Introduction

Painting by: Ms. Hemalata Pradhan

Pleione praecox
1 INTRODUCTION

DNA barcoding is an innovative molecular technique, which uses short and agreed upon DNA sequence(s) from either nuclear or cytoplasmic genome for rapid identification of biological specimen at species level. Based on their study on 200 allied Lepidopteran species, Hebert et al. (2003a) were the first to propose the use of short DNA sequences as “Taxon Barcodes” for species level identification. The DNA sequence which was found to be applicable for barcoding animal species in their pioneering and subsequent studies was the "Folmer region" at the 5' end of cytochrome c oxidase 1 (CO1), present in mitochondrial genome (Hebert et al. 2003a, b, 2004a, b). Based on this initial success with animals, this was projected as the locus that could provide recognition tags to all organisms. Hebert et al. (2003 b) argued that just 15 variable sites in CO1 could provide one billion combinations (4^{15}) of bases giving more than enough possible barcode ‘patterns’ for estimated 10 million organisms. Subsequently, the CO1 region was successfully tested in other animal groups (Barrett and Hebert 2005, Cywinska et al. 2006, Clare et al. 2007). Realizing the importance of this technology, an international “Consortium for the Barcode of Life” (CBOL) was established in 2004. It comprises more than 170 member organizations from more than 45 countries (http://barcoding.si.edu). CBOL’s mission is to promote the exploration and development of DNA Barcoding as global standard for species identification, through rapid compilation of high quality DNA barcodes in a public library of DNA sequences (http://barcoding.si.edu). Another organization, International Barcode of Life (iBOL), based at Guelph, Ontario in fact, is involved in barcoding of different group of plants and animals (http://ibol.org). Its aim is to generate DNA Barcodes for 5 million specimens and 500,000 estimated species existing on planet earth by 2015 (http://ibol.org). DNA barcoding besides helping taxonomists in rapid identification of new/cryptic and polymorphic species (Lahaye et al. 2008, Miwa et al. 2009, Xiao et al. 2010) is also a powerful diagnostic tool in hands of enforcement agencies for checking illegal trade of endangered species of both animals and plants i.e., biopiracy (Eaton et al. 2010, Jeanson et al. 2011, Muellner et al. 2011, Yesson et al. 2011) and an investigative tool for forensic
specialists (Coyle et al. 2005, Ferri et al. 2009). Moreover, it also helps in (i) identifying invasive species (Bleeker et al. 2008, van de Wiel et al. 2009) right at quarantine stage, (ii) plant identification at any stage of life cycle [juvenile or mature] (Gonzalez et al. 2009), (iii) authentication of herbal medicine and their adulterants (Yao et al. 2009, Chen et al. 2010, Asahina et al. 2010, Srirama et al. 2010), (iv) identifying complex food webs by studying species diversity in gut contents of animals (Soininen et al. 2009), (v) analyzing herbivore’s diet components (Valentini et al. 2009), (vi) checking adulterations and substitutions in food products (Jaakola et al. 2010) and (vii) determining the constituent plant species in different honey samples (Valentini et al. 2010). As opposed to the current taxonomic methods which require whole plant preferably in flowering stage for its authentic identification, DNA barcodes once standardized can identify the species even if a minute amount of tissue/fragment is available. DNA barcodes could also act as genetic resource tags and in turn would of help in conservation of genetic diversity (Eaton et al. 2010, Jeanson et al. 2011, Muellner et al. 2011, Yesson et al. 2011). DNA barcoding could also hasten the process of biodiversity inventorization and analysis (Costion et al. 2011) using still smaller fragments of DNA aptly called as minibarcodes (Meusnier et al. 2008). DNA barcoding could also play a significant role in forest biosecurity and bioserveillance of habitats by identifying non-indigenous species from the native species (Armstrong and Ball 2005, Humble and deWaard 2010). Palaeobarcoding, a term coined for DNA barcoding, is used for studying the effect of climate change on biotic diversity. To meet this objective, DNA in permafrost or sedimentary DNA (seda-DNA) is analyzed using metabarcoding (circumventing the need of cells or tissues) technique and correlating it with the known temporal climatic changes that had taken place (Jørgensen et al. 2011, Andersen et al. 2011).

Although, in animals, the mitochondrial gene region ‘CO1’ with the requisite divergence and universality was found to be suitable for species distinction, in plants no such region of genome, cytoplasmic or nuclear, could be identified. Rather, its search was compared with that for the “Holy Grail” (Rubinoff et al. 2006). The plant mitochondrial genes with low nucleotide substitutions and low evolutionary rates were considered unsuitable for barcodes of plants (Chase et al. 2005, Kress et al.
2005, Newmaster et al. 2006). Therefore, nuclear and plastid genes have been the prime focus of research for the identification of the locus/loci which could become species level molecular signatures for the plants. A number of plastid genes have been tested and proposed as probable plant barcodes by different groups. Based on their comparison of the complete plastid genomes of *Nicotiana tabacum* and *Atropa belladonna* followed by determination of raw divergence levels across all genes, introns and intergenic spacers, Kress et al. (2005) identified nine intergenic spacers (*trnK-rps16, trnH-psbA, rp136-rps8, atpB-rbcL, ycf6-psbM, trnV-atpE, trnC-ycf6, psbM-trnD and trnL-F*), which were found to be the most variable regions and thus met the barcode criteria. They proposed the use of ITS (the internal transcribed spacer) from the nuclear genome and *trnH-psbA* spacer from the chloroplast genome as DNA barcodes for flowering plants (Kress et al. 2005). The other loci which were tested for DNA barcoding of plants with some success are *rbcL* (Newmaster et al. 2006, Kress and Erickson 2007), *rpoB*, *rpoC1* (Newmaster et al. 2006, Chase et al. 2005, Sass et al. 2007, Seberg and Petersen 2009) and *matK* (Lahaye et al. 2008). Taberlet et al. (2007) proposed that the *trnL* (UUU) intron and its shorter P6 loop (10-143 bp) could be a suitable plant barcode. Though this region provided low species resolution (67.3%), because of highly conserved primers and very robust amplification it was considered the most suitable barcode to be used in applications other than taxonomy (forensic science, biotechnology and food industry, animal diet). The P6 loop could be retrieved even from highly degraded DNA from the processed food or from permafrost samples. This highlighted its potential use in food industry, forensic science, diet analyses based on faeces and in ancient DNA studies (Taberlet et al. 2007). Beside these single locus barcodes, a two-locus barcode based on a combination of *trnH-psbA* spacer and *rbcL* was proposed (Kress and Erickson 2007). Likewise, Chase et al. (2007) proposed two 3-locus barcode combinations comprising *matK+rpoB+rpoC1* and *matK+rpoC1+trnH-psbA*. Hollingsworth et al. (2009) proposed that some combinations of *rbcL, matK, rpoC1* and *trnH-psbA* could provide a universal plant barcode for plants. Subsequently, the CBOL Plant Working Group (2009) proposed a combination of *matK* and *rbcL* as the core barcode for plants. Li et al. (2011a) too supported the two genes as core barcodes for ferns.
In the nuclear genome, most of the single copy genes and their introns were not considered suitable for barcoding because of the lack of universal primers for their amplification (Kress et al. 2005, Chen et al. 2010). From the nuclear genome, 5.8S ribosomal cistron along with the internal transcribed spacer (ITS) of the nuclear ribosomal DNA, first suggested by Kress et al. (2005) as a possible barcode for plants, has now been extensively analysed for species discrimination. Chen et al. (2010) validated the use of ITS2 region as an effective barcode for identifying medicinal plant species. Following this, Pang et al. (2010) also reported ITS2/ITS as the potential barcode for identification of members of the family Euphorbiaceae. Yao et al. (2010) based on the analysis of ITS2 sequences of 50,790 plants and 12,221 animals downloaded from the GenBank, NCBI recommended ITS2 as the universal barcode for both animals and plants. Recently, the China Plant BOL Group (2011) has strongly advocated the inclusion of ITS in the core barcode for plants along with matK+ndhF. A point to be kept in mind is that all the recommendations for the universal barcode region(s) were based on floristic studies involving distantly related species.

A number of studies, mostly in the recent past, have evaluated the potential of various loci against a taxonomic backdrop by comparing the species discrimination ability within an order, family or genus. Some examples are Caryophyllales (Cuénoud et al. 2002) at the order level; Geraniaceae (Guisinger et al. 2008), Asteraceae (Gao et al. 2010b), Podostemaceae (Kelly et al. 2010), Euphorbiaceae (Pang et al. 2010), Lemnaceae (Wang et al. 2010), Fabaceae (Gao et al. 2011), Caryoteae (Jeanson et al. 2011), Grimmia (Liu et al. 2011b), Meliaceae (Mueller et al. 2011), Cycadaceae (Nicola-Morejón et al. 2011) and Cactaceae (Yesson et al. 2011) at the family level. At the generic level, different barcode loci have been tested for the discrimination of congeneric species of Araucaria, Inga (Hollingsworth et al. 2009), Acacia (Newmaster and Ragupathy 2009), Crocus (Seberg and Petersen 2009), Solanum (Spooner 2009), Carex (Starr et al. 2009), Dendrobium (Yao et al. 2009, Asahina et al. 2010, Singh et al. 2012), Carex and Kobresia (Le Clerc-Blain et al. 2010), Taxus (Liu et al. 2011a), Agalinis (Pettengill and Neel 2010), Picea (Ran et al. 2010), Alnus (Ren et al. 2010), Berberis, Ficus and Gossypium (Roy et al. 2010), Quercus (Piredda et al. 2011) and Holcoglossum (Xiang et al. 2011).
At the time of initiation of the present work in 2007, though few recommendations, based predominantly on floristic studies, about the possible barcode regions to be used for plants were available (Kress et al. 2005, Chase et al. 2007, Kress and Erickson 2007), even the application of DNA barcoding to plants had become suspect because of the lack of agreement about a core barcode, whether based on a single locus or multiple loci. The loci that had yielded different degrees of success in a variety of plants were ITS from the nuclear genome and \textit{matK}, \textit{rbcL}, \textit{rpoB}, \textit{rpoC1}, \textit{trnH-psbA} spacer from the chloroplast genome. With this background, investigations were initiated in the laboratory to evaluate the efficacy of these loci individually or in various combinations against a taxonomic background by comparing species discrimination at the generic and family level. The study aimed at checking the applicability of these barcode loci at the generic level among congeneric species formed the subject of another thesis and was conducted on \textit{Dendrobium}, an orchid, as an illustrative example (Singh 2012). However, it was considered that a holistic test for the DNA barcoding and the suggested loci would be if it was conducted at the family level. Thus, testing of the species resolution capability of the recommended loci among the other species of genera other than \textit{Dendrobium} and also among various genera of the family Orchidaceae became the objectives of the present thesis.

The family Orchidaceae is one of the largest and highly evolved families of angiosperms (Chase 2005). Orchids constitute 9\% of the Indian flora and mainly confined to Himalayas as their main home and also scattered in Eastern and Western Ghats (http://www.orchidsasia.com). The estimate of number of orchid species existing in India varies from 1,141 (Kumar and Manilal 1994) to 1,600 (Medhi and Chakrabarti 2009). Orchids are generally considered promiscuous in nature as the reproductive isolation is mainly based on the specificity of the pollinator (Cameron 2004). Because of this feature, it was envisaged that orchids would offer a stringent test to DNA barcoding concept in general and selected regions in particular. All orchid species with charismatic ornamental flowers and therapeutic properties are highly endangered due to their over exploitation and are thus listed in Appendix II of the Convention of International Trade in Endangered species of Fauna and Flora (CITES);
some are even listed in Appendix I (http://www.cites.org; Sun et al. 2011). This implies that trade of all orchids, could be undertaken only through export permits, while the trade of those listed under Appendix I is totally prohibited (http://www.cites.org). The export of orchids collected from wild is banned (http://www.cites.org). Despite this, the medicinal and ornamental orchids are still illegally traded using their parts or even fragments, which cannot be identified using traditional taxonomic methods. Orchidaceae species constituting a priceless genetic resource are common targets for conservation both in situ and ex situ in the national parks, germplasm banks and botanical gardens (http://www.sfri.org/orchidology.htm). In order to conserve the orchid species and to curb the illicit trade practices of biopiracy, special identification techniques are required which could distinguish the endangered orchid species from other plants even if a part or fragment is used. DNA barcodes of such orchids once available could become powerful tools in the hands of the enforcement agencies responsible for curbing such illegitimate practices. Besides being used as identification tools, DNA barcodes of the endemic orchids could also serve as genetic resource tags. Therefore, the present investigation was initiated with various objectives (i) checking the applicability of DNA barcoding to the members of Orchidaceae and (ii) to identify the locus/loci which, individually or in combination, could provide unique recognition tags to the investigated orchid species. Besides, the universality of primers used was also tested by applying them to species belonging to other families of vascular plants.