support in all clades of phylogenetic trees [33-35]. The ITS was used and combined with other regions in most of the phylogenetic studies thus far [3, 36-40]. The combined data matrices often get better results than individual ones [41-43]. With a higher taxonomic level (section, genus, subgenus, tribe, subfamily), ITS, *matK*, *trnL* and *rbcL* were most often used [44-49], e.g. three subgenera of *Cymbidium* genus could be distinguished clearly from among ten species by its ITS (ITS1+5.8S+ITS2) [50].

In general, the ITS is worth its use for barcoding projects; and in order to increase resolution effects, ITS2 can be an alternative solution, or supplementary markers maybe accompanied as discussed in the combination barcodes latter.

ITS2: Located between ribosomal genes 5.8S and 28S of the ITS regions, recently ITS2 has been highly noticed as a valuable barcode for many plants. The ITS2 spacer provides structural elements necessary for correct pre-rRNA processing and probably has a function in the regulation of the transcription of active ribosomal subunits [51]. As an identifying marker, this sort of ITS fragment does not only receive benefits from the long ITS possess but also overcomes the limits of a full-length region.

In ITS2, the success rates of both PCR and sequencing were very high [3], mostly 100% in *Dendrobium* [51] or 93.8% in a wide range of plants instead of 42.3% for full ITS [52]. This is because the length of ITS2 is quite short at about 248 bp in *Dendrobium* [51], and the design of the universal primers for this sequence is easy due to the availability of flanked conserved regions (5.8S and 26S rRNA). This characteristic was also ideal for barcoding, which should be short enough to recover amplicons from degraded DNA [23, 53] and can overcome the trouble of the universal ITS from contaminated microorganism [53].

The rate of successful identification with the ITS2 is high at 92.7% at the species level and 99.8% at the genus level [52]. Early from 2001, Lau, et al. (2001) [53] discovered the significant ability of ITS2 for differentiating medicinal *Dendrobium* species from one another and also from non-orchids and adulterants. In 2010, the 50790 ITS2 sequences of plants were downloaded from GenBank and were evaluated according to their sequence lengths, GC content, intra- and inter-specific divergence, and efficiency of identification. The study proposed that the ITS2 locus shows significant sequence variability at the species level or lower, and should be used as a universal DNA barcode for identifying plant species. Among them all, the success rates for using the ITS2 region to identify *Orchidaceae* taxa at the species level (%) were variable in different genera *Scaphyglottis* 100.0, *Satyrium* 98.3, *Dendrobium* 91.9, *Dichaea* 81.8, *Disa* 79.7, Masdevallia 79.6, *Paphiopedilum* 76.6, *Telipogon* 76.1, *Cymbidium* 74.1, *Dendrochilum* 71.2, *Cyrtochilum* 69.3, *Phalaenopsis* 65.9, *Oncidium* 65.1, *Maxillaria* 62.9, *Gomesa* 49.1, *Diuris* 31.1, and *Ophrys* 22.7 [54]. In the study on 43 samples of *Dendrobium*, the ITS2 analyses showed a significant divergence between the inter- and intra-specific genetic distances, and the presence of a barcoding gap was obvious [51]. The variability of ITS2 was sufficient enough to distinguish even closely related species. The phylogenetic analysis of ITS2 regions of 64 *Dendrobium* species also showed good results when the cluster analysis mainly supported a relationship between the species of *Dendrobium* established by traditional morphological methods and many previous molecular analyses [51] (Table 2).

trnH-psbA: trnH-psbA is a non-coding intergenic spacer which is a rapidly evolving region. *trnH-psbA* was early on highly evaluated by Kress, et al. (2005) [5] because of its high interspecific variation, high length variation, good priming sites, and it was proposed to be the most viable candidate for a single-locus barcode for land plants identification [55]. Actually, *trnH-psbA* was quite easily amplified with a success rate that either might reach 100% in

large numbers of land plants [5, 56] or at least more than 90% in a wide range of plants [3, 7, 52, 56-58]. In orchids particularly, PCR rates were also very high, up to 100% in *Dendrobium* [59, 60], *Oncidium* [61], *Holcoglossum* [14], and *Cymbidium* [62] (Fig. 2).

trnH-psbA was popularly known to have a high sequence divergence [5, 55, 56] due to its large number of insertions and deletions (indels). The species resolution rates of this region in plants were the highest in many research studies, for examples (82.6%) out of the nine other loci ITS, *rbcL*, *ndhJ*, *matK*, *rpoB*2, *rpoC1*, *ycf5,accD* [55] and (59%) out of the nine other loci *cox1*, 23S rDNA, *rpoB*, *rpoC1*, *rbcL*, *matK*, *atpF-atpH*, *psbK-psbI*] [58]. In the big DNA barcoding project of the CBOL Plant Working Group, *trnH-psbA* also showed the highest species discrimination (69%) in comparison with another six loci (*atpF-atpH*, *matK*, *rbcL*, *rpoB*, *rpoC1*, *psbK-psbI*) from 259 samples of 95 species of 34 genera seed plants [7]. Species resolution rate of *trnH-psbA* in other land plant researches was 67.6%, second high among five screening markers *psbA-trnH*, *matK*, *rbcL*, *rpoC1*, *ycf5*, ITS2, ITS [52].

In *Orchidaceae,* species resolution of this sequence varied from different genera and different studies. Among six markers (*trnH-psbA*, *accD*, *rpoC1*, *rpoB*, *matK* and *ndhJ*), only *trnH-psbA* could resolve 8 of 11 Mesoamerican orchid species (72.7%), which was just lower than *matK* (10/11). Proportion of monophyletic species recovered with UPGMA of *trnH-psbA* were 90.6%, the same as *matK*, and the two highest among eight loci *accD*, *ndhJ*, *matK*, *trnH-psbA*, *rbcL*, *rpoB*, *rpoC1*, and *ycf5* for 172 individuals of 86 species (including 71 individuals of 48 orchid species + other angiosperm species) [57]. On *Dendrobium*, in testing the unique marker *trnH-psbA*, Yao, et al. (2009) calculated the intergenic spacers of all species (0.3 to 2.3%) and the intraspecific variation (0 to 0.1%) [59]. This research's result meant that the barcoding gap was obvious, and so the locus was evaluated as an effective spacer for barcoding *Dendrobium* species and for differentiating *Dendrobium* species from other adulterating species. But in 2011, the species resolution of *trnH-psbA* on *Dendrobium* was 79.3% over 504 samples (lower than *matK* 88.8%, *atpF-atpH* 82.4%, and *rbcL* 79.8% [63], or 8.14%, which is lower than the ITS, ITS2, and *matK* [60]. On *Holcoglossum* (*Orchidaceae*), 52 individuals belong 12 species were analysed for barcode with six markers *rbcL*, *matK*, *atpF-atpH*, *psbK-psbI*, *trnH-psbA* and ITS. Species resolution of *trnH-psbA* was 5/12 species (using Neighbor-Joining algorithm), equal to ITS 5/12 and lower than *matK* 6/12 [14].

In general, *trnH-psbA* was highly evaluated early-on, but less favourable at a later time due to some obvious problems. The most common complaint was about *trnH-psbA*. That it was generally too difficult to align in land plants [5, 55] as well as in *Orchidaceae*. *trnH-psbA* possesses many indels in their sequences [5, 55, 56]. The mononucleotide (A/T) repeats (or known as homopolymers) and/or small tandem repeats (AT) were frequently noted in this intergenic spacer [3, 7, 55, 56]. The existence of homopolymers (AAAA/TTTTTT) for most non-coding regions, including *trnH-psbA*, interrupted the sequence runs and caused problems with obtaining high-quality bidirectional sequences [3, 7, 58, 64], especially with the forward primer, as in *Dendrobium* [16]. The results were both lower and overlapped between bidirectional reads or only allowed partial sequences to be obtained. This is also another major limitation for this locus. In orchids and some other monocots, beside of these similar problems of indels and repeats [16], the genomic rearrangement of the non-homologous inverted repeat has been also found [15, 55], and especially the insertion of well-conserved exon-copies of *rpl*12 and *rps*19 (known as pseudogene) [57, 59, 65]. The indels, repeats and inverted repeats likely in *trnH-psbA* sequences made this region not only significantly complex, but also significantly different in length, and those were the reasons for the difficulty of alignment for analysis. This problem either led to the failure of sequence alignment or the requirement of manual editing [7].

The length of *trnH-psbA* in a wide range of plants is quite short at about 400 bp, to satisfy the criteria of the barcode, which should be short enough for easy amplification [5, 7]. On the other hand, there was research that showed that the short length of *trnH-psbA* makes a lack of information for barcode and phylogenetic analysis [56]. In orchids, the length of *trnH-psbA* are 850 bp in *Dendrobium* (included entire regions *psbA-trnH* 722-785 bp plus regions of *rpl*12 - 279 bp and *rps*19 - 19 bp) [59], 739 bp in *Holcoglossum* [14]. However, the containing of *rpl*12 and *rps*19 in orchids and some monocots caused much longer in size of this region up to > 1000 bp and meets the problem of PCR and sequencing [7, 65]. The high difference in length caused multiple bands in few samples of *Dendrobium* that hard to recognise which one is correct *trnH-psbA* segment, and the bad quality sequence made this locus excluded from the analysis [16].

In another point of view, the difficulty of alignment was not a major obstacle [5] compared to the benefit provided by their sequence information. Indels are useful pieces of information for species discrimination, i.e. they could help to distinguish three species of genus *Solidago,* which cannot be separated due to low sequence divergence, and Kress, et al. (2005) hoped for an improvement of DNA barcoding tools for which to utilize the indel information [5]. Thus, this locus is still valuable for many research projects, to be used as potential barcodes, especially with a combination of supplement barcodes, which will be discussed later.

matK: matK is the gene coding for the maturase K protein. This is also a rapidly evolving gene [7] that potential as an identification molecular marker

that can be used in many barcoding and phylogenetic studies. This region was evaluated at much higher levels of sequence variations for species discrimination [7, 65]. Lahaye, et al. (2008b) [66] suggested that *matK* was the preferred universal barcode for flowering plants, including orchids and re-affirmed best potential of this region in Lahaye. The results of DNA barcode library for 20 endangered *Orchidaceae* species distributed in Mexico using the barcodes *matK* and *rbcL* showed that single *matK* allowed for the identification of the most orchid species [67]. The species resolution of *matK* was 100% for *Paphiopedilum* [6], 5/5 species for five medicinal *Dendrobium* species [68] and 6/12 for *Holcoglossum* [14] in comparing with other studied single regions.

matK has also been proved to have the same problem as *trnH-psbA* with homopolymer runs of mononucleotide repeats for some taxonomic groups and led to low-quality bidirectional sequences [3, 58, 68]. However this rate was not significant, and only a few samples of *matK* amplification in *Dendrobium* gave multiple bands [16]. But *matK* suffered most by a low amplification success rate. From 96 individuals, at 96 species of 48 genera of land plants, the PCR rate of *matK* was just 39.3%, far lower than other screening loci (*trnH-psbA*, *rbcL*, ITS1, *ndhJ*, *rpoB*2, *rpoC1*, *ycf*5, *accD*) [55]. Fazekas, et al. (2008) used 10 primer pairs for sequencing reactions of *matK*, but the success rate could still not covered all the samples (88%) [58]. The poor PCR recovery might due to the nonuniversal primers.

This problem could be overcome through design improvements or by improving the specific primers [7, 65, 68]. Using the specific primer pairs 390F and 1326R from Cuénoud, et al. [69], it could get 100% amplification of *matK* [57]. But particularly in orchids, the amplification rates of *matK* are quite good. These rate could be up to 100% ([61] - Oncidiinae, [62] - Thailand *Cymbidium*, [60] - *Dendrobium*) or at very high range 95.23% ([6] - Indian *Paphiopedilum*), 92.31% ([14] - *Holcoglossum*), and 99.32% ([16] - *Dendrobium*). Then, *matK* is still a good choice for the barcoding of orchids, especially if good primers could be developed.

rbcL: rbcL is the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene in the plastid genome. The single region of *rbcL* was not high favoured because it is too long (1428 bp [5]) and contains highly conserved regions [61]. As with high universality, the amplification rates of *rbcL* were rather good in most of the studies, from 90 to 100% in a broad range of plants [3, 5, 15, 55, 57, 63, 64], and also in orchids [6, 14, 16, 60]. Although *rbcL* could get high-quality bidirectional sequences and well universality [7, 63], this poor variable region could discriminate well at the genus level, but could not show adequate resolution at the species level in plants [3, 5, 64, 70] and in orchids particularly [57]. In contrast with PCR success rates, resolution rates were very low [69.8% - [55], 75% - [64], 79.8% - [63], 58.02% - [15], 26.4% - [3] and not suitable for molecular identification in many *Orchidaceae* species (*Oncidiinae*, *Paphiopedilum*, *Holcoglossum* and *Dendrobium*) [6, 14-16, 25, 61]. However, *rbcL* was highly considered in many combined barcode, as discuss later.

ycf1: Another locus that is "more variable than *matK*" [71] or "any existing plastid candidate barcodes and can serve as a barcode for land plants", that was *ycf1*, proposed as a barcode by Dong, et al. (2015) [15]. Within the plastid genome, *ycf1* spans the small single copy (SSC) and the inverted repeat (IR) regions. The section of *ycf1* in the IR region is short (less than one kilobase long) and conserved. In contrast, the section of *ycf1* in the SSC region has high sequence variability in seed plants [15]. *ycf1* was known to be absent from some genera, but exists in orchids, including *ycf1a* and *ycf1b*. 357 of 420 tree species could be distinguished using *ycf1b* (85%), which was better than any of the *matK*, *rbcL*, and *trnH-psbA* and even slightly better than the core barcode *matK*+*rbcL* (71.31%) [15]. The reason that this valuable region received little attention for DNA barcoding or molecular systematic purposes at low taxonomic that is *ycf1* is too long (5709 bp in *Nicotiana tabacum*) and is too variable to permit the design of universal primers. However, the high variability of *ycf1* indicates its potential value in DNA barcoding of land plants. Phylogenetic application of *ycf1* had been found for *Orchidaceae* [30, 49, 71, 72] and was evaluated as the most variable and parsimony-informative among five other chloroplast genes (*matK*, *rbcL*, *rpoC1*, *rpoC2* and *ycf*2) [71, 73]. *ycf1* should be more tested in molecular identification of orchids.

atpF-atpH: Recently, Kim, et al. (2015) [74] suggested the intergenic spacer *atpF-atpH* as a barcoding marker after concluding a 100% species discrimination ability of this region on 28 individuals of four species of Korean *Cypripedilum-Orchidaceae*. The results obtained from such a set of observations were found, such as sequence variations, species-specific SNPs, indels differences, length variations, and the use of species-specific primers (ARMS method for amplification refractory mutation system). The authors also suggested identification using electrophoresis based on length variations of sequences. *atpF-atpH* has significant length variations among species and was used for molecular identification and phylogenetic study of low taxonomic level plant species, although it does not function independently [15, 21]. However, this region was not suitable for *Holcoglossum* orchid DNA barcoding study [25] and fell for recovery of high-quality bidirectional sequences [7].

rps16-trnQ: Jhong-Yi Lin and his group have explored that the *rps16-trnQ* marker showed the best discrimination power and was considered to be the best DNA barcode in the study. Another 15 of 27 studied cpDNA markers were

also recognised as highly variable among moth orchids, with polymorphic information contents of 8.0 and were suggested to combine with *rps16-trnQ* [75].

trnL-F: Orchidaceae, *trnL-F* regions were evaluated to be effective rabbet DNA barcode marker gene [25]. *trnL-F* was also used in many phylogenetic studies of orchid taxa such as *Orphrys, Angraecinae, Epidendroideae, Arethuseae, Vandae*, *Bulbophyllum*, *Coryciinae*, *Cypripedium*, *Tangtsinia*, and *Orchiade* [2, 29, 40-42, 76-81].

psbK-psbI: Like *trnH-psbA*, the intergenic spacer *psbK-psbI* showed good discriminatory power but had the lowest sequencing success in these trials, and substantial problems generating bidirectional reads [7]. The *Orchidaceae, psbK-psbI* showed the highest mean to be interspecific at a K2P distance (0.1192), followed by *matK* (0.0803), *atpF-atpH* IGS (0.0648), *trnH-psbA* IGS (0.0460) and *rbcL* (0.0248) [44]. By overcoming the obstacles and difficulties of the intergenic spacer as discussed in *trnH-psbA*, this region can be used as a potential molecular marker for orchids.

Multi-locus barcodes

In an effort to find universal barcodes for a wide range of plants as well as orchids, it was clear that no single locus could be sufficient in this role for both universality and resolvability, and multi-locus barcodes seem to be more robust and effective. As Kress, et al. (2005) has suggested, it may need more than one locus for species-level discrimination [5].

In 2007, the two-locus barcode *trnH-psbA+rbcL* was first proposed by Kress and Erickson [55] with the species resolution increased to 85% for angiosperms comparing to the highest 82.6% of single locus *trnH-psbA*. Actually, the combination of *trnH-psbA+rpoB* or *trnH-psbA+rpoC1* was 85% too, but the PCR success rate of *rbcL* in the study was higher than *rpoB* and *rpoC1*, and then the totally result of *trnH-psbA+rbcL* was better.

In the same year 2007, Chase, et al. first suggested the three-locus barcodes either of *matK+rpoC1+rpoB* or *matK+rpoC1+trnH-psbA* [65]. The combination of two or three loci of *matK*, *rpoC1*, *rpoB* and *trnH-psbA* was also recommended for 11 Mesoamerican orchid species in this year and the resolution results showed that *matK*+rpoC1+*rpoB* 100% (discriminate 11/11 species), *matK+rpoC1+trnH-psbA* and *matK+rpoB+trnH-psbA* 90.9% (10/11 species), and the single *matK* 90.9% (10/11 species). In the study of Singh, et al. (2012) on *Dendrobium* (*Orchidaceae*), the combination *matK+rpoC1+rpoB* gave the highest resolution (94.44%) among other three-locus barcodes, just lower than the ITS (100%). From this result, they suggested that "barcodes, if based on the single or limited locus, would be specific taxa" [16]. On the subject *Cymbidium* (*Orchidaceae)*, both the two options 1) The *matK+rpoC1+trnH-psbA*, and 2) The *matK+rpoB+trnH-psbA* have succeeded 100% species resolution of 19 *Cymbidium* in Thailand [62]. However, without *rpoC1* or *rpoB*, the two-locus *matK+trnH-psbA* barcode achieved only moderate improvement (90.9%) in comparing with the single *matK*

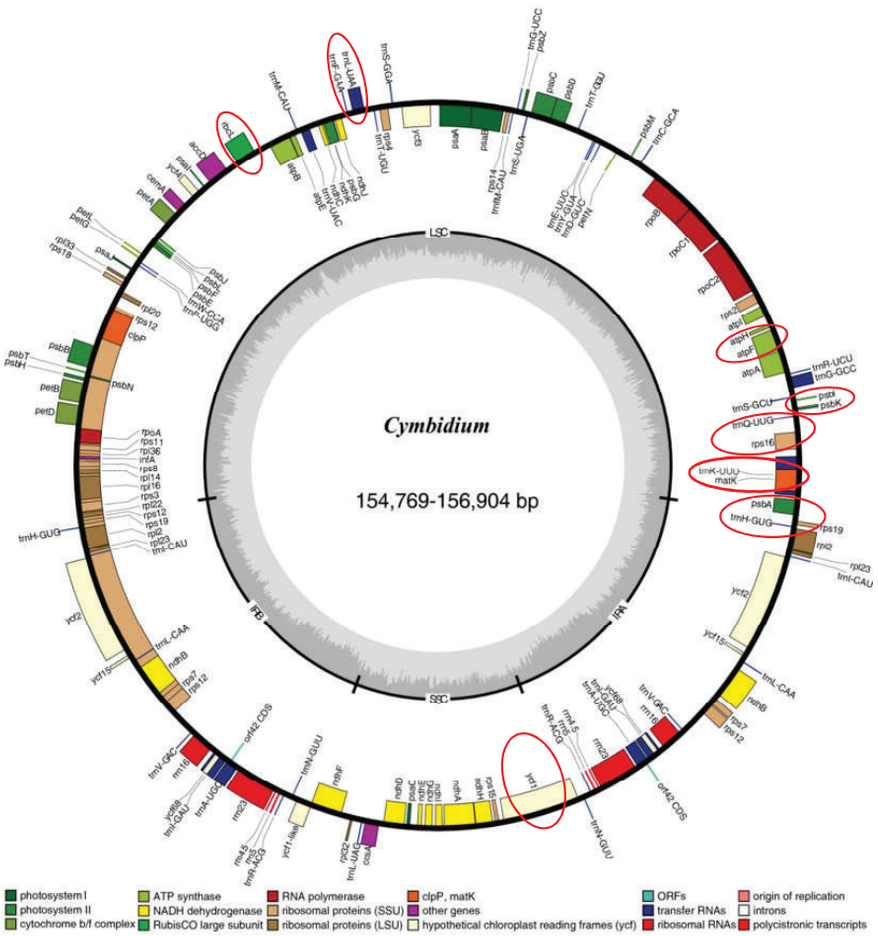

Fig. 2. Gene map of *Cymbidium* chloroplast genome [82] **with notations for potential loci for barcoding** (With Inverted repeat regions (IRa and IRb); small single copy (SSC) and large single copy (LSC) regions. Genes on the outside of the map are transcribed clockwise and genes on the inside of the map are transcribed counterclockwise. Loci in rectangle shapes are potential markers focused in this review).

resolution (90.6% of 48 orchid species plus 38 angiosperm species) [57].

In 2008, Fazekas, et al. selected a more loci combinations for barcoding 92 species of land plants. A multilocus plant barcoding region should have multiple regions chosen from among three of the coding (*rbcL*, *rpoB*, *matK*) and two of the non-coding regions (*trnH-psbA*, *atpF-atpH*) (61-69%). As all combinations assessed using four to seven regions had only marginally different success rates (69-71%); values that were approached by several two- and three-region combinations (61-69%) [58]. This meant that no single combination clearly outperformed all others. This situation was also proved in the study of Hollingsworth, et al. (2009) [21] with some three-locus combination of *rbcL*, *rpoC1*, *matK,* and *trnH-psbA*.

In 2009, the Consortium for the Barcode of Life (CBOL) Plant Working Group first recommended the two-locus combination of *rbcL+matK* as a plant barcode. This combination represented a practical solution to a complex trade-off between universality, sequence quality, discrimination, and cost [7]. Now it is generally agreed that a plant barcode will combine more than one locus (5-7), including a phylogenetically conservative coding locus (*rbcL*) with one or more rapidly evolving regions (partial *matK* gene and the intergenic spacer *trnH-psbA*). Thus Kress, et al. (2009) tested on three loci *rbcL*, *matK* and *trnH-psbA* whether the use of multilocus supermatrices to generate phylogenetic hypotheses at the species-level would improve the resolution power. The results showed that the core-combining barcode proposed by CBOL *rbcL+matK* just discriminated 92%, while *trnH-psbA+rbcL* did 95%. The three-locus combination *rbcL+matK+trnH-psbA* discriminated 98% [64]. *rbcL+matK* was 93.1% species resolution and *rbcL+matK+trnH-psbA* was 95.3% [63]. Fazekas, et al. (2012) also suggested the combination of *rbcL+matK* as the core barcode with another supplement barcode (ITS or *trnH-psbA*) [83].

In 2010, ITS2 and *psbA-trnH* sequences were highly evaluated, with 93.8% and 23.8% PCR success rates in 1,433 species of 551 genera in 135 families from four phyla (Angiosperms, Gymnosperms, Ferns and Mosses), respectively, while its ITS fragments were only successfully amplified in 42.3% of the experiments; and the identification rate of the *psbA-trnH* region was 96.5% at the genus level using the nearest distance method; 72.8%, at the species level. ITS2+*psbA-trnH* was strongly recommended as a core and complementary barcode for a broad series of plant taxa [52].

On the contrast with CBOL (2009) [7], in the BOL project in 2011 on 1,757 species of seed plants in China, the core barcode *rbcL+matK* gave the lowest resolution (49.7%) among other two-locus barcodes, the highest discrimination rate of the three-locus barcode was 81.8% by *matK+trnH-psbA*+ITS. *rbcL+matK* has high species identification power at the species level in just some taxonomic groups (e.g. *Orchidaceae*). The project proposed that the ITS or ITS2 should be incorporated into the core barcode (*rbcL*, *matK*) for seed plants [3]. The combination of *matK*+ITS showed a greater ability to identify species than *matK* or the ITS alone in *Holcoglossum* [14], in *Dendrobium* and *Paphiopedilum* [60].

On the subject *Oncidium* genus, a combination of *trnH-psbA+trnF-ndhJ* was proposed as a potential barcode by correct phylogenetic placement of 13/15 Oncidiinae hybrid varieties [61]. As intergenic spacers were recognized to be high variable, Kim, et al. (2014) [44] also suggested the combination of three intergenic spacers *atpF-atpH+psbK-psbI+trnH-psbA* as the best option for barcoding of the Korean orchid species, the resolution up to 98.8%, among 26 possible combinations of the five regions *rbcL*, *matK*, *atpF-atpH*, *psbK-psbI* and *trnH-psbA*.

In general, the use of combined barcode could give better resolution in most but not all cases depending on taxon specification (Table 2).

Table 2. Summary of studies comparing DNA barcoding regions in plants.

Study	Studied regions	Samples	Amplification (% success)	Sequence separation (%)	Molecular markers recommendation
Kress, et al. (2005) [5]	ITS, *trnH-psbA, atpB-rbcL, psbM-trnD, trnC-ycf6, trnL-F, trnk-rps16, trnV-atpE, rpl36-rps8, ycf6-psbM*	Set 1: 19 species/8 genera/7 families of angiosperm	*trnH-psbA*, *rpl*136-rpf8, *trnL-F* 100% *trn*C-*ycf*6, *ycf*6-*psb*M 90% Other regions = 73-80%	Sequence divergence: - ITS (2.81%) - *trnH-psbA* (1.24%) - *rpl*136-rpf8, *trnL-F* (0.44%) - *atp*B-*rbcL* (0.63%) - *trn*C-*ycf*6 (0.55%)	*trnH-psbA*, ITS
	ITS, *rbcL*, *trnH-psbA*	Set 2: 83 individuals/83 species/72 genera/50 families of angiosperm	*trnH-psbA* = 100%, *rbcL* = 95%, ITS ≤ 88%		*trnH-psbA>>rbcL*

Taberlet, et al. (2006) [84]	*trnL*, P6 loop	more than 100 plant species		*trnL* 67.3% P6 loop 19.5%	
Kress and Erickson (2007) [55]	*trnH-psbA*, *rbcL*, ITS1, *ndhJ*, *matK*, *rpoB*2, *rpoC1*, *ycf5*, *accD* of angiosperms, gymnosperms, ferns, mosses, and liverworts	96 individuals/ 96 species/48 genera/43 families of land plants	*trnH-psbA* = 95.8% *rbcL* = 92.7% *rpoC1* = 83.3% *accD*, *rpoB* ≈ 80% *ndhJ* = 70%, ITS1 = 60.4% *ycf5* = 50% *matK* = 39.3%	*trnH-psbA* (82.6%) ITS (81.5%) *rbcL* (69.8%) Other loci (≤ 70%) *trnH-psbA+rbcL*, *rnH-psbA+rpoB*2, *rnH-psbA+rpoC1* } (85%) Other pairs of two loci (≤ 82.5%)	Two-locus barcode: *trnH-psbA+rbcL*
Chase, et al. (2007) [65]					Three-locus barcode: *rpoC1+rpoB+matK*or *rpoC1+matK+trnH-psbA*
Lahaye, et al. (2008a) [57]	*accD*, *ndhJ*, *matK*, *rbcL*, *trnH-psbA*, *rpoB*, *rpoC1*, *ycf5*	172 individuals/86 species (48 orchid species +38 species from 13 angiosperm families)	All other regions = 95-100% (except *ycf5* and *ndhJ*)	*trnH-psbA*, *matK* } 90.6% *matK+trnH-psbA* (90.9%) Other loci (≤ 87.5%) All barcodes combine (93.1%)	*matK*or *matK+trnH-psbA*
Lahaye, et al. (2008b) [66]	*accD*, *ndhJ*, *matK*, *rbcL*, *trnH-psbA*, *rpoB*, *rpoC1*, *ycf5*, *atpF-atpH*, *psbK-psbI*	101 individuals/38 species			*matK*
Fazekas, et al. (2008) [58]	*cox1*, *matK*, 23S rDNA, *rpoB*, *rpoC1*, *rbcL*, *trnH-psbA*, *atpF-atpH*, *psbK-psbI*	251 individuals/92 species/32 genera of land plants	% sequencing success 23S rDNA, *rbcL* = 100% (2 primer pairs used) *trnH-psbA* = 99% *rpoC1* = 95% (3 primer pairs used) *rpoB* = 92% (5 primer pairs used) *matK* = 88% (10 primer pairs used) *psbK-psbI* = 85% *cox*1 = 72% *atpF-atpH* = 65%	*trnH-psbA* (59%) *matK* (56%) *atpF-atpH*, *psbK-psbI* (45%) *rbcL*, *rpoB* (42-48%) *cox1* (10%) 23S rDNA (7%) *rpoB+rpoC1* (50%) *matK+atpF-atpH+psbK-psbI* (69%) *rbcL+trnH-psbA*, *matK+atpF-atpH* } 64% *rpoB+rpoC1+matK* (61%)	Combinations of 3-4 loci from: *rbcL*, *rpoB*, *matK*, *trnH- psbA*, *atpF-atpH*
CBOL (2009) [7]	*atpF-atpH*, *matK*, *rbcL*, *rpoB*, *rpoC1*, *psbK-psbI*, *trnH-psbA*	907 samples from 550 species genera seed plants	*psbK-psbI* = 77% all others = 90-98%	*rpoC1* (38%), *rpoB* (40%), *atpF-atpH* (50%), *matK* (57%), *rbcL* (58%), *trnH-psbA* (58%) *psbK-psbI* (64%) 2-locus combinations (59-75%) 3-locus combinations (65-76%) All 7 loci combination (73%) *rbcL+matK* (72%)	*rbcL+matK*
Hollingsworth, et al. (2009) [21]	*atpF-atpH*, *matK*, *rbcL*, *rpoB*, *rpoC1*, *psbK-psbI*, *trnH-psbA*				some combination of *rbcL*, *rpoC1*, *matK*, *trnH-psbA*.

Kress, et al. (2009) [64]	*rbcL*, *matK*, *trnH-psbA*	1,035 samples/296 species/181 genera of plants	*rbcL*a = 93% *trnH-psbA* = 94%, but problems sequencing *matK* = 69%	*matK* (99%) *trnH-psbA* (95%) *rbcL*a (75%) *matK*+*rbcL* (92%) *trnH-psbA*+*rbcL* (95%) *rbcL*+*matK*+*trnH-psbA* (98%) When both sequences recovery and correct assignment were taken into account: Of the 286 species *trnH-psbA* (90%) *rbcL*a (70%) *matK* (69%)	*rbcL*+*matK*+*trnH-psbA*
Yao, et al. (2010) [54]	ITS2	50,790 plant and 12,221 animal ITS2 sequences GenBank		Dicotyledons (76.1%) monocotyledons (74.2%) gymnosperms (67.1%) ferns (88.1%) mosses (77.4%) animals (91.7%)	ITS2
Chen, et al. (2010) [52]	*psbA-trnH*, *matK*, *rbcL*, *rpoC1*, *ycf5*, ITS2, ITS	5,905 species/1,010 genera/219 families/7 phyla (Angiosperms, Gymnosperms, Ferns, Mosses, Liver-worts, Algae and Fungi)	*psbA-trnH* = 92.8% ITS2 = 93.8% ITS = 42.3%	At specie level: ITS2 (92.7%) *psbA-trnH* (67.6%) At genus level: ITS2 (99.8%) *psbA-trnH* (> 95%)	ITS2 or ITS2+*psbA-trnH*
BOL (2011) [3]	*rbcL*, *matK*, *trnH-psbA*, ITS	1,757 species/141 genera/75 families/42 orders seed plants	*rbcL* = 94.5%, *matK* = 91.0%, *trnH-psbA* = 90.2%, ITS = 88.0%	ITS (67.2%) ITS2 (54.6%) *rbcL* (26.4%) *trnH-psbA*+ITS (79.1%) *trnH-psbA*+ITS2 (69.7%) *matK*+ITS (75.3%) *matK*+ITS2 (66.1%) *rbcL*+ITS (69.9%) *rbcL*+ITS2 (58.5%) *rbcL*+*matK* (49.7%) *matK*+*trnH-psbA*+ITS (81.8%) *matK*+*trnH-psbA*+ITS2 (75.0%) *rbcL*+*matK*+ITS (77.4%) *rbcL*+*matK*+ITS2 (68.5%) *rbcL*+*matK*+*trnH-psbA* (62.0%) *rbcL*+*matK*+*trnH-psbA*+ITS 82.8% *rbcL*+*matK*+*trnH-psbA* + ITS2 (77.2%)	ITS/ITS2 supplement to core barodes *rbcL*, *matK*
Burgess, et al. (2011) [63]	*rbcL*+*matK*, *rpoC1*, *trnH-psbA*, *atpF-atpH*	2,130 sequences/436 species/269 genera of land plants	*rbcL* = 91.4% *rpoC1* = 74.5%	*matK* (88.8%) *atpF-atpH* (82.4%) *rbcL* (79.8%) *trnH-psbA* (79.3%) *rpoC1* (73.1%) *rbcL*+*matK* (93.1%) *rbcL*+*matK*+*trnH-psbA* (95.3%) Combination 5 loci (97.3%)	*rbcL*+*matK*
Fazekas, et al. (2012) [83]					Core [*matK*+*rbcL*] + supplements (ITS, *trnH-psbA*)
Dong, et al. (2012) [85]	23 loci present in at least three genera				*ycf1-a, trnK, rpl32-trnL, trnH-psbA,* followed by *trnSUGA-trnGUCC, petA-psbJ, rps16-trnQ, ndhC-trnV, ycf1-b, ndhF,*

					rpoB-trnC, psbE-petL, and *rbcL-accD* at genus level
Han, et al. (2013) [23]	ITS , ITS2	Set 1: 91species/5 orders dry medicinal product and herbarium specimens Set 2: 12861 ITS and ITS2 sequences/ 8,313 species/ 8,313 species from 1699 genera, GenBank	ITS2 = 91% ITS = 23%	At the species level: ITS (89.2%) ITS2 (79.2) At the genus level: ITS (97.5%) ITS2 (93.8%)	ITS2
Dong, et al. (2015) [15]	*ycf1* (*ycf1*a, cf1b), *rbcL*, *matK* , *trnH-psbA*	1352 sequences of *matK*, *rbcL* and *ycf1* from 420 species/179 genera/76 plant families.	*rbcL*b = 99.18%, *matK* = 91.43%, *ycf1*b = 85.31%	*ycf1*b (73.97%) *rbcL*b (58.02%) *matK* (57.56%) *rbcL*b+*matK* (71.31%) *ycf1*b+ *rbcL*b (81.39%) *ycf1*b+*matK* (79.83%) *ycf1*b+*rbcL*b+*matK* (86.33%)	*ycf1* (*ycf1*a, *ycf1*b)
		7 relatively well-sampled plant groups		*ycf1*b highest	
Gigot, et al. (2007) [86]	11Mesoamerican orchid species	*trnH-psbA*, *accD*, *rpoC1*, *rpoB*, *matK*, *ndhJ*	All = 100%	*accD* 3/11 (27.3%) *matK* 10/11 (90.9%) *ndhJ* 1/11 (9.1%) *rpoB* 6/11 (54.5%) *rpoC1* 4/11 (36.4%) *trnH-psbA* 8/11 (72.7%) *rpoC1*+*rpoB*+*matK* 11/11 (100%) *rpoC1*+*matK*+*trnH-psbA* 10/11 (90.9%) *rpoB*+*matK*+*trnH-psbA* 10/11 (90.9%)	Combination 2 or 3 of *rpoC1*, *rpoB*, *matK*, *trnH-psbA*
Yao, et al. (2009) [59]	*psbA-trnH*	17 *Dendrobium* species, 1 adulterance	*psbA-trnH* = 100%	Intergenic variation of all species 0.3 to 2.3% Intraspecific variation 0 to 0.1%	*psbA-trnH*
Wu, et al. (2010) [61]	*trnH-psbA*,*matK*, *trnF-ndhJ, ycf1-trnR, accD, rbcL, rpoB, rpoC1*	15 Oncidiinae hybrid varieties	All = 100%	Correct phylogenetic placement of 13/15 varieties	*trnH-psbA*+*trnF-ndhJ*
Parveen, et al. (2012) [6]	ITS, *matK*, *rbcL*, *rpoB*, *rpoC1*	8 species + 3 hybrids *Paphiopedilum*	ITS, *rbcL*, *rpoB*, *rpoC1* = 100% *matK* = 95.23%	*matK* (100%) ITS (50%) *rbcL* (25%) *rpoB*, *rpoC1* (12.5%)	*matK*
Xiang, et al. (2011) [14]	*rbcL*, *matK*, *atpF-atpH*, *psbK-psbI*, *trnH-psbA*, ITS of *Holcoglossum*	12 species	*rbcL* = 100% *matK* = 92.31% ITS = 100% *trnH-psbA* = 100% *atpF-atpH* (low) *psbK-psbI* (low)	*rbcL* lowest *matK* 6/12 ITS 5/12 *trnH-psbA* 5/12 *matK*+ITS 7/12 *matK*+*trnH-psbA* 6/12 ITS+*trnH-psbA* 6/12 *matK*+ITS+*trnH-psbA* 7/12	*matK* or *matK*+ITS/ITS2

Wu, et al. (2012) [20]	ITS	11 *Dendrobium*, 2 adulterant species	ITS = 100%	100%	ITS
Chiang, et al. (2012) [19]	ITS	20 *Dendrobium* species	ITS = 100%		ITS
Singh, et al. (2012) [16]	*matK*, *rbcL*, *rpoB*, *rpoC1*, *trnH-psbA*, ITS	Set 1: 292 individuals/36 species *Dendrobium*	*rpoC1* = 100% *matK* = 99.32% *rpoB* = 99.2% ITS = 98.97% *rbcL* = 96.91%	ITS (100%) *matK* (80.56%) *rpoB* (55.56%) *rbcL* (41.67%) *rpoC1* (38.89%) *matK+rpoB+rpoC1* (94.44%) *matK+rbcL* (86.11 %)	ITS, *matK+rpoB+rpoC1*
		Set 2: 52 species (36 studied species + Genbank)		ITS (100%) *matK* (76.92%) *rpoB* (51.2%) *rpoC1* (42.31%) *rbcL* (38.46%) *matK+rpoB+rpoC1* (92.31 %) *matK+rbcL* (80.77%)	
Siripiyasin, et al. (2012) [62]	*matK*, *rpoB*, *rpoC1*, *trnH-psbA*	19 species *Cymbidium* Thailand	All = 100%	All 100% sepcies resolution	*trnH-psbA+matK+rpoC1* *trnH-psbA+matK+rpoB*
Yukawa, et al. (2013) [26]	4 genetic units in *Grammatophyllumspeciosum* complex	ITS	All = 100%	Discriminate 4 different species of *Grammatophyllum*	ITS
Kim, et al. (2014) [44]	*rbcL*, *matK*, *atpF-atpH*, *psbK-psbI* and *trnH-psbA*	89 species of Orchidaceae	All = 100%	*trnH-psbA* (83.5%) *rbcL* (60.5%) *atpF-atpH+ psbK-psbI+trnH-psbA* (98.8%)	*atpF-atpH+psbK-psbI+trnH-psbA*
Feng, et al. (2015) [51]	ITS2	Set 1: 64 species from *Dendrobium*	ITS2 = 100%	85.9% (by BLAST1 method), 82.8% (by nearest genetic distance method)	ITS2
	trnN-rpl32, petN-psbM, petA-psbJ, trnF-ndhJ, trnE-trnT, accD-psaI, rps15-ycf1, psbA-trnK, atpF	2 species *Phalaenopsisaphrodite* subsp. *Formosana*and *P. amabilis*	All = 100%		*trnN-rpl32, petN-psbM, petA-psbJ, trnF-ndhJ, trnE-trnT, accD-psaI, rps15-ycf1, psbA-trnK, atpF*
Lin, et al. (2015) [75]	*petN-psbM, petA-psbJ, trnT-psbD, trnF-ndhJ, trnN-rpl32, rps16-trnQ, rps16*	19 moth orchids species	All = 100%	petN-psbM 16/19 *petA-psbJ* 16/19 *trnT-psbD* 16/19 *trnF-ndhJ* 16/19 *trnN-rpl32* 16/19 *rps16-trnQ* 19/19 *rps16* 15/19 *trnL* 18/19	*rps16-trnQ*
Kim, et al. (2015) [74]	*rpoC2*, *atpF-atpH*	4 species of *Cypripedium*	*rpoC2* = 100% *atpF-atpH* = 100%	100%	*atpF-atpH*
Xu, et al. (2015) [60]	ITS, ITS2, *matK*, *rbcL*, *trnH-psbA*	Set 1: 184 species *Dendrobium*	All = 100%	ITS (31.93%) ITS2 (22.29%) *matK* (10.48%) *trnH-psbA* (8.14%) *rbcL* (5.56%) ITS+*matK* (76.92%) ITS+*matK+trnH-psbA* (73.13%) ITS2+*matK* (64.84%) *matK+rbcL* (24%)	ITS+*matK*

Some measurements for evaluating effects of molecular markers

Different metrics to evaluate the molecular markers are usually discussed in reference studies. It often suggests that the sequence lengths should be short enough (400-800 bp) for DNA extraction, amplification, and sequencing, but certainly must be long enough to contain sufficient information for sequence divergence [14, 57]. The sequence should possess conserved flanking sites for developing universal PCR primers [55] but routinely retrievable with a single primer pair [7]. Easy alignment is also one of the considered criteria [57] although in the situation of *trnH-psbA* and some other intergenic spacers which are known as so variable that hard to align, the difficulty of alignment is not a major obstacle when comparing with their advantages of variation sites [5].

The most concerned factor to identifying loci is good discriminatory power [3, 14, 16, 62]. This power is either based on sequence divergence or variability [5, 55, 57]. The potential parsimony-informative characters or known as nucleotide substitutions are the ones that much contribute to the divergence between sequences [56, 74, 83]. The one with the most features used to measure significant species-level genetic variability and divergence is "DNA barcoding gap", which is presented between intra- and inter-specific variations. High interspecific, but low intraspecific divergence, are expected to achieve maximal species discrimination sequencing [6, 14, 16, 20, 26, 51, 54, 57, 59, 60, 62, 86]. Indel fragments (insertions and deletions) also contain much useful information for identification work [20], e.g. it can help to distinguish three species of genus *Solidago* despite low sequence divergence [5]. Nucleotide substitutions account for about 70% and indels account for about 30% of all mutations in the chloroplast genome [56]. However, this information is still not used effectively for available bioinformatics tools. The relative amounts of indels need to be further tested [56, 74]. Bioinformatics tools for barcoding should be developed to use indel information [5].

Some studies are also concerned about the GC content of the sequences [20, 54]. GC content is found to be variable with different organisms. DNA with high GC-content is more stable than DNA with low GC-content (Mega net/help). Sequence length variation is also a helpful feature in some cases [54, 74, 75]. Kim, et al. (2015) even suggests using electrophoresis to identify species based on length variation [74]. PCR-based (multiplex and ARMS) method which determine specific SNPs has been used in analyses of sequence taxonomy [19, 74]. The secondary structure of the ITS2 region could provide useful information for species identification and could be considered as a molecular morphological characteristic [54].

To achieve optimal effect for barcoding, the utilization of many different pieces information is necessarily, such as barcode setting gap, length variation, indel variation... DNA barcodes can be very effective in the context of a clearly circumscribed floristic sample or plant community, and that additional data, such as geography and morphology may be required to obtain higher rates of species identification in other contexts [64].

Conclusions

Since the last classification of *Orchidaceae* in 2003, there has been major progress in the determination of relationships, despite that almost all of the problematic placements recognised in the previous classification 11 years ago have now been resolved by molecular methods [9]. However barcoding for the identification of plants, as well as orchid species, still faces many problems and needs improvement. These improvements are now continued to achieve by different ways. New DNA regions more potential and suitable that can overcome the available limits are ongoing investigated. Complete sequenced DNA genomes are used as references to screen for a new locus, from plastid genome, to mitochondrial genome and to nucleus genome.

The combinations of multi-locus barcodes are now highly considered as one of the improvement solutions to obtain the best resolution results. Many factors should be cared about such as how many and which ones would be combined. The final aim is to both maximise the loci to get the best efficiency, and minimise the loci to decrease cost and time (e.g. selection of a 2-locus barcode is based on costs and can prevent further delays in implementing a standard barcode for land plants) [7]. The selections of combining loci depend on the characteristics of each locus. No single locus has shown high levels of universality and resolvability [21], and no single barcoding region has an ability to resolve species to the same degree as nearly any of the multilocus barcoding methods [58]. The combination may include a phylogenetically conservative coding locus (*rbcL*) with one or more rapidly evolving regions (part of the *matK* gene and the intergenic spacer *trnH-psbA*). Chloroplast genome sequences contain regions that are highly variable, and this variability of chloroplast genes differs markedly among genera [75, 85]. However, primers designing for these intergenic regions are the challenges with barcodes which, if based on the single or limited locus, would be specific taxa. So Singh, et al. (2012) [16] have recommended the use of whole chloroplast genome as single locus barcode in future will help. Improve the effective of available potential markers which low amplification rate with specific primers was one of the choices, e.g. *matK* much higher levels of sequence variation and so possess high ability of spices discrimination, but need improve PCR primers [57]; in some cases, single- or multiple-primer sets are necessary [3, 87].

To accurately determine the relationship between either the species or higher taxonomic level,

the molecular markers used should be able to clearly separate studied taxa at first. Then the barcoding markers are closely related to phylogenetic markers. Developing resolution of molecular for authentication of taxa means developing reliability of phylogenetic study; and conversely, many barcoding studies used the phylogenetic tree as one of the metrics to measure the discrimination ability of the molecular regions. As our ranking based on a small statistic from about 50 phylogenetic references in this research, the most used locus is the ITS (80%), following by *matK* (46%), *trnL-F* (28%), *rbcL* (24%), *trnL* (20%), *trnH-psbA* (14%), *ycf1* (8%), *Xdh*, *trnS-G*, *trnK*, *atpI-atpH* (6%), and some other regions (lower than 5%).

REFERENCES

[1] M.W. Chase, K.M. Cameron, J.V. Freudenstein, A.M. Pridgeon, G. Salazar, C. Van Den Berg, A. Schuiteman (2015), "An updated classification of *Orchidaceae*", *Botanical Journal of the Linnean Society*, **177(2)**, pp.151-174.

[2] X.G. Xiang, D.Z. Li, W.T. Jin, H.L. Zhou, J.W. Li, et al. (2012), "Phylogenetic placement of the enigmatic orchid genera *Thaia* and *Tangtsinia*: Evidence from molecular and morphological characters", *Taxon*, **61(1)**, pp.45-54.

[3] BOL China Plant Group (2011), "Comparative analysis of a large dataset indicates that ITS should be incorporated into the core barcode for seed plants", *Proceedings of the National Academy of Sciences of the United States of America*, **108**, pp.19641-19646.

[4] H.M. Kim, S.H. Oh, G.S. Bhandari, C.S. Kim, C.W. Park (2014), "DNA barcoding of *Orchidaceae* in Korea", *Mol. Ecol. Resour.*, **14(3)**, pp.499-507.

[5] W.J. Kress, K.J. Wurdack, E.A. Zimmer, L.A. Weigt, D.H. Janzen (2005), "Use of DNA barcodes to identify flowering plants", *Proceedings of the National Academy of Sciences of the United States of America*, **102**, pp.8369-8374.

[6] I. Parveen, H.K. Singh, S. Raghuvanshi, U.C. Pradhan, S.B. Babbar (2012), "DNA barcoding of endangered Indian *Paphiopedilum* species", *Mol. Ecol. Resour.*, **12(1)**, pp.82-90.

[7] CBOL Plant Working Group (2009), "A DNA barcode for land plants", *Proc. Natl. Acad. Sci. U.S.A.*, **106(31)**, pp.12794-12797.

[8] P.D.N. Hebert, A. Cywinska, S.L. Ball, J.R. De Waard (2003), "Biological identifications through DNA barcodes", *Proc. Biol. Sci.*, **270(1512)**, pp.313-321.

[9] V.S. Shneyer (2009), "DNA barcoding is a new approach in comparative genomics of plants", *Russ. J. Genetics.*, **45(11)**, pp.1267-1278.

[10] K.H. Wolfe, W.H. Li, P.M. Sharp (1987), "Rates of Nucleotide Substitution Vary Greatly among Plant Mitochondrial, Chloroplast, and Nuclear DNAs", *Proc. Natl. Acad. Sci.*, **84(24)**, pp.9054-9058.

[11] G. Drouin, H. Daoud, J. Xia (2008), "Relative Rates of Synonymous Substitutions in the Mitochondrial, Chloroplast and Nuclear Genomes of Seed Plants", *Mol. Phylogenet. Evol.*, **49(3)**, pp.827-831.

[12] Y. Cho, Y.L. Qiu, P. Kuhlman, J.D. Palmer (1998), "Explosive Invasion of Plant Mitochondria by a Group I Intron", *Proc. Natl. Acad. Sci. USA*, **95(24)**, pp.14244-14249.

[13] M.E. Mort, J.K. Archibald, C.P. Randle, et al. (2007), "Inferring Phylogeny at Low Taxonomic Levels: Utility of Rapidly Evolving cpDNA and Nuclear ITS Loci", *Am. J. Bot.*, **94(2)**, pp.173-183.

[14] X.G. Xiang, H. Hu, W. Wang, X.H. Jin (2011), "DNA barcoding of the recently evolved genus Holcoglossum (*Orchidaceae*: Aeridinae): A test of DNA barcode candidates", *Molecular Ecology Resources*, **11(6)**, pp.1012-1021.

[15] W. Dong, C. Xu, C. Li, J. Sun, Y. Zuo, S. Shi, T. Cheng, J. Guo, S. Zhou (2015), "*ycf1*, the most promising plastid DNA barcode of land plants", *Scientific Reports*, **5**, p.8348.

[16] H.K. Singh, I. Parveen, S. Raghuvanshi, S.B. Babbar (2012), "The loci recommended as universal barcodes for plants on the basis of floristic studies may not work with congeneric species as exemplified by DNA barcoding of *Dendrobium* species", *BMC Research Notes*, **5**, p.42.

[17] M.W. Chase, M.F. Fay (2009), "Barcoding of plants and fungi", *Science*, **325(5941)**, pp.682-683.

[18] M.Z. Huang (2010), *DNA Barcoding of Hainan Island Orchid Species Based on matK and ITS Sequences*, http://mt.china-papers.com/1/?p=185765.

[19] C.H. Chiang, T.A. Yu, S.F. Lo, C.L. Kuo, W. Peng, H.S. Tsay (2012), "Molecular authentication of *Dendrobium* species by multiplex polymerase chain reaction and amplification refractory mutation system analysis", *J. Am. Soc. Hortic. Sci.*, **137(6)**, pp.438-444.

[20] C.T. Wu, S.K. Gupta, A.Z.M. Wang, S.F. Lo, C.L. Kuo, Y.J. Ko, C.L. Chen, C.C. Hsien, H.S. Tsay (2012), "Internal transcribed spacer sequence based identification and phylogenic relationship of herba *Dendrobii*", *J. Food Drug Anal.*, **20(1)**, pp.143-151.

[21] M.L. Hollingsworth, A. Clark, L.L. Foerrest, et al. (2009), "Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants", *Molecular Ecology Resources*, **9(2)**, pp.439-457.

[22] I. Alvarez, J.F. Wendel (2003), "Ribosomal ITS sequences and plant phylogenetic inference", *Mol. Phylogenet Evol.*, **29(3)**, pp.417-434.

[23] J.P. Han, Y. Zhu, X. Chen, et al. (2013), "The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS", *BioMed Research International*, pp.1-7.

[24] C.L. Morrison, K. Hovatter, M. Eackles, A.P. Spidle, T.L. King (2005), "Molecular identification of Cypripedioid orchids in international trade", *Selbyana*, **26(12)**, pp.196-216.

[25] H. Hao (2010), *DNA Barcoding and Molecular Phylogeny of Holcoglossum Schltr. (Orchidaceae)*, http://www.dissertationtopic.net/doc/364760.

[26] T. Yukawa, A. Kinoshita, N. Tanaka (2013), "Molecular identification resolves taxonomic confusion in *Grammatophyllum speciosum* complex (*Orchidaceae*)", *Bulletin of the National Museum of Natature and Science*, **39(3)**, pp.137-145.

[27] S. Bernados, M.A. Santos, D. Tyteca, F. Amich (2006), "Phylogenetic relationships of Mediterranean Neottieae and Orchideae (*Orchidaceae*) inferred from nuclear ribosomal ITS sequences", *Acta Botanica Gallica*, **153(2)**, pp.153-165.

[28] J.M. Burke, M.J. Bayly, P.B. Adams, P.Y. Ladiges (2008), "Molecular phylogenetic analysis of *Dendrobium* (*Orchidaceae*), with emphasis on the Australian section Dendrocoryne, and implications for generic classification", *Aust. Syst. Bot.*, **21(1)**, pp.1-14.

[29] D.L. Szlachetko, P. Tukałło, J. Mytnik-Ejsmont, E. Grochocka (2013), "Reclassification of the Angraecum-alliance (*Orchidaceae*, Vandoideae) based on molecular and morfological data", *Biodiv. Res. Conserv.*, **29**, pp.1-23.

[30] G. Sramkó, V.A. Molnár, J.A. Hawkins, R.M. Bateman (2014), "Molecular phylogeny and evolutionary history of the Eurasiatic orchid genus *Himantoglossum* s.l. (*Orchidaceae*)", *Ann. Bot. (Oxford)*, **114(8)**, pp.1609-1626.

[31] K. Srikulnath, S. Sawasdichai, T.K. Jantapanon, P. Pongtongkam, S. Peyachoknagul (2015), "Phylogenetic Relationship of *Dendrobium* Species in Thailand Inferred from Chloroplast *matK* Gene and Nuclear rDNA ITS Region", *The Horticulture Journal*, **84(3)**, pp.243-252.

[32] K.H. Trung, T.D. Khanh, L.H. Ham, T.D. Duong, N.T. Khoa (2013), "Molecular Phylogeny of the Endangered Vietnamese Paphiopedilum Species Based on the Internal Transcribed Spacer of the Nuclear Ribosomal DNA", *Advanced Studies in Biology*, **5(7)**, pp.337-346.

[33] J.A.N. Batista, K.S. Borges, M.W.F. De Faria, K. Proite, A.J. Ramalho, et al. (2013), "Molecular phylogenetics of the species-rich genus Habenaria (*Orchidaceae*) in the New World based on nuclear and plastid DNA sequences", *Molecular Phylogenetics and Evolution*, **67(1)**, pp.95-109.

[34] X.G. Xiang, A. Schuiteman, D.Z. Li, W.C. Huang, S.W. Chung, J.W. Li, H.L. Zhou, W.T. Jin, Y.J. Lai, Z.Y. Li, et al. (2013), "Molecular systematics of *Dendrobium* (*Orchidaceae*, Dendrobieae) from mainland Asia based on plastid and nuclear sequences", *Mol. Phylogenet., Evol.*, **69(3)**, pp.950-960.

[35] W.T. Jin, X.H. Jin, A. Schuiteman, D.Z. Li, X.G. Xiang, W.C. Huang, J.W. Li, L.Q. Huang (2014), "Molecular systematics of subtribe Orchidinae and Asian taxa of Habenariinae (Orchideae, *Orchidaceae*) based on plastid *matK*, *rbcL* and nuclear ITS", *Mol. Phylogenet. Evol.*, **77**, pp.41-53.

[36] W.M. Whitten, M.A. Blanco, N.H. Williams, S. Koehler, G. Carnevali, R.B. Singer, L. Endara, K.M. Neubig (2007), "Molecular phylogenetics of *Maxillaria* and related genera (*Orchidaceae*: Cymbidieae) based on combined molecular data sets", *Amer. J. Bot.*, **94(11)**, pp.1860-1889.

[37] M. Novello, M. Beyer, E.A. Veasey, S. Koehler (2013), *Molecular phylogeny of the brazilian Atlantic forest orchid genus* Brasiliorchis, https://atbc.confex.com/atbc/2013/webprogram/Paper2086.

[38] F. Martos, S.D. Johnson, C.I. Peter, B. Bytebier (2014), "A molecular phylogeny reveals paraphyly of the large genus *Eulophia* (*Orchidaceae*): A case for the reinstatement of *Orthochilus*", *International Association for Plant Taxonomy*, **63(1)**, pp.9-23.

[39] M.H. Li, G.Q. Zhang, Z.J. Liu, S.R. Lan (2014), "Revision of *Hygrochilus* (*Orchidaceae*: Epidendroideae: Aeridinae) and a molecular phylogenetic analysis", *Phytotaxa*, **159(4)**, pp.256-268.

[40] Y. Tang, T. Yukawa, R.M. Bateman, H. Jiang, H. Peng (2015), "Phylogeny and classification of the East Asian *Amitostigma* alliance (*Orchidaceae*: Orchideae) based on six DNA markers", *BMC Evolutionary Biology*, **15**, p.96.

[41] M. Soliva, A. Kocyan, A. Widmer (2001), "Molecular phylogenetics of the sexually deceptive orchid genus *Ophrys* (*Orchidaceae*) based on nuclear and chloroplast DNA sequences", *Mol. Phylogenet. Evol. In press*, **20(1)**, pp.78-88.

[42] G.A. Fischer, B. Gravendeel, A. Sieder, J. Andriantiana, P. Heiselmayer, P.J. Cribb, E.C. Smidt, R.

Samuel, M. Kiehn (2007), "Evolution of resupination of Madagascan species of *Bulbophyllum* (*Orchidaceae*)", *Molecular Phylogenetics and Evolution*, **45(1)**, pp.358-376.

[43] A. Chochai, I.J. Leitch, M.J. Ingrouille, M.F. Fay (2012), "Molecular phylogenetics of *Paphiopedilum* (Cypripedioideae, *Orchidaceae*) based on nuclear ribosomal ITS and plastid sequences", *Botanical Journal of the Linnean Society*, **170(2)**, pp.176-196.

[44] H.M. Kim, S.H. Oh, G.S. Bhandari, C.S. Kim, C.W. Park (2014), "DNA barcoding of Orchidaceae in Korea", *Mol. Ecol. Resour.*, **14(3)**, pp.499-507.

[45] W.M. Whitten, N.H. Williams, M.W. Chase (2000), "Subtribal and generic relationship of Maxillarieae (*Orchidaceae*) with emphasis on Stanhopeinae: Eombined molecular evidence", *American Journal of Botany*, **87(12)**, pp.1842-1856.

[46] H. Topik, P.H. Weston, T. Yukawa, M. Ito (2012), "Phylogeny of subtribe aeridinae (*Orchidaceae*) inferred from DNA sequences data: Advanced analyses including Australasian genera", *Jurnal Teknologi* (Sci. Eng.), **59(1)**, pp.87-95.

[47] M. Moudi, R. Go (2015), "Monophyly of four sections of genus *Dendrobium* (*Orchidaceae*): Evidence from nuclear ribosomal DNA internal transcribe spacer (ITS) sequences", *International Journal of Bioassays*, **4(1)**, pp.3622-3626.

[48] M. Moudi, C.S.Y. Yong, M.N. Saleh, J.O. Abdullah, R. Go (2013), "Phylogenetic analysis among four sections of genus *Dendrobium* Sw. (*Orchidaceae*) in Peninsular Malaysia using *rbcL* sequence data", *International Journal of Bioassays*, **2(6)**, pp.932-937.

[49] L. Li, D.P. Ye, M. Niu, H.F. Yan, T.L. Wen, S.J. Li (2015), "Thuniopsis: A New Orchid Genus and Phylogeny of the Tribe Arethuseae (*Orchidaceae*)", *PLoS ONE*, **10(8)**, e0132777.

[50] S.K. Sharma, J. Dkhar, S. Kumaria, P. Tandon, R.S. Rao (2012), "Assessment of phylogenetic inter-relationships in the genus *Cymbidium* (*Orchidaceae*) based on internal transcribed spacer region of Rdna", *Gene*, **495(1)**, pp.10-15.

[51] S. Feng, Y. Jiang, S. Wang, M. Jiang, Z. Chen, Q. Ying, H. Wang (2015), "Molecular Identification of *Dendrobium* Species (*Orchidaceae*) Based on the DNA Barcode ITS2 Region and Its Application for Phylogenetic Study", *Int. J. Mol. Sci.*, **16(9)**, pp.21975-21988.

[52] S.L. Chen, H. Yao, J.P. Han, et al. (2010), "Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species", *PLoS ONE*, **5(1)**, e8613.

[53] D.T. Lau, P.C. Shaw, J. Wang, P.P. But (2001), "Authentication of medicinal *Dendrobium* species by the internal transcribed spacer of ribosomal DNA", *Planta Med.*, **67(5)**, pp.456-460.

[54] H. Yao, J. Song, C. Liu, et al. (2010), "Use of ITS2 region as the universal DNA barcode for plants and animals", *PLoS ONE*, **5(10)**, e13102.

[55] W.J. Kress, D.L. Erickson (2007), "A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region", *PLoS One*, **2(6)**, e508.

[56] J. Shaw, E.B. Lickey, E.E. Schilling, R.L. Small (2007), "Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetics studies in angiosperms: The tortoise and the hare III", *American Journal of Botany*, **94(3)**, pp.275-288.

[57] R. Lahaye, M. Van Der Bank, D. Bogarin, et al. (2008a) "DNA barcoding the floras of biodiversity hotspots", *Proceedings of the National Academy of Sciences*, **105(8)**, pp.2923-2928.

[58] A.J. Fazekas, K.S. Burgess, P.R. Kesanakurti, et al. (2008), "Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well", *PLoS ONE*, **3(7)**, e2802.

[59] H. Yao, J.Y. Song, X.Y. Ma, et al. (2009), "Identification of *Dendrobium* species by a candidate DNA barcode sequence: The chloroplast *psbA-trnH* intergenic region", *Planta Medica*, **75(6)**, pp.667-669.

[60] S. Xu, D. Li, J. Li, X. Xiang, W. Jin, W. Huang, X. Jin, L. Huang (2015), "Evaluation of the DNA barcodes in *Dendrobium* (*Orchidaceae*) from mainland Asia", *PLoS One*, **10(1)**, e0115168.

[61] F.H. Wu, M.T. Chan, D.C. Liao, C.T. Hsu, Y.W. Lee, et al. (2010), "Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae", *BMC Plant Biol.*, **10**, pp.68-80.

[62] P. Siripiyasing, K. Kaenratana, P. Mokkamul, T. Tanee, R. Sudmoon, A. Chaveerach (2012), "DNA barcoding of *Cymbidium* species (*Orchidaceae*) in Thailand", *Afr. J. Agric. Res.*, **7(3)**, pp.393-400.

[63] K.S. Burgess, A.J. Fazekas, P.R. Kesanakurti, et al. (2011), "Discriminating plant species in a local temperate flora using the *rbcL+ matK* DNA barcode", *Methods in Ecology and Evolution*, **2(4)**, pp.333-340.

[64] W.J. Kress, D.L. Erickson, F.A. Jones, et al. (2009), "Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama", *Proceedings of the National Academy of Sciences*, **106(44)**, pp.18621-18626.

[65] M.W. Chase, R.S. Cowan, P.M. Hollingsworth, et al. (2007), "A proposal for a standardised protocol to barcode all land plants", *Taxon*, **56(2)**, pp.295-299.

[66] R. Lahaye, V. Savolainen, S. Duthoit, O. Maurin, M. Van Der Bank (2008b), "A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system", *Nature Proceedings*, pp.1-23.

[67] V. Sosa, T. Mejía-Saules, M.A. Cuéllar, A.P. Vovides (2013), "DNA Barcoding in Endangered Mesoamerican Groups of Plants", *Botanical Review*, **79(4)**, pp.469-482.

[68] H. Asahina, J. Shinozaki, K. Masuda, Y. Morimitsu, M. Satake (2010), "Identification of medicinal *Dendrobium* species by phylogenetic analyses using *matK* and *rbcL* sequences", *J. Nat. Med.*, **64(2)**, pp.133-138.

[69] P. Cuénoud, V. Savolainen, L.W. Chatrou, M. Powell, R.J. Grayer, M.W. Chase (2002), "Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB* and *matK* DNA sequences", *Am. J. Bot.*, **89(1)**, pp.132-144.

[70] M. Von Crautlein, H. Korpelainen, M. Pietilainen, J. Rikkinen (2011), "DNA barcoding: A tool for improved taxon identification and detection of species diversity", *Biodiversity Conservation*, **20(2)**, pp.373-380.

[71] K.M. Neubig, W.M. Whitten, B.S. Carlsward, et al. (2009), "Phylogenetic utility of ycf1 in orchids: a plastid gene more variable than *matK*", *Plant Systematics and Evolution*, **277(1)**, pp.75-84.

[72] R. Arévalo, K. Cameron (2013), "Molecular Phylogenetics of Mormolyca (*Orchidaceae*: Maxillariinae) based on Combined Molecular Data Sets", *Lankesteriana,* **13(1-2)**, pp.1-11.

[73] Y.Y. Guo, L.Q. Huang, Z.J. Liu, X.Q. Wang (2016), "Promise and challenge of DNA barcoding in Venus Slipper (*Paphiopedilum*)", *PLoS ONE*, **11(1)**, e0146880.

[74] J.S. Kim, H.T. Kim, S.W. Son, J.H. Kim (2015), "Molecular identification of endangered Korean lady's slipper orchids (*Cypripedium*, *Orchidaceae*) and related taxa", *Botany*, **93(9)**, pp.603-610.

[75] C.S. Lin, J.J.W. Chen, Y.T. Huang, M.T. Chan, H. Daniell, W.J. Chang, C.T. Hsu, D.C. Liao, F.H. Wu, S.Y. Lin, C.F. Liao, M.K. Deyholos, G.K.S. Wong, V.A. Albert, M.L. Chou, C.Y. Chen, M.C. Shih (2015), "The location and translocation of *ndh* genes of chloroplast origin in the *Orchidaceae* family", *Scientific Reports,* **5**, 10p, doi: 10.1038/srep 09040.

[76] B.S. Carlsward, W.M. Whitten, N.H. Williams (2003), "Molecular phylogenetics of Neotropical leafless Angraecinae (*Orchidaceae*): Reevaluation of generic concepts", *Int. J. Pl. Sci.*, **164(1)**, pp.43-51.

[77] C. Van den Berg, D.H. Goldman, J.V. Freudenstein, A.M. Pridgeon, K.M. Cameron, et al. (2005), "An overview of the phylogenetic relationships within Epidendroideae inferred from multiple DNA regions and recircumscription of Epidendreae and Arethuseae (*Orchidaceae*)", *American Journal of Botany*, **92(4)**, pp.613-624.

[78] B.S. Carlsward, W.M. Whitten, N.H. Williams, B. Bytebier (2006), "Molecular phylogenetics of Vandeae (*Orchidaceae*) and the evolution of leaflessness", *Amer. J. Bot.*, **93(5)**, pp.770-786.

[79] C. Micheneau, B. Carlsward, M. Fay, B. Bytebier, T. Pailler, M. Chase (2008), "Phylogenetics and biogeography of Mascarene angraecoid orchids (Vandeae, *Orchidaceae*)", *Molecular Phylogenetics and Evolution*, **46(3)**, pp.908-922.

[80] R.J. Waterman, A. Pauw, T.G. Barraclough, V. Savolainen (2009), "Pollinators underestimated: a molecular phylogeny reveals widespread floral convergence in oil-secreting orchids (sub-tribe Coryciinae) of the Cape of South Africa", *Mol. Phylogenet E.*, **51(1)**, pp.100-110.

[81] J.H. Li, Z.J. Liu, G.A. Salazar, P. Bernhardt, H. Perner, Y. Tomohisa, X.H. Jin, S.W. Chung, Y.B. Luo (2011), "Molecular phylogeny of Cypripedium (*Orchidaceae*: Cypripedioideae) inferred from multiple nuclear and chloroplast regions", *Mol. Phylogenet. E.*, **61(2)**, pp.308-320.

[82] J.B. Yang, M. Tang, H.T. Li, D.Z. Zhang (2013), "Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analyses", *BMC Evol. Biol.*, **13**, pp.84-96.

[83] A.J. Fazekas, M. Kuzmina, G. Newmaster Steven, P.M. Hollingsworth (2012), "DNA barcoding methods for land plants", *Methods in Molecular Biology*, **858**, pp.223-252.

[84] P. Taberlet, E. Coissac, F. Pompanon, L. Gielly, C. Miquel, A. Valentini, T. Vermat, G. Corthier, C. Brochmann, E. Willerslev (2006), "Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding", *Nucleic Acids Res.*, **35(3)**, el4.

[85] W. Dong, J. Liu, J. Yu, L. Wang, S. Zhou (2012), "Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding", *PloS ONE*, **7(4)**, e35071.

[86] G. Gigot, J. Van Alphen-Stahl, D. Bogarin, J. Warner, M.W. Chase, et al. (2007), "Finding a suitable DNA barcode for mesoamerican orchids", *Lankesteriana*, **7(1-2)**, pp.200-203.

[87] C.S. Ford, K.L. Ayres, N. Toomey, N. Haider, J. Van Alphen Stahl, L.J. Kelly, N. Wikstrom, P.M. Hollingsworth, R.J. Duff, S.B. Hoot, R.S. Cowan, M.W. Chase, M.J. Wilkinson (2009), "Selection of candidate coding DNA barcoding regions for use on land plants", *Botanical Journal of the Linnean Society*, **159(1)**, pp.1-11.